Editors: Claudia Pahl-Wostl, Sonja Schmidt,
Andrea E. Rizzoli, Anthony J. Jakeman

University of Osnabrück, Germany

iEMSS
INTERNATIONAL
Environmental
Modelling
and Software
SOCIETY

Complexity and Integrated Resources Management

Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society

Vol. 1

14-17 June 2004

Complexity and Integrated Resources Management

Transactions
of the 2nd Biennial Meeting of the
International Environmental Modelling
and Software Society

INTERNATIONAL
Environmental
Modelling
and Software
SOCIETY

Editors
Claudia Pahl-Wostl
Sonja Schmidt
Andrea E. Rizzoli
Anthony J. Jakeman

IEMSs 2004 – 14-17 June 2004,
University of Osnabrück, Germany

Complexity and Integrated Resources Management - Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society

Volume Editors

Claudia Pahl-Wostl
Sonja Schmidt
Institut für Umweltsystemforschung
Universität Osnabrück
Artilleriestr. 34
D 49076 Osnabrück, Germany

Andrea E. Rizzoli
IDSIA Istituto Dalle Molle di studi sull'intelligenza artificiale
Galleria 2
CH 6928 Manno, Switzerland

Anthony J. Jakeman
The Centre for Resource and Environmental Studies (Bldg 43)
The Australian National University
ACT 0200, Canberra, Australia

Each paper in this volume was refereed by an Editor, a member of the Editorial Board and two anonymous referees.

ISBN 88-900787-1-5

iEMSs 2004
Published by the International Environmental Modelling and Software Society (iEMSs)
President: Anthony J. Jakeman.
Address: iEMSs, c/- IDSIA, Galleria 2, 6928 Manno, Switzerland
Email: secretary@iemss.org
Website: http://www.iemss.org

Typeset in Como (Italy) by SEA, Servizi Editoriali Associati

IEMSs 2004 – 14-17 June 2004,
University of Osnabrück, Germany

INTERNATIONAL
Environmental
Modelling
and Software
SOCIETY

Complexity and Integrated Resources Management Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society

Claudia Pahl-Wostl, Sonja Schmidt, Andrea E. Rizzoli, Anthony J. Jakeman *(Editors)*

Co-editors

David Batten
Michel Blind
Felix Chan
Barry Croke
Wolfgang-Albert Flügel
Carlo Giupponi
Romy Greiner
Carlo Gualtieri
Nigel Hall
Matt Hare
Reinout Heijungs
Stefanie Hellweg
Suhejla Hoti
Keith Jeffery
Kostas Karatzas
Markus Knoflacher
Peter Krause
Christine Lim
Ian Littlewood
Michael Matthies
Michael McAleer
Dragutin T. Mihailovic
Les Oxley
Jens C. Refsgaard
Otto Richter
Michela Robba
Dale S. Rothman
Miquel Sànchez-Marrè
Dragan Savic
Huub Scholten
Boris Schröder
Jan Sendzimir
Ralf Seppelt
Achim Sydow
Hilde Passier
David Post
Peter Vanrolleghem

Organizers

The conference has been organized by iEMSs
(the International Environmental Modelling and Software Society) in cooperation with:

Harmoni-CA (Concerted Action on Harmonizing Modelling Tools for River Basin Management)
TIAS (Integrated Assessment Society)
IAHS (International Association of Hydrological Sciences)
BESAI (Binding Environmental Sciences and Artificial Intelligence)
MODSS International Conference on Multi-objective Decision Support Systems for Land, Water and Environmental Management
ISESS International Symposium on Environmental Software Systems
ERCIM the European Research Consortium for Informatics and Mathematics

Local Organizers

The local organization of the conference has been managed by:

DBU (German Environmental Foundation)
USF (Institute of Environmental Systems Research, University of Osnabrück)

Editorial

Dear Reader,

The 2nd Biennial Meeting of the International Environmental Modelling and Software Society (iEMSS 2004) was dedicated to the theme: Complexity and Integrated Resources Management", a very topical theme given the increasing complexity of contemporary resource management problems and the increasing uncertainties from global change. The meeting assembled nearly 300 researchers from all over the globe and from a wide range of disciplines. Presentations discussed latest developments in modelling methodologies and software tools applied to different areas of resources management. Contributions provided evidence of the important role of models to improve our understanding of the complexity of socio-ecological systems and to develop appropriate management strategies. Increasing attention was paid to the role of stakeholders in model development and application and to a new role for models in processes of social learning in participatory resources management.

The conference took place in the facilities of the German Environmental Foundation in Osnabrück. The ambience of these low-energy buildings, designed to minimise their impact on the environment, was well suited to the conference theme and their open and flexible structure facilitated intense discussions and exchange not only during but also between sessions.

I hope that readers will share the excitement of conference participants when browsing through the conference proceedings and reading some of the papers in more detail. Interested readers are advised to consult the journals Environmental Modelling and Software and Ecological Modelling and Advances in Geosciences where special issues emanating from this conference will be published. We also look forward to the third biennial meeting, iEMSs 2006, which will be held in Burlington, Vermont, USA (see http://www.iemss.org/iemss2006).

Claudia Pahl-Wostl

October 2004

The International Environmental Modelling and Software Society acknowledges gratefully the assistance of the following people in realizing the iEMSs 2004 conference:

- Claudia Pahl-Wostl for convening the conference
- Sonja Schmidt for organising the conference
- Andrea Rizzoli for the web-based conference management tool, creating and updating the conference website and for expert and technical advice
- all session organizers and reviewers
- Antje Braeuer and Georg Johann for supporting the organisation whenever and wherever necessary
- all members of the Institute of Environmental Systems Research and the department of resource flow management for supporting iEMSs 2004, especially Ilke Borowski, Frank Hilker, Maja Schlüter and Dominik Reusser
- all members of ZUK "Zentrum für Umweltkommunikation" of the German Environmental Foundation

TABLE OF CONTENTS

iEMSs 2004 sessions *(part one)*

Keynotes

Artificial Intelligence Techniques for Integrated Resource Management

Evolutionary Computing Methods for Environmental Modelling and Software Development

Human Behaviour and Agent-Based Modelling

Participatory and Group Model Building for Natural Resource Management

Information Management in Complex Interactions

Uncertainty in Life-Cycle-Assessment

ERCIM: Information Technologies for Integrated Resource Management

Modeling Biocomplexity - Actors, Landscapes and Alternative Futures

John P. Bolte [a], David W. Hulse[b], Stanley V. Gregory[c], and Court Smith[d].
[a]*Department of Bioengineering, Oregon State University, USA*
[b]*Department of Landscape Architecture, University of Oregon, USA*
[c]*Department of Fisheries and Wildlife, Oregon State University, USA*
[d]*Department of Anthropology, Oregon State University, USA*

Abstract: Increasingly, models (and modelers) are being asked to address the interactions between human influences, ecological processes, and landscape dynamics that impact many diverse aspects of managing complex coupled human and natural systems. These systems may be profoundly influenced by human decisions at multiple spatial and temporal scales, and the limitations of traditional process-level ecosystems modeling approaches for representing the richness of factors shaping landscape dynamics in these coupled systems has resulted in the need for new analysis approaches. Additionally, new tools in the areas of spatial data management and analysis, multicriteria decision-making, individual-based modeling, and complexity science have all begun to impact how we approach modeling these systems. The term "biocomplexity" has emerged a descriptor of the rich patterns of interactions and behaviors in human and natural systems, and the challenges of analyzing biocomplex behavior is resulting in a convergence of approaches leading to new ways of understanding these systems. Important questions related to system vulnerability and resilience, adaptation, feedback processing, cycling, nonlinearities and other complex behaviors are being addressed using models employing new representational approaches to analysis. An emerging application area is alternative futures analyses, the study of how complex coupled human/natural systems dynamically respond to varying management strategies and driving forces. This methodology is increasingly being used to inform decision makers about the implications of policy alternatives related to land and water management, expressed in terms related to human valuations of the landscape. Trajectories of change become important indicators of system sustainability, and models that can provide insight into factors controlling these trajectories are rapidly becoming essential tools for planning. The complexity inherent in these systems challenges the modeling community to provide tools that capture sufficiently the richness of human and ecosystem processes and interactions in ways that are computationally tractable and understandable. We examine one such tool, Evoland, which uses an actor-based approach to conduct alternative futures analyses in the Willamette Basin, Oregon. Actor-based approaches, spatially-explicit landscape representations, and complexity science are providing new ways to effectively model, and ultimately to understand, these systems.

Keywords: complexity; resilience; adaptation; simulation

1. INTRODUCTION

The term "biocomplexity" is used to describe the complex structures, interactions and dynamics of a diverse set of biological and ecological systems, often operating at multiple spatial and temporal scales. The study of biocomplexity reflects an intention to understand fundamental principles governing global behavior of these systems, expressed in terms of biological, physical, ecological and human dimensions, in terms of the interactions and resulting patterns and structures that collectively define system responses (Colwell [1998], Levin [1998], Manson [2001]). Several decades of study and appreciation of the rich nature of the interactions that drive many systems of vital interest to humanity have led to an increasingly sophisticated set of hypotheses on how these systems respond to the many perturbations and cycles that they are exposed to. The scientific community is being asked to bring to bear these advances in our collective understanding of systems impacted by anthropogenic influences to improve management and planning of these systems, resulting in the need for new approaches to incorporating human behavior as an important component of ecological and environmental systems behaviors. As human impacts stress the ability of many systems to deliver the wealth of ecological, social and economic goods and services societies rely on, terms such as "vulnerability" and "resilience" have come into common use as ways to think about system response and the implications of human modification of these systems in maintaining functions perceived as important for human and natural uses. The study of biocomplexity identifies and defines a set of concepts, hypotheses and approaches for understanding and characterizing the rich patterns of interactions and behaviors in these systems, with the goal of providing new insights into important questions related to system vulnerability and resilience, self organization and adaptation, feedback processing, cycling, and nonlinearities. The modeling community is developing new approaches to representation and analysis that are allowing exploration of complex systems in ways that are beginning to answer questions how these systems interact, evolve, and transition to new, often unexpected, behaviors.

The challenges of representing and analyzing biocomplex behavior are resulting in a convergence of approaches leading to new ways of understanding these systems. Recent developments in mathematics related to complex systems analysis have provided a variety of new tools and strategies for exploring complex system dynamics (Bak and Chan [1989], Holland [1995], Kauffman [1969], Fernandez and Sole [2003]) Key insights arising from these analyses focus on questions related to identifying system properties that result in self-organizing or emergent behavior, the nature of interactions that can lead to highly nonlinear behaviors in a range of systems, and the circumstances in which "surprises" in system response may be observed. As these concepts have been expanded from their initial focus on primarily physical phenomena to the examination of increasingly rich ecological, economic and social systems, ecological and environmental modeling efforts have become correspondingly more focused on incorporating biocomplexity considerations in their approaches and analyses. Most of these approaches embody the concept that complex behavior arises from the collective interactions of large numbers of relatively simple entities (Holland [1995], Arthur et al. [1997]). Alternatively, the recently proposed theory of Panarchy (Holling [2001], Gunderson and Pritchard [2002]) proposes an alternative hypothesis that states that complex behavior results from a small number of controlling processes operating at multiple spatial and temporal scales. While full articulation of the underpinnings of these approaches is beyond the scope of this paper, they clearly suggest that new modeling and analysis paradigms are needed, and modelers are beginning to incorporate concepts of self-organization, adaptation, multi-scalar interactions and multiple actors alongside more traditional process-based approaches to develop new classes of models able to more fully characterize and simulate biocomplex systems.

Systems scientists have presented many examples of biocomplexity conceptualizations spanning purely ecological (Walker et al. [1969], Carpenter and Cottingham [1997]), social (Emery and Trist [1965], Bella [1997]), economic (Arthur et al. [1997]) and coupled human/natural systems (Scheffer et al. [2002].) However, these broad conceptualizations have not lent themselves to the modeler's need for reasonably concrete, well-articulated and operational definitions amenable to computation and analysis. For example, a Google search using the phrase "ecosystem resilience" returns on the order of 75000 "hits", most of which discuss resilience of particular systems or classes of systems with broad brush strokes, describing in somewhat vague, fuzzy terms the general concept of a system being robust to change. Examined closely, what constitutes "change" generally becomes somewhat nebulous. In some cases, a change in the composition of the system is implied, without reference to the magnitude of the change in question, or whether the compositional change implies a change of function, e.g. the capacity of the system to provide a particular set of goods and services. In other cases, the focus is on examining system behavior, to better understand circumstances in which perturbations of the system will either be absorbed or send the system off in a new direction.

We are seeing a transition from conceptual to more quantitative methods for describing and analyzing these systems (Carpenter and Cottingham [1997], Carpenter et al. [1999], Lepperhoff [2002], Chattoe [1998]), and a rich literature is emerging in this area. A variety of methodologies building on and extending complex analysis of simpler physically-based systems to quantitatively describe and model biocomplex human and natural system behaviors are emerging, based on more traditional stability analyses applied to nonlinear systems. These analyses examine state spaces defined in terms of stability basin structure, distributions of attractors in state space, and ability of perturbations to move the system into alternate stability domains. Extending these concepts into the biocomplexity realm, we can define operationally useful descriptors of complex behavior that are relevant to management. For example, system resilience can be defined as the capacity of a system to absorb perturbations while continuing to operate within it current stability domain. Models that sufficiently characterize the structure of the state space with respect to attractor basin geometries can provide insight to managers on regions where vulnerability of a system to provide specific productions may be high. However, additional challenges exist: real state spaces may be highly multidimensional, dynamic, and nonlinear or even folded, making analysis of their structure difficult. Further, where management is based on multiple criteria (reflected by multiple model outputs), these outputs may have substantially different state space structures. Nevertheless, these concepts are being used to examine more realistic systems and applied to the management realm.

2. ALTERNATIVE FUTURES

In parallel to the emergence of biocomplexity as an analysis paradigm, a number of studies have recently focused on alternative futures analyses (e.g. Baker et al. [2004], Hulse et al. [2000], Santlemann et al. [2001], Steinitz and McDowell [2001], Voinov et al. [1999], Noth et al. [2000]). This has resulted largely from a need and desire to utilize analytical approaches, generally using process-level models synthesizing multiple landscape elements, to predict a particular set of responses of the target landscape to a particular set of perturbations reflecting alternative landscape management. These efforts generally incorporate stakeholder involvement in determining the nature, pattern and scale of the perturbation(s) considered, and resulting modeled landscapes or landscape trajectories are used to assess the outcome behaviors. While these efforts can be very effective for moving models into the policy and management arena and can provide insight into the implications of specific management strategies, they raise a number of issues related to our ability to effectively model the myriad of potential interactions and behaviors that may (or may not) lead to surprising and unforeseen results. While opening the door for modelers to interject current understanding of important processes and interactions into the management of coupled human/natural systems, alternative futures analyses can place additional burdens on the modeler, particularly related to identifying and incorporating interactions across multiple processes, possibly across multiple spatial and temporal scales. For example, a model-based assessment of stream biological production based on vegetative pattern at a site may generate questionable results when the broader influences of channel migration, wood production from upstream areas, or large, low-periodicity flood events can substantially alter that pattern (Van Sickle et al. [2004]). The utility of incorporating additional complexity in a model is often unclear; particularly in situations that are data limited or mechanisms are not well understood, simpler models may be more reliable predictors of system response (NAS [2001]). Representing human decision making in the landscape may be necessary to incorporate the influence of and feedback to the human component of these systems, and can be accomplished through a stakeholder process (Hulse et al. [2004]) or modeled (Etienne et al. [2003]).

3. ACTOR-BASED APPROACHES TO SIMULATING LANDSCAPE CHANGE

3.1 Overview

Landscape change modeling is at the core of most alternative futures analyses, and the last decade had seen considerable activity in this area (see Parker et al. [2003] for an excellent review). This activity is in part a result of the widespread availability of GIS-based platforms and datasets, complimented by a rapid increase in computing power and sophistication of representational tools for software development resulting from a convergence of approaches derived from individual-based modeling and complexity analysis. In particular, actor-based approaches have become a commonly-used tool for representing human interactions driving landscape change, as well as many other types of systems in which collective behavior arises from collective behavior. Actor-based models typical explicitly represent 1) a landscape as a collection of decision units, defined by spatial properties and attributes relevant to the decision making criteria relevant to the task addressed by the modeler, and 2) entities that make decisions and/or take actions that result in landscape change. While the term “agent” is used commonly in the literature to describe these entities, we prefer the term “actor”, since “agent” has a number of connotations in computer science distinctly different than the usage described here, and “actor” has a clearer semantics consistent with common usage of the term in a non-modeling context.

An appeal of an actor-based approach for landscape change modeling is that modeled actors can be based in large part on actual actors contributing to behaviors of the real system which the model is attempting to capture, increasing the realism of the model. Simulated actors may be based on individual decisionmakers, collections of individuals acting as a homogeneous entity, or as abstractions with no specific real-world counterpart (e.g. organizational structures reflecting collective actions that are not captured in specific real-world organizations). From a modeling perspective, the task of the modeler involves determining an appropriate set of characteristics that represent the attributes of the actor relevant to the model, and a set of actor behaviors that capture the decisions or actions of the actors in the system. While the set of necessary actor attributes is highly dependent on the problem being addressed, behaviors typically consist of some form of decision rules that related site and/or system characteristics to a particular actor action and resulting landscape change. Determining an appropriate set of actors and their corresponding behaviors is a significant modeling challenge, and may involve expert knowledge, surveys, demographic and population behavior analysis, and other methods; this is an active area of research.

Self-organization and adaptation are key aspects of many types of complex behavior generally, and landscape change specifically. Adaptation implies that a system modifies its behavior, or “learns”, through the processing of feedback describing the success of current strategies at achieving desired outcomes. Adaptive mechanisms may occur at multiple scales and may operate though a variety of distinct pathways. At an actor level, adaptation may involve changing decision behavior, reflecting changes in landscape production, actor goal satisfaction and other decision criteria. At a system level, adaptation may manifest as higher-order changes in actor composition, changes in decision spaces and system process reorganization. Relatively few current models explicitly incorporate adaptive processes into their representations; this is another area of active development.

3.2 An Example Alternative Futures Modeling Framework – EvoLand

A number of frameworks for complex systems and alternative futures analyses have been developed (Noth et al. [2000], Sengupta and Bennett [2003], Maxwell and Costanza [1995], Daniels [1999]), each providing a specific set of capabilities for representing and manipulating supported representations of the system of interest. These frameworks can simplify implementation of models and provide standard methods for data management, model integration, and analysis. *EvoLand* (for **Evo**lving **Land**scapes) is an example of a modeling tool that supports development of spatially explicit, actor-based approaches to landscape change and alternative futures analysis. EvoLand provides a framework for representing 1) a landscape consisting of a set of spatial containers, or *integrated decision units* (IDU’s), modeled as a set of polygon-based geographic information system (GIS) coverages containing spatially-explicit depictions of landscape attributes and patterns, 2) a set of actors operating on a landscape, defined in terms of a *value system* that couples actor behavior to global and local production metrics and in part determine policies the actor will select for decision making, 3) a set of *policies* that constrain actor behavior and whose selection and application results in a set of outcomes modifying landscape attributes, 4) a set of autonomous process descriptions that model non-policy driven landscape change,

and 5) a set of *landscape evaluators* modeling responses of various landscape production metrics to landscape attribute changes resulting from actor decision making. EvoLand provides a general-purpose architecture for representing landscape change within a general paradigm incorporating actors, policies, spatially explicit landscape depictions, landscape feedback, and adaptation; application-specific components are “plugged in” to EvoLand as required to model particular processes.

The fundamental organizational structure used in EvoLand is shown in Figure 1. Key elements in this organizational scheme are Policies, Actors, Actions, Policy and Cultural Metaprocesses, Autonomous Landscape Change processors, and

Landscape Evaluators. Definitions for these key elements are provided below. Taken together, these elements provide a basic platform for assembling actor-based models of landscape change. Because many of these elements are “pluggable” software components, the basic EvoLand platform can be used with application-specific actor definitions, policy sets, autonomous process descriptions, and landscape evaluators.

3.2.1 Policies

Policies in EvoLand provide a fundamental construct guiding and constraining actor land use/land management decision making. Policies capture rules, regulations, and incentives and other strategies promulgated by public agencies in response to social demands for ecological and social goods, as well as factors used by private landowners/land managers to make land and water use decisions. They contain information about site attributes defining the spatial domain of application of the policy, whether the policy is mandatory or voluntary, goals the policy is intended to accomplish, and the duration the policy, once applied, will be active at a particular site. As actors assess alternative decisions about land management, they weigh the relative utility of potentially relevant policies to determine what policies they will select to apply at any point in time/space, is any. Once applied, a policy outcome is triggered that modifies one or more site attributes, resulting in landscape change.

Policies are characterized by two types of decision variables: (1) those required to be satisfied before the policy can be considered (also known as noncompensatory attributes or *constraints*), and (2) compensatory factors defining the intention of the policy at addressing specific goals, which can be "traded off" against other objectives in decision making using a multiobjective decision making algorithm. Further, policies may optionally be constrained to operating only with selected actor classes (e.g., all home owners, farmers with streams flowing through their property, forest owners with anadromous fish in adjacent streams.)

1) neighbor influence on decision making, and 2) actor membership in *organizations* promoting a specific value system and capable of promulgating new policies consistent with the organization's values. Neighbor influence is intended to capture the concept of diffusion of innovation: i.e. if an actor observes a neighbor utilizing policies that result in a successful outcome consistent with the actor's value system, the actor is more likely to implement that policy.

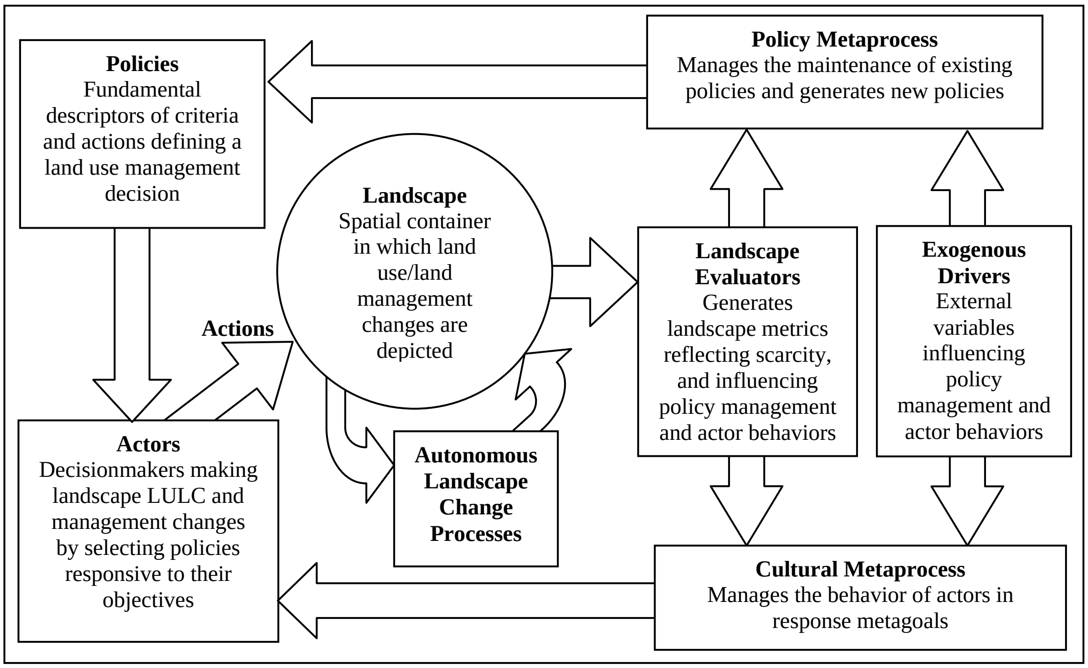


Figure 1. Conceptual Framework for EvoLand

3.2.2 Actors

In EvoLand, actors are entities (individuals or groups) that make decisions about the management of particular landscape units (IDU's) for which they have management authority, based on balancing a set of objectives reflecting their particular values, mandates, and the policy sets in force on the parcels they manage. They do this within the scope of "policy sets" that are operative on particular landscape elements over which they have decision making control. Fundamentally, actors are characterized by the values they express through their behaviors; examples might include ecosystem health, economic production, or property rights. These values in part guide the process actors use to select policies to implement; policies with intentions consistent with the actors' values are more likely to be selected by the actor for application.

In addition to actor values, EvoLand supports interaction between actors via two mechanisms: "Neighbor" is defined in terms of spatial proximity of the two actors; an alternative approach would generalize the location metrics to include non-spatial definitions of proximity, e.g. proximity of values systems.

Actor decision making is based on a stochastic multicriteria model that considers multiple factors to select policies the actor will implement. These factors include the consistency of the policy intention with the actor's values (based on the degree of self-interest the actor exhibits), the alignment of the policy intention with global measures of scarcity of various landscape productions (based on the degree of altruism the actor exhibits), and the degree of actor interaction with other actors successfully employing the policy. EvoLand allow the modeler to experiment with the relative weightings of these factors to exploring their effects on system behavior.

3.2.3 Policy Metaprocess.

EvoLand employs two *metaprocesses* that reflect feedback loops that modify system behavior at a high level. The first of these, the Policy Metaprocess, modifies the policy set that is available to actors. The second, the Cultural Metaprocess, modifies Actor behavior.

The Policy Metaprocess is responsible for generating new policies, modifying existing policies, and removing existing policies that are no longer relevant. An evolutionary model is employed to manage the adaptation and creation of policies responsive to scarcity measures, i.e. a *marketplace of policies* is created, where policies compete for success, defined in terms of measures such as 1) the frequency a given policy is employed, and 2) the utility of the policy at addressing current scarcity issues. The Policy Metaprocess is an example of an adaptive process, using genetic operators (selection, crossover, mutation, and genesis) to evolve new policies based on recombination of successful policies, where the success ("fitness") of a policy is defined via the Landscape Evaluator metrics. This approach doesn't capture actual policymakers in the real system, but focuses more on policy evolution independent of who or what is actually creating those policies.

3.2.4 Cultural Metaprocess.

The Cultural Metaprocess is responsible for adaptively modifying the behavior of actors in the systems. Actor behavior is defined by the value system it uses to guide decision making and its connections to other actors. The Cultural Metaprocess uses output from the Landscape Evaluator to change an actors values in response to shifting societal measures of scarcity, and manages the interactions between actors describe previously.

The specific steps used by the Cultural Metaprocess are similar to that used by the Policy Metaprocess, and focus on allowing actors to adaptively modify their behaviors based on landscape feedback and interactions with other actors. The Cultural Metaprocess may (optionally) adjust actor values in response to changing corresponding to broad societal shifts in values as resources and production become scarce. Alternatively, the Cultural Metaprocess may manifest cultural processes though actor interactions, capturing the concept that as scarcities manifest themselves, the actor population responds through "experiments" that may/may not alleviate the scarcity, and that "successful" experiments spread through diffusion adoption resulting from actors observing successes achieved by other "nearby" actors with similar goals. These experiments are conducted through genesis and evolution of new policies, applied locally; successful policies then have an opportunity to expand globally as an adaptive process. In essence, the system "learns" successful policies through experimentation by individual actors, with successful policies adopted by other actors as landscape attributes and actor interactions allow.

3.2.5 Autonomous Landscape Change Processes

Landscapes change in response to a variety of factors other than human decision making. EvoLand support plug-in components that periodically change the underlying landscape, reflecting autonomous processes that occur independently of human actions. From an alternative futures modeling perspective, this enables EvoLand to incorporate these processes into the simulated trajectories of change. Examples of autonomous processes include vegetative succession, river channel restructuring and meandering in response to flood events, or external population influx and distribution. EvoLand provides a basic framework for incorporating application-specific autonomous processes into a landscape change model, and managing the interactions of these process with policy-driven landscape modifications. Together, these provide a robust representation of change processes that can be adapted to a wide variety of situations.

3.2.6 Landscape Evaluators

These components allow EvoLand to evaluate landscape production of metrics relevant to actor decision making. They are typically spatially explicit models that take a landscape, represented as a attributed coverage of IDU's, as input, and generate a suite of metrics related to a specific type of system production (e.g. ecological population abundances and diversity measures in the case of an ecosystem health-oriented goal; jobs and wealth production in the case of an economically-oriented goal.) The models provide measures of landscape performance and serve as a primary form of feedback considered by EvoLand. They also provide a point at which more traditional approaches to modeling may intersect with actor-based approaches, since these models do not directly interact with actors, but reflect actor influences on landscape change as well as indirectly influencing actor behavior via other mechanisms previously notes. In EvoLand, these models are plug-in components, allowing alternative representations to be readily compared to better understand the implications of specific representations and factors on system behavior, and allowing the extension of EvoLand into additional domains of consideration.

3.2.7 Biocomplexity Analyses

A primary rationale for an alternative futures model, as with any modeling effort, is to provide insights on system behavior. The traditional tools of model analysis (e.g. sensitivity analysis, model verification) are equally applicable to actor-based models. However, the intrinsic complexity typically captured in these models, and the generally long time frames they encompass, suggest a shift in emphasis from rigorous validation to a more exploratory approach to model use. In alternative futures analyses, we are typical more concerned with providing reasonable estimates of the bounds of system behavior than with prediction of specific outcomes, suggesting a Monte Carlo or similar approach focusing on characterizing the likelihood of realizing qualitative distinct system behaviors. From a complexity perspective, the emphasis typically shifts again; analyses focus on system stability, identifying attractors in behavioral space, the nature and strength of these attractors, and the factors that tend to drive the system from one basin of attraction to another characterized by fundamentally different controlling processes, productions and behaviors.

Within EvoLand, we are just starting to experiment with various biocomplexity analyses; our current efforts focus on 1) defining a set of experiments addressing the effects of various mechanism of feedback processing and actor interactions on system behavior, 2) exploring mechanisms of policy evolution and capacity to generate innovative and effective policies as an adaptive process; 3) characterizing the nature of the landscape state spaces to identify dominant attractors that persist under dynamic trajectories of change and the circumstances under which the landscape may move to an alternate attractor basin, and 4) vulnerability of landscapes to change under various policy scenarios.

3.3 Applying EvoLand – A Case Study in the Willamette Basin, Oregon

EvoLand is currently being used to conduct a series of alternative futures analysis in selected areas of Oregon's Willamette River Basin, aimed at better understanding the relationships and interactions between ecological, economic and social drivers of change to improve management of these areas. We are focusing on the confluences of major tributaries along the mainstem of the Willamette River, historically areas of both ecological richness and high anthropogenic impact. The study areas are characterized using spatial datasets incorporating land use and land cover, soils and hydrography, demographic, political and related cultural and physiographic datasets. The IDU's are determined using parcel-level information in combination with other vector coverages relevant to decision making, including floodplain delineations and riparian buffers. Actors are defined primarily through an analysis of demographic patterns; we are currently exploring the use of additional datasets to more richly characterize actor behavior. A number of goals/values are being considered, including ecosystem health, economic production, and property rights protection; each is represented by a landscape model that compute a set of metrics relevant to the particular goal considered. For example, ecosystem health is modeled using a suite of submodels that consider fish abundance and diversity, riparian vegetative structure, and upslope habitat quality; economic production submodels include jobs and wealth production. An initial set of policies are crafted based on current operative policies in the study areas as well as policies that are currently being contemplated; as landscape change modifies landscape production, actor population demographics, and actor behaviors, EvoLand evolves new policies in response to these changing condition as a primary adaptive feedback mechanism. We focus primarily on land use/land cover change, using a 50 to 100 year analysis period and a stochastic analysis approach, using trajectories and patterns of change to determine likely development patterns, vulnerability of specific landscape areas to changes in capacity to provide ecological, economic and social productions of concern. We are using EvoLand to 1) explore the impacts of various feedback loops and interactions on system behavior, expressed though trajectories of change and the nature of the resulting attractor basins of the system productions described above; 2) identify policy characteristics that lead to more or less vulnerable landscapes, and 3) understand the critical linkages between the coupled human/natural systems that collectively generate landscape change.

The explicit consideration of adaptive approaches for policy discovery and actor behavior evolution provide EvoLand the ability to generate new, potentially unexpected behaviors, and suggest a new set of analyses examining effects of alternative feedback mechanisms, actor interactions, and adaptive strategies on policy evolution and trajectories of change. This adaptive response allows EvoLand to create and explore fundamentally new strategies for actor/environment interactions that are not "preprogrammed" into the models. Are ultimate goal is to understand the role that "experimentation", expressed in terms of new strategy generation, propagation of successful strategies through actor networks, and actor rejection of unsuccessful policies, has on the ability of the system to respond to scarcity and changing landscape dynamics. We anticipate being able to test, for example, a hypothesis that resilience of a landscape, defined in terms its state space structure, is related to its capacity for innovation and adaptation, in ways that more traditional models cannot. Using models that significantly capture the rich interactions of actual coupled human/natural systems opens the door to utilizing the insights achieved by these modeling efforts to better understand management of real-world biocomplex systems.

Analyses such as these are fundamentally multiperspective, integrative, and spatially distributed; an actor based approach appears well suited to capturing the rich set of individual behaviors, distributed across a spatially heterogeneous landscape, that collectively result in the system-level patterns these systems display. EvoLand provides a reasonably flexible framework that allows adaptation of existing evaluative models into an actor-based modeling paradigm, and facilitates analysis of feedbacks, adaptive processes, and system behavioral response patterns. Significant issues exist, particularly related to sufficiency of actor characterization, model validation, and interpretation of the rich sets of spatial and temporal information produced by the model. The modeling community has yet to develop a broadly accepted set of approaches to these issues.

4. Future Challenges

There remain many issues and challenges related to the use of actor-based models of biocomplex systems. Despite the wealth of discussions in the literature related to both biocomplexity and actor-based modeling approaches, few concrete examples of the use of actor-based models addressing biocomplexity issues have been presented; still fewer of these incorporate adaptive mechanisms and internal experimentation as fundamental aspects of representation. A key issue at this point whether these approaches represent only a current fad, extensions to previous methodologies, or fundamental new approaches to modeling and understanding complex systems. The complexity analysis models drawing from simpler physical systems have yet to be convincingly demonstrated to have real-word relevance to the more complex adaptive systems ecological and environmental modelers typically address. Our current models, particularly those addressing alternative futures analysis, are difficult to verify in any traditional way and new approaches and datasets are needed to validating these models; this will be a key challenge to allow more widespread acceptance of these models for real-world applications. We have yet to develop well-specified operational definitions of key concepts like resilience, vulnerability, and adaptation, although current models are beginning to make progress in this area (e.g. Carpenter et al. [1999]). Indeed, no widely accepted general theoretical framework for expressing biocomplexity concepts currently exists, much less a common set of approaches for representing this complexity in our models of these systems. However, a number of approaches are being developed and being applied to the analysis of real systems. In particular, actor-based modeling approaches are beginning to emerge and appear to provide a powerful tool for representing the wealth of individual decisions, actions, and interactions that frequently characterize these systems, particularly as adaptive processes become fundamental incorporated and explicitly represented in these models. Models such as EvoLand provide examples for which operational approaches to representing and characterizing actors, adaptive processes, and interpretation of biocomplex responses are being developed.

5. References

Arthur, W.B., S.N. Durlauf, and D. A. Lane, *The Economy as an Evolving Complex System*, Addison Wesley, Reading, MA, 1997.

Bak, P. and C. Chan, Self-Organized Criticality. Physics News in 1988, *Physics Today* January, S-27, 1989.

Baker, J. P., D.W. Hulse, S.V. Gregory, D. White, J. Van Sickle, P.A. Berger, D. Dole, and N.H. Schumaker, Alternative futures for the Willamette River Basin, Oregon, *Ecological Applications* 14(2):313-324, 2004

Bella, D.A., Organized complexity in human affairs: The Tobacco Industry, *Journal of Business Ethics* 16:977-999, 1997.

Carpenter, S. R., and K. L. Cottingham, Resilience and restoration of lakes, *Conservation Ecology* 1(1): 2, URL: http://www.consecol.org/vol1/iss1/art2, 1997.

Carpenter, S., W. Brock, and P. Hanson, Ecological and social dynamics in simple models of ecosystem management, *Conservation Ecology* 3(2):4, URL: http://www.consecol.org/vol3/iss2/art4, 1999.

Chattoe, E., Just how (un)realistic are evolutionary algorithms as representations of social processes?, *Journal of Artificial Societies and Social Simulation*, 1(3), URL: http://www.soc.surrey.ac.uk/JASSS/1/3/2.html, 1998.

Colwell, R., Balancing the biocomplexity of the planet's living systems: A twenty-first century task for science, *Bioscience* 48 (10) 786, 1998.

Daniels, M., Integrating Simulation Technologies with Swarm, *Agent Simulation: Applications, Models and Tools Conference,* University of Chicago, 1999.

Etienne, M., C. LePage, and M. Cohen, A step-by-step approach to building land management scenarios based on multiple viewpoints on multiagent social systems. *Journal of Artificial Societies and Social Simulation*, *6*(4), URL: http://jasss.soc.surrey.ac.uk/6/2/2.html. 2003.

Emery, F. and E. Trist, The causal texture of organizational environments, *Human Relationships*, 18:21-31, 1965.

Fernandez, P. and R.V. Sole, The role of computation in complex regulatory networks, *SFI working paper 03-10-055*, Sante Fe Institute, NM, 19 pp, 2003.

Gunderson, L.H and L. Pritchard, *Resilience and the Behavior of Large-Scale Systems*, Island Press, Washington D.C., 287 pp, 2002.

Hulse, D.W., A. Branscomb, and S.G. Payne, Envisioning alternatives: Using citizen guidance to map future land and water use, *Ecological Applications*, 14(2):325-341, 2004.

National Academy of Sciences (NAS), Assessing the TMDL Approach to Water Quality Management, National Academy Press, Washington DC, 2001.

Holland, J.H., *Hidden Order: How Adaption Builds Complexity*, Helix Books, Addison-Wesley, Reading, Mass, 184 pp, 1995.

Holling, C.S., Understanding the complexity of economic, ecological and social systems, *Ecosystems*, 4:390-405, 2001.

Hulse, S.W., J. Eilers, K. Freemark, C. Hummon, and D. White, Planning alternative future landscapes in Oregon: evaluating effects on water quality and biodiversity, *Landscape Journal*, 19(2) 1-19, 2000.

Kauffman, S.A., Metabolic stability and epigenesis in randomly constructed genetic nets, *J. Theoretical Biology*, 22:437-467, 1969.

Lepperhoff, N., SAM - Simulation of computer-mediated negotiations, *Journal of Artificial Societies and Social Simulation*, 5(4), URL: http://jasss.soc.surrey.ac.uk/5/4/2.html. 2002.

Levin, S.A., Ecosystems and the biosphere as complex adaptive systems, *Ecosystems*, 1:431-436, 1998.

Manson, S.M., Simplifying complexity: A review of complexity theory, *Geoforum*, 32(3) 405-414, 2001.

Maxwell, T. and R. Costanza, Distributed Modular Spatial Ecosystem Modelling, *International Journal of Computer Simulation: Special Issue on Advanced Simulation Methodologies* 5(3): 247-262, 1995.

Noth, M., A. Borning, and P. Waddell, An extensible, modular architecture for simulating urban development, transportation and environmental impacts, *Technical Report 2000-12-01*, Department of Computer Science, University of Washington, 25 pp., 2000.

Parker, D.C., S.M. Manson, M.A. Janssen, M.J. Hoffmann, P. Deadman, Multiagent system models for the simulation of land-use and land-cover change: A review, *Annals of the Association of American Geographers*, 93(2), 2003.

Santlemann, M., K. Freemark, D. White, J. Nassauer, M. Clark, B. Danielson, J. Eilers, R.M. Cruse, S. Galatowitsch, S. Polasky, K.Vache, and J. Wu, Applying ecological principles to land use decisionmaking in agricultural watersheds. In: *Applying Ecological Principles to Land Management*, V.H. Dale and R.A. Haeuber, eds, Spinger Verlag, New York, pp. 226-252, 2001.

Sengupta, R., and Bennett, D. A, Agent-based Modeling Environment for Spatial Decision Support, *International Journal of Geographical Information Science*, 17 (2): 157-180, 2003.

Sheffer, M., F. Westley, W.A. Brock, and M. Holmgren, Dynamic interactions of societies and ecosystems – Linking theories from ecology, economy and sociology, p 195-239 In: *Panarchy: Understanding Transformations in Human and Natural Systems*, L.H. Gunderson and C.S. Holling, eds. Island Press, Washington D.C., 2002.

Steinitz, C. and S. McDowell, Alternative futures for Monrow County, Pennsylvania: a case study in applying ecological principles, In: *Applying Ecological Principles to Land Management*, V.H. Dale and R.A. Haeuber, eds., Springer Verlag, New York, pp. 165-193, 2001.

Van Sickle, J.V, J. Baker, A. Herlihy, P. Bayley, S. Gregory, P. Haggerty, L. Ashkenas, and J. Li, Projecting the biological condition of streams under alternative scenarios of juman land use, *Ecological Applications* 14(2):368-380, 2004.

Voinov, A., R. Costanza, L. Wainger, R. Boumans, F. Villa, T. Maxwell and H. Voinov, Patuxent landscape model: integrated ecological economic modeling of a watershed, *Journal of Ecosystem Modelling and Software*, 14:473-491, 1999.

Walker, B.H., D. Ludwig, C.S. Holling, and R.M. Peterman, Stability of semi-arid savanna razing systems, *Ecology* 69: 473-498, 1969.

Climate and Water Management

Pavel Kabat

a Wageningen University and Research Centre, Wageningen, The NetherlandsPavel.Kabat@wur.nl

Abstract: This paper provides a synthetic overview of the relationship between climate variability and change and the management of freshwater resources. It is divided into three parts. The first part discusses the impacts of climate variability and change on water resources management. The second part outlines some of the steps that can be taken to cope with climate variability and adapt to climate change. The third summarizes some of the implications of the analysis, including the challenges for sustainable development.

1. THE IMPACTS OF CLIMATE VARIABILITY AND CHANGE ON WATER RESOURCES

Natural climate variability has a direct and fundamental bearing on water resources and its management. Finding ways to cope with climate variability lies behind many water resources development and management decisions, and the evolution of hydrology in many ways is a reflection of humankind's struggle to deal with the variation of rainfall and streamflow, both seasonally and inter-annually. Coping with floods and droughts is a major part of the responsibility of water agencies across the world, and their ability to deal with climate variability and its unpredictability is clearly a factor in determining the efficiency and effectiveness of water use.

Importantly, small changes in climate variability can be significantly amplified through the hydrological cycle and have major implications for water resource managers. Data on the drought in West Africa during the 1970s and 1980s, presented in Table 1 (from Servat et al, 1999), show that precipitation decreases of about 25% during this period translated into annual flow reductions of some 50%. Small changes in total or temporal patterns of precipitation may well have considerably larger effects on water resources.

Likewise, climate change is expected to have a major impact on water resources. A significant proportion of the solar energy received by the Earth is utilized in driving the hydrological cycle -- for evaporating the vast quantities of water that are moved and precipitated every year. Since increases in greenhouse gas concentrations in the atmosphere will lead to an increase of the available energy on the surface of the Earth, an 'intensification' of the hydrological cycle will occur. The oceans play a major role in climate, storing and releasing sizeable proportions of the incoming energy. As a consequence, the expected intensification of the hydrological cycle will not be experienced as a smooth linear trend, but rather in the form of oscillations of the variability of climate – with more frequent oscillations and increases in amplitude over some areas. Global increases in temperature will have profound effects on event magnitudes, frequencies and intensities of rainfall as well as its seasonal and geographical distribution and its inter-annual variability (see Table 2, from IPCC, 2001).

Table 1 Decreases of precipitation for countries in West and Central Africa and decreases of river flows in the same region. The averages of the period 1970-1989 are compared with those from the period 1950-1969 (Source: Servat et al., 1998)

Country	Reduction in Precipitation (%)
Cameroon	16
Togo	16
Central African Rep.	17
Benin	19
Ghana	19
Nigeria	19
Guinea	20
Chad	20
Ivory Coast	21
Burkina Faso	22
Guinea Bissau	22
Mali	23
Senegal	25

River	Gauging Station	Reduction of Annual Flow (%)
Comoe	Aniassue	50
Chari	Ndajmena	51
Logone	Lai	39
Niger	Malanville	43
Niger	Niamey	34
Bani	Douna	70
Oueme	Sagon	42
Sassandra	Semien	36
Senegal	Bakel	50
Bakoye	Ouali	66
Black Volta	Dapola	41
Black Volta	Boromo	46
Oubangui	Bangui	30

Table 2 Summary of climate changes and likely impacts on water resources (IPCC, 2001)

Climate change forecast	Climatic change already observed?	To occur in the 21st century?	Effects on water resources
Higher maximum temperatures and more hot days over nearly all land areas	Likely	Very likely	Water resources reduced
Higher minimum temperatures, fewer cold days and frost days, over near all land areas	Very likely	Very likely	Water resources reduced
Diurnal temperature range reduced over most land areas	Very likely	Very likely	
Increase of heat index over land areas	Likely over many areas	Very likely over most areas	Water resources reduced
More intense precipitation events	Likely over many northern hemisphere mid-to-high latitude areas	Very likely over many areas	More frequent and more severe floods
Increased summer continental drying and associated risk of drought	Likely in a few areas	Likely over most mid-latitude continental interiors	More frequent and more severe droughts
Increases in tropical cyclone peak wind intensities	Not observed in the few analyses available	Likely over some areas	More frequent and more severe storm-surge floods
Increases in tropical cyclone mean and peak precipitation intensities	Insufficient data	Likely over some areas	More frequent and more severe floods

The impacts of climate change on freshwater resources must be viewed in the context of a substantial growth during the next 50 years in the net consumption of water – since even if the current trend of a decrease in net consumption in several of the most developed countries continues, water use by 2025 is expected to increase by 25 to 50% (Cosgrove and Rijsberman, 2000). These impacts must also be viewed in the context of the strong natural temporal and spatial variability of the world's freshwater resources. A sizeable proportion of the world's population lives in scattered rural habitats and depends on the availability of water in its natural form for satisfying their basic needs. In differing climate regimes, such 'natural' availability of water is highly variable within a given year and between one year and the next. Any reduction in the amount of water resources or small changes in their distribution in time can lead to serious impacts.

Climate variability and change have been identified as key drivers of ecosystem health and the growth and spreading of water-related diseases. Freshwater is an essential driver of terrestrial ecosystems (such as rainforests) and freshwater ecosystems (such as lakes and wetlands). Since the distribution of terrestrial ecosystems depends on the equilibrium between climate variability and the resilience capacity of these systems to such variability, any change in climate variability or trend in one of its components will lead to latitudinal and altitudinal shifts in the distribution of ecosystems. Likewise, aquatic ecosystems and associated water-related diseases are extremely sensitive, both positively and negatively, to any changes in water temperature, depth, velocity and other parameters resulting from climate change.

Experience has shown that the hydrological cycle is able to accommodate a certain level of external forcing, such as a reduction in precipitation. However, if the forcing is too strong or is maintained for too long, river flows may drop suddenly and sharply, with consequences for both people and ecosystems. In addition, water resources are extremely sensitive to complex non-linear feedback mechanisms due to their non-linearity. This characteristic is illustrated by sediment flows in rivers, which increase exponentially with flood intensity; if climate change leads to more frequent floods, a doubling of reservoir sedimentation rates could easily result.

All of this suggests that climate change and increases in climate variability will exacerbate vulnerability of water resources unless innovative management practices to help cope with, and adapt to, these changes are put in place. The good news is that mechanisms for coping with present-day climate variability in water resources take us a long way down the road towards adapting to any further impacts of climate change.

Coping with climate variability and adapting to climate change

Adaptation and coping strategies are not new. Since the time of Noah's flood, societies and civilisations have adapted to the vagaries of climate variability through various water-related coping mechanisms and adaptation strategies. Over time, however, there has been a substantial evolution in the technologies that allow for more efficient use of water in industry, households and agriculture, as well as more efficient water management systems. This evolution has been accelerated in the last couple of decades by two developments that have the potential to profoundly affect water management adaptation and coping strategies:

The significant improvement in our ability to predict, with some degree of accuracy, climate variability at seasonal and inter-annual scales; and

Our increasing recognition of the likelihood of climate change and its impacts on water resources.

The core of water management has been its historically evolving adaptive capacity and capability. Adaptation and coping strategies for climate variability and change are thus not something that can easily be distinguished from normal water management practices. There are no management options that are uniquely suited for adaptation to climate change that would be measurably different from those already employed for coping with contemporary climate variability. However, there can be a substantive difference in overall policy: whether one adopts a more conventional and incremental 'no regrets' approach, taking measures whose benefits equal or exceed their cost to society and that are worth doing anyway, as opposed to a more anticipatory 'precautionary' approach – one that recognises that the absence of full scientific certainty should not be used as a reason to postpone decisions when faced with the threat of serious or irreversible harm.

Nevertheless, our understanding and implementation of the integrated ensemble of water management measures has changed over time. In particular, Integrated Water Resources Management (IWRM) – "a process which promotes the co-ordinated development and management of water, land and related resources, in order to maximize the resultant economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems" (GWP, 2000) -- has become widely accepted as a key to sustainable freshwater management. A fundamental aspect of IWRM is that it requires integration both within and between the natural and the human system (GWP, 2000). This means integration at various levels – between the lowest level of user to the top policymakers in government; among all the institutions responsible for resources management; among all relevant disciplines; among planning, regulatory, design, operations, maintenance and monitoring functions, and among different stakeholder groups. Most importantly for our purposes, IWRM requires integration at a variety of spatial and temporal scales -- from the global all the way to community and household spatial scales, and from the long-term and decadal temporal scales all the way to inter-seasonal variability and weather events (see Schulze, 2001). It follows that IWRM should also be the encompassing paradigm for coping with natural climate variability and the prerequisite for coping with the highly uncertain consequences of global warming and associated climate change.

2. COPING WITH CLIMATE VARIABILITY

Efforts to cope with climate variability can be significantly advanced if such variations can be predicted, at least to some extent. Although until recently seasonal and inter-annual climate variability were considered impossible to forecast with any degree of certainty, the capacity to predict climate anomalies has grown enormously in the last couple of decades, primarily as a result of the increased ability to forecast the El Niño phenomenon (Bates, 2002).

Clearly, using climate information and skilful forecasts to manage water better has the potential for significant gains, which has generated considerable excitement in some quarters. As a result, the number of efforts to apply climate prediction to water resources management is increasing, ranging from the experience of the Meteorological and Water Resources Agency (FUNCEME) of the State of Ceara in North East Brazil, where the El Niño phenomenon is often associated with devastating drought, to the work of the South Florida Water Management District's experience with incorporating climate forecasts in the development of operating rules for the primary storage component in the State's Everglades ecosystem (Bates, 2002).

Despite these very promising experiences, much work remains to be done. As noted by Bates (2002), the net benefits of using seasonal climate forecasts in water management need to be comprehensively demonstrated. Authoritative, objective, dependable, explicit user-friendly forecasts at temporal and spatial scales tailored to user requirements need to be developed. Scientists need to use appropriate mechanisms for technology transfer, including terminology comprehensible to water managers and tools and methodologies that are acceptable to users. Perhaps most importantly, there is a need for a long-term commitment to dialogue between users and producers of climate forecasts.

As capacity is developed and water managers realize their full potential, short- to medium-term climate forecasts will undoubtedly see increased use in water management applications. Likewise, as forecast skill levels increase, the application of climate forecasting in water resources will increasingly strengthen the case for their value.

3. ADAPTING TO CLIMATE CHANGE

In examining what water managers can do to adapt to climate change, it is important to start by recognizing that at present long-term climate forecasts cannot meet the operational needs of today's water managers. Nevertheless, water managers have two important tools to bring to the table:

- Methods for coping with present-day climate variability in water resources, which take us a long way down the road towards coping for any further impacts of climate change. In particular, further improvement in short-term forecasting could be one of the most important technological breakthroughs to improve our adaptive capacity.

- The full implementation of Integrated Water Resources Management (IWRM) approaches, which will also provide a significant basis upon which to adapt to further impacts of climate change. That this is so is illustrated by Table 2, which shows the IPPC adaptation recommendations for water resource managers. Clearly, these do not differ in any major respect with recommendations that might emerge from an IWRM approach.

Nevertheless, in considering IWRM as the basis for coping with long-term climate change, it is essential that such approaches explicitly address the need for integration at a variety of temporal scales. This dimension of IWRM has not been explored extensively to date. In particular, two issues need consideration. The first (Stakhiv, 1998) is whether water management methods based on the assumption of a stationary climate can be used when that assumption is no longer valid. Some authors, such as Matalas (1997), have argued that stochastic hydrology "can accommodate the uncertainties in water supplies induced by global warming with the operational assumption of stationarity as meaningfully as with the assumption of non-stationarity."

The second issue is whether IWRM – clearly a necessary condition for adapting to climate change – is also sufficient. Gleick (2000) argues that the sole reliance on traditional management responses is a mistake, for four reasons: (1) climate change could lead to hydrologic conditions and extremes of a different nature than current systems were designed to manage; (2) climate changes could produce similar kinds of variability but outside of the range for which current infrastructure was designed and built; (3) relying solely on traditional methods assumes that sufficient time and information will be available before the onset of large or irreversible climate impacts to permit managers to respond appropriately; and (4) special efforts or plans are likely to be required to protect against surprises or uncertainties. Based on the above, Gleick (2000) stresses that "complacency on the part of water managers … may lead to severe impacts that could have been mitigated or prevented by cost-effective actions taken now." But an IWRM policy built on a "no regrets" basis and that properly integrates different time scales would appear able to address these needs and contain the flexibility and robustness to withstand all but the severest climate change scenarios. In addition, sensitivity analyses conducted on watersheds and river basins under a variety of scenarios may help to refine the operation and design of these systems for even greater resiliency (Stakhiv, 1998).

All of the above reinforces the fact that there is an impressive catalogue of specific management measures, both structural and non-structural, that water managers already use routinely to accommodate present-day climate variability that could also serve adaptation to the impact of enhanced climate variability and climate change. Importantly, the expected impacts on water resources should be neither under- nor overestimated, but should be assessed in a scientifically rigorous and realistic manner. Only then can management decisions and appropriate adaptation strategies be formulated.

Adaptation and coping measures are scale-dependent and may vary from individual households to local communities to catchments, as well as from national to international scales. The protective value of natural systems for shielding communities and regions from climate related disasters has been chronically undervalued in decision making. Measures that enhance both ecological and human resilience in the most vulnerable settings are, therefore, crucial for mitigating the growing risk of climate change and climate related disasters.

4. IMPLICATIONS OF THE ANALYSIS

The above discussion of the impacts of climate variability and change on the management of freshwater resources has a number of implications.

The first is that the poor are not only the most vulnerable to the impacts of climate variability and change on water resources, but those with the least capacity to cope with such impacts. Even without climate change, most developing countries will be confronted with serious water problems by the middle of the 21st century. Few of the most vulnerable countries can, therefore, presently afford actions to deal exclusively with impacts of climate change on water resources; win-win actions that address directly the more immediate water management problems while preparing for the consequences of longer term climate changes will usually be the best approach.

The second implication is that enhanced knowledge is needed to better understand the relationship between climate change and water resources -- particularly the processes of the hydrological cycle in relation to the atmosphere and the biosphere. Expectations of water managers that output from GCMs of climate change scenarios will be usable at the relatively detailed spatial and temporal scales at which they operate will not yet be met in the next few years. A global scientific system for assessing the intensification of the hydrological cycle and for predicting and monitoring the effects of climate variability and change on water resources at the water manager's scale of operation is needed.

The third implication concerns politics and institutions. Policy and institutional changes will clearly be needed if coping and adaptive capacity is to be fully developed. One need will be investments in capacity building, especially in the most vulnerable countries. Another will be enhancing co-ordination and drawing up clear divisions of competence, tasks and responsibilities among different agencies acting in watersheds (rather than only in administrative divisions). And ensuring participation of stakeholders in decision-making as well as access to more and better quality information will be key.

A fourth implication is that we are still far from being able to reliably identify areas of special vulnerability. Identifying 'hot spot' areas of potential and actual vulnerability will help in the allocation of scarce scientific and policy resources to priority areas where averting or mitigating climate-related risks is most needed. In order to make progress in identifying such areas, much work is required to clarify the definition of 'hot spot' areas, specify the thresholds of vulnerability, and develop a consistent framework for vulnerability assessment.

And the fifth implication is the need for continuing dialogue between water managers and the climate community of the kind spearheaded by the International Water and Climate Dialogue (www.waterandclimate.org). As climate forecasts continue to become more accurate, the dialogue between climate specialists and water resources managers must keep pace with these advances in order to ensure knowledge is transferred. Promoting cooperation among institutions dealing with water, weather and climate, whether they be operational agencies or research groups, will be crucially important in the years ahead.

5. References

Bates, Bryson, Editor, 2002. Summary Report, "Joint International Workshop on Applicable Methods for Use of Seasonal Climate Forecasts in Water Management", International Research Institute for Climate Prediction (IRI) and Dialogue on Water and Climate, Palisades, New York, November.

Cosgrove, William, and Frank Risjberman, 2000. World Water Vision: Making Water Everybody's Business. Earthscan Publications Ltd. London, UK.

Gleick, P.H. et al., 2000. Water: the potential consequences of climate variability and change for the water resources of the United States. Report for the US. Global Change Research Program, September 2000

Global Water Partnership, Technical Advisory Committee (TAC), 2000. TAC Background Papers No. 4, Global Water Partnership, Stockholm, Sweden, March.

IPCC, 2001. Climate Change 2001, Impacts, Adapation and Vulnerability, Section, Hydrology and Water Resources, Report of Working Group II of the Intergovernmental Panel on Climate Change.

Kabat, P., R.E. Schulze, M.E. Helmuth, and J.A. Veraart, Editors, 2002. "Coping with Impacts of Climate Variability and Climate Change in Water Management: A Scoping Paper", DWC-Report No. DWCSSO-01, International Secretariat of the Dialogue on Water & Climate, Wageningen, Netherlands,

Lenton, Roberto L., 2002. "Managing Natural Resources in the Light of Climate Variability, Natural Resources Forum, JNRF: 26:3, August, 2002

Matalas, N.C. 1997. Stochastic Hydrology in the Context of Climate Change. In Frederick, K.D., Major, D.C. and Stakhiv, E.Z. (eds), Climate Change and Water Resources Planning Criteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Servat, E., Paturel, J.E., Lubes-Niel, H. Kouame, B., Masson, J. M., Travaglio, M., and Marieu, B. 1999. De differents aspects de la variabilite de la pluviometrie en Afrique de l' Ouest et Centrale non Sahelienne. Rev. Sci. Eau, 363-387.

Stakhiv, E. Z., 1998, Policy implications of climate change impacts on water resource management, Water Policy, Vol 1, pp 159-175.

Schulze, R.E., 2002. Modelling as a Tool for Integrated Water Resources Management: Conceptual Issues and Case Study Applications. Water Research Commission, Pretoria, RSA, Report No. 749/01/01.

Economics on the Edge of Chaos:

How does economics deal with complexity and the implications for systems management

Les Oxley
University of Canterbury, New Zealand

We discuss some issues and challenges facing economic modellers when confronted with data generated within a non-linear world. The pitfalls associated with the linearization of inherently non-linear models are raised and the concept of *robustness* defined and proposed as a property of scientifically valid models. The existence of chaos in economic time series is discussed and an example, using financial data, presented.

Keywords: *linearization; robustness, structural stability, chaos, attractor.*

> ***"……. since our premises are necessarily false, good theorising consists to a large extent in avoiding assumptions....(with the property that)....a small change in what is posited will seriously affect the conclusions."*** **Baumol (1958)**

1. INTRODUCTION

Clearly the economic world is nonlinear, so it would appear that focusing on linear dynamics is of limited interest. However economists have typically found nonlinear models to be so difficult and intractable that they have adopted the technique of linearisation to deal with them. Thus linear dynamics have retained a role in economic analysis.

The dynamics which are the focus of this paper cannot arise in a linear dynamical system. Let the vector **x** represent the deviations (possibly in logs) of economic variable from their equilibrium values and A an $n \times n$ matrix. Then the equation:

$$\dot{\mathbf{x}} = A\mathbf{x} \quad (1)$$

represents a linear dynamical system. Any solution path of (1) either:

- converges to the origin (corresponding to equilibrium of the underlying model) or
- diverges to infinity or
- is a cyclical path surrounding the origin (the phase portrait is a centre).

This last possibility is *structurally unstable* in the sense that arbitrarily small perturbations of the matrix A will cause the solution paths to transmute into either divergent or convergent paths. Thus regular and persistent cycles cannot occur in a linear model in a structurally stable way. We return to structural instability below.

When nonlinear models are admitted the dynamic possibilities become much more interesting. Suppose (2) is replaced by:

$$\dot{\mathbf{x}} = F(\mathbf{x}) \quad (2)$$

where F is any (reasonable but possibly nonlinear) function. The dynamic behaviour of (2) includes the following possibilities:

- convergence to points
- divergence to infinity
- convergence to limit cycles
- convergence to strange (fractal) attractors.

The last of these occurs in chaotic systems, and plays an important role in section XX below, which is concerned with empirical chaos.

At this stage it is worthwhile drawing the important distinction between *structural stability* and *robustness.* The former is a property of *whole phase portraits*: we define a structurally stable system as one whose phase portrait transmutes into a qualitatively equivalent (homeomorphic) phase portrait in response to arbitrarily small parameter perturbations. The latter is a property of *individual solution paths*: a robust path is one which transmutes into another path with the same limiting behaviour (as $t \to \infty$) in response to arbitrarily small parameter perturbations (where parameters now include boundary conditions). The closed orbit (centre) phase portrait of a linear model is *structurally unstable*. Saddlepoint models of the type discussed in sections 2 and 3 below are *structurally stable*, and in such a model divergent paths are robust, while convergent paths are not. It will be argued in section 3 that non-robust paths are unobservable and thus, for scientific purposes, can be ignored.

Section 2 below deals with the technique of linearisation and examines its shortcomings. Section 3 develops the notion of robustness and uses it to show that many conventional macroeconomic models can be rejected on methodological grounds alone. Both sections use a simple ‘money and growth’ model to illustrate the argument. This model has certain implications for hyperinflation which are examined empirically in section 4. Section 5 turns to the question of chaos and reports the results of research on the detection of chaos in financial data. Section 6 concludes.

2. LOCAL V. GLOBAL ANALYSIS

Most economic models are inherently nonlinear but, as mentioned in the Introduction, economists have typically found nonlinear dynamics to be intractable and have, accordingly, resorted to the method of local linearisation. This approach relies on a theorem due to Hartman which provides a local linear approximation to a dynamical system such as equation 2 above. The theorem tells us that:

- provided the phase portrait of equation 2 is not a centre, it will be *locally* homeomorphic to the phase portrait of its linearisation (which can be located by evaluating the matrix of first partial derivatives of $F(\mathbf{x})$).

The two vital aspects of this result are:

- The equivalence of phase portraits is *topological* (homeomorphic) only. This equivalence preserves only topological properties, not, for example, distances, angles, curvature or rates of change.
- The equivalence is *local*. It holds only in an arbitrarily small neighbourhood of the point at which the linearisation is calculated. (This could be any point, but in economic models is usually taken to be an equilibrium, i.e. a point $\mathbf{x}^*$ such that $F(\mathbf{x}^*) = \mathbf{0}$.) Outside such an arbitrarily small neighbourhood the equivalence might easily break down.

Both these features of the linearisation process can be misleading when it is applied to economic models, which in many cases, require global rather than local analysis. In this section we develop a simple perfect foresight, market clearing “money and growth” model due originally to Burmeister and Dobell (1971). We compare the local and global dynamics of this model. Local linearisation generates the standard saddlepoint dynamics with a unique long run equilibrium possessing familiar macroeconomic properties. Global analysis reveals that the model has two long run equilibria, a saddlepoint and a stable equilibrium. Convergence to the latter is robust, while convergence to the former is non robust.

Assume a constant returns to scale production function:

$$Y = F(K, L) \tag{3}$$

where Y = output, K = capital stock, L = labour. This may be written in intensive form:

$$y = f(k) \tag{4}$$

where $y = Y/L$ and $k = K/L$. A simple proportionate consumption function is assumed:

$$C = cY \qquad (0 < c < 1) \tag{5}$$

giving a goods market equilibrium condition:

$$Y = cY + \dot{K} + \delta K \qquad (0 < \delta < 1) \tag{6}$$

where δ = depreciation rate. The goods market is assumed to clear continually and thus equation (6) is permanently satisfied. From equations (3) to (6) it is easy to derive the following differential equation:

$$\dot{k} = sf(k) - (\delta + n)k \equiv B(k) \tag{7}$$

where $s = 1 - c$ and n = (exponential) growth of labour force.

Turning now to the money market, a constant exogenous growth rate θ is assumed for the nominal money supply (M):

$$\dot{M} = \theta M \tag{8}$$

A standard demand for money function is assumed:

$$m \equiv M/L = P.G(y, r) \tag{9}$$

where P = price level, r = nominal interest rate and we assume:

$$G_y > 0 \text{ and } G_r < 0 \qquad (10)$$

Equation (9) can be rewritten as:

$$x \equiv m/P = G(y,r) \qquad (11)$$

where x = per capita real money balances. The money market is assumed to clear continually and the conditions of the implicit function theorem are assumed to hold for equation (11). We may then write:

$$r = H(y,x) \qquad (12)$$

for some function H. There are in effect only two assets in the model, money and physical capital. The nominal interest rate may therefore be identified with the expected money yield on physical capital. Assuming the rate of inflation is perfectly foreseen, we may write:

$$r = f'(k) - \delta + \dot{P}/P \qquad (13)$$

Equation (13) effectively identifies the real rate of interest with the marginal product of capital, net of depreciation.

From equations (9) to (14) it is easy to derive the following differential equation in x:

$$\dot{x} = x(f'(k) + \theta - \delta - n - H(f(k), x)) \equiv D(k,x) \qquad (14)$$

Equations (7) and (14) together constitute a dynamical system in k and x, though note that (7) can be solved independently of (11) since the former equation does not involve x. Local linearisation of this model reveals the existence of an equilibrium (k^*, x^*) such that $k^*, x^* > 0$ and:

$$B(k^*) = 0, \qquad D(k^*, x^*) = 0 \qquad (15)$$

In the neighbourhood of this equilibrium dynamics take the saddlepoint form, as illustrated in Figure 1. Paths starting on the stable branch converge to equilibrium. All other paths are divergent and thus treated as economically meaningless (we return to this point in section 3 below, where the "transversality condition" is discussed). Note, however, that *all* solution paths satisfy the perfect foresight and market clearing conditions. In long run equilibrium the level of per capita real money balances is constant and hence the steady state rate of inflation is equal to the (exogenous) rate of money growth minus the (exogenous) rate of population growth.

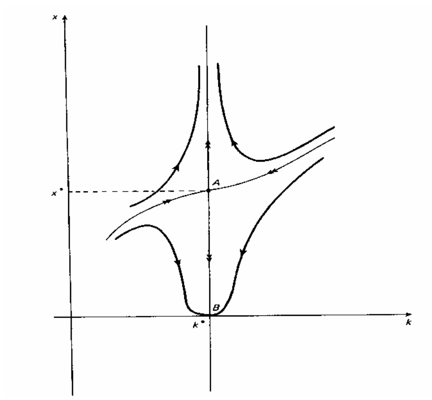


Figure 1

Thus controlling money growth is necessary and sufficient for the (long run) control of inflation.

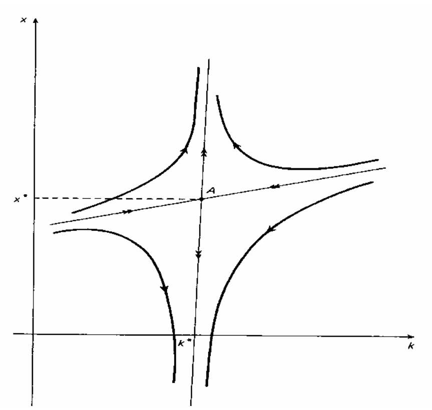


Figure 2

The global dynamics of this model are quite different from the local dynamics (see figure 2). In addition to the saddlepoint discussed above (point A in figure 2), there is a globally stable equilibrium, (k^*,0) (point B in figure 2). Note that Hartman's Theorem does indeed hold in the neighbourhood of point A. Paths with initial conditions lying above the stable branch entail real per capita money diverging to infinity and might therefore be rejected. However, paths with initial conditions lying below the stable branch entail real per capita money tending to zero. There is no reason to reject these paths: they all satisfy the market clearing and perfect foresight conditions and (as will be discussed below) most transversality conditions. Since real per capita money tends to zero along such a path, the price level must eventually grow faster than the rate of growth of nominal money minus the population growth rate. Using standard functional forms for the production function and demand for money function, it is easy to show that the inflation rate diverges to infinity along such a path, and hence the control of the nominal money supply is *not* sufficient for the control of inflation. These paths represent hyperinflationary bubbles, but they are wholly rational, equilibrium bubbles, with no inbuilt tendency to burst. In section 3 below we argue that they represent the major scientifically valid implication of the model.

The model developed above illustrates clearly the dangers of local linearisation. Hartman's theorem only holds locally, so that treating the linearisation as globally valid can lead to to the neglect of dynamics which may have important economic implications. Thus a global approach to dynamics is essential in most economic applications.

3. ROBUSTNESS

It is widely accepted that the demarcation between scientific and non-scientific assertions should be based on the requirement that a scientific assertion be susceptible, at least in principle, to refutation by empirical data. An observation which refutes an implication of a theory also refutes at least one of that theory's underlying assumptions. It follows that the underlying assumptions of a scientific theory need not be open to empirical test, provided that testable implications can be drawn from them. In practice it is generally the case that the assumptions underlying a theory cannot be *exactly* true, but rather, it is to be hoped, are close approximations to reality. Under these circumstances it is essential that the implications of a theory are robust to small variations in its underlying assumptions. Without this property, empirical testing of a theory is impossible.

Consider, for example, a chemical theory which predicts the outcome of a particular chemical reaction under conditions of constant ambient temperature. Whatever care the experimental chemist takes, he will not be able to hold the ambient temperature *exactly* constant: it is bound to fluctuate slightly during the course of the experiment. Suppose now that the outcome of the experiment is substantially different from what the theory predicted. Is the theory refuted? The theoretical chemist can always reply that the ambient temperature was not *exactly* constant as her theory required, and that the experiment does not, therefore, constitute a refutation of her theory. This defence would not be possible if the robustness property were required *ab initio*.

We go further than this, and claim that non-robust implications of a theory are scientifically meaningless and should be ignored. It is a necessary (though possibly not sufficient) condition for scientific validity that a theory make robust predictions.

We may define robustness more precisely as follows:

> ***Definition.*** *Any property of a model will be called robust if the set of parameter values for which it occurs is of strictly positive Lebesgue measure.*

A non-robust property is one which occurs for a parameter set of zero Lebesgue measure and can thus be thought of as having a zero probability of occurring. Of course it is a well-known conundrum of probability theory that although a non-occurring event has probability zero, the converse does not hold. An event with zero probability could occur, but we consider it appropriate to label such events as "unobservable". Note that it is not enough for robustness that the relevant parameter set is dense, because dense sets can easily have zero measure. For example, the set of rational numbers is dense in the set of real numbers, but it is countable, and therefore certainly of measure zero. For the case of dynamic models, the definition of robustness given in the *Introduction* is a special case of the definition above, where the property at issue is the limiting behaviour of a solution path as $t \to \infty$. In a saddlepoint model, such as the linearised model of section 2, convergence is non-robust while divergence is robust. Since the "monetary control of inflation" result depends upon convergence, it too is non-robust, and must therefore be considered an unobservable event. In the global version of the model convergence to the saddlepoint equilibrium (point A) is non-robust, while convergence to the stable equilibrium (point B) is robust and hence so are the hyperinflationary bubbles that this convergence entails, According to the global dynamics of the model then, the "monetary control of inflation" property is unobservable, while hyperinflationary bubbles are observable. In general we conclude that *any* assertion which depends upon convergence in a saddlepoint model can be safely dismissed as non-scientific. Thus a large part of macrodynamics can be dismissed on methodological grounds alone.

A favourite technique for avoiding the non-robustness problem in saddlepoint models is to add auxiliary assumptions to the model in an attempt to rule out divergent paths. Suppose, for example, that the economy is controlled by an agent choosing per capita consumption (z) to maximise a utility function such as:

$$\int_0^\infty U(z,x)e^{(n-\rho)t}\,dt \qquad (16)$$

subject to the constraints of equations (7) and (14). To ensure convergence of the integral, take $n < \rho$. Applying Pontryagin's Maximum Principle to this problem yields the transversality condition:

$$\lambda k + \mu x \to 0 \quad \text{as} \quad t \to \infty \qquad (17)$$

where λ and μ are the costate variables for equations (7) and (14) respectively. In the linearised saddlepoint model of section 2, only convergent paths satisfy this transversality condition. Hence the addition of a utility maximising assumption shrinks the set of solution paths down to the set of convergent paths, thus avoiding the non-robustness problem.

There are a number of difficulties with this argument:

1. Short of a centrally planned economy it is hard to see exactly who the controlling agent is supposed to be. The utility maximising approach seems to involve teleology.
2. The transversality condition is *not*, in general, a necessary condition for the solution of the Pontryagin problem. (e.g. see Halkin, 1971)
3. In the global version of the model in section 2, the transversality condition (17) is satisfied on paths converging to the stable equilibrium (point B), for a wide class of utility functions U.
4. In the linearised version of the model, it is not clear how the economy is supposed to get on to the stable branch if it should become displaced from it (for example by a shift in policy parameters).

Problem no.4 is "solved" by appeal to the notion of a "jump variable". If the price level jumps by exactly the right amount, real per capita balances can jump by just the amount required to place the economy on the stable branch, and thus ensure convergence. Again, short of a centrally planned economy, it is hard to see what mechanism is supposed to bring this jump about.

In summary then, the non-robustness problem which necessarily arises with saddlepoint dynamics, can be avoided by (a) imposing a transversality condition and (b) appealing to the notion of a "jump variable". As discussed above both are unsatisfactory assumptions, and both take the model a long way from its (relatively) innocuous origins in perfect foresight and market clearing. We propose that the robustness property be required of any model claiming scientific status, and that *ad hoc* auxilliary assumptions, designed to ensure robustness, be disallowed. On this basis, virtually all the macroeconomics based on saddlepoint dynamics can be rejected on methodological grounds alone.

In the next section we will pursue the issue of empirical observability and support for nonlinear models via an example (other examples can be found in George and Oxley 2000). In section 4 we investigate the existence of chaos using 16127 daily observations from the Standard & Poor's Composite Price Index (this example is taken from Harrison et al. 1999).

4. CHAOS IN THE S&P 500

Chaos is widely found in the fields of physics and other natural sciences, however, the existence of chaos in economic data is still an open question. Various contributions have been made to this economic literature, including Barnett et al. (1994), Barnett and Chen (1988), Chen (1996), Brock and Sayers (1989), and Ramsey, Sayers and Rothman (1990). In addition, a new international journal, which is exclusively devoted to, and entitled, *Studies in Non-linear Dynamics and Econometrics,* has recently been founded which is testimony to the interest in this area.

Some of the main problems, which pervade the area of economic time series evidence on chaos, are the effects of noise, trend, and more general structural change. Of these noise and time evolution appear to be the most problematic with the latter often modelled via ARCH/GARCH processes that allow changing means and variances. These problems are frequently compounded by the paucity of available data. In attempting to answer questions relating to the existence of nonlinearities and chaos in economic data, researchers have normally used either the Hinich bispectrum test, the BDS test of Brock, Dechert and Scheinkman (1996), White's (1989), test or more recently Kaplan's test, to identify nonlinearities and, when considering chaos, have used tools based on phase space reconstruction developed and used successfully in the physical sciences. The most commonly used of these chaos tests are the Lyapunov exponents test and the Grassberger-Procaccia (GP) correlation dimension test. While these tests have revealed an abundance of previously unexplained non-linear structure and yielded a deeper understanding of the dynamics of many different economic time series, the case of deterministic chaos in these types of series is yet to be clarified. The main problem we identify in this study is that of noise which degrades these measurement techniques. The use of conventional filtering methods such as low pass filtering using Fourier transforms, moving averages etc., and also singular spectrum analysis based on singular value decomposition commonly used in economics, can lead to distortion of the dynamics.

4.1 The Testing Approach

Before economic data can be analysed for the existence of deterministic chaos, the twin problems of growing time trends and noise require consideration. The main contribution of this paper will be to the latter where new non-linear noise reduction (NNR), techniques will be applied to the data. However, the following general methodology will be followed. Firstly, the (log) data will be adjusted to remove systematic calendar effects and trend effects by differencing. Secondly, in order to reconstruct a chaotic attractor in phase space, two basic parameters, the embedding dimension *m,* and delay time *h,* must be correctly determined. The embedding theorem (*m=2d+1*, where *d* is correlation dimension) provides a sufficient condition for reconstructing an attractor from a scalar time series. An efficient method to determine an acceptable minimum *m*, from experimental time series is the so-called *false nearest neighbour* (FNN), recently developed using a geometrical construction. It monitors the behaviour of near neighbours under changes in the embedding dimension from $m \to m+1$. When the number of the false nearest neighbours arising through projection is zero in dimension *m*, the attractor has unfolded in this embedding dimension *m*. This technique is robust to the noise and a correct region of the embedding dimension can be determined in the presence of noise, which is important for the type of data used here. An estimate of the value of the delay time *h,* is provided by the autocorrelation function (ACF).

The Lyapunov exponents test and the Grassberger-Procassia correlation dimension method are well-documented methods used in the quantitative analysis of time series data as tests for chaos; see for example, Abarbanel et al. (1993). Here we concentrate on the latter.

The geometrical features of an attractor can be specified using the Grassberger-Procaccia correlation *dimension.* Suppose we have a scalar time series $x_i \; (i = 1,2,\ldots N)$ of a dynamical variable sampled at an equal time interval Δt from which the *K* vectors $\mathbf{Y}_j$ *(j=1, 2,...K)* in the *m*-dimensional phase space can be reconstructed using the time delay technique. Then the *correlation dimension* D_2 is defined and calculated as:

$$D_2 = \lim_{\varepsilon \to 0} \frac{\log_2 C_m(\varepsilon)}{\log_2 \varepsilon}, \qquad (18)$$

where $C_m(\varepsilon)$ is known as the correlation integral and can be computed as

$$C_m(\varepsilon) = \lim_{K \to \infty} \frac{1}{K(K-1)} \sum_{ij}^{K} \theta(\varepsilon - \|\mathbf{Y}_i - \mathbf{Y}_j\|), \quad (19)$$

Where $\theta(x)$ is the Heaviside step function and $\|\mathbf{Y}_i - \mathbf{Y}_j\|$ is the distance between the vectors $\mathbf{Y}_i$ and $\mathbf{Y_j}$. Thus, the sum $\sum_{ij}^{K} \theta(\varepsilon - \|\mathbf{Y}_i - \mathbf{Y}_j\|)$ is equal to the number of pairs *(i,j)* whose distance $\|\mathbf{Y}_i - \mathbf{Y}_j\|$ in the reconstructed phase space is less than the distance ε. For a chaotic attractor, D_2 is a non-integer, the value of which determines whether the system is low- or high dimensional.

The use of this approach must, however, be applied with caution since it describes a kind of scaling of behaviour in the limit as the distance between points on the attractor approaches zero and therefore is sensitive to the presence of noise. Indeed our numerical experiments have shown that a noise level as small as 2~5% of the time series content can make these measurements inaccurate and inconclusive. Moreover, noise can also prevent precise prediction. Here we use NNR algorithms based on finding and extracting the approximate trajectory, which is close to the original clean dynamics in reconstructed phase space from the observed time series. The implementation of the algorithms involves three basic steps: i) to reconstruct the underlying attractor from the observed series, ii) to estimate the local dynamical behaviour choosing a class of models and fitting the parameters statistically, and iii) to adjust the observations to make them consistent with the clean dynamics. The technique can reduce noise by about one order of magnitude. If some standard techniques are employed to pre-process the data, such as band-pass filtering, filtered embedding and singular value decomposition, significantly larger amounts of noise can be reduced since the local dynamics are enhanced.

These non-linear noise reduction algorithms have been developed under the assumption that the noise is additive rather than dynamic. In practice, the two may not be distinguishable, being based on data only and both of them can be reduced, as long as the exact dynamics can be reconstructed.

While the correlation dimension measurement is often accepted as 'proof' of chaos, it is not a definitive test against time series data with certain types of coloured noise. This issue will be resolved using *surrogate* techniques. The correlation dimension method must be applied together with a surrogate technique to reliably discriminate between chaos and noise from a time series, so as to avoid claims of chaos when simpler explanations (such as linearly correlated noise) are adequate. The consideration of surrogate data is based on the following. 1) Statement of a null hypothesis that shall be tested for consistency with the recorded original data. 2) Generation of a number of *surrogate* data sets; an algorithm is to randomise the phase of the raw data so that the surrogated set has the same Fourier spectra as the original. 3) Calculation of the value of interest, e.g., correlation dimension, Lyapunov exponents etc., for the original and all the surrogates. 4) Calculate mean and spread of the results obtained from the surrogates to determine whether the difference to the original, if any, is statistically significant.

4.2 Some results based upon the S&P500

Although several economic data sets have been considered, we will report only one based upon the Standard and Poor's Composite Price Index, which comprises 16127 daily observations on the logarithmic price change $x_t = 100[\log(p_t) - \log(p_{t-1})]$. For details of the data set see Gallant et al. (1993). The time series x_t has been adjusted to remove systematic calendar and trend effects and is taken to be jointly stationary. A representative window of the raw series taken after 1947 is shown as Figure 3(a).

Figure 3

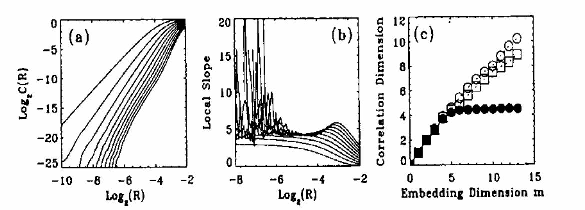


Figure 2. (a) Logarithmic plot of correlation function $C_m(R)$ versus correlation distance R for embedding dimension $m = 3, 4, ..., 13$, (b) Local slope as a function of $\text{Log}_2(R)$ derived from (a), and (c) Correlation dimension D_2 versus embedding dimension m: □ – data before NNR, ● – data after NNR, and ○ – surrogate set from the data after NNR.

Our analysis, based upon the GP correlation dimension measurement, in conjunction with the non-linear noise reduction filtering and surrogate technique provides strong evidence in favour of chaos in this data. Some results are detailed below. As shown in Figure 4 (c) (open squares), the raw data gave no saturated correlation dimension on increasing embedding dimension, suggesting that the data may be noise-dominated. Figure 4(b) presents the data after NNR of the raw data.

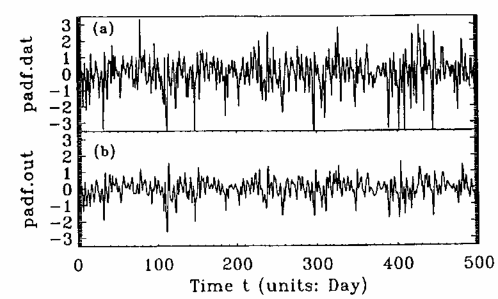


Figure 1. Two time series of $x(t)$ (a) raw data and (b) data after NNR.

Figure 4

Comparison of this with the raw data shows that the effect of noise is manifested as relatively large amplitude random fluctuations, masking the overall deterministic patterns in which the noise level is estimated to be ~90% of the clean signal. Correlation dimension analysis of this noise-filtered data revealed a clear (scaling) saturation region on increasing the embedding dimension as shown in Figure 4 (solid circles in (c)), indicative of deterministic chaos. In confirming that the convergent correlation dimension is a result of chaotic dynamics, both the raw and noise-filtered series were surrogated to randomise the phase and so destroy the deterministic structure. The results show the dimension to diverge as shown in Figure 4(c) (open circles), consistent with stochastic behaviour, from which we confirm that the saturated correlation dimension of the noise-filtered data in Figure 4(c) arises from an endogenous deterministic mechanism.

5. CONCLUSIONS

The paper had several aims. Firstly, we wish to emphasise the need to consider nonlinearities in economic models and to this end highlight some of the dangers of linearisation and using 'tricks' to force behaviours on models which, in general cases, do not arise. Secondly, we argue in favour of "robustness" being considered a necessary property of economic models and make the important distinction between structural stability and robustness. Finally, we use an example to consider the role of nonlinearities in economic behaviour. The example considers whether evidence of deterministic chaos exists in the Standard and Poor's Composite Price Index. When the noise is removed via nonlinear noise reduction techniques, evidence of deterministic chaos is, in fact, found. Nonlinear models and nonlinear methods have a major role to play in economics and econometrics and although they may involve more sophisticated modelling and estimation techniques, the rewards are enormous.

6. ACKNOWLEDGEMENTS

I would like to thank Donald George and Bob Harrison for permission to draw upon our published research and the New Zealand Marsden Fund whose generous support helped completion of the paper.

7. REFERENCES

Abarbanel, H., R. Brown, J. Sidorowich, & L. Tsimring, The analysis of observed chaotic dynamics in physical systems, *Review of Modern Physics,* 65, 1331-92, 1993.

Barnett, W., & P. Chen, The aggregation-theoretic monetary aggregates are chaotic and have strange attractors, in W. Barnett, Berndt, E. and H. White, (eds.), *Dynamic Econometric Modelling.* Cambridge University Press, Cambridge, 1988.

Barnett, W., R. Gallant, M. Hinich, J. Jungeilges, D. Kaplan, & M. Jensen, A single-blind controlled competition among tests for nonlinearity and chaos, *Working Paper 190,* Washington University, St. Louis, Missouri, 1994.

Baumol, W.J., (1958) Topology of second-order linear difference equations with constant coefficients. *Econometrica,* 26, 258-287.

Brock, W., W. Dechert, & J. Scheinkman, A test for independence based on the correlation dimension, *Econometric Reviews,* 15, 197-235, 1996.

Brock, W., & C. Sayers, Is the business cycle characterised by deterministic chaos? *Journal of Monetary Economics,* 22, 71-80, 1989.

Burmeister, E. and Dobell, R. *Mathematical Theories of Economic Growth.* Macmillan, London, 1971.

Chen, P., A random walk or color chaos on the stock market? Time frequency analysis of S&P index, *Studies in Nonlinear Dynamics and Econometrics,* 1, 87-123, 1996.

Gallant, A., Ossi, P., & G. Tauchen Non-linear dynamic structures, *Econometrica,* 61, 871-907, 1993.

George, D., & Oxley, L., Robustness and local linearization in economic models, *Journal of Economic Surveys,* 13, 529-550, 1999.

Halkin, H., Necessary conditions for optimal control problems with infinite horizons. *Econometrica,* 42, 267-272, 1971.

Harrison, R., Yu, D., Oxley, L., Lu, W., & D. George, Non-linear noise reduction and detecting chaos: Some evidence from the S&P Composite Price Index, *Mathematics and Computers in Simulation,* 48, 497-502, 1999.

Ramsey, J., C. Sayers, & P. Rothman, The statistical properties of dimension calculations using small data sets: Some economic applications, *International Economic Review,* 31, 991-1020, 1990.

White, H., An additional hidden unit test for neglected nonlinearity in multilayer feed forward networks, *Proceedings of the International Joint Conference on Neural Networks,* Washington DC, vol. 2, 451-5, 1989.

The Implications of Complexity for Integrated Resources Management

C. Pahl-Wostl

Institute of Environmental Systems Research, University of Osnabrück, Germany

Abstract: Integrated environmental resources management is a purposeful activity with the goal to maintain and improve the state of an environmental resource affected by human activities. In many cases different goals are in conflict and the notion "Integrated" indicates clearly that resources management should be approached from a broad perspective taking all potential trade-offs and different scales in space and time into account. However, we are yet far from putting integrated resources management taking fully into account the complexity of human-technology-environment systems into practice.The tradition of resources management and of dealing with environmental problems is characterized by a command and control approach. The increasing awareness for the complexity of environmental problems and of human-technology-environment systems has triggered the development of new management approaches. The paper discusses the importance to focus on the transition to new management paradigms based on the insight that the systems to be managed are complex adaptive systems. It provides arguments for the role of social learning processes and the need to develop methods combining approaches from hard and soft systems analysis. Soft systems analysis focuses on the importance of subjective perceptions and socially constructed reality. Soft systems methods and group model building techniques are quite common in management science where the prime target of management has always been the social system. Resources management is still quite slow to take up such innovations that should follow as a logical consequence of adopting an integrated management approach. Integrated water resources management is used as example to provide evidence for the need to implement participatory and adaptive management approaches that are able to cope with increasing uncertainties arising from fast changing socio-economic conditions and global and climate change. Promising developments and future research directions are discussed. The paper concludes with pointing out the need for changes in the scientific community to improve the conditions for interdisciplinary, system-oriented and trans-disciplinary research.

Keywords: Complexity, Mental models, Group model building; Adaptive Management, Soft systems analysis, Complex Adaptive Systems, Social Learning

1. INTRODUCTION

Integrated environmental resources management is a purposeful activity with the goal to maintain and improve the state of an environmental resource affected by human activities. Management should guarantee services provided by the resource (e.g. water for irrigation, fisheries), prevent damages (e.g. flooding) and maintain the state of the resource for the use of future generations (e.g. preserve groundwater resources) but respect also the maintenance of the integrity of ecosystems as a goal in itself (e.g. maintenance of a good ecological state of rivers). In many cases these different goals are in conflict and the notion "Integrated" indicates clearly that resources management should be approached from a broad perspective taking all potential trade-offs and different scales in space and time into account. However, we are yet far from putting integrated resources management taking fully into account the complexity of human-technology-environment systems into practice. Experiences in managing environmental problems and resources provide partly success stories but when judged from a long-term perspective many policies showed unexpected side-effects. To name just a few:

- Flood control efforts such as levee and dam construction have led to more severe floods by preventing the natural dissipation of excess water in flood plains. The cost of flood damage has increased as the flood plains were developed by people who believed they are safe.
- Pesticides and herbicides have stimulated the evolution of resistant pests and weeds, killed of natural predators, and accumulated up the food chain to poison fish, birds and possibly humans.
- Programs to increase the capacity of roads designed to reduce congestion have increased traffic, delays, and pollution by attracting more people to drive with the car and by providing incentives for a

spatial segregation of workplace, residential and shopping areas.

- Policies of fire suppression have increased the size and severity of forest fires. Rather than frequent, small fires, fire suppression leads to the accumulation of dead wood and other fuels leading to larger, hotter, and more dangerous fires, often consuming the oldest and largest tress that previously survived fires unharmed.
- High security standards for heavily subsidized water supply systems designed to meet maximum daily demand have lead to quite expensive and inflexible systems and to exaggerated expectations of the public regarding the provision of services at no cost.

In all these cases policy makers, resource managers and engineers underestimated the importance of feedback effects, non-linearities, time delays and changes in human behaviour as a consequence of policy interventions. Human actors typically tend to reduce the complexity and dimensions if they are confronted with a problem to be tackled (Sterman, 2000; Vennix, 1996). What may be more appropriately described as a messy problem situation is often compressed into a description of a well defined problem with simple cause-effect relationships. Open loop structures that behave quite “benign” are assumed instead of feedback cycles. The problem of traffic congestion and the corresponding dissatisfaction of car drivers for example have been attributed to a lack of road capacity. However, such a simplifying approach is misleading. It helps to handle the problem, may be successful in the short-term but the negative effects of long-term consequences may often outweigh short-term benefits. What has been neglected in the traffic example was the fact that more and better roads provide an incentive for people to use the car more often and abandon public transport, to move to a place where they depend on the car etc. Hence it is crucial to develop methods that allow exploring all possible scenarios of the co-evolutionary development of human-technology-environment systems that may result from policy interventions, methods that allow to categorize different policy problems and that allow to develop and apply appropriate management strategies.

The insight emerging from these examples that environmental problems should be addressed from a wider perspective taking into account complexities, non-linearities and the limits of control is not really new. As already Ludwig et al. (1993) pointed out in the case of fisheries management, it seems to be more appropriate to think of resource managing of humans than the converse. They make a strong argument against the illusion of control of environmental problems. The idea of adaptive management has been introduced in resources management for quite some time (Holling, 1978; Walters, 1986; Pahl-Wostl, 1995; Lee, 1999). It is based on the insight that the ability to predict future key drivers as well as system behaviour and responses are inherently limited. But the implementation of consequences for policies is quite slow. It is argued here that the major reason for the slow pace in change is not the absence of alternative management strategies but rather the obstacles encountered in the transition process towards new management paradigms. To better understand the nature of the transition process it is useful to contrast current management paradigms with alternative approaches and to investigate the importance of learning processes at different scales both for new management styles and for the transition towards them.

The operations research and engineering approach to management has been mainly characterized by a control paradigm. Some assumptions that are important to implement management as control in a system are:

- A system can exist in a finite set of states and each state can be uniquely characterized by observation.
- Based on this characterization one can devise a unique set of control measures to move the system from one state to another state.
- Uncertainties in the state transition functions can be quantified by probabilities.
- Risks are quantified by multiplying the probability of an event with the magnitude of the expected damage.

Technical systems are constructed such that they can be controlled. However, human-technology-environment systems are more appropriately described as complex adaptive systems where different paradigms have to be used. What are the consequences if one takes into account that one deals with complex adaptive systems both regarding the systems to be managed and the learning and decision making processes that are the essence of the management process? Complex adaptive systems are characterized by self-organization, adaptation, heterogeneity across scales and distributed control. The state space is not closed and predictable but open and evolving. Development may be path and context dependence, the system attempts to escape external pressures by adaptation in changing its internal structure. The system itself is in constant change. Regarding the assumptions of a control paradigm one can note the following deviations:

- *A system can exist in a finite set of states and each state can be uniquely characterized by observation* – the state of a system depends on history and context, systems are hierarchical.
- *Based on this characterization one can devise a unique set of control measures to move the system from one state to another state* – systems may escape attempts of external control by adaptation and human beings may behave differently than anticipated.
- *Uncertainties in the state transition functions can be quantified by probabilities* – for some extreme states it may be impossible to quantify transition probabilities, non-linear developments may render probabilistic judgements exceedingly difficult.
- *Risks are quantified by multiplying the probability of an event with the magnitude of the expected damage* – some risks are related to ethical issues and require risk dialogues and people judge risks differently based on their perception of being able to influence the risk (e.g. dying in a car accident versus dying in an airplane crash).

Remaining within the concepts of dynamic systems and optimization one can illustrate the difference between the paradigms using the metaphor of a fitness landscape where hills refer to desirable states and valleys to states to be avoided. The control paradigm is based on finding optimal solutions in a constrained and rigid state space. The learning and evolutionary paradigm is based on finding methods to support navigation in a fitness landscape that is in continuous change. Rather than sticking to one paradigm it is important to develop and apply methods to choose the appropriate approach for the management problem to be tackled.

2. COMPLEXITY AND THE IMPORTANCE OF LEARNING PROCESSES

The increased awareness for the complexity of systems and for management as learning rather than control seems to be an overall trend in different fields (Senge, 1990; Pahl-Wostl, 1995, 2004; Levin, 1998; Hartvigsen at al, 1998; Berkes et al, 2002). On one hand the systems to be managed and the problems to be tackled have become indeed more complex. The pace of change in socio-economic conditions and technologies is tremendous. Uncertainties arising from global change in general and climate change in particular pose major challenges for the management of environmental resources. On the other hand the awareness for the need to take the complexity of problems fully into account has increased and the frame of analysis has partly changed. One may talk of socially constructed problem domains. The frame of reference determines how a problem is conceptualized (Shakley at al, 1996). Such a socially constructed problem domain stabilizes itself. Institutions are developed, technologies are implemented based on a shared paradigm. Hence, any transition to a new management regime requires collective learning processes and new methods are required that allow to analyse the origins and importance of socially constructed reality and the impediments for a change.

Regarding the social construction of reality and the origin of subjective perceptions, it is useful to introduce here two concepts – frames and mental models. More than one definition exists for frames and mental models and sometimes the distinction is blurred (Doyle and Ford, 1998; Sterman, 2000; Craps et al, in press). In the current paper the following distinction is made: A mental model refers to a specific mental representation of information about reality. A frame refers to the context into which such a mental model is embedded and which gives sense and meaning to it. Differences in framing are one of the key reasons for problems in communication among actors. Two people may engage in a conversation – one acts in a power frame (goal to dominate the conversation) the other in a cooperation frame (goal to engage in a collaborative relationship). They will interpret each others arguments very differently and hold contradicting expectations about each others behaviour. People make judgements about motives other actors hold. Hence the framing of the goal of a negotiation process – the role of different actors, their position, their views on what is at stake are key factors and determine entirely the outcome of a process.

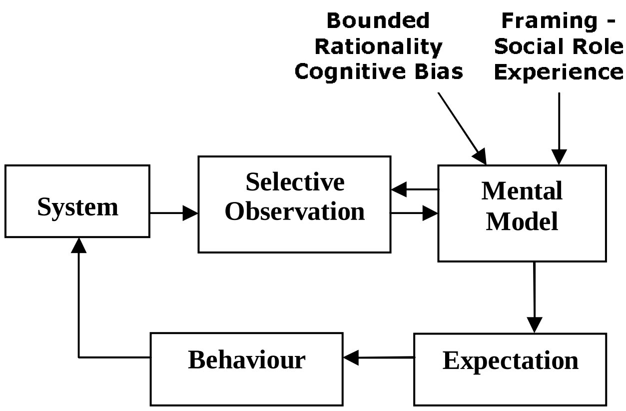


Figure 1 The role of mental models in processing information

Figure 1 represents the role of mental models and frames in the processing of information. People hold internal representations, mental models of reality. Mental models are assumed to be quite enduring structures of the internal representation of a real system (Doyle and Ford, 1998, 1999; Sterman, 2000). Such mental models may be shaped by the role of actors in a social system, their previous experience and cognitive biases that result from heuristics that allow human beings to survive and act in a very complex and partly unpredictable world. Mental models determine the processing of information which is selective. Experience may help to construct a context from few pieces of information, to draw analogies to previous situations and select a type of response and behaviour that is deemed to be appropriate based on previous experience. Sometimes selective information processing may prevent learning and the adaptation to a changing environment – this applies for individuals, for enterprises or for scientific organizations.

Human beings have a confirmation bias – they search for and process selectively information confirming their beliefs (Evans, 1990). Sometimes beliefs may be proved to be wrong by factual knowledge. Sometimes beliefs about the social environment may support the construction of social reality and influence it (e.g. expectations about the behaviour of others). If one believes for example that other actors in a negotiation process are not willing to cooperate one is full of distrust which may trigger a corresponding behaviour from the other side.

Mental models should be corrected if they are factually wrong – this requires first an agreement among actors on the soundness of the factual knowledge that is provided by empirical analyses or modelling exercises. Mental models may be linked to normative assumptions, values and preferences which determine the interpretation of knowledge. In this case a change of mental models requires processes of reflection and negotiation. Hence we need to combine hard and soft systems approaches and put strong emphasis on the role of different types of learning in management processes.

Table 1 illustrates the differences between hard and soft system approaches in systems science.

	Hard	**Soft**
objective	given	problematic
focus	reality how to do it	perceptions what and how
models	of X relevant to Y	of pure purpose to structure a debate
paradigm	optimising goal seeking	learning
expert	external expert	participative (facilitator)
system	exist in the world	in the process of inquiry

Table 1 Comparison between Hard and Soft systems approaches (after Checkland, 1989)

Whereas hard systems approaches emphasize factual knowledge and the role of the analyst as external observer, soft system approaches emphasize subjective perceptions and the role of the analyst as participant in a process of social learning. Similarly the role of models is different in the two approaches. Whereas in the hard systems approach models serve to represent the relationships of variables in the real world, in the soft systems approach models serve to structure the debate.

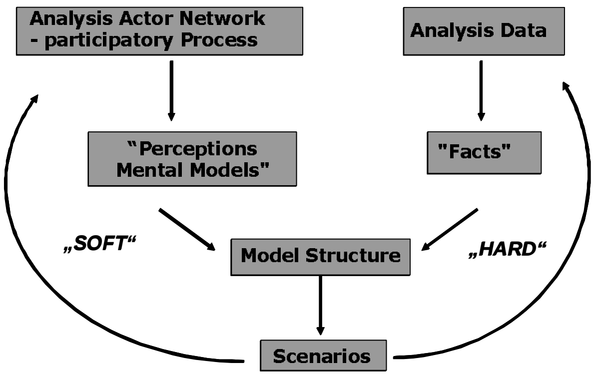


Figure 2 shows the overall approach of combining subjective perceptions and factual knowledge in a participatory group model building process

Figure 2 represents schematically a combination of hard and soft systems approaches in model and scenario development. Pioneers in this field of participatory model and scenario development come mainly from management science (de Geus, 1992; Lane, 1992; Vennix, 1995; Van der Heijden, 1996). The parentheses indicate that the distinction between facts and subjective perceptions is a gradual transition rather than a distinction between two polar and well defined categories.

In such a process, mental models that are factually wrong, should be corrected. Actors may hold erroneous and divergent views on the magnitudes of effects, causal relationships and conclusions drawn from statistical inference. The requirement for learning is the acceptance of factual knowledge. A group of actors has to agree on the soundness of facts provided by the analyst. The soundness of the scientific method used for deriving the data should be the prime criterion – this may not always be guaranteed. Finding agreement is easier for empirical data than finding agreement for results derived from simulation models. The latter contain already embedded assumptions that may be questioned. Methods to improve the sound use of factual knowledge in a stakeholder group are for example the elicitation of mental models by different techniques (e.g. mental mapping, system dynamics approaches) and the subsequent comparison of such models with results derived from factual analyses. Such an elicitation process can be the first step of a group model building processes (Vennix, 1996; Sterman, 2000; Pahl-Wostl, 2002b). Developing models in a group model building process is of particular importance if uncertainties and decision stakes are high and more than one interpretation can be derived from model results.

More demanding than correcting mental models that are factually wrong are those situations when mental models determine and stabilize a socially constructed reality in a group. Examples may be the perception of a messy problem situation or norms and rules of good practice shared in a group of practitioners (e.g. water managers). People may hold for example a mental model of the role of a scientist or engineer. Such mental representations shape the social exchange in a group, determine expectations and behaviour.

Methods to facilitate learning in such situations include behavioural simulations or group model building exercises combined with role playing games (e.g. Barreteau et al, 2001; Den Exter, 2003; Duijn et al 2003; Pahl-Wostl, 2002b). In such gaming approaches the social interactions between the participants are the driving force for the simulations. By adopting another role than in real life, actors may start to improve their understanding for perspectives of other actors. The games enable the participants to reflect on the way in which the decisions are taken and identify needs for change.

The methods outlined above are quite common in management science where the prime target of management was always the social system. Resources management is still quite slow to take up such innovations that should follow as a logical consequence of adopting an integrated management approach. But developments are promising as illustrated for the example of integrated water resources management.

3. THE EXAMPLE OF INTEGRATED WATER MANAGEMENT

3.1 The role of participation in integrated water management

Water management has traditionally been characterized by a control paradigm that is now slowly changing. Such change is partly attributable to the need to implement IWRM and to the insights that water management faces increasing uncertainties from climate change and fast changing socio-economic boundary conditions. Integrated water management should provide a framework for integrated decision-making, where we strive to: (1) assess the nature and status of the water resource; (2) define short-term and long-term goals for the system; (3) determine objectives and actions needed to achieve selected goals; (4) assess both benefits and costs of each action; (5) implement desired actions; (6) evaluate the effects actions and progress toward goals; and (7) re-evaluate goals and objectives as part of an iterative process. This sequence sounds quite logic and straightforward to being implemented. However, integration and new approaches to manage risks in the light of increasing uncertainties require transformation processes in institutional resource regimes and management style. Technical solutions are not anymore sufficient to tackle the intricate problems we face today. Equally important are issues of good governance, with the human dimension in a prominent place. Scaling issues need to be explored to understand the complex dynamics of institutional resource regimes and to improve the match between biophysical and actor based scales. The strong tradition of local and regional water resources management has to be combined with integrative river basin approaches and an embedding of them into a perspective of global change. This necessitates linking research areas that have up to now developed rather independently with little exchange among them and social learning of different stakeholder groups.

Currently the concept of social learning is under investigation in the European FP5 project HarmoniCOP (www.harmonicop.info). The major objective of the HarmoniCOP project is to increase the understanding of participatory river basin management in Europe. It aims to generate practically useful information about and improve the scientific base of social learning and the role of ICT tools in river basin management and support the implementation of the European Water Framework Directive.

Elements of Social Learning for river basin management can be summarized as

- Build up shared problem perception in a group of actors and the ability to communicate about different point of view.
- Build trust for self-reflection - recognition of individual mental frames and images and how they pertain to decision making.
- Recognize mutual dependencies and interactions.

- Reflect on assumptions about the dynamics and cause-effect relationships in the basin.
- Reflect on subjective valuation schemes.
- Engage in collective learning- and decision processes.

The notion of social learning has been used in quite different meanings to refer to processes of learning and change of individuals and social systems. In the influential work of Bandura (1977) social learning refers to individual learning based on observation of others and their social interactions within a group e.g. through imitation of role models. It assumes an iterative feedback between the learner and their environment, the learner changing the environment, and these changes affecting the learner.

This approach is too narrow to embrace all the learning processes of relevance in resources management. Of major interest in this respect is the concept of "communities of practice" developed by Wenger (1998) emphasizing learning as participation. Individuals engage in actions and interactions that have to be embedded in culture and history. Such interactions are influenced by and may change social structure and, at the same time, the individual gains experience situated in a context. Such learning processes confirm and shape the identity of the individual in its social surroundings. They confirm and change social practice and the associated interpretation of the environment.

Such a broad understanding of social learning that is rooted in the more interpretative strands of the social sciences characterizes also the approach adopted by the HarmoniCOP project. Figure 3 represents the framework for social learning developed in the HarmoniCOP project to account for learning processes in water resources management. The concept of social learning was developed in HarmoniCOP that has two pillars. They relate to the processing of factual information (content management) and engaging in processes of social exchange (social involvement). Social involvement refers to essential elements of social processes such as the framing of the problem, the management of the boundaries between different stakeholder groups or the type of negotiation strategies chosen.

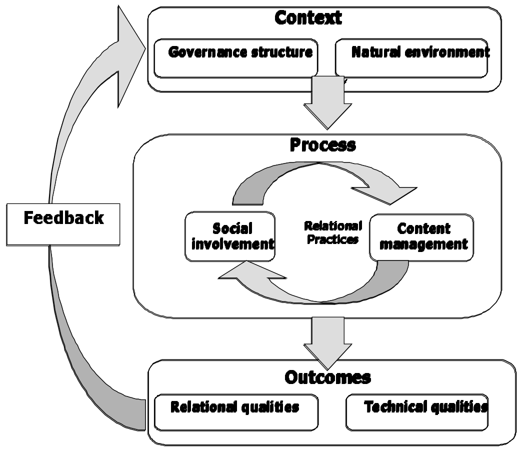


Figure 3 Conceptual framework for social learning in resources management (Craps et al, 2003

ICT tools play a key role (Maurel et al, 2004) in promoting relational practices. They may elaborate and provide well balanced information for the debate in ways that are relevant for the stakeholders and that allows collective learning, helps to elicit perspectives and behaviours of stakeholders, to make them explicit to the others and facilitate relational practices – e.g. participative mapping, role playing games, behavioural simulations

Currently the importance of social learning and the role of ICT tools are investigated in nine case studies on participatory water management related to the implementation of the European Water Framework Directive in nine European countries. The goal is to investigate if social learning takes place, how it is promoted and what its implications are on the goal of developing river basin management plans and of managing the river basin in a more sustainable way. Case studies focus on different scales (local to trans-boundary) and in particular on the interaction between scales. Social learning includes processes at the level of local committees up to negotiation processes in trans-boundary basins and their mutual influence. The knowledge about the interactions between scales and the type of institutional settings that are required to promote them is still quite limited. First results indicate clearly the importance of culture, regional context and the reigning management paradigm on social learning processes. Preliminary results support findings from the MANTRA East project (Timmerman and Langaas, 2003) providing strong evidence that water management in most European countries is not yet based on a participatory approach but on expert knowledge guiding management decisions. Stakeholders are mainly informed or engaged in consultation processes. Involvement and co-decision-making is far from being realized in practice which is a certain impediment to implementing new water policies.

The European Union is particularly active in the area of Integrated Water Resources Management regarding the implementation of innovative water policies (European Water Framework Directive and the European Water Initiative). During FP5 more than have been spend to fund projects related to IWRM and the development of integrated catchment models. However, most of the projects include “endusers” at the end of the process of tool development. The HarmoniCOP projects described in the last section is one exception by starting from the perspective of social learning and stakeholder perspectives and by investigating how ICT tools can be used to support such processes. The HarmoniCA concerted action tries to bridge the gap between science and policy with specific emphasis on the implementation of the European Water Framework Directive. A number of interactive workshops provided evidence that the perception of model developers on the importance of models and the perception of policy makers on the current role of model on water management diverge considerably (Hare, 2004). One conclusion to improve the role of models in IWRM was to establish a closer link between stakeholder participation processes and model development. This is also reflected in the development of an overall methodology for participatory water management – IMA-PIP.

3.2 The transition to adaptive water management

Two new EU projects (AquaStress and NeWater) currently in the phase of implementation under the umbrella of the 6th framework programme of the European Union are based on a new water management paradigm. NeWater – New approaches to adaptive water management under uncertainty – focuses on the transition to adaptive water management building on the concept that management is a learning process in complex adaptive systems. Adaptive management can more generally be defined as a systematic process for continually improving management policies and practices by learning from the outcomes of implemented management strategies. The most effective form of adaptive management employs management programs that are designed to experimentally compare selected policies or practices, by evaluating alternative hypotheses about the system being managed. As it is defined in the approach promoted in this paper adaptive management has as another target - its goal is to *increase the adaptive capacity of the (water) system.* It is aimed at integrated system design. The problem to be tackled is to increase the ability of the whole human-technology-environment system to respond to change rather than reacting to undesirable impacts of change. Hence it is a pro-active management style. Increasing the ability for change includes for example increasing the use of small-scale technology or combing formal regulations with informal institutional settings (Pahl-Wostl et al, in prep).

The focus on the transition to adaptive water management reflects the insight that understanding the transition is the most crucial point for adaptive water management. The adaptive water management regime to be achieved will depend on the path chosen. Given the interdependent nature of social, technical and environmental processes change must be based on a collective learning process. The approach for social learning introduced in the previous section strongly suggests that the social capital and governance structure generated depends on the quality of the learning process implemented in the transition phase.

The NeWater project has a strong methodological component. New methods will be tested in a number of case studies in Europe, Africa and Central Asia. Much emphasis will be given to assess key drivers of global change and the vulnerability of river basins. The practitioners in a basin will play a crucial role in guaranteeing that the methods developed meet the demands from the practitioners and take into account concerns and expertise in a basin. They will benefit from being able to direct research efforts to the issues of most relevance to them. Based on a joint assessment, suitable methods and tools for improved basin management will be developed and tested. More details on the overall concept are given elsewhere (Pahl-Wostl et al in prep.).

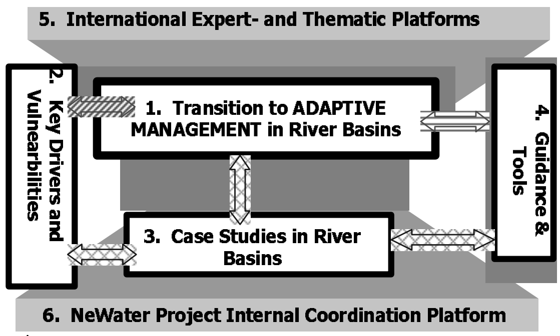


Figure 4 Major building blocks of the NeWater project.

The structure of the project represented in Figure 4 illustrates as well the opportunities provided in the 6th framework programme of the European Union.

The major advantages which are important to pursue the type of research described in the previous sections can be summarized as:

- Possibilities for interdisciplinary projects where disciplines can be chosen to meet the demands of the complex problems under investigation instead of being constrained by the disciplinary structure characterizing many funding agencies.
- Strong stakeholder participation and participatory action research.
- Direct combination between basic and applied research and tool development for practitioners
- New opportunities for public-private partnerships.
- Possibility to include case studies from Europe, Africa, and Central Asia.

4. CONCLUSIONS AND OVERALL DEVELOPMENTS

The paper emphasized the need to take complexity into account in resources management and to develop appropriate methods for different situations. We need approaches that allow characterizing messy problems and finding solutions to deal with them in an adequate manner. These are situations in which there are large differences regarding the perceptions of the nature of the problem, the need for action and what type of action should be done. Such differences arise on one hand from uncertainties in the factual knowledge base and on the other hand from ambiguities in problem framing and diversities in the perception of the nature of the problem. It is important to have a sound base for using the appropriate methods since participatory processes are resource intensive. Duijn et al (2003) suggested a categorization of different problem situations along two dimensions. If there is little consensus about knowledge and the values and aims involved, policy making as interactive learning process is of particular importance. In the case of a high degree of consensus about knowledge and the values and aims involved, policy making can proceed as management in the classical sense. However, as pointed out before the framing of the problem in a stakeholder group may not correspond to the real nature of the problem situation. Hence, a sound analysis for categorizing a problem situation and the stakeholders involved, their interests is highly recommended for any environmental management problem.

The research questions to be tackled in understanding the complex dynamics and the management of human-technology-environment systems are highly intellectually challenging and new insights are in major demand from the policy side. Hence, we need the best and highly skilled people to address these burning questions. At the same time the incentives in the scientific community to go in this direction are quite small. The opportunities for doing basic research tackling interdisciplinary questions are very slim. Hence we need a transition in science as well!

The research agenda outlined in the previous sections is based on new partnerships between science and society. We need more emphasis to promote interdisciplinary research. The prevailing structure of the disciplinary scientific community is an impediment to the development of cutting-edge pioneering research in this field. Most faculties at universities are still organized in disciplinary structures. Funding agencies are not well prepared to handle proposals crossing disciplinary boundaries. However, scientific breakthroughs and the development of new fields occur when different disciplines meet. A certain progress has been made (e.g. the life-sciences) - the deep divide between the social and the natural sciences has yet to be overcome. However, new innovative approaches are urgently needed to support sustainable strategies for dealing with complex socio-environmental problems.

A number of promising developments can be noted. New societies start to emerge and prosper. TIAS, The Integrated Assessment Society, was founded recently (www.tias-web.info). Its aim is to promote research on methods for the integration of knowledge on a problem domain from different sources and for understanding complex societal learning and decision making processes required to deal with the problem. Ecological Economics has established itself with much success as a society at the interface between natural and social systems (www.ecologicaleconomics.org).

The 2^{nd} biannual meeting of the International Modeling and Software Society on “Complexity in Integrated Resources Management” will make a further contribution to promote progress in this field.

5. ACKNOWLEDGEMENTS

The work reported from the HarmoniCOP project has been financially supported by the European Commission under the contract number EVK1-CT-2002-00120.

6. REFERENCES

Bandura, A., Social Learning Theory. Ehglewood Cliffs, Prentice-Hall, N-J.,1977.

Barreteau, O., F. Bousquet and J.-M. Attonaty, Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems. *Journal of Artificial Societies and Social Simulation*, 4: http://www.soc.surrey.ac.uk/JASSS/4/2/5.html, 2001.

Berkes F, L. Colding J, and C. Folke (eds.), Navigating Social-Ecological Systems: Building Resilience for Complexity and Change. Cambridge University Press, Cambridge, MA., 2002.

Craps, M. editor. Social Learning in River Basin Management. Report of workpackage 2 of the HarmoniCOP project (www.harmonicop.info), 2003.

De Geus, A. Modeling to predict or to learn? *European Journal for Operations Research* 59, 1–5, 1992.

Dewulf,A., M. Craps., R. Bouwen, T. Taillieu,T. and C. Pahl-Wostl, in press. Integrated Management of Natural Resources: Uncertainy, Ambiguity and the Role of Framing and Reframing.

Den Exter, K. and A. Specht, Assisting Stakeholder Decision Making using System Dynamics Group Model-building: http://www.regional.org.au/au/apen/2003/papers/108denexterk.htm, 2003.

Douglas , M and A. Wildavsky. Risk and Culture. Berkely, 1983.

Doyle, J.K. and D.N. Ford. Mental models concepts for system dynamics research. *System Dynamics Review*, 14, 3-30, 1998.

Doyle, J.K. and D.N. Ford, Mental models concepts revisited: some clarifications and a reply to Lane. *System Dynamics Review*, 15, 411-415, 1999.

Duijn,M., F.A. Immers, F.A. Waaldijk, and H.J. Stoelhorst, Gaming Approach Route 26: a combination of computer simulation, design tools and social interaction. *Journal of Artificial Societies and Social Simulation*, 6: http://www.soc.surrey.ac.uk/JASSS/6/3/7.html, 2003.

Evans, J., Bias in Human Reasoning: Causes and Consequences. Psychology Press, UK, 1990.

Funtowicz, S. and J. Ravetz, Science for the Post-Normal Age, *Futures*, 25, 735-755, 1993.

Hare, M.P., and C. Pahl-Wostl, Stakeholder categorisation in participatory integrated assessment processes. *Integrated Assessment*, 3 (1), 50-62, 2002.

Hartvigsen, G, A. Kinzig, and G. Peterson, Use and Analysis of Complex Adaptive Systems in Ecosystem Science: Overview of Special Section,. *Ecosystems*, 1, 427-430, 1998.

Holling,C.S. editor. Adaptive Environmental Assessment and Management. John Wiley and Sons, New York, 1978.

Lane. D.C., Modelling as Learning: A consultancy method for enhancing learning in management teams. *European Journal of Operations Research*, 59, 64-84, 1992.

Lee, K.N., Appraising Adaptive Management. *Conservation Ecology*, 3, 3.-16, 1999.

Levin, SA., Ecosystems and the biosphere as complex adaptive systems, *Ecosystems*, 1, 431-436, 1998.

Ludwig,D., R. Hilborn and C. Walters, Uncertainty, resource exploitation, and conservation: Lessons from history, *Science*, 260: 17-18, 1993.

Nilsson, S., The role and use of information in Transboundary Water Management. PhD Thesis, Department of Land and Water Resources Engineering, Royal Institute of Technology, Stockholm, Sweden, 2003.

Pahl-Wostl, C., The dynamic nature of ecosystems: Chaos and order entwined. Wiley & Sons, Chichester, 1995.

Pahl-Wostl, C., Participative and stakeholder-based policy design and modeling processes, *Integrated Assessmen,t* 3(1), 3-14, 2002a.

Pahl-Wostl , C., Towards sustainability in the water sector - The importance of human actors and processes of social learning, *Aquatic Sciences*, 64(4), 394-411, 2002b.

Pahl-Wostl, C., Self-Regulation of Limnetic Ecosystems. in: The Lakes Handbook. O'Sullivan, P.E.and Reynolds,C.S. (eds.) Blackwell Science Ltd. pp. 583-608, 2004.

Pahl-Wostl, C. and M.P. Hare, in press, Processes of Social Learning in Integrated Resources Management. Journal of Community and Applied Social Psychology.

Senge, P., The fifth discipline, Doubleday, New York, 1990.

Shackley,S., B., Wynne, and C. Waterton, Imaging Complexity: The past, present and future potential of complex thinking. *Futures*, 28, 201-225, 1996.

Sterman, J.D., Business Dynamics: System Thinking and Modeling for a Complex World. Mac Graw Hill, 2000.

Timmerman,J. and S. Langaas, Environmental Information in European Transboundary Water Management, IWA Publishing, UK, 2003.

Van der Heijden, K., Scenarios, the Art of Strategic Conversation. Wiley, Chichester, UK, 1996.

Walters, C., Adaptive Management of Renewable Resources. Macmillan, New York, 1986.

Vennix, J., Group Model Building. Wiley, 1996.

Vennix, J., Group Model-building: tackling messy problems. *System Dynamics Review*, 15(4), 379-401, 1999.

Wenger, E. Communities of Practice; Learning, Meaning, and Identity, Cambridge University Press, 1998.

Economic hotspots: Visualizing Vulnerability to Flooding

Anne van der Veen

Department of Business, Technology and Public Policy
University of Twente
P.O. Box 217
Enschede
The Netherlands

Abstract: We simulate a large-scale flooding in the Province of South-Holland in the economic centre of the Netherlands. In traditional research, damage due to flooding is computed with a unit loss method coupling land use information to depth-damage functions. Normally only direct costs are incorporated as an estimate of damage to infrastructure, property and business disruption. We extend this damage concept with the indirect economic effects on the rest of the regional and national economy on basis of a bi-regional input output table. We broaden this damage estimation to the concept of vulnerability. Vulnerability is defined as a function of dependence, transferability and susceptibility. Susceptibility is the probability and extent of flooding. Dependency is the degree to which an activity relates to other economic activities in the rest of the country. Input Output multipliers form representations of this dependency. Transferability is the ability of an economic activity to respond to a disaster by deferring, using substitutes or relocating. We measure transferability as the degree of centrality of an economic activity in a network. The more central an activity is, the less it encounters possibilities to transfer production and the more vulnerable it is for flooding. Vulnerability of economic activities is then visualised in a GIS. Kernel density estimation is applied to generalise point information on inundated firms to sectoral information in space. We apply spatial interpolation techniques for the whole of the province of South Holland. Combining information of sectoral data on transferability and, we are able to create maps of economic hotspots. Our simulation of a flood in the centre of Holland reveals the vulnerability of a densely populated delta.

Keywords: Risk, economic damage, vulnerability, GIS.

1. INTRODUCTION

People and the environment are suffering increasingly from the effects of natural disasters due to high population growth and density, migration and unplanned urbanization, environmental degradation and possibly global climate change (UNEP, 2002). The number of people affected by disasters rose from an average of 147 million a year in the 1980s to 211 million a year in the 1990s. The consequences of climate change for low-lying countries like the Netherlands are potentially formidable (IPCC, 1998). From the seaside, the Netherlands faces an increase in the probability of being flooded, but also from the riverside, the delta has to deal with an increase in the discharge of water. On top of that, the Netherlands encounters subsidence. Physical processes due to ground water extraction and gas extraction challenge Dutch water policy from a new and different perspective.

Sea level rise in combination with subsidence will thus lead to a relative sea level rise, which is a danger to Dutch society. Policy measures concentrate on the coast, and along the rivers Rhine and Meuse. For many years, the standard answer by Dutch Rijkswaterstaat was to raise the dikes along the rivers and the strengthening of weak positions in the dunes along the North Sea. However, this policy of raising dikes is now part of a nation wide discussion. The search is for alternative measures, which have the same effect. Accordingly, Rijkswaterstaat (2000) developed policy measures, which are aimed at giving water more space in the floodplain of the Dutch rivers. The reasoning behind this proposal is that allowing the river to expand in times of high water will decrease the pressure on the existing dikes and thus will decrease the risk to Dutch society.

Behind the concept of risk is the issue "What is it we are protecting?" A first and quick answer to this question is the value to society of the damage after an inundation. Probability times effect then is an indicator of risk to society.

In the following, we will shortly concentrate on the concept of risk. We will argue that the current measures of risk to society mainly focus on direct economic effects and do not cover indirect economic damage. Secondly, by concentrating on risk we refrain from the resilience of society after a disaster and the ability of society to adapt. Otherwise stated, the question is "How vulnerable are we for disasters, can we cope with it and how do we measure such vulnerability?" Finally, can we visualize vulnerability? The latter question leads us to the detection of economic hotspots. We will discuss the concepts of risk and vulnerability and illustrate the empirical content of vulnerability on basis of a simulation of a large-scale flood in the province of South Holland in the Netherlands.

2. RISK AND VULNERABILITY

Risk and vulnerability are words that have gone through a certain process changing its meaning and connotation. Out of a huge literature (see e.g. Blaikie et al., (1994)), Green (2003) defines risk as the probability of a negative perturbation to a system multiplied by the effect on that system. He refers to the more engineering mode of dealing with the question how vulnerable society is for disasters. At the other extreme we find the literature on risk and uncertainty as proposed by Funtowicz and Ravetz (1993).

Common practice in the flooding (engineering) literature is to visualize risk and thus the underlying effect by counting unit losses (Parker *et al.*, 1987). With different flood-depths, depth-damage data is used to asses flood losses. The current state of this type of models (Vrisou van Eck and Kok, 2001) is that data on land cover is

$$S = \sum_{i=1}^{n} \alpha_i m_i S_i$$

In this formula:

S = Total damage
α = Damage factor
m = Number of entities in damage class i
S_i = Damage value for class i
n = Number of damage classes i

collected and downloaded into a GIS environment. Damage assessment then counts the number of units of a certain type in the affected area and multiplies this with a damage factor. The latter is a relationship empirically derived from surveys, in which a relationship is established between depth and damage. The damage factor is the heart of the method and thus plays an important role in estimating damage. In standard research on flood management, the value of damage is based on a replacement value. However, as discussed in (Van der Veen *et al.*, 2004) this might not reflect the economic value of the goods at risk ((Cochrane, 1997; Cole, 1998; Rose and Benavides, 1998; MAFF, 2000; Freeman *et al.*, 2002; Rose and Lim, 2002). This annoying matter is caused by a few misunderstandings:

1. There is no agreement on the economic points of departure. Financial appraisals are mixed up with cost-benefit analyses (CBA). In the latter, the usual concept is economic cost, which relates to opportunity costs in welfare economics, whereas a financial appraisal is often base for investigating the sum of money to be recovered from insurance companies.
2. There is confusion on temporal and spatial scales: Financial appraisal limits itself to a single organisation, whereas CBA requires wider borders, like a region, a nation, or the European Union.
3. There is confusion on the definition of direct costs: As stated by (Cochrane, 2003), " It is commonly asserted that total damage is the sum of direct damage (damage to building and contents) and lost *value added*. Double counting exists here because value added includes the services of capital, whereas direct damage should reflect the cost of replacing the undepreciated portion of such capital".
4. Stock concepts are confused with flow concepts.
5. The borderline between direct and indirect costs is not well defined.

The distinction between direct and indirect costs is highlighted in Figure 1. (Cochrane, 1997). If factory B is flooded suppliers of goods and services are hit, as well as firms that purchase goods B. In the end, final demand of consumption, investment, export and government spending is touched. Part of a risk concept thus implies taking into account forward and backward linkages in a regional or national economy.

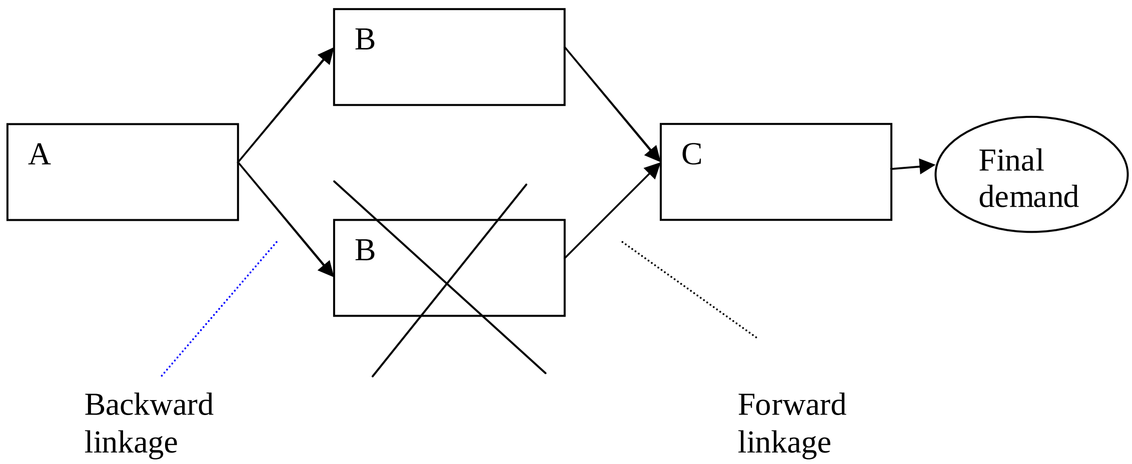


Figure 1. Forward and Backward Linkages in an Economy, when Factory B is Damaged

Note however that this extension does not allow for redundancy in an economy: if there is a second firm B that is able to take over the production, an economy is less vulnerable. See (van der Veen *et al.*, 2003) for an extensive discussion on the problem of computing an aggregate indicator of the total economic costs of a large-scale disaster, incorporating the dynamics of an economy. In the next section we will define the concepts of vulnerability and redundancy more precisely.

2.1 Vulnerability

In this paper we want to discuss the concept of vulnerability and pay attention to a method to visualize vulnerability in a disaggregated way in a GIS. By extending the concept of risk to a vulnerability concept we have to include the coping capacity of a region/nation to deal with floods. What is this coping capacity of society after a disaster?
As a point of departure we take the concept of vulnerability as introduced in a seminal paper by Parker, Green en Thompson (1987). Vulnerability V is presented with the following formula:

V = f (S, D, T), where

S = *Susceptibility, defined by the probability and extent to which the physical presence of water will affect inputs or outputs of an activity.*

D = *Dependence, reflecting the degree to which an activity requires other activities to function normally.*

T = *Transferability, the ability of an activity to respond to a disruptive threat by deferring or using substitutes or relocating.*

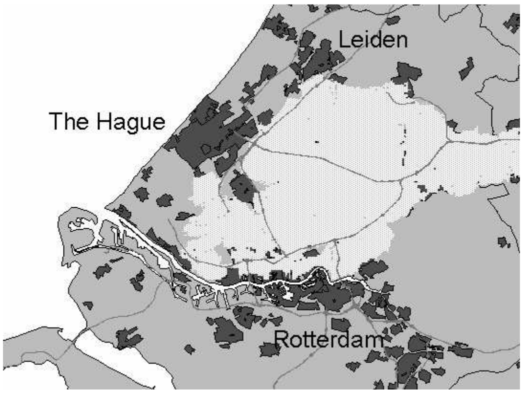
Susceptibility refers to the geo-location of a site that is under investigation. Some sites are more prone to flooding and may encounter more often flooding. Susceptibility therefore relates to the geo-concept of damage. Dependency and transferability relate to the characteristics of the economic system. Dependency and transferability (nowadays we would say redundancy) are concepts best understood by representing the economic system as a network of interrelated activities. Within such a network, there are certain functions and sectors that are important for the functioning of the network as a whole. The first concept refers to how dependent we are upon output produced at a site and the latter refers to the local redundancy in the network. Both concepts are highly interrelated.

Note that by introducing concepts like dependency and redundancy we relate to the concept of economic costs in Cost-Benefit Analysis as discussed in EPA (2000). The concept of economic costs is a dynamic accounting for adaptations in an economic structure.

Note also that the concepts of vulnerability and redundancy in recent literature move to the term resiliency (Perrings, 2001). However, we feel more confident with the term vulnerability; as Green (2003) states: "vulnerability can be defined as: "*the time varying status of some desirable or undesirable characteristic(s) of the system in question*". So, in turn, the resilience of a system is: "*the dynamic response of vulnerability over time to the perturbations to which the system is subjected"*. A *resilient* system is then one which bends under stress but does not break, and which returns to a desirable state after the perturbation has passed". End of quotation. Consequently, it is vulnerability, which has an empirical content.

In the following section, we will illustrate the concept of vulnerability based on a case in the Netherlands.

3. SIMULATING A LARGE-SCALE FLOOD: THE CASE OF THE NETHERLANDS



The case we present is a large-scale flood due to a dike failure near Krimpen aan de IJssel, a small village in the Netherlands near Rotterdam. In 1953, the dramatic year of big floods in North Western Europe, a dike in this village broke. However, by sailing in a small ship in the breach, a disaster was prevented. We now simulate this flood and try assessing the economic consequences. Delft Hydraulics supplied hydrological data and Rijkswaterstaat made GIS data available on firms, economic sectors and employment per grid (Vrisou van Eck and Kok, 2001). The blue area in the card represents the effect of the inundation after 10 days. In the middle of the area, water depth is -6 meter. Moreover, we have at our disposal a bi-regional Input-Output Table (Eding *et al.*, 1995) representing the economic structure of the Province of South Holland and the Rest of the Netherlands. We distinguish 28 sectors in our economic model.

3.1 Susceptibility

Susceptibility is the physical characteristic of the location that makes an activity vulnerable. For our case study, we apply a simulation model, so we skip this element.

3.2 Dependency

Starting from the definition of dependency "reflecting the degree to which an activity requires other economic activities as an input to function normally", we refer to the idea of an economy as a network of linkages of interrelated industries. In doing so we can apply standard concepts in economics to estimate how much we depend on certain sectors for our regional and national welfare. From input-output analysis, (Leontief, 1952; Leontief, 1986) we apply the concept of multipliers[1] (Miller, 1985; Rose and Lim, 2002) to account for forward and backward linkages in the whole economy when a disaster hits a certain sector (Figure 1).

[1] Input-Output analysis can be described as an economic method that focuses on the trade pattern between the constituent parts of the economy. These parts can be defined in very general terms, such as producers, consumers, the labour force, governmental agencies and foreign countries. Classifications may vary depending on the level of aggregation. For example, on the production side one may distinguish various types of industries or sectors such as agriculture, industry, several types of services, etc. The economic structure itself is modelled in terms of an input-output or inter-industry table, which represents the sales from one sector to another. The sales can be between industries, but also between firms and households or between workers and firms. The pattern of trade is interpreted in terms of regularities in the table, which can be analysed using various mathematical techniques. In conventional notation, X is referred to as an output vector and F – vector of final demand. The relationship between the two is established by

$$X=(I-A)^{-1}F$$

where A is the symmetric matrix of (constant) input coefficients, and $(I-A)^{-1}$ is the economy multiplier matrix (see Miller and Blair, 1985). Knowledge of the table enables us to calculate the effects of shifts in demand via multiplier analysis.
The tables are compiled at specialized institutes. In the Netherlands Statistics Netherlands (CBS) publishes these tables. In addition, other places such as universities sometimes compile specialized versions.

3.3 Redundancy

Redundancy was defined by the ability of an activity (or system) to respond to a disruption by overcoming dependence by deferring, using substitutes or even relocating. Unused capacity, inventories and possibilities to import determine the coping capacity of the economy after a disaster (FEMA, 1999). See Figure 2.

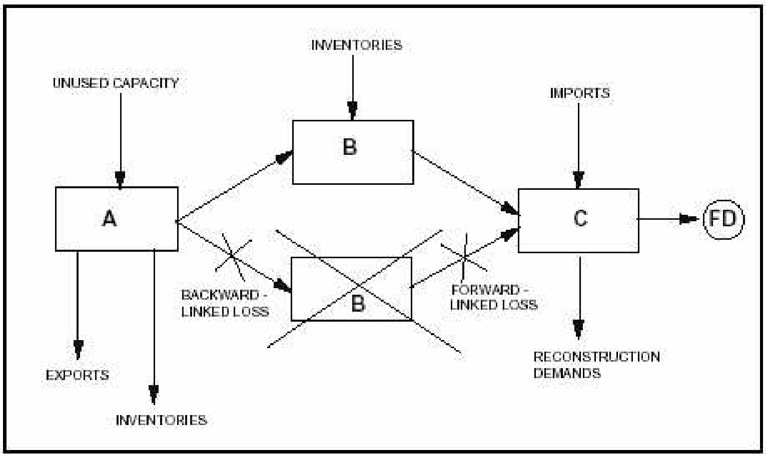


Figure 2. Transferability: choice between alternatives (FEMA, 1999)

The choice between alternatives in order to cope with the consequences of a disaster is elaborated in (FEMA, 1999). See Figure 3.

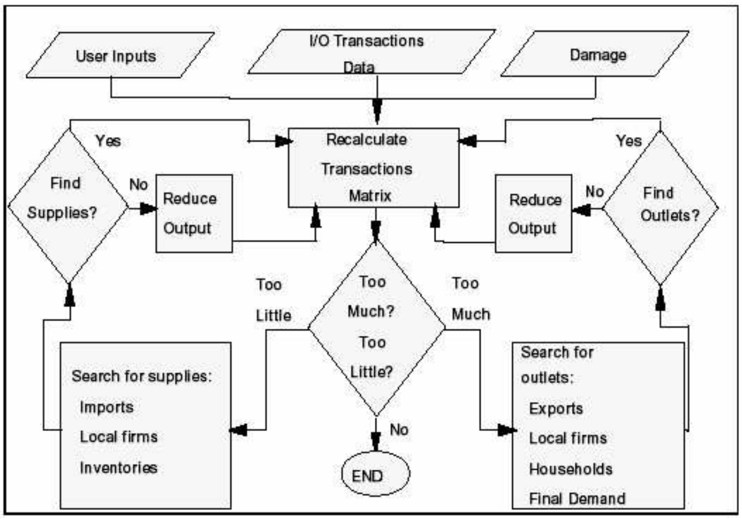


Figure 3. Determining redundancy in an economy (FEMA, 1999)

However, data requirements for such a scheme are very demanding. In our case study, we have not the empirical detail at our disposal to copy the US FEMA approach. For the Dutch situation, we have to change over to a shortcut that reflects the detailed redundancy concept. We search for an alternative approach by looking at redundancy from a network perspective. Here the A matrix in economic input-output analysis can be viewed as a picture of the network of an economy. Our hypothesis is the following:

The more central the role of a sector is in a certain regional economy, the more difficult it will be to transfer production or to substitute production by a sector in the rest of the Netherlands.

Comparing a measure of centrality of economic sectors in the province of South Holland with a centrality measure for the same sectors in the Rest of the Netherlands will indicate redundancy in the flooded provincial economy. Indeed, this concept of centrality is an aggregate of the concept used by FEMA; however, given our data limitations, we will work with it.

Centrality is a well-known concept in regional economics, measuring structural properties that can explain the performance of a network (Frank, 2002). It shows what sectors or entities play an important role and what sectors are strategic. Conventional measures of centrality relate to the identification of key sectors in an economy (Hazari, 1970). In more recent research Kilkenny and Nalbarte (1998) apply social network analysis (Leinhardt, 1977) to recognize central key sectors in a community.

Tallberg (2000) describes three measurements of centrality:

- The proportion of indirect contacts via a certain actor. This measurement is an important explanatory variable in studies of actor attributes.
- The number of direct contacts with other actors, being an indicator of the popularity of an activity.
- The distances and paths in a network, focusing on structural properties of a network.

We prefer to apply the second indicator measuring the popularity of activities of sectors in an economy. The Input-Output A matrix (See footnote 1) can be viewed as a description of all relationships in an economy. By measuring the number and content of all contacts between economic sectors, we obtain a centrality indicator for individual sectors.
For our purpose, we like to extend the use of network analysis and the measure for centrality by incorporating the redundancy in the network. Our interest is in the importance of an economic sector relative to its bi-regional counterpart in the rest of the country. If the bi-regional counterpart is more central, we assume that this sector can replace production after a disaster. As a measure, we compute the quotient of the degree centrality of an economic sector in the Province of South Holland and the same economic sector in the Rest of the Netherlands.

4 Visualizing vulnerability

So far, we presented vulnerability as an aggregate concept. Now the question is, is it possible to disaggregate vulnerability and to relate it to a GIS environment? What are economic hotspots? The question thus is how to disaggregate redundancy and dependency to grids?

4.1 Hot spots

In the literature, hotspots are defined as:

"Typically named hot spots, these are concentrations of incidents within a limited geographical context that appear over time" (Levine, 2002).

A hotspot thus measures the intensity of an incident in space. Examples can be found in applied research on crime and the formation of town centres (Thurstain and Goodwin, 2000; Thurstain *et al.*, 2001; Levine, 2002). For our project, we want to detect the intensity of economic production in space measured by the economic concept of value added. Grids in space with high value added are labelled as hotspots.

4.1.1 From points to grids

As a first step in our empirical work, we have to generalise point information on firms to sectoral information in space. The data per grid are derived from employment data for individual firms per zip code, transformed into value added on basis of provincial data on value added and on employment. For details see van der Veen *et al.*, (2003). We distinguish 28 economic sectors, ranging from agriculture and several types of industry to sectors of services.

We apply spatial interpolation techniques for the whole of the province of South Holland. The result is a contour map of the phenomenon of value added. The available techniques offer several possibilities to cluster spatial information. Analogous to the work of (Thurstain and Goodwin, 2000) on town centres, we apply Kernel Density estimation (Levine, 2002) producing a circular area (kernel) of a certain bandwidth around an indicator[2].

[2] One of the critical steps in using kernel-density estimation is to define the bandwidth in the formula. As an illustration we choose a bandwidth of 5000 m; for the rest of our empirical research we apply 2000 m. To test this ecological fallacy we have to perform an extensive sensitivity analysis in future research.

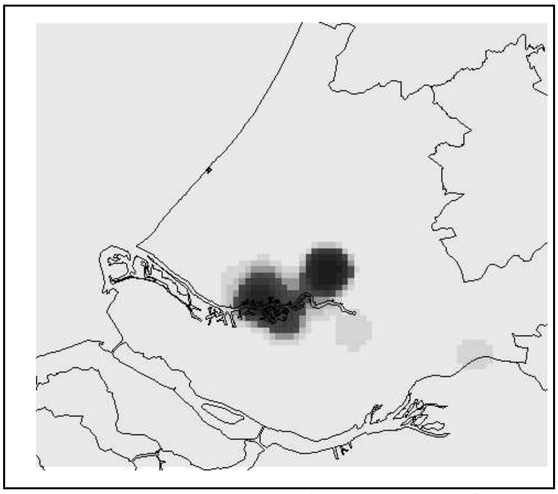


Figure 5. Clustering of value added of the petroleum industry in the province of South Holland.

As an example we show in Figure 4 the clustering of the Petroleum Industry in the province of South Holland. Here we measure the concentration of value added in the Petroleum Industry for all grids in the province.

4.1.2 Combining information

In a second step we made similar figures as for the Petroleum Industry for the other 27 sectors in our provincial economy and applied an overlay procedure to combine and aggregate the data of the 28 sectors to a higher level. All sectors are combined with no additional weights.

Results for the added value of the whole economy are presented in Figure 5. As can be seen, Rotterdam is a hotspot, but we also discover other important points in space.

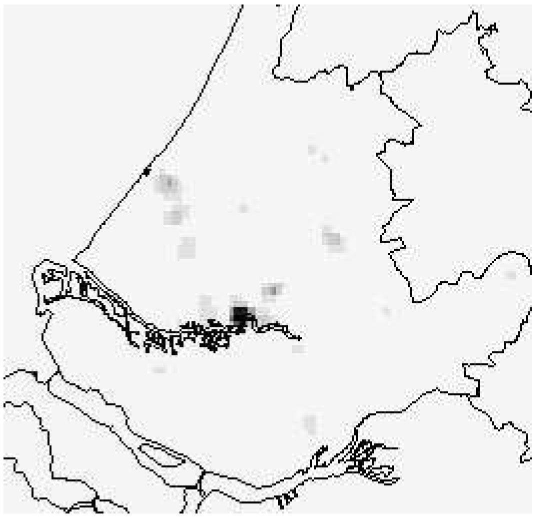


Figure 6. Economic hotspots in South Holland; Combining information on value added of 28 economic sectors.

4.1.1 Adding dependency and redundancy indicators

Finally, density surfaces are combined with concepts of redundancy and dependency within ArcView GIS 3.3 software, by using a weighted overlay procedure. It is a technique for applying a common scale of values to diverse and dissimilar inputs in order to create an integrated analysis. This procedure enables us to add indicators of importance to layers[3]. The weights we use relate to multipliers in the Input-Output network and to the indicators of centrality.

We follow a step procedure to present economic hotspots. First, we insert information on multipliers as weights in the aggregating procedure in Figure 5: If a sector has indirect links to the rest of the Dutch economy the importance of the spot is increased. Figure 6 shows the result. New spots are added[4], to show that there are strong links within the Dutch economy increasing the vulnerability to floods.

[3] As an alternative we suggest to use an arrhythmic overlay procedure. The main difference between these two procedures is in the way 'weights" are dealt with in the procedure. In the case of a weighted overlay, all weights are related to each other and to the total value of the weights.

[4] Note that some spots disappear due to relative scaling.

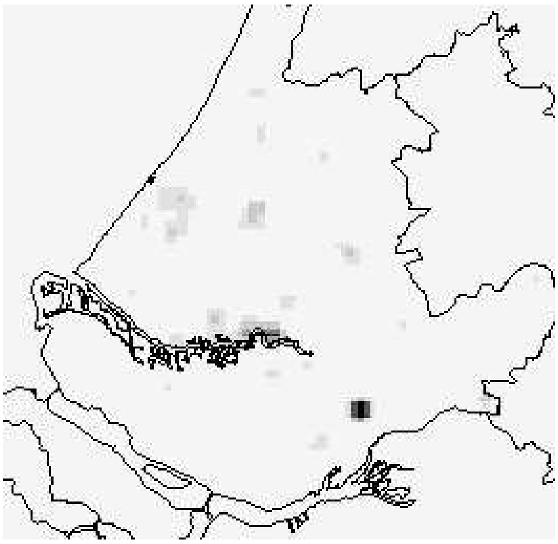


Figure 7. Economic hotspots in South Holland; Combining information on value added of 28 economic sectors with information on multipliers.

Secondly, we insert information on redundancy in Figure 5: In section 3.3 we discussed an indicator of centrality. We computed the quotient of the two indicators for the Province of South Holland and for the Rest of the Netherlands. A high weight implies less redundancy in the rest in the economy. Figure 7 presents the results: Indeed, there is a slight decline in the number of hot spots; we observe that there is scope in the rest of the Dutch economy to take over some production.

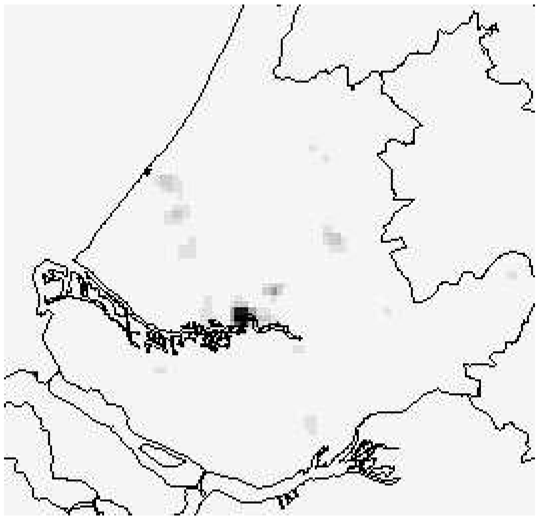


Figure 8: Economic hotspots in South olland; Combining information on value added of 28 economic sectors with information on redundancy in the rest of the Dutch economy

.

In Figure 8 we combine Figure 5 with information on dependency and redundancy. The weight we use for each of the economic sectors is the average of the multiplier and the centrality index. In the result we see the mutual influence of direct and indirect effects in an economy and the coping capacity of the Dutch economy to deal with a disaster like a flood.

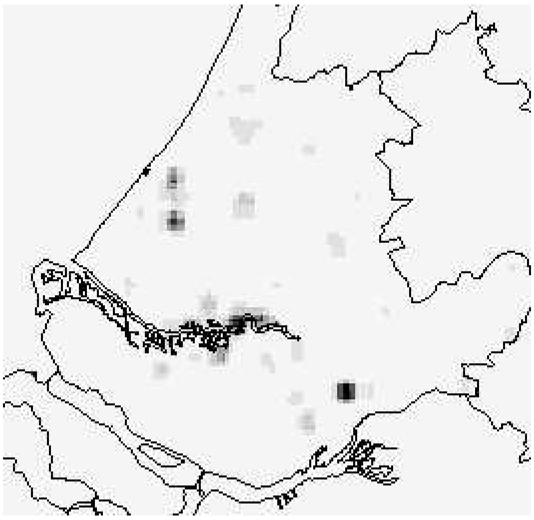


Figure 9.: Economic hotspots in South Holland; Combining information on value added of 28 economic sectors with information on multipliers and information on redundancy in the rest of the Dutch economy.

Finally, in Figure 9 we reproduce Figure 8 in combination with the borderlines of the flood that was simulated. Moreover, we insert the main highways in the area.

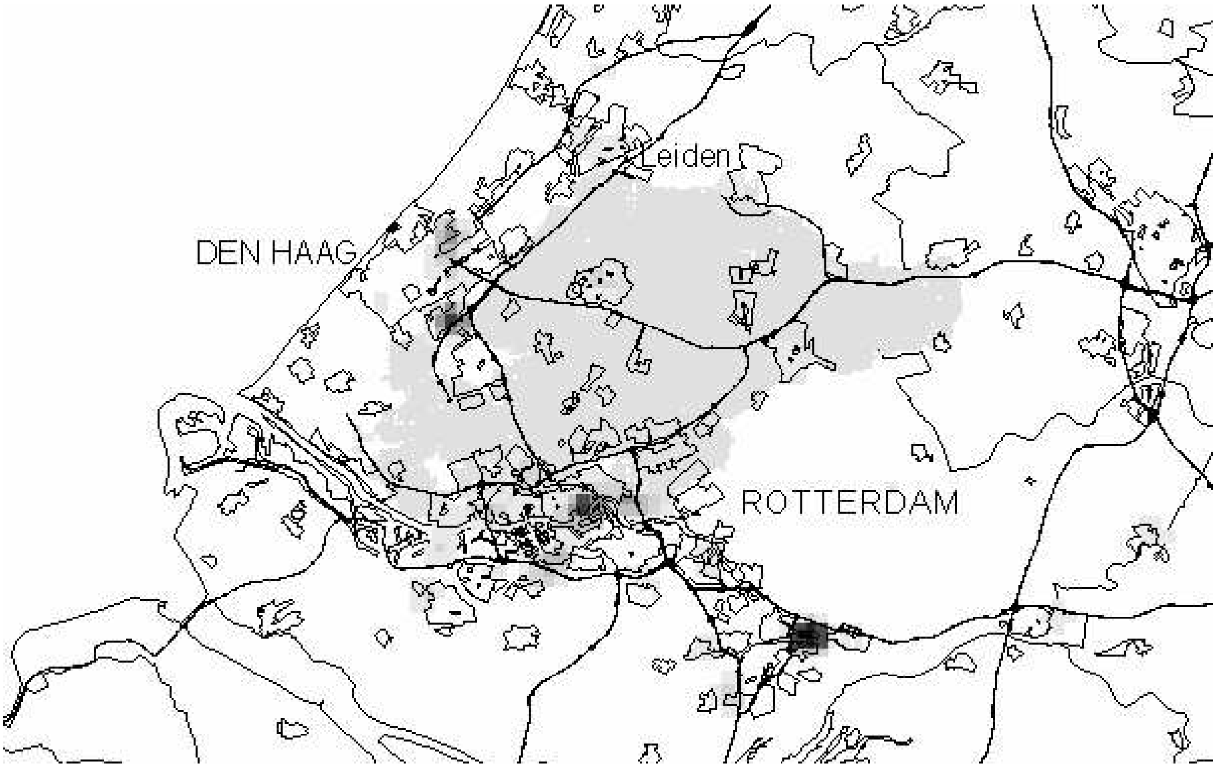


Figure 10. Economic hotspots in the province of South Holland due to a simulated flood near Krimpen aan de IJssel

Two interesting topics are revealed:

1. In the area of inundation, some important spots are highlighted.
2. Other economic hot spots are situated outside the flooded area, but are heavily hit because the infrastructure of highways passes the flooded area. Especially the latter phenomenon asks for additional research on the role of economic lifelines (highways, cable infrastructure and electricity and gas lines) in the assessment of risk and vulnerability.

5. CONCLUSIONS

Disasters like flooding have an enormous impact on local and national economies. Reasoning along the classical lines of risk misses the impact on the rest of the economy, but also does not take into account redundancy in an economy. A challenge is to visualize vulnerability: Where are the economic hotspots? By referring to the economy as a network we were able to develop a new and innovative tool visualising vulnerability. Additional research is necessary to test this instrument and the underlying indicators. Moreover, we envisage important new research around the role of economic lifelines in the concept of vulnerability.

6. REFERENCES

Blaikie, P., Cannon, T., Davis, I. and Wisner, B.:1994. *At Risk: natural hazards, people's vulnerability, and disasters,*. London, Routledge.

Cochrane, H. C.:1997. *Forecasting the Economic Impact of a Midwest Earthquake*. Buffalo, National Center for Earthquake Engineering Research.

Cochrane, H. C.:2003, Economic Loss: Myth and Measurement, A. Van der Veen, A. L. Vetere Arellano and J.-P. Nordvik, Joint NEDIES and University of Twente Workshop: *In search of a common methodology on damage estimation*, EUR Report 20997 EN, Office for Official Publications of the European Communities, European Commission, Bruxelles.

Cole, S.:1998. Decision support for calamity preparedness: socioeconomic and interregional impacts. *Engineering and* *socioeconomic impacts of earthquakes*. M. Shinozuka, A. Rose and R. T. Eguchi. Buffalo, MCEER.

Eding, G., Stelder, T. M., Vos, E. R. and Oosterhaven, J.:1995. *Bi-regionale Interactie*. Groningen, REG, Stichting Ruimtelijk Economie Groningen.

EPA:2000, Guidelines for preparing economic analyses, EPA 240-r-00-003, U.S. Environmental Protection Agency,

FEMA:1999, Hazus: Technical Manual, Volume 3, FEMA, Washington, USA.

Frank, O.:2002. "Using centrality modeling in survey networks." *Social Networks* **24**: 385-394.

Freeman, P. K., Martin, L. A., Mechler, R., Warner, K. and Hausmann, P.:2002, Catastrophes and Development; Integrating natural catastrophes into development planning; Working papers series no. 4, The World Bank, Washington.

Funtowicz, S. O. and Ravetz, J. R.:1993. "Science for the Post-Normal Age." *Futures* **25**(7): 739-755.

Green, C.:2003, Evaluating vulnerability and resilience in flood management, A. Van der Veen, A. L. Vetere Arellano and J.-P. Nordvik, Joint NEDIES and University of Twente Workshop: *In search of a common methodology on damage estimation*, EUR Report 20997 EN, Office for Official Publications of the European Communities, European Commission, Bruxelles.

Hazari, B. R.:1970. "Empirical Identification of Key Sectors in the Indian Economy." *The Review of Economics and Statistics* **52**(3): 301-305.

IPCC:1998. *The Regional Impacts of Climate Change: An Assessment of Vulnerability*. Cambridge, Cambridge University Press.

Kilkenny, M. and Nalbarte, L.:1998. *Keystone sector Identification*, TVA, Tennesee Vally Authority.

Leinhardt, S., Ed. *Social Networks, a Developing Paradigm*. New York, Academic Press.

Leontief, W.:1952. "Some Basic Problems of Structural Analysis." *The Review of Economics and Statistics* **34**(1): 1-9.

Leontief, W.:1986. Input-Output Economics. *Leontief, Wassily.; Input output economics. Second edition New York and Oxford: Oxford University Press 1986; 3 18. Previously published: 1951.*

Levine, N.:2002, CrimeStat II : A Spatial Statistics Program for the Analysis of Crime Incident Locations, Crimestat Manual, Ned Levine & Associates, TX, Natinonal INstitue of Justice, Washington DC.

Lomborg, B.:1998. *The Skeptical Environmentalist*. Cambridge, Cambridge University Press.

MAFF:2000, Flood and Coastal Defense Project Appraisal Guidance: Economic Appraisal, Ministry of Agriculture, fisheries and food, London, UK.

Miller, R. E. B., Peter D.:1985. *Input-Output Analysis: Foundations and Extensions.* Englewood Cliffs, Prentice-Hall, Inc.

Parker, D. J., Green, C. H. and Thompson, P. M.:1987. *Urban flood protection benefits, a project appraisal guide*. Aldershot, Gower.

Perrings, C.:2001. Resilience and Sustainability. *Frontiers of environmental economics. Cheltenham, U.K.* H. Folmer and et al. and Northampton, Mass., Elgar; distributed by American International Distribution Corporation, Williston, Vt.**:** 319-41.

Rijkswaterstaat:2000, Ruimte voor rivieren,, Ministerie van Verkeer en Waterstaat, Directoraat-Generaal, Rijkswaterstaat., Den Haag.

Rose, A. and Benavides, J.:1998. Regional economic impacts. *Engineering and socioeconomic impacts of earthquakes*. M. Shinozuka, A. Rose and R. T. Eguchi. Buffalo, MCEER.

Rose, A. and Lim, D.:2002. "Business interruption losses from natural hazards: conceptual and methodological issues in the case of the Northridge earthquake." *Environmental hazards* **4**: 1-14.

Tallberg, C.:2000, Comparing Degree-based and Closeness-based Centrality Measures, Department of Statistics, Stockholm University, Stockholm.

Thurstain, M., Batty, M., Hacklay, M., Lloyd, D., Bodt, R., Hyman, A., Batty, S., Tomalin, C., Cadell, C., Falk, N., Sheppard, S. and Curtis, S.:2001, Producing Boundaries and Statistics for Town Centres, CASA, Center for Advanced Spatial Analysis, London.

Thurstain, M. and Goodwin, D. U.:2000, Defining and Delineating the Central Areas of Towns for Statistical Monitorin Using Continuous Surface Representations, Paper 18, Centre for Advanced Spatial Analysis, London.

UNEP:2002. *Global Environmental Outlook 3*. London, Earthscan.

van der Veen, A., Steenge, A. E., Bockarjova, M. and Logtmeijer, C. J. J.:2004, Structural economic effects of large scale

inundation: a simulation of the Krimpen dike breakage, *Disaster Prevention and Management*, to appear.
Van der Veen, A., Vetere Arellano, A. L. and Nordvik, J.-P.:2003, Joint NEDIES and University of Twente Workshop *In search of a common methodology on damage estimation*, EUR Report 20997 EN, Office for Official Publications of the European Communities, European Commission, Bruxelles.
Vrisou van Eck, N. and Kok, M.:2001, Standaard Methode Schade en Slachtofers als gevolg van overstromingen, HKV & Dienst weg en waterbouwkunde, Lelystad & Delft.

Implementation of the STREAMES environmental decision-support system

David Cabanillas[1], **Esther Llorens**[2], **Joaquim Comas**[2], **Manel Poch**[2], **Luigi Ceccaroni**[1] **and Steven Willmott**[1]

[1]Universitat Politècnica de Catalunya,
LSI department, Campus Nord, Jordi Girona 1-3, E-08034 Barcelona
Catalonia, Spain
e-mail: {david / luigi / steve}@lsi.upc.es

[2]Universitat de Girona,
Campus Montivili, E-17071 Girona
Catalonia, Spain
e-mail: {quim / esther / manel}@lequia.udg.es

Abstract: As part of the European-Commission--funded STREAMES project, a system is being developed with the objective of capturing knowledge from water managers and environmental-science experts, regarding nutrients-excess effects in streams and of combining this knowledge into a user-friendly tool to assist water managers in evaluating streams' nutrient-retention capabilities. In this paper, we summarize the decision-support knowledge components which have been identified in previous work and, based on these, present an implementation of a prototype of an environmental decision-support system. The decision support provided by the system to water managers consists of: (1) diagnosis: inferring possible stream problems, assessing the alteration degree of the stream, and evaluating the source and magnitude of nutrient loads; (2) actions: offering alternative, ranked courses of action to solve possible problems; (3) forecast: providing several scenarios to simulate the effect of the different actions proposed as solutions.

Keywords: Environmental decision-support system, implementation, river, rule-based expert system, water management

1. INTRODUCTION

In the context of European Commission's Fifth Framework Programme (1998-2002) and Water Framework Directive (WFD) (2000/60/EC), the STREAMES project[1] aims to analyze nutrient cycles in a particular, human-altered environment: the river ecosystem, with special emphasis on the Mediterranean region. The decision-making process involved in altered-rivers management requires extensive human expertise from people directly implicated in day-to-day stream problems (water managers and environmental-science experts), knowledge from different science fields and complex calculations over large amounts of numerical and symbolic data. Therefore stream optimal management requires an integrated and interdisciplinary approach. To face this complexity, the STREAMES project aims to develop and implement a knowledge-based system, which will contribute achieving a good ecological state in rivers with bad water quality. This system manages general knowledge extracted from different literature sources as well as specific knowledge acquired by processing empirical data collected from the project's study-sites and from interviews and meetings with human experts.

[1] STream REAch Management, an Expert System. Human effects on nutrient cycling in fluvial ecosystems: The development of an ES to assess stream water quality management at reach scale. [http://www.streames.org, EVK1-CT-2000-00081].

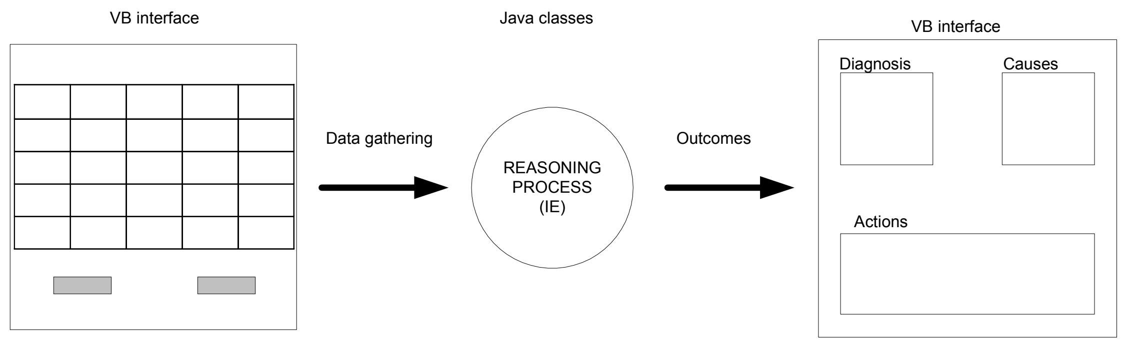


Figure 1. From data to outcomes (simplified)

In the system, artificial intelligence techniques are applied to the water-management field in the form of an environmental decision-support system (EDSS).

EDSSs are a subset of decision support systems (DSSs), which in turn are a subset of computer-based information systems (CBIS). Some examples of EDSSs developed recently and applied to the water domain are described by, among others, Chang et al. [1997], Davis et al. [1998], De Marchi et al. [1999], Rousseau et al. [2000], Rodríguez-Roda et al. [2002], the Great Lakes Commission for the Great Lakes States and Provinces [2003], Matthies et al. [2003] and Ceccaroni et al. [2004].

1.1. EDSS development

The development of the STREAMES EDSS has been carried out following a methodology composed of a series of phases, each with its own inputs, activities and outputs (modified from Poch et al. [2002]):

1. environmental problem analysis
2. data collection and knowledge acquisition
3. system analysis and design
4. *problem-solving method* (PSM) selection[2]
5. PSMs integration
6. system implementation
7. validation
8. maintenance

In the case of the STREAMES EDSS, most phases from *problem analysis* to *PSM selection* are described in a previous work by Comas et al. [2002]. With respect to PSM selection, the following ones were chosen:

PSM1. A *rule-based system*, to resolve stream management problems, whose diagnosis and solution involves qualitative data and knowledge processing.

PSM2. *Numerical and statistical models*, to estimate point and non-point nutrient inputs and self-purification capacity.

In the STREAMES project, these PSMs are complemented with *geographical information systems* and the rules of the expert system are grouped into four modules (see 2.3).

2. IMPLEMENTATION

In this paper, we analyze the implementation of the *rule-based expert system* (RBES) and of the graphical user interface. RBESs are mainly composed of a knowledge base (KB) and an inferential engine.

2.1. Inferential engine

The inferential engine (IE) works with rules (see PSM1 in section 1.1) and provides the reasoning mechanism. In our case the inference is backward chaining. With the objective of being able to reuse the RBES, we developed an IE shell that can be adapted and customized. In case of reuse in different domains, the KB would need to be redeveloped with new knowledge-components (as described in section 2.2). The current implementation of the IE is Java-based (platform independent) and is integrated with a friendly user-interface in Visual Basic (VB) (see Figure 1).

Once the system is started the user has to fill in different forms of data-input that the system presents via the user-interface. In the IE, the rules correspond to the decision trees described in section 2.2.1 and the facts correspond to the data introduced by the user.

[2] The term *PSM* corresponds to the term *model* used by Poch et al [2002].

Decision tree ID	Decision tree name	Represented problems
DT1	Nitrogen	- Excess of ammonium - Excess of nitrate - Excess of nitrite
DT2	Phosphorous	- Eutrophication
DT3	Organic matter	- Excess of organic matter - Anoxia
DT4	Suspended solids	- Excess of suspended solids - Clogging
DT5	Salts	- Anthropogenic alteration of salinity
DT6	Stream characteristics	- Low river--self-purification

Table 1. Decision trees and related diagnosed problems

Then the inference process starts, trying to find out if the facts match some of the antecedents of each one of the active rules. If a rule is triggered, new facts can be introduced into the facts base, as a result of the inference. This process finishes when the IE has tested all the facts with all the active rules. Afterwards, the IE delivers the results to the interface component that parses and shows the results to the user in an appropriate format.

2.2. Knowledge components

For building and validating the KB of a decision-support system for our given practical domain, four knowledge components (KC) are needed:

- KC1. a domain ontology for nutrient cycles in river ecosystems (to formally describe terminology and processes);
- KC2. a decision-support ontology to formalize the output of the system (see section 2.4);
- KC3. a library of decision trees (see section 2.2.1) or an equivalent rule representation scheme;
- KC4. a set of domain requirements that are used to select a suitable set of elements of KC3.

In Comas et al. [2002, 2003], KC3 and part of KC4 were made explicit; we developed the two ontologies and the remaining part of KC4.

2.2.1. Decision trees

STREAMES' KB is codified by means of rules, which are sets of *conditions and conclusions*. As a prior step to build the KB, knowledge is structured and represented in decision trees (DTs) [Comas et al., 2003]. Every DT refers to a set of specific problems (shown in Table 1) and is composed of two modules: one for problem diagnosis and one for cause detection.

The developed DTs correspond to those problems for which water managers and environment experts expressed a greater interest and preoccupation. Six DTs have been developed: one for nitrogen-related problems, one for eutrophication[3], one for organic-matter problems (which include part of the anoxia problems[4]), one for suspended solids and clogging, one for salinity problems and one for alterations of the stream ecosystem. While the first five ones are related to physico-chemical elements in the water, the last one is focused on the physical, biological and morphological characteristics of the river ecosystem (riparian zone and streambed), which can affect the river's functionality and self-purification capacity. The self-purification capacity is in turn an important aspect to be taken into account in water pollution problems.

The module of cause detection of the DTs includes a set of pre-defined causes. For example, if a low river--self-purification due to a physical alteration of the system is detected, causes such as the following ones are evaluated: riparian banks destruction, dredging, morphological alteration of the riverbed by human activities, modification of flow regime.

2.3. Rule modules

In the STREAMES KB, rules are grouped into four modules, or *steps*: the first one, *symptom discovery*, is derived from KC4, while the following three ones codify the DTs (KC3). The sequence of the steps is:

1. *Symptom discovery*. If certain symptoms are detected, this meta-rules module activates one or more DTs.
2. *Problem diagnosis*. This module represents the knowledge necessary to diagnose the problem corresponding to

[3] Eutrophication problem is evaluated by means of the N:P molar-ratio calculation.

[4] Oxygen depletion may be due also to ammonium oxidation and eutrophication problems; these situations are not considered by the current version of the EDSS.

the symptoms. A specific problem and its possible side-effects are confirmed and communicated to the user.
3. *Cause detection*. Different, possible causes of the problem under analysis are deduced and evaluated.
4. *Actuation*. A set of actions, corresponding to the causes, is proposed to solve the problem.

For a full understanding of the KB implementation and functioning, as well as the interaction with the user, we present a complete use case of the EDSS.

The process starts with the selection by the user of one of the following two options:
1. evaluation of possible stream problems;
2. assessment of the alteration degree of the stream.

In the following, we consider the first option because it is the one related to the implementation of the RBES.

2.3.1. Symptom discovery

The system begins to gather data, asking questions to the user about groups of significant descriptors (DS), or quality elements for the classification of ecological status. Some of these DS are in accordance with the WFD; other ones have been defined by the authors according to their experience and the knowledge acquired from diverse sources (e.g., EPA manuals by Barbour et al. [1999]):
1. River basin DS. These elements are related to the location of the river in its river catchment, to the characterization of the basin and to the identification of diffuse pollution sources (e.g., geology, predominant land use).
2. Streambed characterization DS. These elements are to estimate the quality of the river in relation to the riverbed. We distinguish two classes:
    a. Biological and habitat DS. These are related to the micro-scale aspects, e.g.: color of sediments, presence of bio-film, fishes, algae, macro-invertebrates.
    b. Streambed DS. These are related to larger-scale aspects, e.g.: types of streambed, channel sinuosity.
3. Hydromorphological DS supporting the biological DS. Examples of these elements are: stream width, water velocity and, in general, the hydrological regime and the river continuity.
4. Water quality DS. Examples of these elements are: nitrogen and phosphorous data, water odor, conductivity, water color, water temperature, pH.
5. Point nutrient-source DS. Identification, location and characterization of the existing point sources of nutrients in the river catchment, e.g.: input of wastewater, ammonium.
6. Riparian DS. These elements characterize the riparian zone and help to estimate the quality of the river in relation to it. Examples are: types of riparian vegetation, soil permeability.

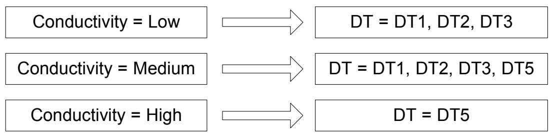


Figure 2. Symptom-discovery meta-rules

These data and a set of meta-rules representing domain requirements are used to select the DTs to be activated (see Figure 2 for an example of these rules).

2.3.2. Problem diagnosis

When, for instance, DT2 (*phosphorous*) is selected, its problem-diagnosis module is activated. Part of the problem-diagnosis rule-inference is shown in Figure 3.

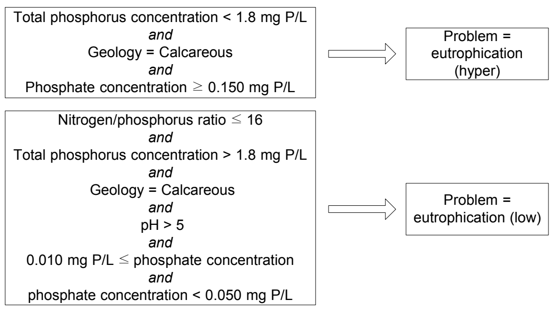


Figure 3. Problem-diagnosis rules for the *phosphorous* decision tree.

In the same way, inference is carried out in the rest of DTs activated by the meta-rules.

2.3.3. Cause detection

For each problem diagnosed, the cause-detection module of the corresponding DT is activated. Part of the cause-detection rule-inference is shown in Figure 4.

2.3.4. Actuation

Once the system executed all triggered rules in activated DTs, it shows the user a set of

<diagnosis, cause> pairs (DCPs), for him to analyze.

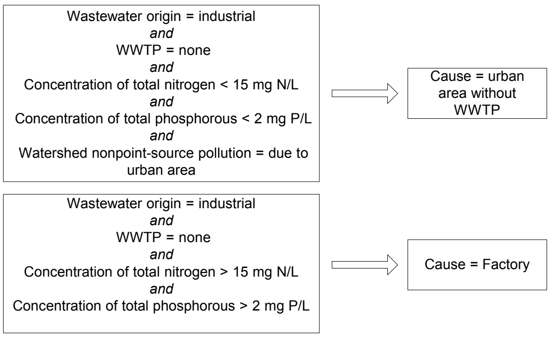


Figure 4. Cause-detection rules for the *phosphorous* decision tree.

The user chooses the DCPs he is interested in and, for each one, the actuation category (hydromorphology, chemistry, biota, best practices, hydrology) and the actuation geographical-scope (river basin, riparian zone, river body). With these data, the system is able to offer an ordered list of recommended courses of action to carry out (see an example in Figure 5), as well as, when possible, a series of complementary parameters, such as: chances of success, feasibility, response time, effort vs. environmental benefit, references.

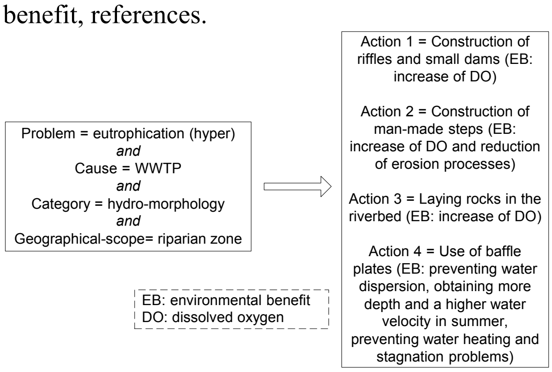


Figure 5. Recommended actions in the *actuation* step (simplified).

2.3.5. Forecast

The system forecasts what improvements would take place in the river if one of the actions suggested were carried out. As outcome, the system shows the user a comparison of the current problematic state versus the state after the application of the action, as well as a measure of the improvement in the quality of water.

2.4. Decision support

In summary, the decision support supplied by the system consists of providing:

1. Diagnosis: inferring possible stream problems, assessing the alteration degree of the stream, and evaluating the source and magnitude of nutrient loads.
2. Actions: offering alternative, ranked courses of action to solve possible problems.
3. Forecast: providing several scenarios to simulate the effect of the different actions proposed as solutions.

An example of the outcome of the system is as follows. The EDSS detects that the stream undergoes a *hyper-eutrophication* problem. Also, the EDSS has been able to infer that the cause related to this diagnosis is a point source (a WWTP without nitrogen removal). According to this diagnosis and cause, the EDSS proposes several actuations: restoration of riparian vegetation, optimization of the nitrification/de-nitrification process, nitrogen removal. Furthermore, the EDSS allows estimating the effect of the actuations proposed for stream improvement. If, for example, *nitrogen removal* were implemented, the nutrient loads into the river would decrease, the problem would be partially solved and the prediction would be *low eutrophication*.

3. CONCLUSIONS AND FUTURE WORK

Recently, attention has been focused on providing decision support for evaluating streams' nutrient-retention capabilities. Such support is needed to guide a water manager in planning actions regarding relief from nutrients-excess effects in streams. This paper contributes to the efforts for building and validating decision-support knowledge-bases (KBs) for the streams domain. We identified four knowledge components explicitly required to develop the KB of an environmental decision-support system (EDSS). These include, in the case of the river domain: (1) a domain ontology for nutrient cycles in river ecosystems (to formally describe terminology and processes); (2) a decision-support ontology to formalize the output of the system; (3) a library of decision trees or an equivalent rule-representation scheme; (4) a set of domain requirements that are used to select a suitable set of decision trees. We summarized the knowledge components which have been identified in previous work and, based on these, presented an implementation that exploits rule-based expert systems to aid water managers in planning practical and effective courses of action in response to early symptom discovery.

Future work includes: (1) integration with other technologies and models, such as the nutrient emission model MONERIS and geographical information systems, to improve the EDSS; (2) introduction of more powerful rules, using fuzzy sets and new operators; (3) automatic rule generation and validation; (4) comparison with other knowledge-based systems, such as case-based and model-based reasoning systems.

4. ACKNOWLEDGMENTS

The research described in this paper is partly supported by the EC projects STREAMES (EVK1-CT-2000-00081) and @LIS TechNET (ALA/2002/ 049-055). The opinions expressed in this paper are those of the authors and are not necessarily those of the STREAMES project's partners. The authors would like to thank the anonymous reviewers of this paper in its draft form, who have provided many constructive comments and suggestions.

5. REFERENCES

Barbour, M.T., Gerritsen, J., Snyder, B.D. and Stribling, J.B., Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C., 1999.

Ceccaroni, L., Cortés U., and Sànchez-Marrè, M., OntoWEDSS: augmenting environmental decision-support systems with ontologies. *Environmental Modelling & Software, in press*, 2004.

Chang, N. B., Chen, Y. L. and Chen, H.W., A fuzzy regression analysis for the construction cost estimation of wastewater treatment plants. *Journal of Environmental Science and Health. Part A - Environmental Science and Engineering and Toxic and Hazardous Substance Control*, *32*(4), 885-899, 1997.

Comas, J., Llorens E., Martí, E., Puig, M. A., Riera, J. L., Sabater, F. and Poch, M., Knowledge acquisition in the STREAMES project: the key process in the Environmental Decision Support System development*, in AI Communications 16*, 2003.

Comas, J., Llorens, E., Poch, M., Markakis, G., Battin, T., Gafny, S., Maneux, E., Martí, E., Morais, M., Puig, M. A., Pusch, M., Riera, J. L., Sabater, F., Solimini, A. G. and Vervier, P., The STREAMES Project: Linking Heuristic And Empirical Knowledge Into An Expert System To Assess Stream Managers*. 1st International Conference on Environmental Modelling and Software (iEMSs2002)*, *3*, 444-449, 2002.

Davis, J. R., Farley, T. F. N., Young, W. J., and Cuddy, S. M., The experience of using a decision support system for nutrient management in Australia, *Water Science and Technology*, *37*(3), 209-216, 1998.

Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy, *Official Journal L327*, *2000-10-23*, *P0001*.

Great Lakes Commission for the Great Lakes States and Provinces. *Toward a Water Resources Management Decision Support System for the Great Lakes-St. Lawrence River Basin*. Great Lakes Commission, Eisenhower Corporate Park, 2805 S. Industrial Hwy, Suite 100, Ann Arbor, MI 48104-6791, 2003. The full report is available online at [http://www.glc.org/ wateruse/wrmdss].

De Marchi, C. , Ivanov, P., Jolma, A., Masliev, I., Smith, M. G. and Somlyody, L., Innovative tools for water quality management and policy analysis: DESERT and STREAMPLAN. *Water Science and Technology*, *40*(10), 103-110, 1999.

Matthies, M., Berlekamp, J., Lautenbach, S., Graf, N. and Reimer, S., Decision Support System for the Elbe River Water Quality Management. In Proceedings of the *International Congress on Modelling and Simulation (MODSIM 2003)*, Townsville, Australia. 2003.

Rodríguez-Roda, I., Comas, J., Colprim, J., Poch, M., Sànchez-Marrè, M., Cortés, U., Baeza, J. and Lafuente, J., A hybrid supervisory system to support wastewater treatment plant operation: implementation and validation, *Water Science and Technology*, *45*(4-5), 289-297, 2002.

Rousseau, A. N., Mailhot, A., Turcotte, R., Duchemin, M., Blanchette, C., Roux, M., Etong, N., Dupont, J., and Villeneuve, J. - P., GIBSI – An integrated modelling system prototype for river basin management. *Hydrobiologia* 422/423: 465–475, 2000.

Poch, M., Comas, J., Rodríguez-Roda, I., Sànchez-Marrè, M. and U. Cortés, *Ten Years of Experience in Designing and Building Real Environmental Decision Support Systems. What Have We Learnt?, 1st International Conference on Environmental Modelling and Software (iEMSs2002)*, *3*, 22-33, 2002.

Knowledge Discovery in Environmental Data Bases using GESCONDA

Karina Gibert[1], **Xavier Flores**[2], **Ignasi Rodríguez-Roda**[2], **Miquel Sànchez-Marrè**[3]

[1]*Dep. Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Barcelona, Catalonia, EU (karina@eio.upc.es)*

[2]*Laboratori d'Enginyeria Química i Ambiental (LEQUIA). Universitat de Girona, Girona, Catalonia, EU ({xavi | ignasi}@lequia.udg.es)*

[3]*Knowldege Engineering and Machine Learning Group (KEMLG). Universitat Politècnica de Catalunya , Barcelona, Catalonia, EU (miquel@lsi.upc.es)*

Abstract: In this work, last results of the research project "Development of an Intelligent Data Analysis System for Knowledge Management in Environmental Data Bases (DB)" are presented. The project is focussed on the design and development of a prototype for Knowledge Discovery (KD) and intelligent data analysis, and specially oriented to environmental DB. It is remarkable the high quantity of information and knowledge patterns that are implicit in large DB coming from environmental domains. In this project, several environmental DB such as meteorological phenomena, wastewater treatment plants (WWTP), or environmental emergencies were used for testing. KD is a prior and mandatory step to get reliable Intelligent Environmental Decision Support Systems. Although in the literature other KD tools exists (WEKA, Intelligent Miner…) none of them integrate, like GESCONDA, statistical and AI methods, the possibility of explicit management of the produced knowledge in Knowledge Bases (KB) (in the classical AI sense), mixed techniques that can cooperate among them to discover and extract the knowledge contained in data, dynamical data analysis… in a single tool, allowing interaction among all the methods. The purpose of the paper is to present the final architecture of GESCONDA, as well as some of the methods incorporated in last phases. Later, an application to discover knowledge patterns from an environmental DB (a WWTP) is detailed. The DB has been mined using several methods available in GESCONDA. First of all, statistical filtering approaches were applied for data preparation. Afterwards, a hybrid clustering technique (clustering based on rules) was used to discover the structure of the target phenomenon. Finally, clustering results were used as input for rule induction making new knowledge explicit. Results and feedback from validation steps show that the tool seems to be useful and efficient for KD.

Keywords: Knowledge Acquisition and Management, Data Mining, Machine Learning, Environmental Databases, Statistical Modelling, Rules Induction, WasteWater Treatment Plant.

1. INTRODUCTION

An Environmental Decision Support System (EDSS) can be defined as an intelligent information system for decreasing the decision-making time and improving consistency and quality of decisions in Environmental Systems.

An EDSS is an ideal decision-oriented tool for suggesting recommendations in an environmental domain. The main outstanding feature of EDSS is the knowledge embodied, which provides the system with enhanced abilities to reason about the environmental system in a more reliable way. A common problem in their development is how to obtain that knowledge. Classic approaches are based on getting knowledge by manual interactive sessions with environmental experts. But when there are available databases summarising the behaviour of the environmental system in the past, there is a more interesting and promising approach: using several common automated techniques from both Statistics and Machine Learning fields. The conjoint use of those techniques is usually named as data mining (Gibert et al. 1998).

All this information and knowledge is very important for prediction, control, supervision and minimisation of environmental impact either in Nature and Human beings themselves. The project is involved with building an Intelligent Data Analysis System (IDAS) to provide the support to these kind of environmental decision-making.

In this paper, last methods incorporated in system are introduced. A real application is presented in order to illustrate how the system supports KD in a real-world environmental database.

3. FINAL ARCHITECTURE

GESCONDA is the name given to the IDAS developed within the project. On the basis of previous experiences, it was decided that it would have multi-layer architecture of 4 levels connecting the user with the environmental system or process.

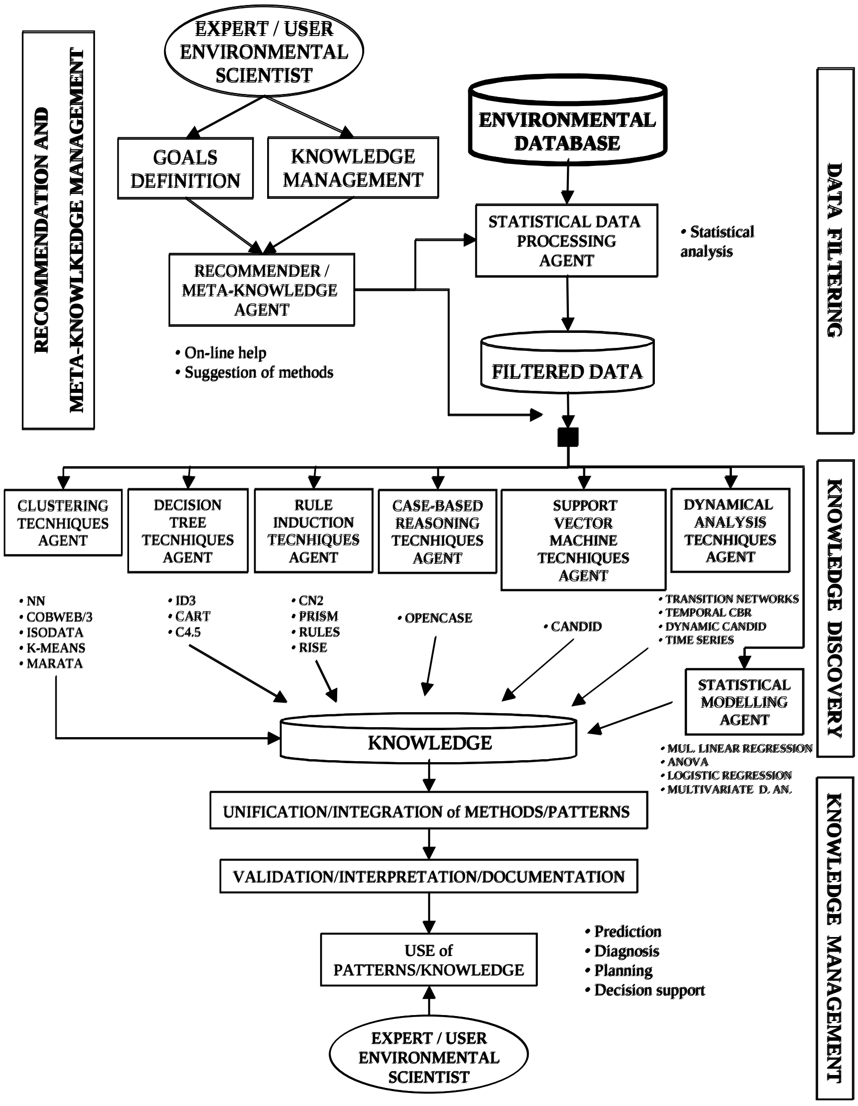


Figure 1. Architecture of GESCONDA

These 4 levels are the following:

- Data Filtering: a) Data cleaning; b) Missing data analysis and management; c) Outlier data analysis and management; d) Statistical one-way analysis; e) Statistical two-way analysis; o Graphical visualisation tools; f) Attribute or Variable transformation
- Recommendation and Meta-Knowledge Management: a) Problem goal definition; b) Method suggestion; c) Parameter setting; d) Attribute or Variable Meta-Knowledge management; e) Example Meta-knowledge management; f) Domain theory knowledge elicitation.
- Knowledge Discovery: a) Clustering (ML and Statistical); b) Decision tree induction; c) Classification rule induction; d) Case-based reasoning; e) Support vector machine; f) Statistical Modelling; g) Dynamic analysis.
- Knowledge Management: a) Integration of different knowledge patterns for a predictive task,

or planning, or system supervision; b) Validation of the acquired Knowledge pattern; c) Knowledge utilisation by end-users; d) User interaction.
Fig. 1, depicts the architecture of the system.

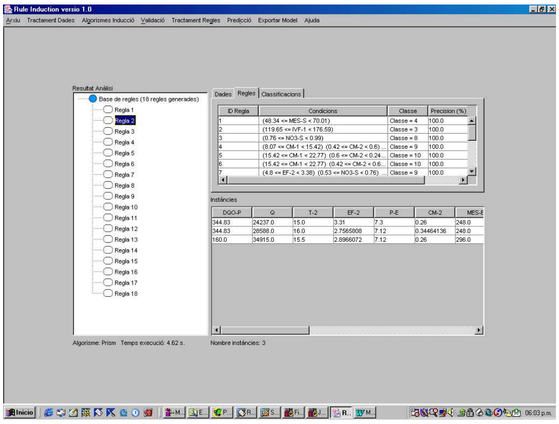


Figure 2 A set of induced rules

The GESCONDA system will provide a set of mixed techniques that will be useful to acquire relevant knowledge from environmental systems, through available databases. This knowledge will be used afterwards in the implementation of reliable EDSS. The portability of the software will be provided by a common Java platform.

In next section there is a more detailed description of the rules induction agent and the statistical modelling one.

4. AGENTS

In previous works (Sànchez-Marrè et al 2002) data-filtering agent, clustering agent and decision-tree induction agent, which were developed the first were already presented. In this paper, details on agents added to GESCONDA in a second development period are presented.

4.1 Rules induction

The rule induction agent is the responsible to induce classification rules directly from supervised databases. Resulting rule-bases can be the input for building a Knowledge-Based classifier system.

Induced rules have two components: the left-hand side is a boolean expression built as a conjunction of *selectors* or conditions on the attributes (*Length = high*, *Diameter* $\in$ *[3.4, 5.8]...)*; the right-hand side is the label of the class to which the instances satisfying the left-hand side belong (*Class4*, …).

These techniques induce a set of rules covering as much as possible all the instances within the training set. In general, not all the instances can be classified by the induced rule set. Resulting rules are provided with a predictive accuracy, estimation rate and with coverage estimation rate. Some validation techniques, such as simple validation or cross-validation, have been implemented to test the reliability of induced rules. After validation, the rules set can be used as a model to predict the class of new unseen instances. The knowledge base can be exported to a text file or a CLIPS file, for a later use within a knowledge-based system.

Fig. 2 depicts a set of rules induced from a certain database by GESCONDA. Several different algorithms such as RULES (Pham&Aksoy, 1995), PRISM (Cendrowska, 1987), CN2 (Clark & Niblett, 1989), and RISE (Domingos, 1996) have been selected and implemented in the system. The agent also supports a tuning of the resulting rules, as well as the possibility of removing very low accurate rules either manually or automatically.

First three algorithms are selector-based. RULES computes induced rules starting from an initial rule with empty left-hand side. Step by step, it adds one selector each time until it obtains a 100% accurate rule, where all covering instances are correctly assigned to one class. If some instances are not classified yet, it builds a new possible rule.

PRISM is based on a similar principle as RULES. Main difference is that induced rules are computed separately for each class. First, only instances labelled with the first class are considered for the inductive process, and so on until the last class.

The very popular CN2 is based on a heuristic search for the best combination of selectors, which are known as *complex* in the algorithm terminology. Only k complex are maintained and explored. The best complex is selected as the basis for new rules. The right-hand side of the rule is set to the more frequent class from instances covered by the complex. When all instances are covered or no more complex can be formed it stops.

RISE is very different from the previous ones. It is an instance-based algorithm, which starts considering each instance within the training set as a possible rule. Iteratively, it generalizes the most similar instances (i.e., rules) making new rules, always more general. A similarity measure is needed. Several similarity measures have been implemented. It progresses until no more accuracy gain is obtained with a generalization step.

4.2 Statistical Modelling agent

The multivariate descriptive techniques allow studying the structure of a given domain. Further, it is convenient to properly formalize the model in order to use it in the future, either with descriptive or predictive purposes. When those models are formalized under a mathematical paradigm and taking into account the uncertainty, statistical modelling is on. It establishes algebraic relations between a response variable and a set of explicative variables (attributes, regressors) in such a way that knowing the values of the explicative variables for a certain instance, the response value can be determined with a known precision.

The statistical model produces quantitative and formal results about relationships between variables and it complements the qualitative and non-formal results from descriptive analysis.

This agent is in charge of building different statistical models, depending on the cases:

• Multiple linear regresion, which allows to relate a numerical response with a set of numerical or categorical (qualitative) regressors;
• ANOVA, involved with explaining a categorical response on the basis of a set of regressors;
• Logistic regression, explaining a dicotomic variable by a set of regressors.

For each method, the system is implementing the following steps:

- Parameters estimation for the model
- Providing goodness of fitting coefficients
- Providing tools for a graphical residuals analysis, in order to validate the model.

5. AN APPLICATION

5.1 The data

The main goal of wastewater treatment plants is to guarantee the outflow water quality (referred to certain legal requirements), in order to restore the natural environmental balance, which is disturbed by industry waste or domestic wastewater.

The process used to achieve this goal is really complex and delicate; on the one hand, because of the intrinsic features of wastewater; on the other hand, because of the bad consequences of an incorrect management of the plant (Gime98).

Data analyzed in this paper comes from a Waste Water Treatment Plant in Catalonia (in Spain). Here is a brief description of the plant performing (see fig. 3): the water flows sequentially through several processes; in the *pretreatment*, an initial separation of solids from wastewater is performed; *primary treatment* consists of leaving the wastewater in a settler for some hours; solids will deposit down the settler and could be sent out; *secondary* treatment occurs inside a biological reactor: a (biomass) population of microorganisms degrades the organic matter dissolved in the wastewater; in the studied particular plant there are two separate biological reactors with a second settler between them (double stage activated sludge plant or AB process); in the *advanced treatment* another settler is used to separate the water from the biomass and water is clean and ready to exit the plant. The settler output (solids or biomass) produces a kind of mud which is the input of another process called *sludge line*.

Database is a sample of 149 observations taken between January and May 2002. Each observation refers to a daily mean. The state of the Plant is described through a set of 18 variables (or attibutes) which can be grouped as (see fig. 3):

• Input (measures taken at the plant entrance): *Q-E:* Inflow wastewater (daily m^3 of water); *DQO-E:* Oxigen chemical demand (mg/l); *MES-E:* Suspended Solids (mg/l); *P-E:* Phosphates (mg/l).
• After Settler (measures taken when the wastewater comes out of the first settler): *DQO-P:* Oxigen chemical demand (mg/l); *MES-P:* Suspended Solids (mg/l);
• Biological treatment 1 (in 1st biological reactor): *MLSS-1:* Mixed liquor suspended solids (mg/l); *IVF-1:* Volumetric index (ml/g); *CM-1*: Organic load (Kg DBO/ Kg MLSS).
• Biological treatment 2 (in 2nd biological reactor): *MLSS-2*: Mixed liquor suspended solids (mg/l); *IVF-2*: Volumetric index (ml/g); *CM-2*: Organic load (Kg DBO/Kg MLSS); *T-2*: Temperature (C°);

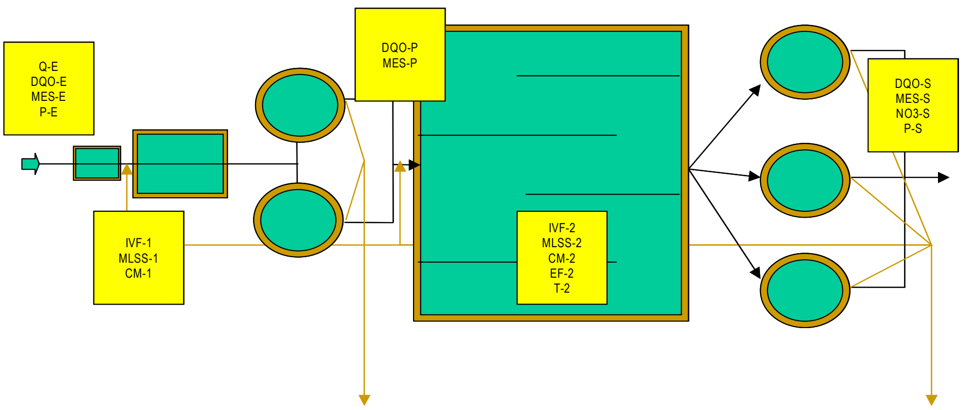


Figure 3. A wasetwater treatment plant chart

EF-2: Sludge Residence time (days).

- Output (when the water is meeting the river): *DQO-S*: Oxigen chemical demand (mg/l); *MES-S*: Suspended Solids (mg/l); *NO3-S*: Nitrates (mg/l); *P-S*: Phosphates (mg/l).

5.2 Descriptive analysis and data filtering

Facilities of Statistical Data Processing Agent were used for identifying outliers and missing data, and properly manage them. Descriptive analysis was also useful to identify some probabilistic models for some variables, which will determine the later statistical model construction.

5.3 Clustering

In order to identify the characteristic situations presented in the plant a clustering process was performed using Clustering based on rules (ClBR) (Gibert 1996), which is an ascendant hierarchical method that permits, among others, conjoint considering of numerical and categorical variables to identify the clusters. As usual in hierarchical clusters, the number of classes is determined a posteriori upon the hierarchy built by the method. As a result 12 classes were identified.

With the descriptive analysis agent, conditional distributions of different variables through classes could be studied. Fig. 4 shows first that of MLSS-1 (solids concentration in the reactor) along the classes: Classes were grouped in two main blocks, that where first step is operating as a properly biological reactor (with values of MLSS-1, greater than 3000 mg/l, classes 1 to 6), and that where it is working as preaeration (with low values of MLSS-1, classes 7 to 12), according to the special performing conditions of the studied plant, which indeed has a first stage that can operate in those two ways. Going further with such graphics for the rest of the variables (for instance column CM-1, fig. 4), differences among classes 1 to 6, on the one hand, and/or 7 to 12 on the other can be studied, what makes possible a first interpretation of classes. However, it is interesting to see how an induction rule method can also discover the meaning of the classes in an automatic way, confirming what it can already be seen here using only descriptive techniques (see 5.4).

5.4 Rule induction

After determining and evaluating the reliability of the clustering process, a new step to induce classification rules was performed. The rule induction agent used the class identifier obtained by the clustering agent as the input label of the instances. In this application several inductive methods were ran in order to induce rules that could be later used for recognizing the class of a new instance. Also, by analysing the meaning of obtained rules, interpretation of classes will be clearer. Some parameters were tested and tuned, such as the number of intervals in the discretization of the continuous attributes, done by trial and error.

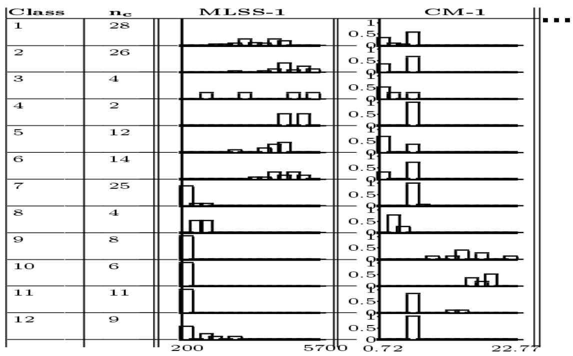


Figure 4 Multiple boxplot

Upon the experts criteria, the best set of rules, in terms of making explicit the knowledge about plant performing, is that produced by Prism. So, a final set of 18 rules is proposed as the best. Fig. 2 shows GESCONDA's rule induction process. It is worth noticing that some rules addressed concrete classes: r2 identified class 3, which corresponds to an initial bulking situation, and r3 identified cluster 8, which corresponds to nitrifying situation, although there are classes predicted by more than one rule. Thus, experts concluded that the inductive rules and the clusters previously obtained were a really useful and coherent knowledge discovered from the available data.

Rules generated by PRISM are totally accurate, but the coverage is not so good, and some instances could not be classified with the rule set. However, experts found the rules representative enough.

5.5 Final interpretation of the classes

As said before, in the first six classes the first stage of the plant works as biological reactor, while in the others as preaeration. Here is the final interpretation of the classes, which was built together with the experts, by combining the conditional distributions of all the variables along classes (fig. 4) with the rules induced by Prism:

- First stage operating as biological reactor:
  - *Class 1:* it is the common situation. This class is labelled as normal or correct plant operation.
  - *Class 2:* it is quite similar to class 1 but with optimal operation (better effluent quality).
  - *Class 3:* abnormal situation owing proliferation

of filamentous bacteria (bulking) in first stage, difficulting sludge settleability.
  - C*lass 4*: refers to those days with higher loading rates (organic overloading).
  - *Classes 5, 6*: bulking (filamentous organisms proliferation) episodes in the second stage
- First stage operating as preaeration:
  - *Classes 7 and 12*: common situation.
  - Class 8: periods of partial nitrification due to the growth of autotrophic biomass.
  - *Class 9*: rainy and stormy days, with a nutrient desequilibrium suitable for viscous bulking .
  - *Classes 10, 11*: viscous bulking, not associated to exceeding proliferation of filamentous organisms. Mainly related to *Zooglea Ramigera*. This commonly occurs after rainy periods, associated with nutrient disequilibrium

6. CONCLUSIONS

The main conclusion is that GESCONDA is an Intelligent Data Analysis System, which offers a common interface to the user for using a set of different tools that helps his/her decision-making processes. From different real applications, it has been seen that this is a very promising approach and the previous partial experiences on this line suggested great benefits making it. Currently, the statistical data-filtering agent, the clustering agent, the decision trees induction agent and the rules induction agent are completed; the statistical modelling agent is including multiple linear regression and ANOVA (on and two way), and logistic regression is in progress.

Main agents are already built and the schedule of the project was correctly followed. At present, renovation for the next three years is pending, in order to face the remaining agents development (support vector machines or dinamical analysis). Validation of the current version of the system using real databases is on. The close assessment of the environmental engineers of LEQUIA, and SOREA people guarantee usefulness of the system.

From the presented application, it can be said that WasteWater Treatment Plants constitutes a complex domain which requires complex analysis using different approaches in order to extract useful knowledge. ClBR appeared to be a good method for identifying typical situations in that domain (like bulking or storming days). Exploratory techniques, such as displaying conditional distributions of the variables vs classes are of great help to understand the meaning of the classes. However, the possibility offered by GESCONDA of combining the results of a clustering process with rules induction, made much more easier the interpretation and allowed consolidation of the discovered knowledge. An integrated tool like the proposed one allows facing the analysis of phenomena, like WWTP, were knowledge is not well-stablished yet and permits knowledge discovery in a friendly way.

From this results, an initial case base has already been built to be included in the supervisory system of a real plant which is using case based reasoning. As a complementary study, statistical models for predicting the class of a new observation on the basis of the variables identified as relevants in the rules induction process will provide a quantitative model useful for bounding the prediction error rate.

6. ACKNOWLEDGEMENTS

The authors wish to thank the partial support provided by the Spanish CICyT project TIC2000-1011, and by the EU project A-TEAM (IST 1999-10176). Thanks to Francesc Coll and M. Eugenia Garcia for partially developing a piece of software.

7. REFERENCES

Cendrowska,. J 1987. PRISM: an algorithm for inducing module rules. *Int'l Journal of Man-Machine Studies* 27(4):349-370.

Clark & Niblett, 1989. P. Clark & T. Niblett. The CN" induction algorithm. *ML* 3:261-283.

Comas J., Dzeroski S. Gibert, K., Rodríguez-Roda I. and Sànchez-Marrè M. 2001 Knowledge Discovery by means of inductive methods in WWTP data. *AI Comm* 14(1):45-62.

Domingos, P 1996 Unifying Instance-Based and Rule-Based Induction, *ML* 24(2):141-168.

Gibert K. 1996 The use of symbolic information in automation of statistical treatment of ISD. *AI Communications* 9(1) 36—37 IOS

Gibert K., T. Aluja and U. Cortés, 1998 Knowledge Discovery with Clustering Based on Rules. Interpr... *LNAI* 510:83-92, Springer

Pham & Aksoy, 1995 D.T. Pham & M.S. Aksoy. RULES: a simple ruler extraction system. *Expert Systems with Applications* 8(1).

Rodríguez-Roda I., Comas J., Colprim J., Poch M., Sànchez-Marrè M., Cortés U., Baeza, J. & Lafuente (2002) *J. Water Science & Technology*, 45(4-5), pp. 289-297

Sànchez-Marrè M., Gibert, K., Rodríguez-Roda, I., et al. (2002) Development of an Intelligent Data Analysis System for Knowledge Management in Environmental Data Bases. First Int'l Conf. of Int'l EMS Society. Vol III 420-425. Lugano, Suiza.

Reduction of complex models using data-mining and nonlinear projection techniques

Bernhardt, K. [a], Wirtz, K.W.

[a]Institute for Chemistry and Biology of the Marine Environment (ICBM)
Carl-von-Ossietzky University Oldenburg
P.O. Box 2503
26111 Oldenburg, Germany

Abstract: Complex models of environmental systems typically depend on a large amount of uncertain parameters. Therefore, they are often difficult to handle and do not provide an insight into effective modes of the underlying system's dynamics. Unlike earlier analytical attempts to find more effective model representations, we present a new combination of methods that only relies on data generated by complex, process-based models. These methods are taken from the field of data-mining and enable the recognition of patterns in measured or modelled data by unsupervised learning strategies. As these methods do not directly lead to a better understanding of the systems' driving processes, we suggest the linkage between pattern recognition and process identification by a multi-stage approach. In a first step, a large data-base was produced by a mechanistic model for species competition in a virtual ecosystem for a range of parameter settings. Using Vector Quantisation and nonlinear projection techniques such as Self-Organising Maps and nonlinear Principal Component Analysis, typical states of the complex model's dynamics as well as major pathways connecting these states were then identified. The visualisation of the results points to the existence of nonlinear transformations of former model state variables and parameters to few effective variables. Effective variables built this way preserve most of the model's dynamic behaviour, while they are nonetheless easier to use and require much less parameterisation effort.

Keywords: model reduction; data-mining; Self-Organising Map; nonlinear projection; multispecies competition

1 INTRODUCTION

Process-based models are widely used for modelling key mechanisms ruling ecosystem dynamics. The vast number of potentially relevant interactions and adaptations in ecosystems thereby increases the corresponding model complexity. Secondly, process identification is rarely unique, i.e. data can be reproduced with a variety of models or parameterisations (see e.g. Beven [2001]). Given this high complexity and the sparseness of data, parameter uncertainty is difficult to handle in these models.
An alternative way to reproduce and extrapolate data is the use of methods taken from the field of data-mining, such as Neural Networks (NN), clustering methods or (non-)linear projection techniques which are able to 'learn' distinct features of a dataset (Fodor [2002]). No knowledge of the underlying system is required using data-driven methods. New understanding of underlying key mechanisms however, can not be gained and generic models with a large domain of applicability are not provided.
The aim of this work is the construction of a new type of deterministic but reduced and efficient model from a classic complex mechanistic model by only using information contained in the data generated by the complex model. A nonlinear statistical analysis and reduction of this data should reflect the overall dynamics even for uncertain model parameterisations and should yield interpretable information on dominant internal modes. We propose a multi-step analysis using data-mining and nonlinear projection techniques to extract these modes or 'effective variables' (Wirtz and Eckhardt [1996]). Those variables have been shown to successfully replace complex descriptions of adaptive processes in biosystems (e.g.Wirtz [2002]). Up to now they had to be built using intuitive modelling knowledge which is a major impediment for a broader use.
The recombination of the effective variables resulting in a reduced-form deterministic model consequently combines the benefits of the process-

oriented as well as data-mining approaches. The existence of such a reduced representation is supported by the finding, that even huge ecosystem models have a limited number of internal dynamical modes (Ebenhöh [1996]).
In principal, the proposed reduction scheme can be applied to any deterministic process-based model. In this study, we present the extraction of effective variables using a combination of Vector Quantisation algorithms such as the Self-Organising Map (Kohonen [1997]) and nonlinear Principal Component Analysis (Kramer [1991]).
We have chosen the reduction of a prominent model of multispecies competition with rich dynamics including chaotic behaviour (Huisman and Weissing [1999]) as a test case.

2 A MODEL OF SPECIES COMPETITION

The model analysed in this study was proposed by Huisman and Weissing [1999]. It describes competition for resources like phytoplankton species concurring for nitrogen and phosphorus.
Consider n_P phytoplankton species and n_N nutrients. Let state variables $\hat{P}_i$ and $\hat{N}_j$ be the population abundance of species i and the availability of resource j respectively. The dynamics of species i follows

$$\frac{d\hat{P}_i}{dt} = \hat{P}_i \cdot (\mu_i - \omega_i) \qquad i = 1, \ldots, n_P \tag{1}$$

where ω_i are model parameters describing the mortality. The growth rate μ_i is controlled by the most limiting resource via a minimum of Monod functions with K_{ji} denoting the half-saturation constant for resource j of species i and g_i the maximal growth rate:

$$\mu_i = \min_v \left(\frac{g_i \hat{N}_v}{K_{vi} + \hat{N}_v} \right) \qquad v = 1, \ldots, n_N. \tag{2}$$

The time evolution of the abiotic resource j is described as

$$\frac{d\hat{N}_j}{dt} = D \cdot \left(S_j - \hat{N}_j \right) - \sum_i c_{ij} \cdot \mu_i \cdot \hat{P}_i \qquad j = 1, \ldots, n_N \tag{3}$$

where D is a constant factor describing the nutrient turnover rate, S_j is the supply concentration and parameters c_{ij} quantify the content of nutrient j in species i.
For different choices of model parameters the system can be driven into attractors with different topologies containing fixed-point dynamics (no changes in species abundances for one or more species), limit cycles (fluctuating coexistence of species) or chaotic behaviour. For further details on the parameter settings see Huisman and Weissing [1999, 2001].
To keep the analysis simple, we numerically integrated (1) and (3) to produce 16 time series of 2000 points each for a model configuration with five species ($n_P = 5$) and three abiotic resources ($n_N = 3$) by varying only two of the half-saturation constants (k_{21} and k_{25}). The other model parameters (D, S_j, ω_i, g_i and c_{ij}) were kept at the fixed values used in Huisman and Weissing [2001]. Time series modelled this way show all sorts of dynamics described above (see Figure 1).

3 THE SELF-ORGANISING MAP

The Self-Organising Map (SOM) algorithm was introduced by Kohonen [1997]. It resembles a neural network variant consisting of topologically ordered nodes on a grid of predefined dimensionality.
A SOM is able to 'learn' structures of high-dimensional input data-vectors and to project them onto a lower-dimensional output space. It is therefore often used for Vector Quantisation (VQ) where a reduced representation of complex datasets is built by replacing the data-vectors with a smaller subset of so-called prototype vectors. Additionally, the existence of the typically two-dimensional output grid simplifies the visual inspection of the dataset and helps to identify patterns inherent to the data.
The algorithm transforms a dataset consisting of vectors $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T \in \Re^n$ with discrete-time coordinate $t = 0, 1, 2, \ldots$, e.g. measurements of n variables over time. In this case, each $\mathbf{x}(t)$ is a ten-dimensional vector with entries $\hat{P}_i$, $\hat{N}_j$, k_{21} and k_{25} ($n = n_P + n_N + 2$). The SOM-network consists of a z-dimensional array of k nodes associated with prototype-vectors $\mathbf{m}_k \in \Re^n$ with orthogonal or hexagonal neighbourhood relationships between adjacent nodes.
The data-vectors are iteratively compared with all $\mathbf{m}_k$ by using euclidean distances to find the best-matching node denoted by c. The updating procedure for prototype s then follows

$$\mathbf{m}_s(t+1) = \mathbf{m}_s(t) + h_{cs} \cdot [\mathbf{x}(t) - \mathbf{m}_s(t)], \tag{4}$$

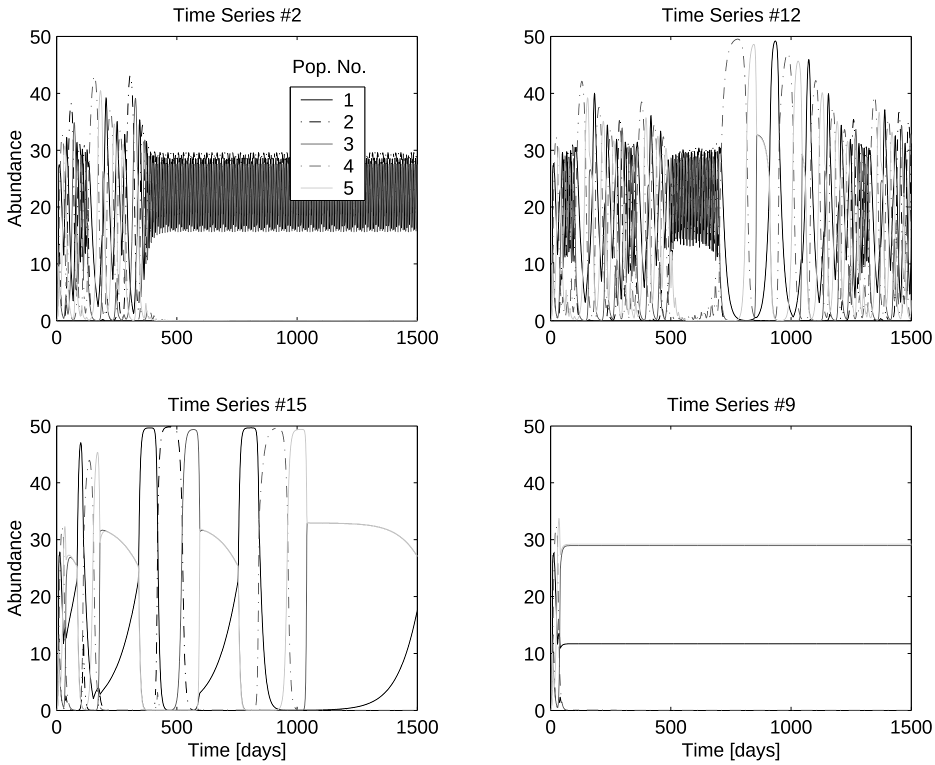


Figure 1: Time series of $\hat{P}_i(i = 1, \ldots, 5)$ created by parameter variation of the test model. Parameter settings $\{k_{21}, k_{25}\}$ are $\{0.2, 0.325\}$ for time series #2, $\{0.275, 0.4\}$ for time series #12, $\{0.325, 0.35\}$ for time series #15 and $\{0.275, 0.275\}$ for time series #9.

where h_{cs} is a neighbourhood function that asserts the convergence of the algorithm for $h_{cs} \rightarrow 0$ when $t \rightarrow \infty$. Mostly, the Gaussian function in dependence of $\|r_c - r_s\|$ is used, where $r_c \in \Re^z$ and $r_s \in \Re^z$ are the location vectors of nodes c and s. Additionally, h_{cs} is multiplied by the learning-rate factor $\alpha(t) \in [0, 1]$ that decreases monotonously over time to prevent the distortion of already ordered parts of the map at later time steps.

In measuring the quality of the SOM-mapping (see e.g. Bauer and Pawelzik [1992]; Villmann et al. [1997]) a compromise has to be made between an optimised reproduction of the data vectors and the minimisation of the topological distortion by neighbourhood violations. In this work the SOM-Toolbox 2.0 package (Vesanto et al. [1999]) was used that calculates the average quantisation error and the topographic error (Kiviluoto [1996]). The best network of different map configurations was assumed to minimise the sum of these two measures. This procedure tends to find solutions overfitting the dataset but this drawback was accepted as the details of the VQ step were found to be of minor importance for the following analysis.

3.1 Vector quantisation of the dataset

In advance of the analysis the data-matrix was standardised by mean and standard deviation of the individual variables ($\hat{P}_i \rightarrow P_i$). To incorporate information about control parameters of the competition model into the learning procedure, the constant time series of k_{21} and k_{25} were added as additional variables to the training dataset.

SOM networks of different map configurations were trained and the best network with 50 x 50 prototype vectors was found to explain 96.2% of the data variance.

4 NONLINEAR PROJECTION

Even though the SOM itself represents a kind of nonlinear projection technique it is not very well suited for the extraction of distinct modes of the underlying dynamics as the vectors spanning the SOM network can not be interpreted in terms of variable model entities. This limitation also exists for other unsupervised learning strategies that construct relevant topological constraints directly from the data (e.g. Martinetz and Schulten [1991]; Baraldi and

Alpaydin [2002]). Hence, the need for finding 'directions' along which features of the system vary continuously remains. A promising technique to extract these effective variables is nonlinear principal component analysis (NLPCA) put forward by Kramer [1991].
The NLPCA relies on an autoassociative feedforward neural network as depicted in Figure 2. It projects data-vectors $\mathbf{x}(t)$ onto a so-called bottleneck layer u and compares the decoded vectors $\mathbf{x}'(t) = (x'_1(t), x'_2(t), \ldots, x'_n(t))^T$ with the input data to minimise the cost function $J = \langle \|\mathbf{x}(t) - \mathbf{x}'(t)\|^2 \rangle$.

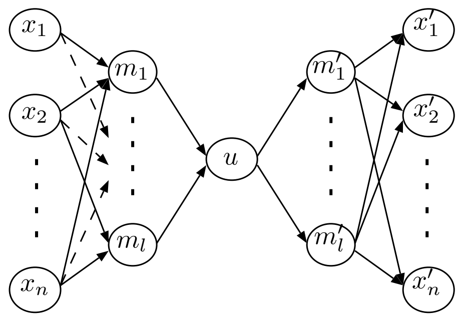


Figure 2: Example for an autoassociative neural network with l nodes $m_1, \ldots, m_l$ in the first and $m'_1, \ldots, m'_l$ in the second hidden layer.

The mappings $\mathbf{x} \rightarrow \mathbf{m}$ and $\mathbf{m}' \rightarrow \mathbf{x}'$ are typically performed by nonlinear transfer functions (e.g. the sigmoidal function), whereas mappings from and to the nonlinear principal component $\mathbf{u}$ use the identity function. The number of nodes in the hidden layers of the network determines the approximation quality of the data.
Typical problems arising during neural networks training are overfitting and local minima in the cost function J. In our analysis we employ the NeuMATSA (Neuralnets for Multivariate and Time Series analysis) package (Hsieh [2001]) where multiple runs and penalty terms for the network weights smooth the nonlinear responses of the transfer functions to obtain results less sensitive to local minima. Only by the data reduction of the preceding SOM analysis the NLPCA step is made applicable. Thus, an immense speed-up of the minimisation of the neural networks' weights is gained and the already smoothed SOM representation additionally accounts for the avoidance of local minima in the cost function.

4.1 NLPCA of the SOM-filtered data

To prevent the NLPCA from overfitting, 20% of the SOM-filtered dataset were chosen randomly as test-dataset and ensembles of 25 runs were selected for configurations of nodes in the hidden layers ranging from one to five. The analysis was terminated when the quality of the mapping as quantified by the mean squared error (MSE) for the test set decreased subsequently to an initial rise.

After extraction of the first nonlinear PCA, further components were iteratively found by subtracting earlier solutions from the SOM dataset and by repeating the analysis using the residuals.
The first nonlinear mode found this way explained 61%, whereas the second and third mode accounted for 16.5 and 2.7% of the SOM networks' variance, respectively. Thus, the dataset can be assumed to be essentially two-dimensional and the first two nonlinear modes extracted by NLPCA can be interpreted as effective variables (EV) of the underlying model. Figure 3 shows the first two modes in state space $\{P_1, P_2\}$. Clearly, variation of the original model variables is constrained indicating implicit model trade-offs and the existence of a reduced EV representation.

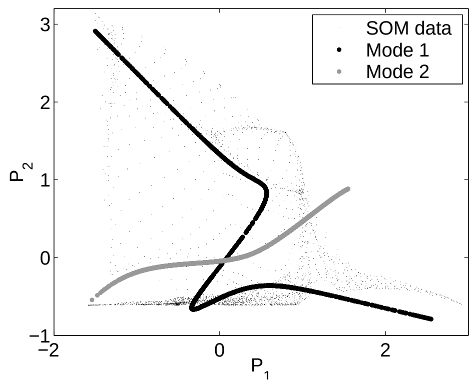


Figure 3: Nonlinear PCA modes 1-2 in a VQ state subspace with dots representing SOM-filtered model data.

4.2 Interpretation of modes

Figure 4 shows successive segments of an example time series projected into the EV space. The typical cycles with varying periods found in the dataset (see Figure 1) are clearly separated from each other and resemble the form of limit-cycles in the complex model's phase-space. Successive changes between these cycles illustrate the ability of the method to separate dynamical states.

To further investigate and interpret the effective nonlinear modes in terms of former model variables, the projected data was aggregated into bins of equal size with a minimum of five datapoints per class. Figure 5 shows the class distributions of the first mode.

The smooth course of the distributions together with the relatively small inner class variability even for densely covered bins (e.g. small positive values of the first nonlinear PCA) may enable a meaningful interpretation of the nonlinear modes. For example, a comparison of the chaotic transition (from small negative to small positive values of PCA 1 in Figures 4 and 5) for P_i provides an insight into the particular case when species coexist. This type of coexistence can thus be imagined as an occupation of 'dynamically separated' niches.

5 DISCUSSION

Improvements of the methodical parts of this work, as discussed in section 4 for the SOM algorithm, can be thought of. As outlined in Malthouse [1998], NLPCA solutions for the projection problem are only suboptimal and alternatives like the Principal Curves approach of Hastie and Stuetzle [1989], for example, can be tested as well. In this work however, the projection discrepancy does not constrain the usefulness of NLPCA as smooth solutions following mean features of the dataset are explicitly requested.

First outcomes of this work show that a combination of Vector Quantisation and nonlinear projection can already provide valuable insights into the dynamics underlying process-oriented models. The extraction of relevant nonlinear modes describing a model on a higher or aggregated level is a first step towards effective variable models that are easier to use and better to interpret than their complex model equivalents.

The results shown here point to the existence of non-linear but nonetheless simple transformations of former model state variables and parameters to effective variables. The projections of quantised model data along the first two nonlinear modes, from which only a subset is presented in Figure 5, already support a piecewise linear transformation from the original space of model entities to new aggregated variables. In future studies reduced-form models will be formulated using effective variables as provided by the approach put forward in this study. We will thereby rely on results and techniques presented here comprising (i) the simultaneous incorporation of model outcomes and varied model coefficients into the analysis, (ii) internal trade-offs between model variables for different attractors of the model dynamics and (iii) the smoothness of the projections of a small set of effective variables to the original model space. The extraction of these nonlinear transformations constitutes an analytical means to interpret the nonlinear principal components in terms of simulated processes as an essential step towards a reduced-form representation of complex, mechanistic models.

REFERENCES

Baraldi, A. and E. Alpaydin. Constructive feedforward art clustering networks - part ii. *IEEE Transactions on Neural Networks*, 13(3):662–677, 2002.

Bauer, H.-U. and K. Pawelzik. Quantifying the neighborhood preservation of self-organizing feature maps. *IEEE Transactions on Neural Networks*, 3(4):570–579, 1992.

Beven, K. On modelling as collective intelligence. *Hydrological Processes*, 15:2205–2207, 2001.

Ebenhöh, W. Stability in models versus stability in real ecosystems. *Senckenbergiana Maritima*, 27: 251–254, 1996.

Fodor, I. A survey of dimension reduction techniques. Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory, 2002.

Hastie, T. and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84(406): 502–516, 1989.

Hsieh, W. Nonlinear principal component analysis by neural networks. *Tellus*, 53A:599–615, 2001.

Huisman, J. and F. J. Weissing. Biodiversity of plankton by species oscillations and chaos. *Nature*, 402:407–410, 1999.

Huisman, J. and F. Weissing. Fundamental unpredictability in multispecies competition. *The American Naturalist*, 157(5):488–493, 2001.

Kiviluoto, K. Topology preservation in self-organizing maps. In *Proceedings of IEEE International Conference on Neural Networks*, volume 1, pages 294–299, 1996.

Kohonen, T. *Self-organizing maps*. Springer, Berlin, 2nd edition, 1997.

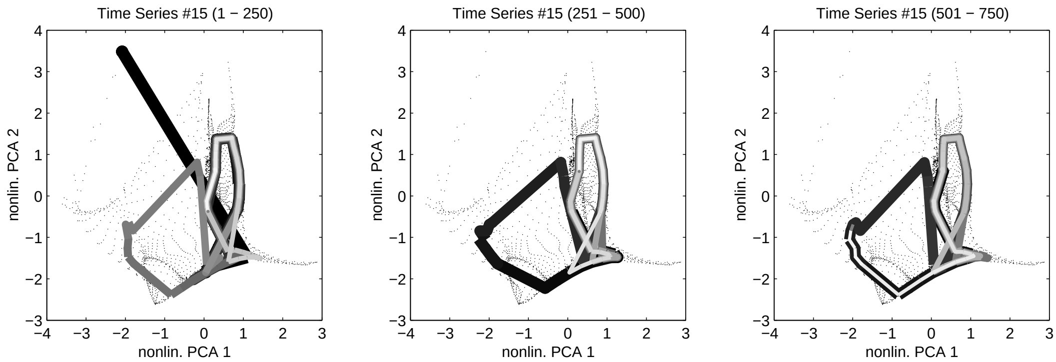


Figure 4: Plot of time series examples in EV space (spanned by the first two NLPCA modes). Shown are the first 750 time steps of series #12. Earlier time steps are drawn in thick dark and later ones in thin light lines.

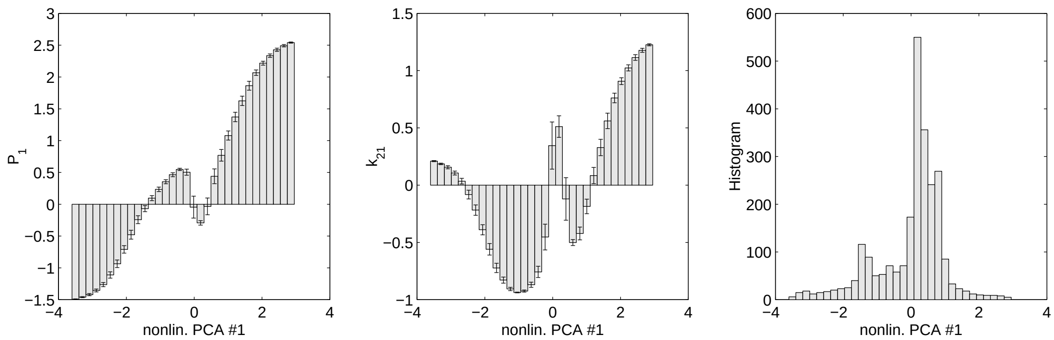


Figure 5: Aggregation of the projected data along the first nonlinear mode together with a histogram of bin occupancy.

Kramer, M. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2):233–243, 1991.

Malthouse, E. Limitations of nonlinear pca as performed with generic neural networks. *IEEE Transactions on Neural Networks*, 9(1):165–173, 1998.

Martinetz, T. and K. Schulten. A ”neural-gas” network learns topologies. In *Artifical Networks*, pages 397–402. North-Holland, Amsterdam, 1991.

Vesanto, J., J. Himberg, E. Alhoniemi, and J. Parhankangas. Self-organizing map in matlab: the SOM toolbox. In *Proceedings of the Matlab DSP Conference*, pages 35–40, Espoo, Finland, 1999.

Villmann, T., R. Der, M. Herrmann, and T. Martinetz. Topology preservation in self-organizing feature maps: exact definition and measurement. *IEEE Transactions on Neural Networks*, 8(2): 1291–1303, 1997.

Wirtz, K. A generic model for changes in microbial kinetic coefficients. *Journal of Biotechnology*, 97:147–162, 2002.

Wirtz, K. and B. Eckhardt. Effective variables in ecosystem models with an application to phytoplankton succession. *Ecological Modelling*, 92: 33–53, 1996.

Circadian Patterns Recognition in Ecosystems by Wavelet Filtering and Fuzzy Clustering

Stefano Marsili-Libelli and Simone Arrigucci

Dept. of Systems and Computer Engineering, University of Florence
Via S. Marta, 3 - 50139 Florence, ITALY
Email:marsili@ingfi1.ing.unifi.it

Abstract: This paper presents a method for extracting representative patterns from a set of data representing circadian cycles. The analysis is based on a combination of wavelet filtering and fuzzy clustering. The data are first processed with a discrete wavelet decomposition in order to filter out the noise and isolate the relevant circadian cycle. It is shown that the second level decomposition yields the best cycle approximation, filtering out measurement noise and other artefacts and preserving the main cycle features. From the filtered data the following discriminating features are extracted: minimum and maximum daily values, and the slope of the line passing through these extreme points. These features are then processed with a fuzzy clustering algorithm, in order to isolate significant behaviours. The Fuzzy Maximum Likelihood Estimates (FMLE) method was used for its variable metric, being able to conform the cluster shape and volume to the data. This combined algorithm is applied to the physico-chemical data from the Orbetello lagoon with the aim of detecting ecologically meaningful behaviours. The results show that relevant daily patterns are indeed isolated and in particular combination of variables leading to the dystrophic crisis are correctly interpreted. The relevance of the selected patterns is confirmed by their distribution over the calendar day, which corresponds to a clear seasonal patterns.

***Keywords**:* Wavelet filtering; Fuzzy clustering; knowledge-based systems; artificial intelligence;

1. INTRODUCTION

Ecosystems are subject to fluctuations with a wide range of periods, from the short-term random variations to daily cycles and seasonal changes. Of all these, the diurnal cycle is the most important, being related to the day-night sequence which drives most ecological process. Daily patterns are a major feature of ecosystems and their seasonal changes may reveal important information regarding its functioning and, if properly interpreted, can represent a valuable tool in ecosystems forecasting and management.

A previous study of the circadian cycles (Marsili-Libelli, 2004) used a fuzzy method for the automatic pattern recognition, but its weak point was the heuristic choice of the model patterns. To complement the previous algorithms, this paper proposes a new method to construct a consistent, objective knowledge-base of significant patterns. Circadian cycles are daily fluctuations in biological activities.

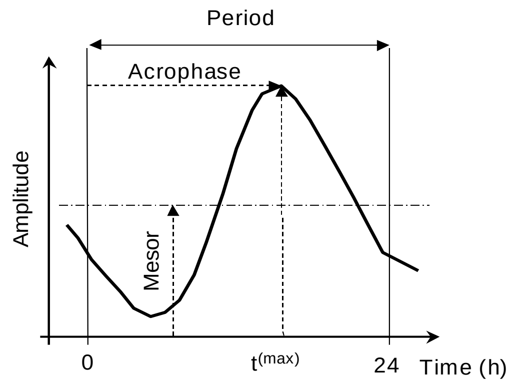


Fig. 1. Basic features of the circadian cycle.

They are the consequence of the daily variations of the solar radiation and follow the general pattern of Figure 1. The important parameters of a circadian cycle are the maximum value (acrophase), its timing and the mean daily value (mesor). The basic idea is to extract the underlying circadian pattern from noisy data using wavelet filtering and partition these approximations using a fuzzy clustering method. The result of this combined procedure is a number of "typical" patterns, which can be used as the required knowledge-base.

The reasons for applying this two-step procedure are now examined:

1. Data are normally affected by noise in many ways. In order to extract the information related to the circadian pattern, a wavelet decomposition is performed. This filtering technique uses windows of variable size and is capable of performing a joint time-frequency analysis revealing aspects of data that other signal analysis techniques miss, especially if the data record is short. It can also de-noise the data without appreciable degradation, separating components of different scales: low-frequency, which are usually ecologically meaningful, and high-frequency disturbances.
2. Once the basic patterns are isolated, significant features are extracted and grouped into a number of meaningful behaviours through fuzzy clustering. In particular the Fuzzy Maximum Likelihood Estimates (FMLE) algorithm is used for its ability to produce clusters of varying volume and shape (Babuska, 1998). This attractive feature is a consequence of being based, like the well-known Gustafson and Kessel method, on an adaptive distance norm derived from the fuzzy covariance matrix.

The paper may be regarded as a sequel to a previous one (Marsili-Libelli, 2004) where the knowledge-base was constructed with a mixed statistical-fuzzy technique. The method is demonstrated with reference to daily variation of dissolved oxygen (DO) data from a eutrophic lagoon, described in the next section.

2. PATTERN DETECTION IN THE ORBETELLO LAGOON DATA

The Orbetello lagoon, schematically shown in Figure 2, is located along Italy's west coast. It consists of two shallow coastal reservoirs with a combined surface of approximately 27 km^2, an average depth of 0.8 m. Two water-quality monitoring stations, indicated by the two circles in Figure 2, transmit hourly physico-chemical data to the Orbetello Lagoon Managerial Office headquarters. These data include Dissolved Oxygen (DO), Oxido-Reduction Potential (ORP), pH and temperature.

The submersed vegetation is composed of macroalgae *(Chaetomorpha linum, Cladophora vagabunda, Gracilaria verrucosa, Ulva rigida)* and macrophytes *(Ruppia maritima).* Given the large availability of nutrients and the limited water renewal, when the macroalgae decompose after an excessive growth, an oxygen imbalance may occur, causing anoxia.

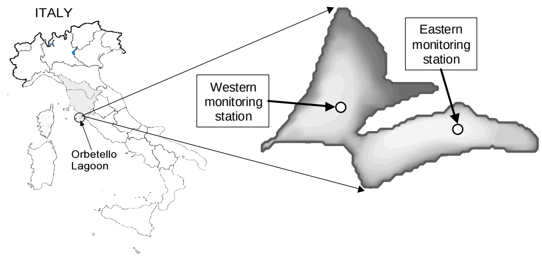


Fig. 2. The Orbetello lagoon with the location of the two monitoring stations.

This kind of dystrophic crisis is well studied (Christian *et al.*, 1998). During the normal growth phase, macroalgae represent a sink for dissolved inorganic nitrogen and oxidising processes prevail: day-time dissolved oxygen (DO) has a well defined afternoon peak, often well above the saturation level. When the growth phase ends, the fast anoxic decomposition enriches the sediment with reduced organic nitrogen (Christian *et al.*, 1998). These reducing conditions can be detected by low, almost constant DO and low daytime oxidation-reduction potential (ORP) in the water.

Thus the daily cycle of these variables contains important season-dependent information, which this method attempts to extract from noisy data using the two-step procedure outlined in Section 1. The algorithm is organised in a hierarchical structure, where DO is the primary variable from which the clustering features are extracted. The secondary variables ORP, pH and temperature are processed at a later stage, depending on the DO results and are used to reveal finer details in a cluster structure.

3. WAVELET PREPROCESING

Frequency-based techniques have been used extensively for data filtering, but their drawback is that the time-information is completely lost. This is a serious shortcoming if the signal is non-stationary, because trends, bursts and specific one-time events may be missed. This is due to the use of infinite functions, usually sinusoids, as the basis functions. The question then arises why not choose a basis function that has a finite duration, instead of choosing an infinite-duration one? Wavelets (Strang and Nguyen, 1996; Torrence and Compo, 1998) are finite-duration signals which can be used to replace sinusoids as basis functions for filtering. In this way time and frequency analysis can be combined through a variable windowing technique and local analysis of short-duration events can be performed without losing the power of frequency analysis.

Given a wavelet function $\psi(a,b,t)$ where a and b represent the scaling and the shifting factors, the continuous wavelet transform (CWT) is defined as the integral of the signal *s(t)* multiplied by the scaled wavelet $\psi(a,b,t)$

$$C(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t)\psi\left(\frac{t-b}{a}\right)dt \qquad (1)$$

Given the finite time duration of $\psi(a,b,t)$ the integration limits in eq. (1) can in practice be limited. In CWT a and b vary continuously. If the wavelet is stretched $(a >> 1)$ it contains mostly low frequencies and gives a global view, whereas compressed wavelets $(a << 1)$ contain mostly high frequencies and give details view of a small portion of the signal *s(t)*. Thus by varying a the analysis can be concentrated on global or specific features of the signal. However, computing Eq. (1) for any combination of *(a,b)* would be greatly time-consuming with little insight into the process generating the signal. Restricting the CTW to dyadic scaling, i.e. $\left(a^{-\frac{T_s}{2}}, a^{-\frac{T_s}{4}}, a^{-\frac{T_s}{8}}, ..., a^{-\frac{T_s}{2n}}\right)$ the Discrete Wavelet Transform (DWT) is obtained, based on powers of two of the base window T_s, which in our case represents the circadian period of 24 h. DTW yields a signal hierarchical decomposition into *Approximations*, grouping the high-scale, low-frequency components of the signal, and *Details*, retaining the low-scale, high-frequency parts.

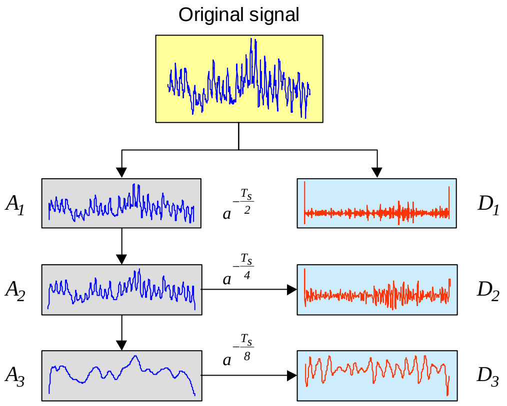


Fig. 3. Discrete-time wavelet multilevel decomposition tree.

This process can be iterated in the scale $a^{-\frac{T_s}{2k}}$, as shown in Figure 3, halving the window T_s at each step in order to get increasingly low-frequency, long-time approximations and higher-frequency details. In performing this process of *Multilevel Decomposition* (MD) it should be considered that at each step the available data are halved and the time scale is doubles with respect to the basic sampling time T_s, as shown in Figure 4.

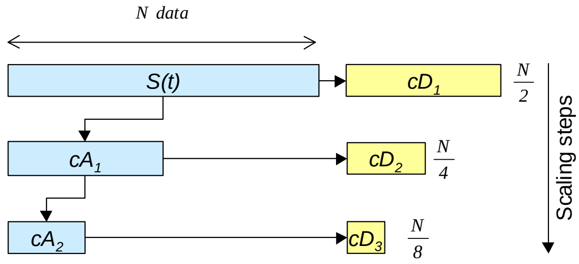


Fig. 4. Data halving at each scaling step of MD.

Therefore the iteration should be stopped before the data become so scarce that edge effect become important and the time-scale of the transformation exceeds the range of interest.

3.1. Settings for wavelet analysis of circadian patterns

For the analysis of circadian patterns a Meyers wavelet (Meyers et al., 1993) was selected for its properties of symmetry, orthogonality and finite support. The ratio between time window and sampling interval is important. Since the object of the study is the circadian pattern, it seems obvious to select a $T_s = 24\ h$ window which, given a sampling interval $\Delta = 1\ h$, includes 24 values. With these conditions, the multilevel decomposition cannot be iterated beyond the second level or the edge effects would seriously deteriorate the results. Moreover, restricting the analysis to Level 2 has a clear physical meaning:

- ***Level 1*** implies *2 h* differences, hence the approximation A_1 represents the de-noised signal, filtering out the measurement noise, assuming that a random 1 sample fluctuation is the combined effect of the DO probe error and short-term environmental changes (wind gusts, passing clouds, etc.). Likewise, the detail D_1 represents the noise contribution due to these short-lived effects;

- ***Level 2*** is performed at *4 h* intervals, producing a signal free of climatic artefacts. Thus approximation A_2 represents the true circadian cycle, rich in ecological information (algae photosynthesis and respiration, decom-position processes, etc.). Likewise, detail D_2 accounts for the environmental variability on a 4 h horizon.

Thus it can be concluded that a Level 2 Approximation is sufficient to extract the required information from the daily signal. The problem of edge effects is now considered. Unfortunately 24 is not a power of 2, hence the dyadic scaling does not provide optimal data scaling. One possible answer to this problem could be the processing of three consecutive days and consider the decomposition (A_2, D_2) of the central day only. However, Figure 5 shows that the advantage of this method is minimal *vis à vis* the increased computational burden and the increased fraction of excluded days because of missing data (triplets would be needed instead of single days).

A quantitative way to appreciate the degradation due to edge effects in processing each single day can be obtained by computing the percentage of energy corresponding to the approximation E_a and the sum of the percentages of details energy E_d. The variation of E_a in processing single days or triplets is less that 0.1 %, whereas the E_d variation may be as high as 50%. This confirms the choice of A_2 as a stable decomposition, truly representative of the circadian cycle.

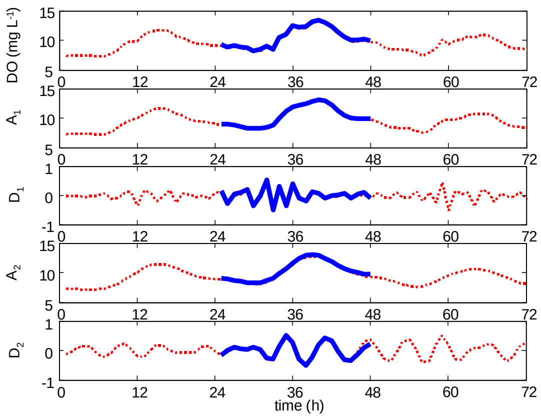


Fig. 5. Comparison between Level 2 MD of a triplet (dotted line) and the central day alone (thick line).

3.2. Features extraction from filtered data

The characteristics of wavelet filtering for this application can be summarised as follows: a second level Discrete Walevelet Transform (DWT) is performed on single days composed of 24 hourly samples and the second approximation (A_2) is considered to be a good denoised reconstruction of the underlying circadian cycle. From this filtered signal the following features are extracted for clustering: minimum and maximum DO values (DO_{min}, DO_{max}) and the slope α of the line connecting these values, as shown in Figure 6. Therefore the data points $\boldsymbol{x}$ are triplets of the following kind

$$\boldsymbol{x}_k = \left[DO^k_{min} \quad DO^k_{max} \quad \alpha^k \right] | \; k = 1,2,...,N \; .$$

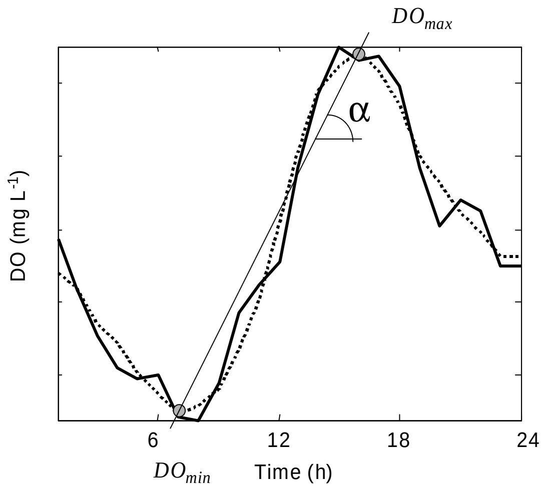


Fig. 6. Features of a DO circadian cycle.

4. FUZZY CLUSTERING FOR CIRCADIAN PATTERN RECOGNITION

These three features are then processed using the Fuzzy Maximum Likelihood Estimates (FMLE) fuzzy clustering algorithm (Babuska, 1998). The fuzzy clustering algorithm arranges the data $\{\boldsymbol{x}_k \,|\, k = 1,...,N\}$ into c clusters through the minimisation of the partition functional

$$J(\mu_{i,k}, m) = \sum_{i=1}^{c} \sum_{k=1}^{N} \left(\mu_{i,k}\right)^m d^2_{ik_A} \qquad (2)$$

where the distance $d^2_{ik_A}$ depends on the norm-inducing matrix $\boldsymbol{A}$, which in the FMLE case is a function of the fuzzy covariance matrix $\boldsymbol{R}$.

$$\boldsymbol{R} = \frac{\sum_{k=1}^{N} \mu_{ik} (\boldsymbol{x}_k - \boldsymbol{v}_i)(\boldsymbol{x}_k - \boldsymbol{v}_i)^T}{\sum_{k=1}^{N} \mu_{ik}} \qquad (3)$$

$$d^2_{ik_A} = \frac{(det\ \Sigma_i)^{\frac{1}{2}}}{\left(\frac{1}{N}\sum_{k=1}^{N} \mu_{ik}\right)} exp\left(\frac{1}{2}(\boldsymbol{x}_k - \boldsymbol{v}_i)^T \boldsymbol{R}^{-1} (\boldsymbol{x}_k - \boldsymbol{v}_i)\right) \qquad (4)$$

The resulting membership $\mu_{i,k}$ represents the degree of membership of the *k-th* data point $\boldsymbol{x}_k$ to the *i-th* cluster with center $\boldsymbol{v}_i \,|\, i = 1,...,c$. The effectiveness of the partition can be evaluated in information-theoretic terms using the normalised partition entropy H_n, defined as (Babuska, 1998).

$$H_n = -\frac{1}{1 - \frac{c}{n}} \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{i,k} \log\left(\mu_{i,k}\right) \qquad (5)$$

This quantity was used to decide the appropriate number of clusters c in the partition, as the one which minimises the H_n value for $c > 2$. Processing the daily DO data in the years 2001, 2002 and

2003 resulted in a number of clusters with a clear ecological meaning. The number of cluster for each year, determined with the partition entropy of eq. (5), varied from $c = 3$ for 2001 to $c = 6$ in 2002 and $c = 4$ for 2003.

As an example, the clusters obtained for 2001 are shown in Figure 7.

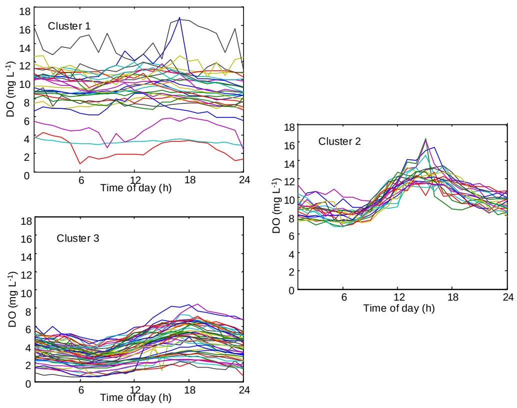


Fig. 7. DO clusters obtained from the 2001 data.

For the year 2001 three basic behaviours were isolated, with cluster 2 representing the typical spring cycle with a strong afternoon peak, due to intense photosynthesis, and cluster 3 typical of the summer situation with low oxygen due to high temperatures and declining algal activity. Cluster 1 groups atypical behaviours, possibly induced by adverse weather conditions, preventing the development of the normal circadian cycle. The distribution of these patterns over the year, together with that of unclustered patterns is shown in Figure 8.

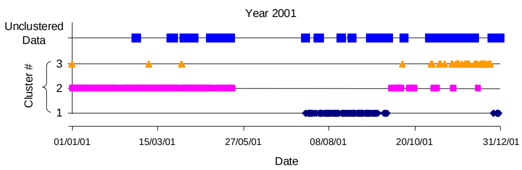


Fig. 8. Distribution of clustered and unclustered patterns over 2001.

The same analysis for 2002 produced six clusters whose distribution over the year is shown in Figure 9. This year presented a wider variability, which resulted in a larger number of clusters. As to their characteristics, cluster 3 was similar to cluster 2 of 2001 (high algal activity) and cluster 4 was similar to cluster 3 of 2001 (summer low activity).

The data from 2003 produced the most interesting structure: the basic DO clustering resulted in the four clusters of Figure 10. However, a finer structure can be extracted from cluster 2, observing that these DO patterns correspond to very different ORP cycles. Tracing back these

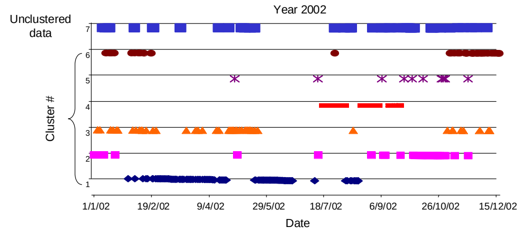


Fig. 9. Distribution of clustered and unclustered patterns over 2002.

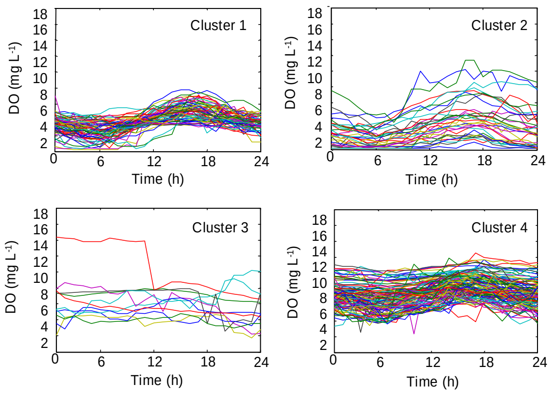


Fig. 10. DO clusters obtained from the 2003 data.

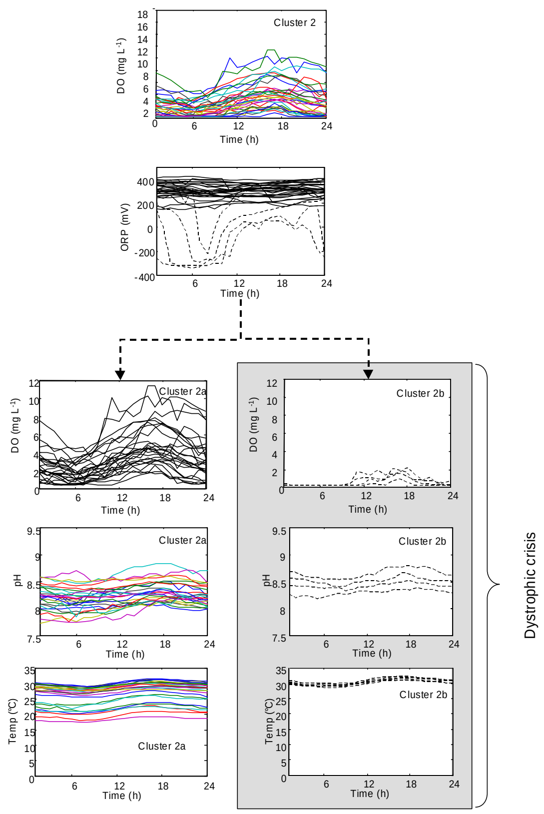


Fig. 11. Finding a finer structure in cluster 2.

patterns to the original DO cycles in cluster 2, this can be further decomposed into two subsets, one of which is clearly representative of the dystrophic crisis, as shown in Figure 11. This separation is confirmed by the two other variables, pH and temperature, yielding the composite cluster structure for 2003 shown in Figure 12.

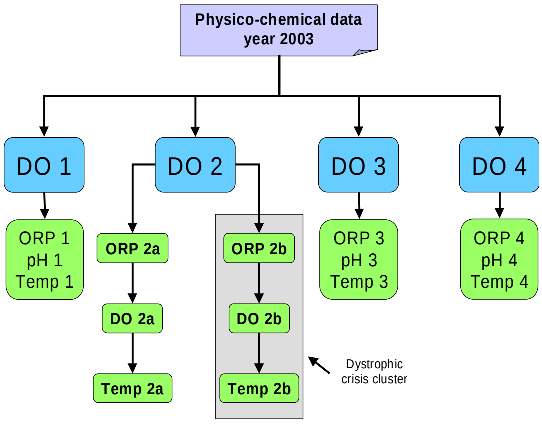


Fig. 12. Cluster structure for the year 2003.

It is also interesting to observe how the clustered and unclustered data are distributed over the year, as shown in Figure 13, with cluster 2 prevailing during the warm season, and in particular with cluster 2b, representative of the dystrophic crisis, placed exactly when the crisis actually occurred, as shown in Figure 14.

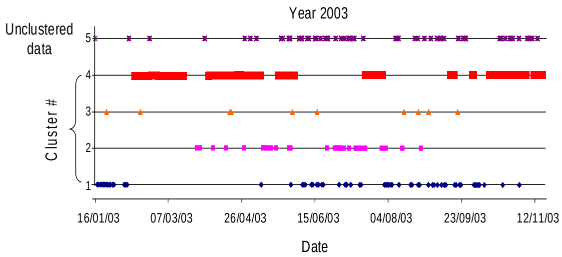


Fig. 13. Distribution of clustered and unclustered cycles over the calendar day in 2003.

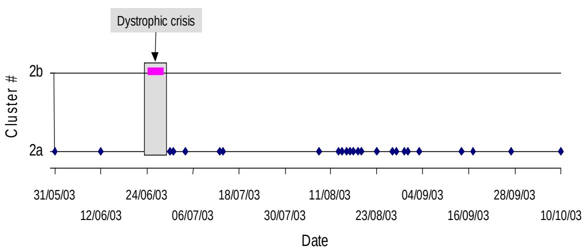


Fig. 14. Placement of cluster 2b during the dystrophic crisis.

5. CONCLUSION

This paper has presented a method for extracting representative prototypes of circadian cycles. It is based on a combination of wavelet filtering and fuzzy clustering. The data were first processed with a discrete wavelet decomposition in order to isolate the relevant circadian cycle, which is well represented by the second level approximation A_2. From the filtered cycles three relevant features were extracted: maximum and minimum values and the slope of their connecting line. The cycles were then classified on the basis of these features using the Fuzzy Maximum Likelihood Estimates (FMLE) clustering algorithm, which was preferred for its variable metric, being able to conform the cluster shapes to the data.

The application of the algorithm to the physico-chemical data from the Orbetello lagoon resulted in a consistent classification of relevant circadian cycles and in particular it isolated the patterns corresponding to the dystrophic crisis observed in extreme eutrophic conditions. Though the pattern structure varied over the three years of the study (2001 - 2003), the method allowed positive cycle identification, whose distribution over each year was consistent with the observed behaviours and seasonal variability.

6. ACKNOWLEDGEMENT

The Orbetello Lagoon Managerial Office is acknowledged for supporting this research under contract n. 96/988 of 24.07.2003.

7. REFERENCES

Babuska, R., *Fuzzy modeling for control.* Kluver Publ. Co., Amsterdam, 260 pp., 1998.

Christian, R.R., Naldi, M. and Viaroli, P., 1998. Construction and analysis of static, structured models of nitrogen cycling in coastal ecosystems. in Koch A.L., Robinson J.A. Milliken G.A. (eds.), *Mathematical Modelling in Microbial Ecology*, Chapman & Hall, New York.

Marsili-Libelli, S., Fuzzy pattern recognition of circadian cycles in ecosystems. *Ecol. Model.*, **174,** 67 - 84, 2004.

Meyers, S.D., Kelly, B.G., and J.J. O'Brien. An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of Yanai waves. *Mon. Weather Rev.* **121**, 2858–2866, 1993.

Strang, G. and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1996.

Torrence,C. and A., Compo, A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, **79**, 61 –78, 1998.

Towards Sustainable Management of Wastes: Results of a Modelling Case Study, and the Emerging Environmental Decision Support System Based on the Energy Footprint

V. Krivtsov, P. Dacombe, C. Banks, S.H.Spiller and S. Heaven

Department of Civil Engineering and the Environment, University of Southampton, Southampton, UK
Email for correspondence: e96kri69@netscape.net

Abstract: Here we present an emerging EDSS, based on the energy and materials flow model for evaluation of alternatives for processing domestic and commercial waste. The input information for the model is obtained through a literature search, consultations with stakeholders, and also using a specially designed expert system (ES) - The 'Smart Waste Questionnaire'. The ES implements an IT-based interrogative data acquisition methodology built using a custom-designed expert system software and integrated databases. In this paper we present the results of an ongoing case study of the model's application to the management of specific waste types in the area of Southampton (England, UK). The model has been designed in Microsoft Excel and Visual Basic, which were chosen due to their user-friendly interface and common availability. In the model calculations, the input data on material flows, waste quantities and mass balances are combined with information on the energy requirements for different types of collection and processing systems for re-use, recovery, recycling and disposal, and on the energy benefits (e.g. production of heat by incineration or gas by anaerobic digestion, pyrolysis, etc.) of these options. The output shows the energy balance of the current practice, and also allows a comparison with a number of alternative scenarios. The results are directly applicable only for the study area. However, following certain modifications the methodology used may be easily applied elsewhere. Examples of the collected data and model simulations are given, problems with data collection and availability outlined, and limitations and implications of the study discussed.

Key Words: energy footprint, industrial ecosystem, waste management, glass, paper, expert system

1. INTRODUCITON

Complex integrative modelling studies help to enhance our knowledge of industrial ecosystems and are, therefore, an important prerequisite for sustainable development of Mankind (Korhonen 2001; Krivtsov *et al.* 2004a). Current progress in waste management is hampered by the lack of methods capable of evaluating sustainable practices. The tools used for comparison of alternative scenarios, involving collection, separation, and processing of waste fractions, must have a rigorous conceptual basis, and account not only for economic considerations, site-specific logistics and governmental guidelines, but also for environmental issues. In other words, society needs evaluation tools able to optimise (within given local, regional and global constraints) the existing waste management practices by minimising the required energy budget and both present and future environmental impacts.

Here we present an overview of the integrative research into the energy footprint of Southampton wastes, conducted by the University of Southampton in collaboration with Sothampton City Council, Hampshire County Council, and a number of relevant waste management companies and other Universities. This research integrates an extensive data collection programme, with the examination of data obtained through statistical analysis of a database, and with comprehensive simulation modelling analysis.

To analyse the characteristics of the overall energy footprint of Southampton wastes, the information obtained in different lines of investigation may be integrated by means of statistical and simulation modelling. In the analysis of waste management energy footprints, it is paramount to consider all relevant processes starting from the point where materials become wastes, until disposal and/or reprocessing. In particular, it is crucial to include both the energy consumed during processing and/ or disposal, and the energy consumed in transportation.

This paper, therefore, brings together the information obtained within the comprehensive analysis of data related to specific stages of the Southampton's waste cycle, and the results of simulation modelling, hence integrating the information separately obtained on various sub-systems in the analysis of the emerging interaction patterns observed. It sets the basis for an emerging environmental decision support system, i.e. EDSS (Rizzoli & Young 1997), aimed at facilitating the optimization of current waste management practices.

Integration of the software is often an important prerequisite for successful environmental analysis (Krivtsov *et al.* 2004b), and there is clear evidence of the ability for integrated environmental modelling to aid understanding and assist decision making in resolving complex environmental and natural resource management problems (Argent 2004). In particular, integration of an expert system and a database technology, equipped with a suitable interface may be useful for practical environmental applications, including those in waste and resource management (Liao 2003; Lukasheh *et al.* 2001).

In the research presented here, the standard Microsoft Office Suite of programs was chosen to handle most of the data due to the predominance of the software, and, importantly, the facility to integrate the different applications (e.g. database, spreadsheet) without altering data. Integration of functions available in Microsoft Excel with a customised software written in Microsoft Visual Basic and Microsoft VBA allowed us to perform the analysis in an interactive mode within the model specifically compiled to represent the local conditions. This is achieved through the integration of standard and customised software, thus supplementing the capabilities of commonly available commercial tools by custom-designed tools reflecting the needs of particular tasks.

2. DATA COLLECTION

For the purposes of the research presented here, data collection involves using the information obtained through literature search, and extensive consultations with stakeholders, including, e.g., Southampton City Council (SSC), Hampshire County Council (HSC), Onyx Environmental, etc. A new extension of the research also incorporates the application of a specially designed expert system 'Smart Waste Questionnaire', intended to facilitate collection of the information related to commercial and industrial wastes. It should be noted that the estimates of waste quantities, transport distances, vehicle and material characteristics represent the current assumptions incorporated in our modelling analysis. These estimates are based on the information available, and are continuously being improved to increase the predicting and interpretative power of our research.

3. MODEL DESCRIPTION

The model presented here considers all relevant processes starting from the point where materials become wastes, until disposal and/or reprocessing. Most importantly, the model accounts not only for the energy consumed during processing and/ or disposal, but also takes into consideration the energy consumed in transportation. The model has been compiled in Visual Basic linked to Microsoft Excel. Structurally, the model consists of interlinked submodels, with each sub model simulating a specific stage in the overall process. The most important submodels include

- Refuse collection and landfill transfer
- Kerbside collection
- Stage one transport (i.e., from consumers to collection points)
- Stage two transport (i.e., from collection point to processing plant)
- The processing plant
- Processed material transfer
- Manufacture

In the case of paper/card, however, an important consideration should be given to the incineration process, as removing of paper/card from the waste stream may result in a decrease of the wastes' calorific value. Although currently Southampton wastes are not incinerated, an incinerator is being built in Marchwood, and is due to become operational in the very near future. Therefore, an additional submodel describing incineration process has been incorporated into the model structure for the paper/card case study.

The subprograms representing separate submodels are called from the main program, and the total energy consumption is obtained by adding up the energy consumptions of the specific components returned by the subprogrammes. It should be noted that, as much as possible, parsimonious code is achieved by the reuse of submodels (Rizzoli & Young 1997). For example, one of the most reused submodels is the submodel calculating vehicle energy consumption from data on the vehicle load, journey distance, waiting and handling time, fuel consumption per mile traveled, and the average speed.

Integration of the separate lines of investigation within the model's calculations is briefly demonstrated below on the example of two case studies assessing energy footprints of Southampton glass and paper/card. Further details of the study and the results obtained are given in separate publications (Krivtsov *et al.* 2003; Krivtsov *et al.* 2004a).

4. EXPERT SYSTEM Smart Questionnaire

Application of a specially designed Expert System 'Smart Questionnaire' provides an exciting possibility to expand the scope of the study to include commercial and industrial wastes. The concept of a 'smart' questionnaire is that it adapts the questions asked according to the user's responses, thus avoiding irrelevant and confusing information. Existing in-built databases can be used for verification to determine whether the answers given are within normal bounds, and if necessary the same question can be asked in different ways as a self-checking procedure. Problems such as the units in which waste is quantified are overcome by using internal conversion factors, thus allowing the respondent to specify quantities in terms with which he/she is familiar. A further advantage is that the information is already in digital database format, alleviating the need for manual data entry with the errors this entails. The databases within the questionnaire are originally being populated using best available data from the Environment Agency's national waste survey which relates waste generation to business type and size, as well as data from other relevant sources. Business classification in the model is by means of a specially designed coding system, which could then be mapped to SIC or European waste classification codes using embedded databases.

In some cases, the results obtained using the expert system may provide a level of detail greater than that required for consideration of processing and infrastructure needs, but the database output could be tuned to provide information in the most useful format - for example, to group wastes according to the 12 master categories suggested by the Zero Waste campaign, if that is required. The aim is to overcome the limitation of data collection as a rigid and inflexible process governed by the needs perceived at the time. At the same time it is necessary to keep a balance between detail and the practicability of gaining information, and accuracy and replicability, which can be achieved by periodic review and updating of data for tracking and predicting purposes. It is intended to improve continually both the data gathering (input) and reporting (output) of the 'smart' questionnaire using feedback from business, waste management contractors and other interested parties such as SEEDA (Southeast England Development agency) and SIEnA (Solent Industry & Environment Association).

Application of the smart questionnaire has a number of stages, the first of which is collation of existing data on C&I wastes and on the business profile in Hampshire. The data are being used to develop a sampling framework based on location, business type and key resource streams.

It should be noted that the system is currently under development, and the initial trial version presented here lacks, for example, sophisticated learning capabilities. Initially the system's learning will be confined to the update of the embedded databases, which are subsequently used to determine whether certain responses are within the normal bounds for any particular variable, and, where appropriate, how far are they from the average. Treatment of the uncertainty is also not very sophisticated, and is currently being confined to comparing of the response obtained with the normal bounds for any particular variable. However, the system is being continuously improved, and it is envisaged that the initial limitations will be duly addressed in the future.

5. RESULTS

Example 1 - Southampton glass

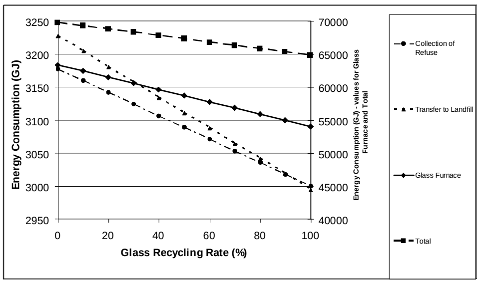


Figure 1. The dependence between the Southampton glass wastes energy footprint (major components) and the recycling rate.

By running a number of ' what if' scenarios it was found that the energy consumption related to handling and processing Southampton glass wastes (NB: that includes manufacture of necessary yearly supply, which is included within the system boundaries) ranges between 60,000 and 70,000 GJ per year. For the current situation, the estimate of the overall yearly energy footprint was approximately 68,600 GJ. The bulk of this energy consumption is in the manufacturing process. Therefore, energy savings in manufacture achieved through the increased use of cullet [i.e. due to increased recycling] would outweigh the increased energy consumption related to the stages of collection, transportation, and processing at the processing plant [Figures 1-2].

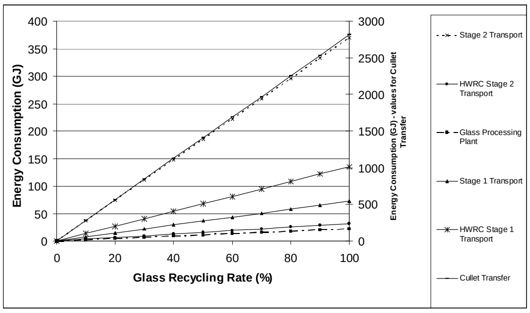


Figure 2. The dependence between the Southampton glass wastes energy footprint (secondary components) and the recycling rate.

It is worth pointing out that although within the assumptions incorporated in the model, the energy consumption per tonne of recovered material invariably decreases with an increase in the recycling rate, the rapid decrease is demonstrated only at relatively low recycling rates, after which the curve levels out [Figure 3]. This, in the author's view, provides an exciting possibility for optimizing the management decision strategies accounting for this and a multitude of other relevant factors.

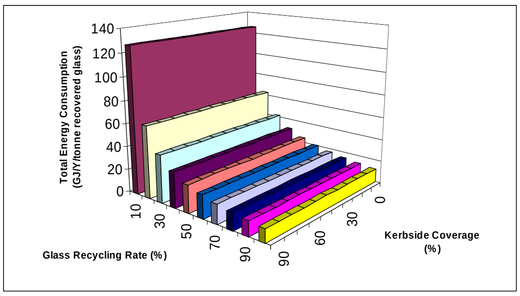


Figure 3. Energy spent per tonne of glass waste recovered.

Example 2 - Southampton paper/card

Analysis of energy footprint associated with Southampton paper wastes revealed a similar (i.e. to the glass case study) relationship between the overall energy spendings and the recycling rate [Figure 4]. As with glass, the bulk of energy consumption is in the manufacturing process, and the savings made here through increased use of recycled paper offset any increases in transport and processing energy consumed elsewhere. Our analysis have also revealed that recycling via kerbside collection has more potential (e.g. in terms of quantities collected) than via bring-sites. It should be noted, however, that in reality the exact preference to any particular option should, of course, be determined by a number of factors (i.e. in addition to the energy footprint), including, e.g., infrastructure/layout of road network, vehicle characteristics and availability, availability of human resources to run the scheme, population density, etc..

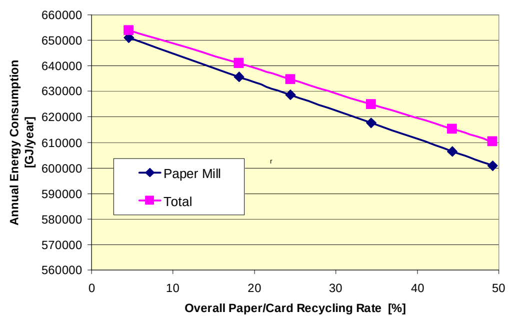


Figure 4. The dependence between the Southampton paper wastes energy footprint and the recycling rate.

6. DISCUSSION

Analysis of multicomponent systems is greatly aided by application of simulation modelling techniques (Krivtsov *et al.* 2000; Krivtsov 2001). Complex interplay among system components have previously been taken into account in a number of waste management and industrial ecology studies (Abou Najm *et al.* 2002a, b; Adamov *et al.* 1999; Bjorklund *et al.* 1999; Clift 1998; Cosmi *et al.* 2000; Korhonen *et al.* 2001; Krivtsov *et al.* 2004a). The results of the research presented here are in good agreement with the previous studies.

In general, the integration of the standard commercial and customised software has been very useful in the investigations of the energy footprint of Southampton wastes presented here. Our integrative data analysis and modelling approach included collection of a wide range of data, examination of these data using a database, and analysis by statistical tests and simulation modelling.

The key element in the data acquisition stage of the project is the development and deployment of an expert system - 'smart' interactive questionnaire. This can later be developed into an informative and immediate feedback mechanism through which participating companies can obtain information on the best and most cost-effective ways of dealing with their waste in a particular geographical area. It is also hoped that the scheme will subsequently be expanded into a web-based tool accessible to the wider business community of Hampshire.

With some adaptation, the integrative data collection and modelling approach presented here may be applied to other types of wastes in Southampton. With further adaptation, it may also be applied to other geographical locations. Hence the research presented here may be regarded as a generic template for analysis of energy footprints associated with processing of wastes.

7. FURTHER WORK

The first stage of the project has involved analyses of household wastes. Currently, however, the structure of the emerging EDSS is being enhanced, which will allow us to process data on commercial and industrial wastes using an expert system 'Smart Waste Questionnaire' (described above). This is an exciting development of the research as it provides a possibility to quantify the waste streams generated by commercial, including small- and medium-sized, enterprises, which are notoriously difficult to assess.

Clearly, the success of the data collection survey is dependent upon the degree of cooperation offered by the business community: to maximise this, the survey work is being designed so that it offers real and tangible benefits to the business community. To achieve this, the 'smart' interactive questionnaire will be enhanced so that it not only allows accurate information gathering, but also gives feedback to individual businesses on how to maximise their waste management opportunities. This interactive system will ultimately be offered as a web-based tool accessible to all businesses in Hampshire, and ultimately across the South East.

It is worth pointing out that the emerging EDSS system presented here is still under development. However, the system is being continuously enhanced, and it is intended that further developments will include incorporation of spatial representation using GIS, improvements to the Expert System 'Smart Waste Questionaire' as regards learning capabilities and uncertainty treatment, incorporation of a wider range of artificial intelligence techniques, and design of a graphical user-friendly interface to allow the system use by stakeholders and decision makers. It is also intended to address the issues of software and hardware compatibility, and to make the versions available to run on a wider range of platforms and operating systems (the current prerequisite is Windows XP).

It should also be noted that the analysis presented here relate specifically to Southampton. However, the existing software could be applied for analysis of waste management energy footprints elsewhere, although the relevant constants and pathways may have to be changed by the system analyst to reflect the local conditions. Extension of the scope of this research beyond Southampton area remains, however, subject for further work.

8. CONCLUSIONS

An integration of all the separate submodels into a comprehensive overall model has allowed us to assess the overall energy footprint related to glass and paper/card fractions of Southampton wastes. In future, we intend to use the overall model for the extensive comparative analysis of direct and indirect interactions within the industrial ecosystem studied (*sensu* Krivtsov et al., 2000; see also Krivtsov, 2001, 2002 and references therein).

The major source of energy savings from recycling may best be achieved through increased use of recycled material in the manufacturing process. With maximum recycling, energy savings of up to ~11% for both the glass and paper/card fractions are feasible. Therefore, recycled materials should be used to make more new materials, rather than for alternative purposes (e.g. glass as a replacement for aggregates).

The information gathered within the current and the next stage of the project will generate a resource map of C&I waste resources in Hampshire and will provide an opportunity to rethink fundamentally what information is needed from waste audits and how it can best be used. The results will be used in planning the County's infrastructure needs, in identifying opportunities for new businesses to develop using the resources reclaimed from waste, and in mapping the way forward for the sustainable management of C&I waste in the UK.

9. ACKNOLEDGEMENTS

The authors wish to thank Biffaward, BOC foundation, Hampshire County Council, Onyx Environmental, and Southampton City Council for providing data and funding for this study, and for valuable discussion. Constructive comments of two unknown reviewers were helpful for improving the final version of the paper.

10. REFERNECES

Abou Najm M., El-Fadel M., Ayoub G., El-Taha M. & Al-Awar F. (2002a) An optimisation model for regional integrated solid waste management I. Model formulation. *Waste Management & Research* 20: 37-45.

Abou Najm M., El-Fadel M., Ayoub G., El-Taha M. & Al-Awar F. (2002b) An optimisation model for regional integrated solid waste management II. Model application and sensitivity analyses. *Waste Management & Research* 20: 46-54.

Adamov E. O., Ganev I. K., Lopatkin A. V., Orlov V. V. & Smirnov V. S. (1999) Self-consistent model of nuclear power development and fuel cycle. *Atomic Energy* 86: 337-344.

Argent R. M. (2004) Concepts, methods and applications in environmental model integration. *Environmental Modelling & Software* 19: 217.

Bjorklund A., Dalemo M. & Sonesson U. (1999) Evaluating a municipal waste management plan using ORWARE. *Journal of Cleaner Production* 7: 271-280.

Clift R. (1998) Engineering for the environment: The new model engineer and her role. *Process Safety and Environmental Protection* 76: 151-160.

Cosmi C., Cuomo V., Macchiato M., Mangiamele L., Masi S. & Salvia M. (2000) Waste management modeling by MARKAL model: A case study for Basilicata Region. *Environmental Modeling & Assessment* 5: 19-27.

Korhonen J. (2001) Industrial ecosystems - some conditions for success. *International Journal of Sustainable Development and World Ecology* 8: 29-39.

Korhonen J., Wihersaari M. & Savolainen I. (2001) Industrial ecosystem in the Finnish forest industry: using the material and energy flow model of a forest ecosystem in a forest industry system. *Ecological Economics* 39: 145-161.

Krivtsov V., Corliss J., Bellinger E. & Sigee D. (2000) Indirect regulation rule for consecutive stages of ecological succession. *Ecological Modelling* 133: 73-82.

Krivtsov V. (2001) Study of cause-and-effect relationships in the formation of biocenoses. *Russian Journal of Ecology* 32: 230-234.

Krivtsov V., Banks C. J., Heaven S. & Dacombe P. (2003) Sustainable waste management - the energy question. *Wastes Management, Chartered Institute of Wastes Management* January 2003: 34-35.

Krivtsov V., Wäger P. A., Dacombe P., Gilgen P. W., Heaven S., Hilty L. M. & Banks C. J. (2004a) Analysis Of Energy Footprints Associated With Recycling Of Glass And Plastic - Case Studies For Industrial Ecology. *Ecological Modelling* 174, 175-189.

Krivtsov V., Walker S. J. J., Staines H. J., Watling R., Burt-Smith G. & Garside A. (2004b) Integrative analysis of ecological patterns in an untended temperate woodland utilising standard and customised software. *Environmental Modelling & Software* 19: 325-335.

Liao S. H. (2003) Knowledge management technologies and applications - literature review from 1995 to 2002. *Expert Systems with Applications* 25: 155-164.

Lukasheh A. F., Droste R. L. & Warith M. A. (2001) Review of Expert System (ES), Geographic Information System (GIS), Decision Support System (DSS), and their applications in landfill design and management. *Waste Management & Research* 19: 177-185.

Rizzoli A. E. & Young W. J. (1997) Delivering environmental decision support systems: software tools and techniques. *Environmental Modelling & Software* 12: 237-249.

Eliciting Sources of Uncertainty in Ecological Simulation Models

V. Brilhante [a], J. L. Campos dos Santos [b]

[a]Computing Science Department, Federal University of Amazonas – UFAM, Manaus, Brazil
virginia@dcc.ufam.edu.br

[b]National Institute for Research of the Amazon – INPA, Manaus, Brazil
lcampos@inpa.gov.br

Abstract:

Uncertainty is an intrinsic feature of complex ecological models. Given that it is not possible to rid the models from uncertainty, we are left with taking notice of it for consideration in model-based decision making. Traditional ecological modelling methods and tools do not support explicit accounts of model uncertainty. This work gives a contribution towards making known, or bringing to the surface, sources of uncertainty that are embedded in ecological models. The sources of uncertainty are related to the models' supporting data and equations. A metadata standard is used to specify data-related sources of uncertainty, such as creator and coverage. In the technique developed, models are described and simulated using logic, which allows the sources of uncertainty to be easily represented, and later propagated and combined during simulation. The combined sources of uncertainty can then be presented to the user who can assess their impact on model outputs and tune up his confidence in the model for decision making.

Keywords: Uncertainty elicitation; logic-based ecological modelling; metadata.

1 Introduction

Artificial Intelligence has long been pursuing the successful handling of uncertainty by its systems [Cohen, 1985], largely motivated by the many intrinsically uncertain tasks we perform in everyday life – decision-making, argumentation, learning – and by how well we manage.

The domain of ecological modelling offers a challenging context for uncertainty handling, where well-defined mathematical and simulation techniques are applied to problems that are not totally understood. A significant part of existing functional relationships in an ecosystem can be ignored in a model, for being too complex, not well understood or simply unknown. Moreover, supporting field or experimental data is usually incomplete, partly due to logistic problems and inherent difficulties in data collection and accurate measurement and sensing in the natural environment. In the light of that, it does not seem promising to try and build ecological models that are uncertainty free.

We can, nevertheless, live in better terms with it by getting *to know* the uncertainty that is embedded in the models, which may lead to better informed model-based decision making. Conventional modelling techniques and tools do not provide for representation or reporting of model uncertainty. This work presents a computational technique where models' sources of uncertainty are explicitly represented, propagated throughout the model during simulation, and shown to users for assessment of their impact on simulation results and decision making that may follow.

2 Symbolic Representation of Uncertainty

The approach followed for uncertainty representation in this work is symbolic. This has been little explored in Artificial Intelligence as opposed to the more popular numerical approaches, such as degrees of belief handled by Bayesian methods [Parsons and Hunter, 1998]. Numerical approaches seem to perform well on domains such as the stock market, where quantitative data is available to characterise problems, and degrees of belief are quantifiable with clearly defined semantics. On the other hand, symbolic approaches seem to fit better in large complex domains such as environmental applications or medicine, where the nature of data and its interpretations cannot be purely quantitative. A number representing a probability, a degree of belief, an evidence or any other so called uncertainty measurement may be an overly concise representation that obscures the reasons that one took into account to reach that number.

2.1 Endorsements Theory Revisited

Our non-numerical and declarative representation of uncertainty is based on Cohen's theory of endorsements [Cohen, 1985]. The theory advocates explicit representation of uncertainty related to domain information handled by a reasoning system, allowing users to reason about uncertainty directly, instead of implicitly through some numerical calculus, and assess how much to believe in the system's outcomes.

In modelling, a *source of uncertainty* is any information that can suggest to model users reasons for strengthening or weakening their belief in the model's results. Sources of uncertainty and their largely domain dependent values are represented by data structures called *endorsements*, which can be attached to data, rules, conclusions, tasks and resolution procedures. Building an endorsement-based system involves: 1. identification and naming of the sources of uncertainty, or endorsements, in a domain; 2. specification of how these sources interact, so that they can be combined; and 3. specification of rules for ranking combinations of sources of uncertainty so that decisions can be made.

Our identified sources of uncertainty (Section 3.2) are attached to logical clauses that define elements of system dynamics models, namely, state variables, intermediate variables, parameters, flows and links [Ford, 1999]. An automated mechanism has been implemented to combine the endorsements (Section 5). Step 3 above was not carried out in this work because of the subjective kind of analysis that the identified endorsements lend themselves to in the domain of ecological modelling (Section 6).

2.2 Logic as the Modelling Language

Logic-based approaches for ecological modelling have been proposed in [Robertson et al., 1991] on the grounds of language accessibility for modellers and representational power. Logic is adequate to express declaratively domain knowledge and model assumptions, which in turn enable some forms of modelling automation [Brilhante, 2003] and more informed model analysis. In this work we use Horn clauses with negation under the closed world assumption [Apt, 1997] to specify models and their sources of uncertainty. Adopting this well-known representational formalism has the advantage that we can straightforwardly build systems that reason upon the represented knowledge using an off-the-shelf implementation language which is Prolog.

Particular forms of clauses are used for each kind of model element. These clauses describe a model's static structure that is equivalent to its structure denoted by the system dynamics diagrams, as illustrated by the fragment from a forest Carbon cycle model in Figure 1. The figure shows just one state variable representing the stock of Carbon in the woody litter lying on the forest floor, regulated by the incoming flow of woody litter production and the outgoing flows of Carbon to soil organic matter and the atmosphere. Intermediate variables (denoted by circles) and parameters (denoted by squares) appear with fictitious names.

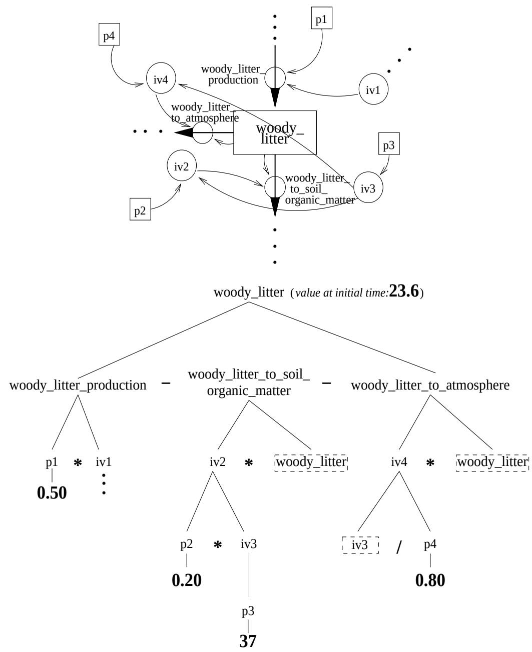


Figure 1: Fragment of a model in system dynamics notation and its visualisation as a tree structure.

Running a model consists of solving differential or difference equations which regulate changes in the values of the model variables as simulation time progresses. We can visualise a system dynamics model simulation at one tick of the simulation clock as a set of tree-like structures, one to each state variable, having nodes representing variables which are interrelated by mathematical operations. Their roots are state variables to be solved, intermediate nodes are flows and intermediate variables, and leaves are the model parameters. The idea is depicted in Figure 1.

We have implemented the sources of uncertainty elicitation technique in a working logic-based system which includes an interpreter to run simulations of system dynamics models and a mechanism to combine instances of sources of uncertainty associated with model elements. The interpreter operates over the static structure of the model specified as a Prolog program, and is able to calculate the value of any model element at any simulation time. It goes into action when given goals of the form $goal(E, V, T)$ where E is a model element and V is the value it holds at time T. The interpreter works recursively backwards in time. Time T given in the top-level goal is successively decremented by 1 until the simulation gets to the initial time 0. A subgoal in the proof of $goal(E, V, T)$ is $goal(E, V_{\mathrm{p}}, T_{\mathrm{p}})$, where V_{p} and T_{p} stand for the previous value of the model element at the previous simulation time point.

Using logic programming has allowed us to represent model elements together with their endorsements under the homogeneous representational framework of logical clauses. The simulation interpreter, in turn, applies Prolog's built-in proof procedure to solve goals. This is a system development approach that contrasts with devising model-element specific data structures and a how-to-simulate procedural program.

3 METADATA IN MODEL UNCERTAINTY ELICITATION

As we have discussed, sources of uncertainty buried in ecological models can be many, related to procedures of data collection, to data integrity and to the modelling process itself. The approach we have taken for bringing some of this uncertainty to the surface is to "see the forest for the trees": to identify uncertainty related to model components and later combine it to provide an overall reading of uncertainty in the model.

Within this approach, the prime model components for sources of uncertainty identification are parameters and state variables. Model simulation starts from initial values assigned to these model elements. As shown in Figure 1, if we visualise the model structure as a tree of influences, the initial values are at the tree leaves (influenced by no other element in the model). The parameters' values remain constant, while the state variables' values can (and are usually expected to) change during simulation. These initial values are measures, rates, averages, constants, coefficients, percentages etc. that in the best case come from a modelling dataset built from field experiments and statistical data treatment. Sometimes, the initial values come from the literature or are mere guess estimations.

Our endorsements constitute information about the modelling data, and as such, can be specified as *metadata*.

3.1 Why Metadata?

Metadata is, quite simply, data about data. It describes the attributes and contents of a document, work or dataset [Duval et al., 2002]. The scope of questions it can cover in this way include: **Who** collected and who distributed the data? **What** is the subject (e.g., a dataset)? **When** was the data collected? **Why** was the data collected (its purpose)? **How** was the data collected? **How** should it be used? **How** much does it cost? Answers to these questions in the form of metadata can span over a quality spectrum – from raw to quality assured metadata. The better the quality, the more able users will be to evaluate, with less uncertainty, whether or not the data is useful to them.

Difficult and time-consuming to produce as it may be at first, metadata can help decision makers and researchers to find and use data, and can also benefit the primary creators of the data by maintaining its value and assuring its continued use over a span of years. From all its benefits, the one that is most relevant to the application of metadata in this work, in particular, is that metadata *allows for data understanding*, which is essential in data modelling and sharing. The source of uncertainty elicitation technique we present, harnesses metadata to extend such benefit to simulation models. Models annotated with metadata become more re-usable and sharable as users become less dependent on the modeller and the data provider.

3.2 Applying a Metadata Standard

The formalisation and use of metadata standards for data description is necessary to make it more precise and accessible to an audience that is as wide as possible [Campos dos Santos, 2003]. Standards provide a common set of terms naming key data attributes, which, if consistently used as recommended, facilitate data understanding to human users and the development of computer applications that handle metadata. Standard metadata terms can easily be translated to encoding formalisms such as XML, RDF, etc. In recent years we have seen an increase in the adoption of the Dublin Core Metadata Standard (DCMS) [Dublin Core Metadata Initiative, 2003]. The standard is considered to comprise a set of attributes that is simple and effective for describing a wide range of data resources.

We have identified sources of uncertainty by considering characteristics of models' supporting data that may strengthen or weaken one's understanding of or reliance on the model. The DCMS attributes lent themselves well to representation of such sources of uncertainty as metadata. A non-exhaustive list of these attributes is shown below.

Source – Where the data comes from. The range of possibilities includes: *literature*, *field experiments*, *model* (in case of a model element whose values are generated by the model itself), *modeller's definition*, and *modeller's assumption*. Every parameter and initial value has at least this source of uncertainty.

Creator – Who is responsible for making the data available, either through a scientific publication or other forms of communication.

Date – When the data was published or provided.

Identifier – An identifier of the place of publication of the data, e.g., journal, proceedings, book, URL, etc.

Coverage – Spatial location from where the data has been sampled; e.g., a research station, a geographical area like 'central Amazonia', etc.

Description – The state of the system from which the data was collected, by the time of collection. For example, a forest in equilibrium, or a forest that has suffered logging, burning or cultivation.

Description – Sampling information such as which data has been sampled (e.g., nutrients content in litter) and in which sampling campaigns (e.g. before logging, after logging), sampling design used (e.g. census, at random), number of samples, sampling frequency (annually, weekly), etc.

The 'Description' attribute is iterated, as DCMS allows, to distinguish the two different categories of descriptional information. References to and separate manipulation of the iterations' content can be resolved at the implementation level of the sources of uncertainty elicitation technique.

4 Model Equations and Sources of Uncertainty

Besides metadata, we also associate sources of uncertainty with model equations, which regulate the changing values of model elements. Parameters, the leaves of the models' tree-like structures, are constants and as such do not have regulating equations. It is through resolving the equations that the data and metadata associated with the parameters and initial values of state variables are fed into and propagated throughout model simulation. Two sources of uncertainty related to model equations are:

Equation Source – Similarly to data source, this can be *literature*, *field experiments*, *modeller's definition*, and *modeller's assumption*.

Equation Description – Explains what an equation means. E.g., woody_litter_production = 0.50 * biomass_mortality means that woody litter production is derived from the forest biomass mortality, assuming a 50% conversion of biomass to Carbon.

The identified sources of uncertainty – metadata or equation-related – are instantiated to values according to the specific model given. Hereafter, we shall use the unifying term 'endorsements' to refer to instantiated sources of uncertainty.

5 Combining Sources of Uncertainty

The interpreter produces the proof tree of each $goal(E, V, T)$ simulation. The proof tree is a data structure containing information about how the goal has been proved [Apt, 1997] over the model description enriched with the endorsements attached to the model components involved in the proof. It is gradually constructed as the interpreter operates recursively resulting in a nested data structure with hierarchy levels correspondent to the levels of recursion. This data structure is then processed for the endorsements to be combined.

Each run of the interpreter solves a given $goal(E, V, T)$ for a specific model element E. This model element will bear uncertainty that encompasses the uncertainty of all other model elements which directly or indirectly influence it. In system dynamics model diagrams, such influence is represented by the network of flows and links interconnecting model elements. In the numerical simulation, the values of all other model elements connected to a model element are operands in calculating its value (see Figure 1). To provide for this, the proof tree of $goal(E, V, T)$ contains the endorsements of all such model elements connected to E. The endorsements are combined by means of a combination function that takes two sets of values for a source of uncertainty and finds the *union* and *intersection* of these two sets. The combination function is applied progressively, combining values of sources of uncertainty[1] attached to pairs of model elements in the goal's proof tree until it is fully parsed. We call the resulting intersection and union sets of each source of uncertainty its *lower bound* and *upper bound*, respectively.

Let us now see an example of application of the combination function given a certain model. Suppose the goal $goal$('leaf_litter_production',5,V) is given to the interpreter. Also suppose that 'leaf_litter_production' is a flow element in the model, directly influenced by the state variable 'above_ground_vegetation' and the intermediate variable 'leaf_litter_production_coefficient', which, in turn, is directly influenced by the parameters 'measured_leaf_litter_production' and 'measured_biomass'. Figure 2 depicts the example, with the endorsements being propagated and combined bottom up. The Ⓒ symbol in the figure stands for the endorsements combination function. The final lower and upper bounds of each source of uncertainty associated with model elements involved in the simulation are shown at the top, next to the 'leaf_litter_production' flow, the model element given in the simulation goal.

6 Interpreting Combined Sources of Uncertainty

The examples in Figures 1 and 2 are small excerpts from a much larger model of the cycle of Carbon in a logged forest to which we have applied the sources of uncertainty elicitation technique. The fall of leaves from the trees is one of the ways in which Carbon flows through forest systems. In the model, the rate in which leaf litter is produced (represented by the 'leaf_litter_production' flow) is calculated using a coefficient based on measured volumes of leaf litter and biomass, reported in relevant literature. The measurements have been taken from areas nearby the modelled logged forest, which is desirable, however from forests that were, at the time of sampling, in different states: leaf litter was measured in a forest in equilibrium, and biomass in a forest in post-burning state.

[1] Except values of the source of uncertainty 'equation reasoning'.

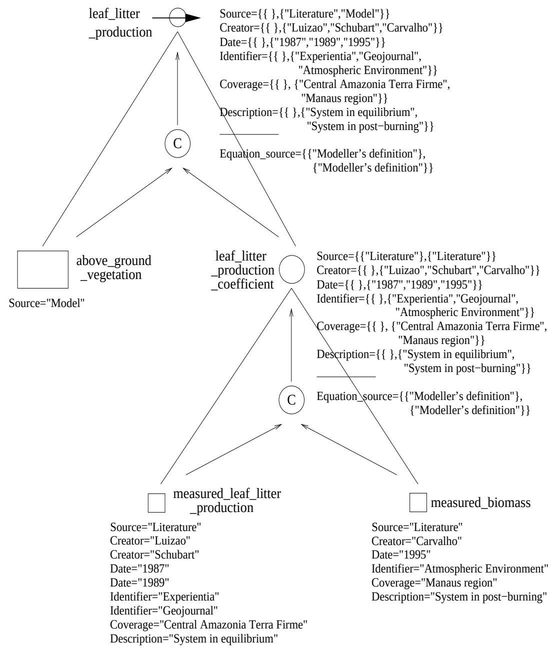


Figure 2: An example of endorsements propagation and combination.

The generic combination function explained in Section 5 calculates endorsements bounds in order to give a user access to a condensed account of information about the origins of parameters that calibrate the model, which can be relevant to the user's assessment of how adequate is that model to his purposes. The interpretation of the endorsements bounds is a subjective, non-deterministic task that is left to the user. For instance, in the 'leaf_litter_production' example, the bounds found for the Coverage source of uncertainty could weaken one's confidence in the model. The values calculated for the flow could be considered more applicable if all the data used had come from the same site as the modelled system or from sites with similar characteristics. If homogeneous endorsements are desirable, the user will be looking for coinciding upper and lower bounds. On the other hand, for the Identifier source of uncertainty in the 'leaf_litter_production' example, diverse and numerous places of publication could be preferable – the more widely published the data the better. In this case, the user would be looking for an upper bound set with many elements and an empty lower bound set associated with the source of uncertainty.

7 An Application Scenario

The technique has been applied to a large system dynamics model, originally implemented in Stella

II[2], of a tropical forest ecosystem in *terra firme*[3] areas of Central Amazonia, with emphasis on nutrient cycling and DBH[4] growth of trees of commercial and non-commercial species. The model was built at INPA (Instituto Nacional de Pesquisas da Amazônia) in Manaus, Brazil, to simulate logging strategies and predict their effect on the forest's sustainability, supporting the design of guidelines for sustainable timber exploitation in the region.

The model was taken as a representative example of complex models which require a varied and wide range of supporting data. Its dataset contained roughly 250.000 data records, regarding vegetation, litter, mesofauna, micro-biology, topsoil chemistry, hydrochemistry, soil physics, and other ecological entities and phenomena. Given such a diverse dataset, in both content and methods, in most simulations, for various model elements in the simulation goal, we obtained empty lower bounds and broad upper bound sets for the model's sources of uncertainty, as it was intuitively expected. This is particularly apparent if in the goal we have a model element that is influenced by many others, which causes the combination mechanism to try and pull together the diversity of the model's endorsements.

8 RESULTS AND CONCLUSIONS

We have presented a computational technique by way of which sources of uncertainty that would otherwise remain implicit in ecological simulation models are identified, propagated by simulation, combined and made available to a model user. This was achieved through a novel application of metadata, namely, its integration with executable simulation models. A modelling feature such as this can enlarge the understanding of an ecological model, as well as enhance its usage as a decision-making or research tool. The feature can also assist modellers in incremental development of models, not only helping to identify gaps in knowledge (as any modelling activity does), but also in pointing out *uncertain* knowledge.

9 FUTURE WORK

The range of models' sources of uncertainty can certainly be widened. We have identified only a sample of them to which applying simple Dublin Core has sufficed. Widening the range of sources of uncertainty will lead to consideration of using qualified or more specific Dublin Core encoding schemes. Moreover, combination heuristics could be tailor-made having in mind specific sources of uncertainty. For the coverage source of uncertainty, for example, combination heuristics could be based on spatial relations between locations, such as distance, subsumption, similarities and differences of environmental conditions in the locations, etc.

We now have the opportunity of exploring the technique within the LBA – Large Scale Biosphere-Atmosphere Experiment in Amazonia – project. We would also like to make the research-prototype system we have into a tool for others to use, possibly with an interface to main-stream ecological modelling graphical tools.

[2] High Performance Systems, Inc.
[3] Area that is not fboded when a river's water level rises.
[4] Diameter at Breast Hight of tree trunks.

ACKNOWLEDGEMENTS

The authors wish to thank FAPEAM (Fundação de Amparo a Pesquisa do Estado do Amazonas) for its sponsorship through the research project Metadata, Ontologies and Sustainability Indicators integrated to Environmental Modelling.

REFERENCES

Apt, K. R. *From logic programming to Prolog*. Prenctice Hall, 1997.

Brilhante, V. *Ontology and Reuse in Model Synthesis*. PhD thesis, School of Informatics, University of Edinburgh, 2003.

Campos dos Santos, J. L. *A Biodiversity Information System in an Open Data/Metadatabase Architecture*. ITC Printing Departament, 2003.

Cohen, P. *Heuristic Reasoning About Uncertainty: an Artificial Intelligence Approach*. Pitman, London, 1985.

Dublin Core Metadata Initiative. Dublin Core Metadata Standard. http://dublincore.org/, 2003. Accessed on 20 Feb. 2004.

Duval, E., W. Hodgins, S. Sutton, and S. Weibel. Metadata principles and practicalities. *D-Lib Magazine*, 8(4), 2002.

Ford, A. *Modeling the environment: an introduction to system dynamics models of environmental systems*. Island Press, 1999.

Parsons, S. and A. Hunter. A review of uncertainty handling formalisms. In *Applications of Uncertainty Formalisms*, pages 8–37, 1998.

Robertson, D., A. Bundy, R. Muetzelfeldt, M. Haggith, and M. Uschold. *Eco-logic: logic-based approaches to ecological modelling*. MIT Press, Cambridge, Massachusetts, 1991.

Assessing the ecological impacts of salinity management using a Bayesian Decision Network

A. Sadoddin[a], **R.A. Letcher**[b] **and L.T. H. Newham**[b]

[a] *Centre for Resource and Environmental Studies,* [b] *Integrated Catchment Assessment and Management Centre, The Australian National University, Canberra ACT 0200, Australia (amir.sadoddin@anu.edu.au)*

Abstract: This paper outlines a component of a study currently being undertaken to provide a new tool for the holistic management of dryland salinity. The Little River catchment in the upper Macquarie River basin of New South Wales (NSW), Australia, is used as a case study. The model uses a Bayesian Decision Network (BDN) approach to integrating the various system components – biophysical, social, ecological, and economic. The method of integration of the system components is demonstrated through an example application showing the impacts of various scenarios on terrestrial and riparian ecology. The paper outlines these scenarios and demonstrates the way in which they are spatially incorporated in the model. The ecological impacts of management scenarios have been assessed using a probabilistic approach to evaluating ecological criteria for a range of management actions compared with the present situation.

Keywords: Salinity management, Bayesian decision networks, terrestrial and riparian ecology

1. INTRODUCTION

Salinisation is a major environmental problem affecting land and water resources in Australia. Employing a holistic approach to consider all components in a catchment system, in a cause and effect context, is essential to address this deteriorating situation. This paper presents a method for investigating some of the ecological impacts of salinity management options in the Little River Catchment as a component of an integrated model of salinity management at the catchment scale. The model uses a Bayesian Decision Network (BDN) approach to integrating the various system components – biophysical, social, ecologic, and economic.

2. CASE STUDY: THE LITTLE RIVER

The Little River is a tributary of the Macquarie River lying southwest of Wellington in central western NSW and is part of the headwaters of the Murray-Darling Basin. The catchment covers an area of 2310 km^2. Approximately 80% of the vegetation communities in the catchment have been disturbed for agricultural purposes (Seddon *et al.*, 2002) and there are severe salinity outbreaks in some parts of the catchment. It is estimated that approximately 12% of the salt load of the Macquarie River at Dubbo originates from the catchment (IVEY & DPMS, 2001: 6.6). Assessment of saline sites in the catchment between 1988 and 1998 estimates that the spatial extent of saline lands increased by a factor of 4.6 in this period of time (Nicholson and Wooldridge, 2001).

3. A CONCEPTUAL MODEL FOR SALINITY MANAGEMENT

In this study a Bayesian decision network approach is applied to consider the influence of management options on environmental, physical, social, and economic outcomes. Sadoddin *et al.*, (2003) describes the development of the BDN approach and advantages of the use of them in more detail. Figure 1 shows the current conceptual framework underlying the BDN being developed for the catchment. This framework incorporates ecological, physical, economic and social aspects of the salinity problem. This paper focuses on evaluating the links between management decisions and terrestrial and riparian ecological impacts. Each set of salinity management actions corresponds to spatial land cover patterns across the catchment, and in turn has potential impacts on terrestrial and riparian ecology. Several criteria have been set up to assess the ecological consequences of salinity management using a probabilistic approach. The links between salinity management and terrestrial and riparian ecology in the system are a key component of the integrated model (see ecological subset marked out by bold boxes in Figure 1). In order to construct these links in the integrated model, conditional probabilities tables must be derived linking the vegetation management options with spatial land cover patterns and then with the impacts on terrestrial and riparian habitats. These estimates, along with joint probability distributions for the variables in the ecological subset, provide required components to calculate total probability distributions for the state variables of the BDN model.

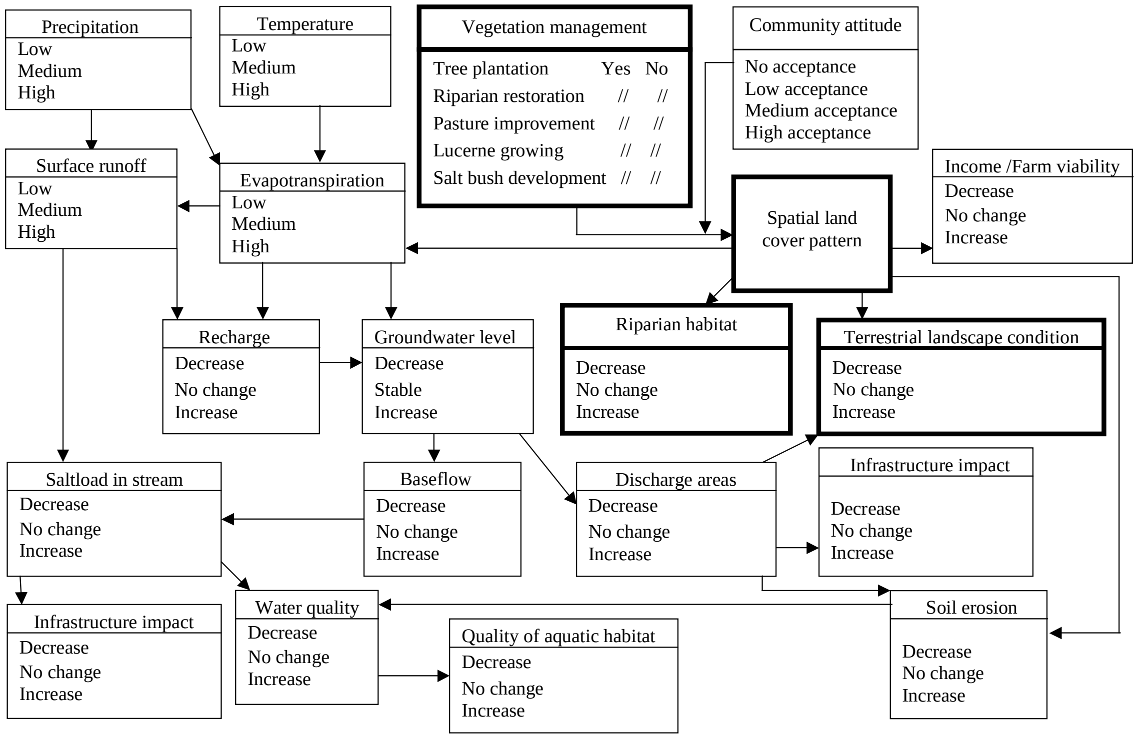


Figure 1. BDN conceptual framework for the Little River Catchment adapted from Sadoddin *et al.*, (2003)

4. SCENARIOS

As shown in Figure 1 a specific node named “spatial land cover pattern” has been incorporated into the conceptual model reflecting the influence of different combinations of management actions on land cover spatially throughout the catchment. To simulate potential spatial land cover patterns under different management options, a significant effort has been made to determine the areas in the catchment suitable for each of the land cover options. Table 1 summarises the scenario rules for each of salinity management actions in the catchment.

Only areas with annual rainfall greater than 700 mm are suitable for commercial tree plantation in the region (Hall *et al.,* 2003). Only a very narrow strip (approximately 6% of the catchment) in the south of the catchment meets this criterion. Hence, implementation of commercial tree planting in the catchment has little economic justification due to the small area and the large distance to sawmills and markets (Hall *et al.,* 2003). Therefore, in this study, salinity management by tree plantation action is assumed to consist of local native trees rather than commercial species. The potential riparian area has been predicted by Seddon *et al.,* (2003) using a combination of geology, soil, elevation, slope and topographic position layer maps using a statistical analysis. Pasture improvement and lucerne establishment have been considered as the management actions applicable in areas currently under native pasture and improved pasture respectively. The area potentially suitable for planting saltbush has been determined by applying the FLAG model in the catchment (Dowling, 2000). FLAG model is an approach incorporating terrain analysis that uses a number of topographic indices derived from elevation map to provide a wetness index map indicative of potential groundwater discharge and salinity.

Considering all five management actions given in Table 1, the total number of different combinations of the actions gives 31 scenarios (2^5-1) in addition to the base case scenario (current situation). Equation 1 can calculate the number of each individual scenario considered in this paper.

$$S = 1 + B + 2L + 4I + 8R + 16F \quad (1)$$

where S is scenario number, B, L, I, R, F are different management actions (see Table 1). B, L, I, R, F equal 1 if Yes, otherwise they equal 0.

Table 1. Scenario rules for salinity management

Management action	Rules for distribution	% Of suitable areas
Non-commercial tree plantation (F)	Entire catchment except: -areas currently under forest - riparian areas	10
Riparian restoration (R)	Only in riparian zone, not in areas currently with trees	50
Pasture improvement (I)	Only in areas currently under native pasture	50
Lucerne establishment (L)	Only in areas currently under improved pasture	10
Saltbush development (B)	Only in potential waterlogged areas, not in areas currently under trees	50

For each scenario, the management actions described in Table 1, have been implemented in the following order of land allocation: 1) tree plantation, 2) riparian restoration, 3) saltbush development, 4) pasture improvement, and 5) lucerne establishment. Using spatial datasets in raster format, land allocation for each scenario was determined on a grid basis with cell size of 10 hectares. This grid cell size was selected to establish appropriate habitat size and considering realistic on-ground management interventions (Williams *et al.,* 2002). In order to lay out the frequency distribution of outcomes of salinity management actions, some 50 samples of each scenario option have been randomly synthesised. This has been carried out by using established GIS datasets including current land cover, and the five maps of potential areas for each management option. Land cover scenario maps have been generated in ARCINFO using ARC Macro Language (AML) code. The ecological indicators described in the next section were also evaluated for each of the 50 samples of each scenario to derive a probability distribution of ecological impacts for each scenario.

5. MODELLING IMPACTS ON TERRESTRIAL AND RIPARIAN ECOLOGY

Biodiversity is a broad and complex ecological concept that can be focused in different ways, and at different organizational levels (for example. genetic, species, population), and also with different degrees of complexity (for example Chevalier *et al.,* 1997). Measuring or modelling biodiversity is extremely difficult because of these problems. As such, the model considered in this paper uses several indicators of impacts to demonstrate conditions that are likely to affect biodiversity and ecosystem health rather than attempting to model biodiversity impacts directly. Indirectly, landscape diversity and forest fragmentation are important concepts in revegetation and salinity management. Since there is a relationship between the spatial configuration and composition of landscape elements and biological diversity, this concept is addressed in the framework instead of targeting biodiversity directly (McGarigal *et al.,* 1994). A significant number of mathematical indices have been developed and appeared in the literatures that allow the description of different aspects of landscape diversity. Fragmentation indices can be applied to assess the condition of ecosystem processes and quality of habitat for a significant percentage of all mammal, reptile, bird, and amphibian species that are found in forest habitats (Riitters *et al.,* 2002). There are however, relatively few metrics sufficient to capture landscape pattern (Lausch and Herzog, 2002). This section describes the indicators of terrestrial and riparian ecosystem health used to assess ecological impacts in the integrated model. Four indices have been chosen to represent the impact of salinity management on terrestrial and riparian ecosystems in the catchment. For all indices the index i denotes the scenario number (1, …, 32) while j denotes the sample (1, …, 50). All indices are dependent on i and j. The criteria have been evaluated for the current situation and also for synthesised land cover maps corresponding to different management actions across the catchment as a whole. The impact of management change is measured as a percentage change from the base case situation. That is, the probability distribution of impact for each indicator is calculated from $Y_{i,}$ where

$$Y_{i,j} = \left(\frac{I_{i,j} - I_1}{I_1}\right)*100 \qquad (2)$$

$I_{i,j}$ is index value for each sample j (j=1, …, 50) of each scenario i (i =2, …, 32)

I_1 is index value for the base case, and $Y_{i,j}$ is percentage change from the base case.

a) Weighted Mean Patch Size Index (WMPSI $_{ij}$)

The weighted mean patch size index measures the direct impact of each management scenario on patch sizes for each land cover type across the catchment. It also reflects the biodiversity conservation value of each land cover. It has been selected because in a patchy landscape, patch size is an important criterion in determining what species of animals are able to survive. The negative effect of fragmentation increases where the patch size is smaller. Also there is a direct correlation between the patch size and the positive influence of ecotone development. Thus, in a small patch the positive effects of ecotone development are less than for a bigger patch (Odum, 1993). The equation used to calculate WMPSI is:

$$WMPSI = \sum_{m=1}^{7} \frac{\alpha_m}{n_m} \sum_{k=1}^{n_m} P_{k,m} \qquad (3)$$

where m is type of land cover (see Table 2)
n_m is number of patches of land cover m type, $P_{k,m}$ is size of each of the patches, (k = 1, …, n_m), α_m is the weight value for each land cover type m.

The weight values used for different land covers are given in Table 2. These subjective values have been derived from expert ecological knowledge and are subject to change in different contexts. In particular, the weight values are sensitive to the history of management. An increase in WMPSI denotes an improvement of biodiversity conservation value of the region.

Table 2. Weight values for different land covers

Management action	m	α_m
Trees	1	1
Riparian	2	1
Lucerne	3	0.1
Improved pasture	4	0.3
Native pasture	5	0.6
Crops	6	0.05
Saltbush	7	0.3

b) Weighted Land Cover Area Index (WLCAI$_{ij}$)

The weighted land cover area index is an aggregated measure of the extent of natural versus modified landscapes in the catchment. It has been selected because measuring the area of different land covers and considering a corresponding weight provides an estimate of the degree of naturalness in the catchment. In a biodiversity conservation context, measuring the degree of 'naturalness' provides useful information to contribute to broader conservation value assessments (Parkes *et al.,* 2003). Additionally, a comparison between different land cover types is another essential process that is required for assessing native vegetation quality ((Parkes *et al.,* 2003). Pre-clearing distribution of vegetation communities in the Little River Catchment predicated by Seddon *et al.,* (2003) shows that the entire catchment was covered by six native tree communities. This can be seen as a benchmark representing the average characteristics of mature and long- undisturbed stands of vegetation communities in the catchment (Parkes *et al.,* 2003). The equation used to calculate WLCAI is:

$$WLCAI = \sum_{m=1}^{7} \alpha_m \sum_{k=1}^{n_m} P_{k,m} \qquad (4)$$

Variables are as defined previously and the same weight values for each land cover (see Table 2) are used as for WMPSI. An increase in WLCAI denotes improving the catchment situation in terms of degree of naturalness.

c) Forest Connectivity Index (FCI$_{ij}$)

The FCI measures the spatial pattern of forested areas. Since a given amount of forest can be arranged in many patterns and the spatial pattern has significant effect on fragmentation characteristics, an index measuring forest connectivity has been used. When the spatial pattern of forest changes, the wellbeing of forest dependant organisms and competitive arrangements among populations will be affected (O'Neill *et al.,* 1988 cited in Riitters *et al.,* 2002). Fragmentation also increases the energy cost/benefit ratio of movement due to contortion in movement pattern (Gardner *et al.,* 1991 cited in Riitters *et al.,* 2002). The FCI has been measured on raster land cover map. In order to calculate forest connectivity index, at first each pixel edge needs to be labelled according to the cover types of the two adjacent pixels. Then FCI is calculated as a ratio of the number of pixel edges in the landscape that border two forest pixels to the total number of pixel edges that have a forest pixel on at least one side (Riitters *et al.,* 2002). With measuring FCI, the degree of isolation or integration of forest can be quantified. The equation used to calculate FCI is:

$$FCI = e_{pf,pf} / (e_{pf,pn} + e_{pf,pf}) \qquad (5)$$

where $e_{pf,pf}$ is number of edges between two forest pixels, $e_{pf,pn}$ is number of edges between forest pixels and non-forest pixels, p_n is a non-forest pixel or land cover with m=3, …, 7. p_f is a forest pixel or land covers with m = 1 & 2 (see Table 2). An increase in FCI denotes a higher connectivity of forest pixels indicating a higher degree of integration of forest.

d) Riparian Proportion Index (RPI$_{ij}$)

Riparian zones can be considered as a boundary between terrestrial and aquatic ecosystems. Forests along waterways, also known as riparian forests, are an important resource that function to maintain the integrity of the stream channel, reduces the impact of pollution sources and supply food and habitat resources to wildlife (Newsom *et al.,* 2001). The proportion of the riparian zone that is forested is a useful indicator of ecosystem health. One of the salinity management actions investigated in this study is reforestation along streams in the catchment. Water quality and habitat benefits have direct relationships with riparian proportion along stream networks (Newsom *et al.,* 2001). Thus, the riparian proportion index is calculated as:

$$RPI = \frac{r_2}{\sum_{m=2}^{7} r_m} \qquad (6)$$

where r_m is number of grid cells with land cover m along waterways, r_2 is number of grid cells with riparian forests.

The influence of the various management actions on the ecological endpoints in the BDN model framework have been estimated through calculation of the change in the indices from the base case scenario. For each ecological index, the values of Y_i (see Equation 2) for all management scenarios (2, …, 32) across 50 samples have been grouped using five class intervals. Probability distributions for each scenario have then been extracted.

6. RESULTS

The resultant probability distributions for each ecological indicator were placed into several categories according to the level of effects of the management scenarios on the indicators. The categories were ranked from best to worst ecological outcomes. A lower category number indicates less degradation in relation to WMPSI and FCI and/or greater improvement in relation to WLCAI and RPI.

The results of the scenarios classification for all indices are illustrated in Figures 2 to 5. The change from the base case for WMPSI, WLCAI, FCI, and RPI covers a range of 31.3, 20.7, 5.5, and 127.5 percent respectively. The results show that the response of the four indices to a management scenario does not occur in the same direction. In general, applying management actions that increase the number of patches in the catchment decreases the values of mean patch size and forest connectivity indices. This is particular as so for tree plantation action, because the weight value for trees is greater than for other land covers (see Table 2). In contrast, the land cover area and riparian proportion indices improve under management scenarios associated with tree plantation. The influence of implementing the scenarios associated with improved pasture and lucerne on land cover area index is not large. This is because of the lower weight values for improved pasture and lucerne.

WMPSI and FCI are sensitive to the number of patches, while the other indices are not. The general trends in the data show that there is a positive relationship between the scenario number and management scenario 'category' for both mean patch size and forest connectivity indices (Figures 2 and 4). The reverse trend can be seen for weighted land cover area, and riparian proportion indices (Figures 3 and 5).

The mean values of the indicators WLCAI, FCI, and RPI over 50 samples clearly indicate that four groups of management scenarios can be identified. Table 3 gives the four scenario groups. In particular, for FCI and RPI, the variation of the means inside each group is negligible. While, for WMPSI two distinct groups can be identified and there is significantly a continuous change in the value of WMPSI in the second group.

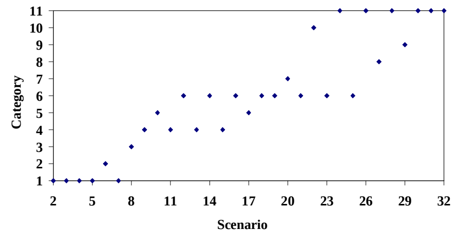


Figure 2. Rank of management scenarios (WMPSI)

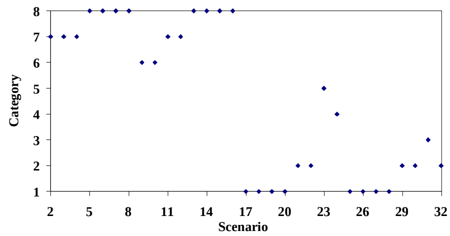


Figure 3. Rank of management scenarios (WLCAI)

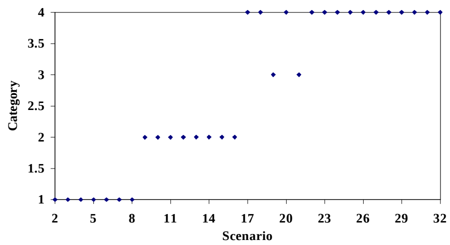


Figure 4. Rank of management scenarios (FCI)

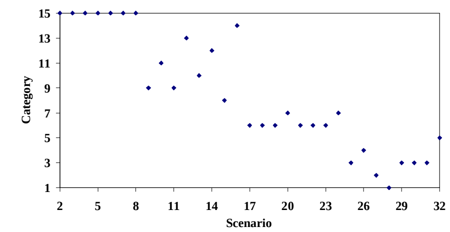


Figure 5. Rank of management scenarios (RPI)

Table 4 gives the range of standard deviation values for each ecological index and the number of class intervals with non-zero probability values.

Table 3. Groups of management scenarios

Group	Scenario number	Key attribute
1	2-8	No tree plantation in terrestrial or riparian areas
2	9-16	No tree plantation in terrestrial area
3	17-24	No tree plantation in riparian area
4	25-32	Tree plantation in terrestrial and riparian areas

Table 4 shows that the values of indicators over the 50 samples are clustered together. In addition, the maximum numbers of class intervals for which non-zero probability values exist is two, reflecting relative certainty in the derived probability distributions.

Table 4. Statistical information for indices

Indicator	No. of non-zero class intervals	Min. st.dev (%) for all scenarios	Max. st.dev (%) for all scenarios
MPSI	2	0.11	0.51
LCAI	2	0.02	0.32
FCI	2	0.00	0.11
RPI	2	0.00	7.10

7. DISCUSSION AND CONCLUSIONS

The implementation of management scenarios for dryland salinity can result in many different spatial patterns of land cover. A BDN is an appropriate approach to deal with the spatial variability associated with different land cover patterns from management implementation. It is also an appropriate tool to systematically represent the uncertainties associated with the different components in the model. This research uses four indicators to assess ecological consequences of salinity management actions using a probabilistic approach. Changes in the indices do not occur in the same direction across all scenarios. Thus, a method for interpreting these changes into "better" or "worse" ecological outcomes must also be developed. This is required so that managers are given clear direction on the impacts of scenarios in the model. This will be achieved through consultation with ecologists on appropriate weights to recombine changes in these indices to achieve a qualitative measure of ecological impact.

8. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance of Dr Barry Croke, Dr Sue Briggs, Rassoul Mahini, Ioan Fazey, and Joern Fischer in developing the system understanding and also Paul Carlile, Julian Seddon and Trevor Dowling for providing their models and spatial data.

9. REFERENCES

Chevalier, J., J. Cracraft, F. Grifo and C. Meine, Biodiversity, science and the human prospect, Center for Biodiversity and Conservation, American Museum of Natural History, 1997

Dowling, T., FLAG analysis of catchments in the Wellington region of NSW, CSIRO Land and Water Consultancy Report, 2000

Hall, N., W. Watson and M. Oliver, Farm economic analysis: Little River catchment, Integrated Catchment assessment and Management Centre, The Australian National University, Report No. 2003 TARGET 6, 2002

IVEY & DPMS, Wellington Dubbo Landcare Management Regional Plan, Stage One Report, 2000

IVEY & DPMS, Mid-Macquarie Landcare Regional Plan, Stage Two Report, 2001

Lausch, A. and F. Herzog, Applicability of landscape metrics for the monitoring of landscape change: issues of scale, resolution and interpretability, *Ecological Indicators*, 2, 3-15, 2002

Mcgarigal K. and B. Marks, FRAGSTATS: spatial pattern analysis program for quantifying landscape structure, United State Department of Agriculture, 1995

Newsom, A., C. Hershner and D. Schatt, Riparian forest buffer restoration targeting for the Rappahannock River watershed, Virginia Institute of Marine Science, Technical Report No. 01-2, 2001

Nicholson, A. and A. Wooldridge, Dryland salinity in the Lachlan catchment, Cowra Workshop Summary Report, 2001

Odum, E. P., *Ecology and our endangered life-support systems*, Sinauer Publication, 1993

Parkes, D., G. Newell and D. Cheal, Assessing the quality of native vegetation: The 'habitat hectares' approach, *Ecological Management & Restoration*, 4, S29 - S38, 2003

Riitters, K. H., J. D. Wickham, R.V. O'Neill, K. B Jones, E. R. Smith, J. W. Coulston, T. G. Wade and J. H. Smith, Fragmentation of continental United States forests, *Ecosystems*, 5, 815-822, 2002

Sadoddin, A., R. Letcher and A. Jakeman, A Bayesian decision network approach for salinity management in the Little River catchment, NSW., International Congress on Modelling and Simulation, 953-959, 2003

Seddon, J., S. Briggs and S. Doyle, Little River catchment biodiversity assessment, TARGET Project Report, 2002

Williams, J., R. Bird and K. Junor, Native vegetation 3: revegetating the farm, *On farm Magazine*, 2002

A Statistical Input Pruning Method for Artificial Neural Networks Used in Environmental Modelling

G. B. Kingston, H. R. Maier and M. F. Lambert
Centre for Applied Modelling in Water Engineering, School of Civil and Environmental Engineering, University of Adelaide, Adelaide, SA, 5005, Australia. Email: gkingsto@civeng.adelaide.edu.au

Abstract: Artificial neural networks (ANNs) provide a useful and effective tool for modelling poorly understood and complex processes, such as those that occur in nature. However, developing an ANN to properly model the desired relationship is not a trivial task. Selection of the correct causal inputs is one of the most important tasks faced by neural network practitioners, but as knowledge regarding the relationships modelled by ANNs is generally limited, selecting the appropriate inputs is also one of the most difficult tasks in the development of an ANN. Many of the methods available for assessing the significance of potential input variables do not consider the uncertainty or variability associated with the input relevance measures used and, consequently, this important factor is neglected during hypothesis testing. In this paper a model-based method is presented for pruning ANN inputs, based on the statistical significance of the relationship between the input variables and the response variable. The approach uses Bayesian methods to estimate the input relevance measure such that the uncertainty associated with this parameter can be quantified and hypothesis testing can be carried out in a straightforward and statistical manner. The proposed methodology is applied to a synthetically generated data set and it is found to successfully identify the 3 relevant inputs that were used to generate the data from 15 possible input variables that were originally entered into the ANN.

Keywords: Artificial neural networks; Input selection; Pruning; Bayesian; Environmental modelling

1. INTRODUCTION

Artificial neural networks (ANNs) provide a useful and effective tool for modelling the complex and poorly understood processes that occur in nature, as they are able to extract functional relationships between model inputs and outputs from data without requiring explicit consideration of the actual data generating process. However, in order to achieve a good representation of the data-generating relationship, an ANN needs to contain all information relevant to the problem. Therefore, selection of the correct causal inputs is one of the most important tasks faced by neural network practitioners.

Knowledge about exact environmental relationships is generally lacking and, consequently, it is difficult to select the correct set of inputs that are relevant to the process. Often, little consideration is given to this task as it has been assumed that, because ANNs are a data driven approach, the relevant inputs will be determined automatically during the modelling process (Maier and Dandy 2000). However, the number of potential inputs can be large for complex environmental systems, particularly when the process under study is dynamic and requires the inclusion of time-lagged input variables. Presenting all potential inputs to an ANN increases the size and complexity of the network, which slows training and increases the amount of data required to estimate the free parameters, or weights, of the network. Moreover, the inclusion of irrelevant inputs can confuse the training process, resulting in spurious correlations in the data being modelled, rather than the actual underlying process.

To help ensure that a good representation of the underlying process is obtained, it is necessary to consider methods for assessing the statistical significance of potential inputs. This is particularly important when the model is used to acquire knowledge about the system, rather than being used solely for predictive purposes. In this paper a model-based method is presented for pruning ANN inputs, based on the statistical significance of the relationship between the inputs and the response variable. This approach uses Bayesian methods to estimate the input relevance measure such that the uncertainty associated with this parameter can be quantified and hypothesis testing can be carried out in a straightforward manner. The method is applied to a synthetically generated data set in order to demonstrate its application.

2. BACKGROUND

2.1 Input Significance Testing

According to Refenes and Zapranis (1999), determining the significance of a potential ANN input involves the 3 following stages:

1. Defining the relevance of the input to the model.
2. Defining the variance of the relevance measure.
3. Testing the hypothesis that the input is irrelevant to the model.

There have been a number of methods proposed in the literature for addressing the first stage of this problem. These include sensitivity analyses (Lek et al. 1996), assessing the weights of the trained network (Garson 1991), and stepwise methods where the importance of an input is determined by the change in predictive error when it is added to or subtracted from the network (Maier et al. 1998). Although these methods provide a means of determining the overall influence of a potential input, they are generally based on the single-valued weights of a trained ANN and, therefore, do not facilitate the further two stages of the problem. Consequently, inputs are included or excluded from the model in a subjective manner, depending on their effect on the output or model error, as there is no way to statistically test their significance.

Olden and Jackson (2002) introduced a randomization method for statistically assessing the importance of an input based on the comparison of the input's overall connection weight (*OCW*) with a statistical measure of irrelevance. The overall connection weight of an input is the sum of the products of the weights between an input and the output. With reference to Figure 1, the *OCW* of input 1 can be calculated by determining $c_{A,1}$ and $c_{B,1}$, which are the contributions of input 1 via hidden nodes A and B, respectively, and summing them to obtain OCW_1 as follows:

$$\begin{aligned} c_{A,1} &= w_{A,1} \times w_{O,A} \\ c_{B,1} &= w_{B,1} \times w_{O,B} \\ OCW_1 &= c_{A,1} + c_{B,1} \end{aligned} \qquad (1)$$

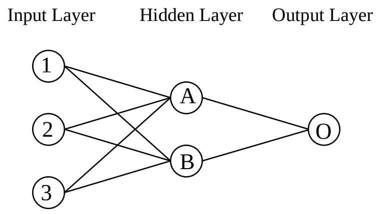


Figure 1. Example ANN structure

Under this paradigm the statistical measure of irrelevance is determined by removing any functional structure between the model inputs and outputs and using a bootstrap procedure to obtain a probability density function (pdf) of the input's *OCW* when there was no remaining relationship between it and the output. Inputs were considered irrelevant if the original *OCW* of the input was not significantly different from the *OCW* when the relationship had been removed.

Although the method of Olden and Jackson (2002) addresses each of the 3 stages of input significance testing, its success is reliant on finding a single set of optimal weights that correctly approximate the underlying function. Due to complications during training and the inherent variability of the underlying process itself, it is unlikely that a single optimal weight vector will be found, particularly when irrelevant inputs are included in the model. It is therefore important to consider a distribution of the network weights such that the uncertainty associated with finding an optimal weight vector can be incorporated into the input significance tests. By describing the weights as distributions a range of possible weight values is considered, preventing one, possibly incorrect weight vector, from completely dominating the calculated *OCW*s, which are fundamental in testing the relevance of the inputs.

2.2 Bayesian Weight Estimation

Bayesian methodology was first applied to estimate the weights of an ANN by Mackay (1992) and Neal (1992). It provides an approach for explicitly handling uncertainty in the weights by considering the weight vector, **w**, as a random variable. Using Bayes' Theorem, the posterior weight distribution, P(**w**|**y**,**x**), may be inferred from the data as follows:

$$P(\mathbf{w}|\mathbf{y},\mathbf{x}) = \frac{P(\mathbf{y}|\mathbf{x},\mathbf{w})P(\mathbf{w}|\mathbf{x})}{P(\mathbf{y}|\mathbf{x})} \quad \text{or} \qquad (2)$$
$$P(\mathbf{w}|\mathbf{y},\mathbf{x}) \propto P(\mathbf{y}|\mathbf{x},\mathbf{w})P(\mathbf{w}|\mathbf{x})$$

where **w** is a vector of ANN weights, **y** is a vector of N observations and **x** is a set of N input vectors. In (2), P(**w**|**x**) is the prior weight distribution, which describes any knowledge of **w** before the data were observed. P(**y**| **x**,**w**) is known as the likelihood function. This function uses information obtained by comparing the model predictions to the observed data to update the prior knowledge of **w**. By assuming that each observation is independently drawn from a Gaussian distribution, the likelihood function can be described by:

$$P(\mathbf{y}|\mathbf{x},\mathbf{w}) = \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{y_i - f(\mathbf{x}_i,\mathbf{w})}{\sigma}\right)^2\right) \quad (3)$$

where $f(\mathbf{x}_i,\mathbf{w})$ is the ANN output for the *i*th input vector and σ is the standard deviation of the model residuals.

2.2.1 The Metropolis Algorithm

The high dimensionality of the conditional probabilities in (2) makes it difficult to calculate the posterior weight distribution analytically. Consequently, methods have been introduced to approximate (2). Neal (1992) introduced a Markov chain Monte Carlo (MCMC) implementation to sample from the posterior weight distribution such that P(**w**|**y**,**x**) could be evaluated numerically.

The Metropolis algorithm is a commonly used MCMC approach, which generates samples from the posterior distribution of an unknown variable, e.g. ANN weights. As it is difficult to sample from the complex posterior distribution directly, this method uses a simpler, symmetrical distribution (a multinormal distribution was used in this study), known as the proposal distribution, to generate candidate weight vectors. By employing an adaptive acceptance-rejection criterion the random walk sequence of weight vectors converges towards the posterior distribution over many iterations. Details of the computational implementation of the Metropolis algorithm can be found in Thyer et al. (2002).

The covariance of the proposal distribution has important implications on the convergence properties and efficiency of the Metropolis algorithm. Poor selection of this parameter may result in an insufficient number of generated samples to adequately represent the posterior distribution. Haario et al. (2001) introduced a variation of the Metropolis algorithm that was developed to increase its convergence rate. In this algorithm the proposal distribution continually adapts to the posterior distribution by updating the covariance at each iteration based on all previous states of the weight vector. The adaptation strategy ensures that information about the posterior distribution, accumulated from the beginning of the simulation, is used to increase the efficiency of the algorithm. This algorithm is known as the adaptive Metropolis algorithm.

3. METHODS

The proposed input selection method is a model-based pruning approach, where the initial ANN includes all potential inputs and "irrelevant" inputs are eliminated, or pruned, from the network throughout the process. The method addresses the 3 stages of input significance testing in a systematic and consistent manner by using the Bayesian framework to estimate distributions of the network weights.

The overall connection weight (*OCW*) measure, used by Olden and Jackson (2002), is employed to quantify the input variables' relevance to the model. The *OCW* of an input measures the strength and direction of the relationship between that input and the output. If this measure is approximately equal to zero there is no relationship between the input and the response variable.

The adaptive Metropolis algorithm is used to generate samples from the posterior weight distribution. The corresponding *OCW* values are then calculated for each sampled weight vector, producing empirical distributions of the *OCW*s, which capture the variation in these relevance measures. In this study, a uniform prior distribution over the range [-3,3] was assumed for each weight. After a warm-up period of 30,000 iterations, 100,000 weight vectors were sampled from the posterior weight distribution and the corresponding *OCW*s calculated.

By having distributions of the input relevance measures, Bayesian probability intervals can be formed in order to test the hypothesis that an input is irrelevant to the model (*OCW*=0). The probability intervals are initially formed around the mode of the pdf such that 100(1-*a*)% of the distribution is contained within the interval, where *a* is the significance level. If zero lies within these bounds, the hypothesis that the input is irrelevant to the model is true.

The weights of an ANN are generally small values centred around zero, thus it is likely that the *OCW*s are also relatively close to zero. Moreover, it is expected that the initial *OCW* distributions will be quite variable due to the inclusion of irrelevant inputs in the model. Therefore, it is likely that a number of the initial *OCW* distributions will contain zero regardless of whether the input is

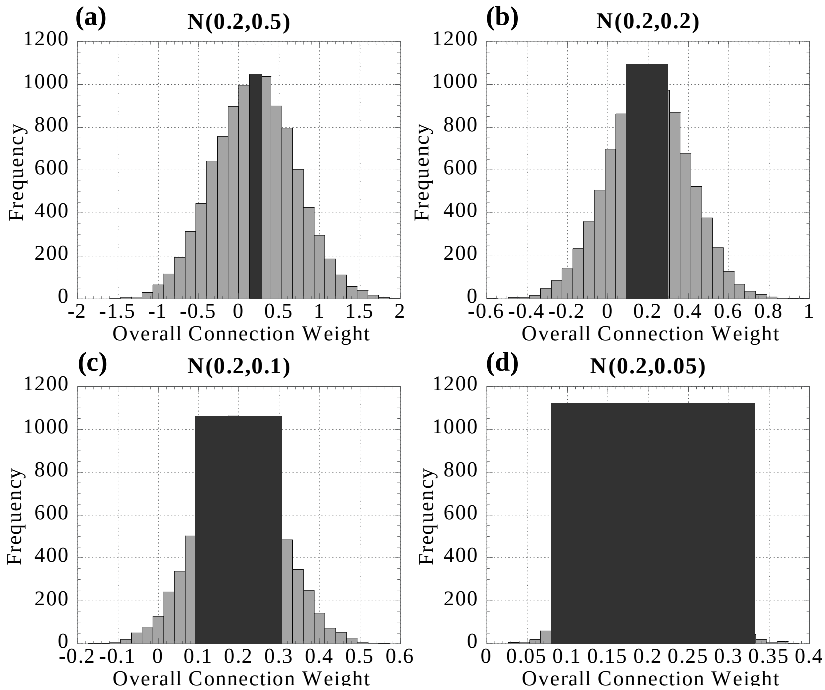


Figure 2 Increasing width of probability intervals (shaded region) at different stages of the pruning process. The standard deviation of the *OCW* reduces from 0.5 in (a) to 0.05 in (d) when the relationship between the input and the output becomes well defined. Eventually the input is tested for its dissimilarity to 0.

relevant to the model or not. To ensure that important inputs are not pruned in these initial stages when the relationship is poorly defined, inputs are only pruned from the network when their *OCW*s are statistically similar to zero at a high significance level (e.g. 95%). As the process continues and irrelevant inputs are pruned from the network, the relationship between inputs and outputs becomes better defined and the variance in the *OCW*s reduces. This means that the pdf of a significant input's *OCW* is less likely to contain zero as more irrelevant inputs are pruned. Therefore, the significance level at which inputs are tested for their similarity to zero may be reduced gradually throughout the process. Eventually inputs are tested for their statistical dissimilarity to zero (*OCW*=0 at 5% ⇔ *OCW*≠0 at 95%), which ensures that only inputs having a significant relationship with the output are included in the model. This process is illustrated in Figure 2 where the pdf of an *OCW* is Gaussian with a mean of 0.2. The standard deviation of the distribution decreases from 0.5 in Figure 2 (a) to 0.05 in Figure 2 (d) as indicated by the scale on the x-axis. As the variance decreases it becomes more evident that the *OCW* is significantly different from zero. Even though the probability intervals are eventually widened to include 95% of the distribution, zero is never included within this range, indicating that the input is statistically significant. However, if the bounds were set wider when the variance was large this input would have been considered irrelevant.

The following process is carried out until all inputs remaining in the model are statistically significant:

1. Sample 100,000 weight vectors from the posterior weight distribution and calculate the corresponding *OCW*s for each input, forming empirical distributions of the *OCW*s.
2. Test the hypothesis that the inputs are irrelevant (beginning at the 95% significance level) by constructing probability intervals around the mode *OCW* value for each input. If zero is included within these intervals the input is considered irrelevant and is pruned from the network. If no inputs are irrelevant at the current significance level, widen the bounds to include a greater proportion of the distribution (e.g. decrease the significance level by 5-10%).
3. Repeat steps 1 and 2 until the only remaining inputs have *OCW*s that are statistically different from zero at a high significance level (e.g. 95%) or, in other words, that the *OCW* is equal to zero at a low significance level (e.g. 5%).

4. CASE STUDY

4.1 Data

Autoregressive (AR) models are commonly used to model natural systems (e.g. hydrological time series data). The autoregressive model, AR(9), was used to generate a set of synthetic time series data according to:

$$x_t = 0.3x_{t-1} - 0.6x_{t-4} - 0.5x_{t-9} + \varepsilon_t \quad (4)$$

where ε_t is a random noise component with distribution N~(0,1). This model was selected for demonstrating the proposed input selection method as it depends on more than one input variable and has known dependence attributes. Moreover, the use of synthetic data enabled the generation of as much data as was required. 400 data points were generated as this number was considered to represent a realistic data set size for environmental data, which are generally limiting.

4.2 ANN Model

Although the response variable x_t only depends on inputs x_{t-1}, x_{t-4} and x_{t-9}, 15 inputs from x_{t-1} to x_{t-15} were included in the ANN in order to determine whether the proposed input selection method could identify the 12 irrelevant inputs that needed to be pruned from the model. An ANN with 1 hidden layer with 2 hidden layer nodes was used to model the data. It should be noted that due to the large number of inputs included in the model, and thus the large number of free parameters, it is likely that the model would overfit to noise in the data in the initial stages of the pruning process. This amplifies the need to only prune those inputs that have *OCW*s statistically similar to zero at a high significance level in the initial stages. Initially, testing the hypothesis of input irrelevance began at the 95% significance level (i.e *OCW*=0 with 95% probability). However, there were no irrelevant inputs at this level and the significance was decreased in increments of 5% until there were one or more irrelevant inputs. This occurred at the 85% significance level.

5. RESULTS & DISCUSSION

The results of the input selection process are given in Table 1. The final inputs remaining in the model (relevant at the 95% significance level) were x_{t-1}, x_{t-4} and x_{t-9} which are the correct causal inputs for the AR(9) data. Therefore, the proposed method was able to properly identify the irrelevant inputs such that they could be pruned from the ANN. It can be seen in Table 1 that 7 runs were required to achieve the final model.

Plots of the *OCW* distributions of inputs x_{t-3} and x_{t-9} are shown in Figure 3. Figures 3 (a) and (b) show the *OCW* distributions of x_{t-3} after run 1 and run 6 respectively, while Figures 3 (c) and (d) give the same plots for x_{t-9}. It can be seen that the variances

Table 1 Results of the pruning process. The remaining inputs were x_{t-1}, x_{t-4} and x_{t-9}.

Run no.	No. of initial inputs	Significance level[a]	Irrelevant inputs
1	15	85%	x_{t-6}, x_{t-13}, x_{t-14}
2	12	80%	x_{t-8}
3	11	80%	x_{t-7}
4	10	75%	x_{t-5}
5	9	5%	x_{t-2}, x_{t-10}, x_{t-11}, x_{t-12}, x_{t-15}
6	4	5%	x_{t-3}
7	3	5%	-

[a]with which the *OCW*s of the pruned inputs were similar to 0

of the *OCW* distributions are quite large when 15 inputs were included in the model (Figures (a) and (c)). Additionally, it appears that x_{t-3} is significant to the model at this stage, which indicates that the underlying relationship has been incorrectly approximated due to the inclusion of irrelevant inputs. This demonstrates that an ANN will not necessarily determine which inputs are relevant to the output automatically and highlights the need for analytical methods for this purpose. When the model contained only 4 inputs the relationships between inputs and outputs became better defined as indicated by the reduced spread of the distributions in Figures 3 (b) and (d). Here it has been correctly identified that x_{t-3} is irrelevant to the model and x_{t-9} is relevant.

6. CONCLUSIONS

Selection of the correct causal inputs is vital for ensuring that an ANN model gives a good representation of the underlying function. This is particularly important when the model is used to gain knowledge of the system and an interpretation of the network function is required.

A number of methods have been proposed in the literature for assessing the relevance of potential input variables in predicting the response variable, but few have considered the variability in the relevance measure or the uncertainty in the network weights, both of which are fundamental for assessing input significance. In this paper an input pruning method has been presented which considers both of these factors by using Bayesian methods to estimate the network weights. When the method was applied to a synthetically generated data set it was able to correctly identify the 12 irrelevant input variables that were initially included in the ANN such that these were pruned

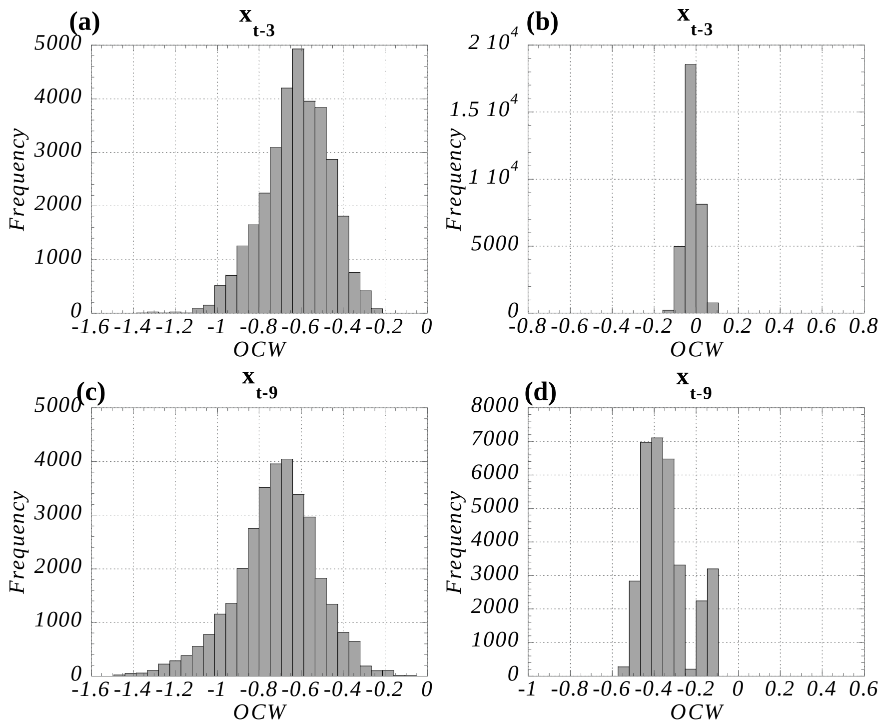


Figure 3 OCW distributions of inputs x_{t-3} and x_{t-9}. (a) and (c) are the distributions obtained after 1 run of the pruning process (15 inputs included), while (b) and (d) are the distributions obtained after 6 runs (4 inputs included)

and the final model only included the 3 correct causal inputs.

A limitation of the proposed method is that the network architecture needs to be specified and this may have implications on the relationship modelled. Future research will consider a method for pruning inputs and hidden nodes concurrently.

7. REFERENCES

Garson, G. D. 1991. Interpreting neural network connection weights. AI Expert **6**:47-51.

Haario, H., E. Saksman, and J. Tamminen. 2001. An adaptive Metropolis algorithm. Bernoulli **7**:223-242.

Lek, S., M. Delacoste, P. Baran, I. Dimopoulos, J. Lauga, and S. Aulagnier. 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecological Modelling **90**:39-52.

Mackay, D. J. C. 1992. A practical Bayesian framework for backpropagation networks. Neural Computation **4**:448-472.

Maier, H. R., and G. C. Dandy. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling and Software 15:101-124.

Maier, H. R., G. C. Dandy, and M. D. Burch. 1998. Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray, South Australia. Ecological Modelling **105**:257-272.

Neal, R. M. 1992. Bayesian Training of Backpropagation Networks by the Hybrid Monte Carlo Method. Department of Computer Science, University of Toronto.

Olden, J. D., and D. A. Jackson. 2002. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecological Modelling **154**:135-150.

Refenes, A.-P. N., and A. D. Zapranis. 1999. Neural model identification, variable selection and model adequacy. Journal of Forecasting **18**:299-332.

Thyer, M., G. Kuczera, and Q. J. Wang. 2002. Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation. Journal of Hydrology **265**:246-257.

Artificial neural networks prediction of PM10 in the Milan area

M. Cecchetti, G. Corani [a], G. Guariso[a]

[a]Dipartimento di Elettronica ed Informazione
Politecnico di Milano
e-mail:corani(guariso)@elet.polimi.it

Abstract: PM10 constitutes a major concern for Milan air quality. We presents a series of results obtained applying different neural networks approaches to the PM10 prediction problem. The 1-day ahead prediction shows a satisfactory level of accuracy, which may be further improved if a proper deseasonalization approach is adopted, thus transferring some a priori knowledge in the data pre-processing step. Then, we tackle the problem of the 2-days ahead prediction; in order to optimize the neural network architecture identification procedure, we try a pruning approach besides the usual trial and error one. Prediction performances are very close between the two models, and denote a significant decrease of accuracy with respect to the 1-day case, even though some meteorological improper (i.e. future measures) input is added to the model structure.

Keywords: feedforward neural networks, PM10, time series forecast

1 INTRODUCTION

The Milan urban area is located at the center of the Po Valley, the most industrialised and populated district in Italy. According to the Municipal Report on the State of the Environment (AMA [Municipal Environment and Mobility Agency]), the yearly average of pollutants such as SO2, NOx, CO, TSP has decreased respectively by about 90%, 50%, 65%, 60% during the period 1989-2001; no exceedings of alarm and attention thresholds have been observed since 1997 for SO_2 and *TSP*, and since 1999 for CO. The situation is just slightly worse for NO_x which, though clearly decreasing, showed, on the average, 8 yearly exceedings of the attention threshold (none of the alarm) over the same period. The yearly averages of micropollutants such as benzene and lead are also largely under the thresholds established for human health protection. These significant results are due to various actions, such as improved formulations of fossil oils for industrial activities, large adoption of methane for residential heatings, and the renewal of the fleet of circulating vehicles. The most severe health issue is now constituted by the high levels of particulate matter (PM10), a category of pollutants including solid and liquid particles having an effective aerodynamic diameter smaller than $10\mu m$. PM10 can be a health hazard for several reasons; it can harm lung tissues and throat, aggravate asthma and increase respiratory illness. Indeed, high PM10 levels have been correlated to increase in hospital admissions for lung and heart disease (Ostro et al. [1999]).

Health-based standards for PM10 have been established by a European directive (99/30/CE). Accordingly, the Regional law (DGR 19/10/2001) fixes at $50\mu g/m^3$ the attention threshold for the daily average; if the threshold is exceeded for 5 consecutive days, the "*attention state*" is declared [1]. Differently from the other pollutants quoted above, the yearly average of PM10 has been substantially stable (about $45\,\mu g/m^3$) since the beginning of monitoring in 1998. On average, PM10 exceeds the attention threshold in Milan for about 100 days/year, and about 20 "*attention state days*" are declared every year (AMA [Municipal Environment and Mobility Agency]).

A system able to predict PM10 concentrations could provide a useful anticipation to Public Authorities, in order to plan an increase in the public transport, warn people to avoid exposures to unhealthy air, or alert them of possible traffic blocks.

[1]The most recent regional law [DGR 13858 29/7/2003] preventively decrees the block of the pre-Euro vehicles for some hours during winter days, leaving the definition of other emergency actions to the responsibility of local authorities

Neural networks for air pollution forecast collected a general consensus over the last years, as pointed out by the review of Gardner and Dorling [1998]. In this paper, we train neural networks to forecast the PM10 daily average concentration, assuming to exploit all the data available until 9 a.m. of the current day *t*. As a second modelling step, we extend the prediction horizon to the following day. To this end, we add some further meteorological input variables, measured at ground level over both the day *t* and *t+1*. Although such an approach is, technically speaking, improper since it uses inputs which are unavailable in real time operations, it constitutes an interesting subject of investigation since the obtained performances may be considered as an upper bound of what can be achieved by adding forecasts of such meteorological variables in the model.

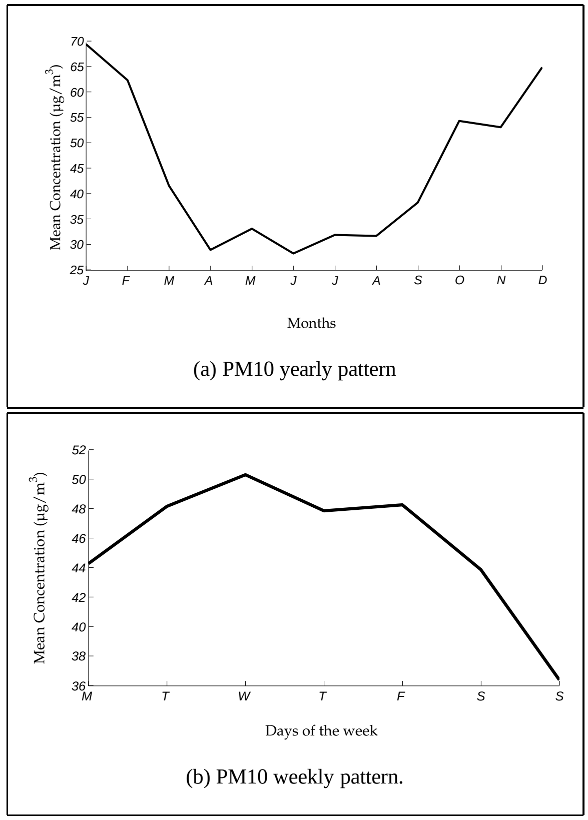
Since PM10 time series underlies clear periodicities at yearly and weekly level, we also investigate different deseasonalization approaches in the data preprocessing step.

A final issue addressed by this study regards the methodology adopted for the identification of the optimal neural network architecture. For the 1-day prediction, we found satisfactory results by selecting the architecture by trial and error, as usually done in neural networks applications. For the 2-days prediction, indeed much more difficult, we try also an alternative approach, based on the Optimal Brain Surgeon pruning algorithm (Hassibi and Stork [1993]).

2 TIME SERIES ANALYSIS

The data used in this work have been collected from a monitoring station located in a residential area of the city and refer to the years 1999-2002. The dataset is constituted by hourly time series; missing values range between 5% and 10% depending on the considered monitor. Data are splitted into training (1999-2000), validation (2001) and testing sets (2002).

To be used as input variable for the predictor, each monitored parameter has to be grouped from the original hourly series to a daily time series; this has been accomplished selecting the most suitable grouping operator (mean over 24 hours, mean over a certain time window, maximum) for any given input, by means of an extensive correlation analysis. The input variables set finally adopted comprises an autoregressive PM10 term, NO_x (used to track the road traffic emissions), SO_2 (a proxy for heatings



(a) PM10 yearly pattern

(b) PM10 weekly pattern.

Figure 1: PM10 average profiles.

emissions), and a wide set of meteorological variables, such as temperature, humidity, wind velocity, solar radiation, atmospheric pressure and Pasquill stability class.

Simply plotting the PM10 monthly averages, it is possible to notice cyclic patterns with high peaks in winter and much lower concentrations during summer (Figure 1a). Such an effect is due to the combined action of the unfavorable dispersion conditions which are encountered during winter (e.g., reduced mixing layer height) and to the increase of anthropic emissions (heavier traffic volumes and residential heatings). The average PM10 weekly profile also shows a typical pattern, with a decrease of about 25-30% during the weekends (Figure 1b). Indeed, the spectral analysis performed on the daily values clearly supports such evidence, as one can see from the periodogram shown in Figure 2: two peaks can be clearly detected, corresponding respectively to the yearly and weekly frequencies. The same kind of periodicities have been detected (analyzing both time and frequency domains) also on SO_2 and NO_x.

Such recognized periodicities suggest that a proper

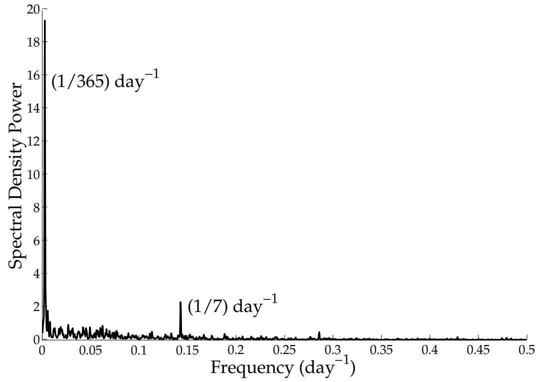


Figure 2: PM10 periodogram

data-deseasonalization approach may be useful in order to introduce some *a priori* knowledge in the data pre-processing step.

3 PREDICTION METHODOLOGIES AND RESULTS

We identify the predictor as a traditional feed forward neural network (Bishop [1995]). The network is constituted by a hidden layer, collecting a series of neurons having hyperbolic tangent as transfer function. The output layer contains just one neuron, having an identity transfer function, which actually returns the prediction. The optimal network architecture is found via trial and error, training 20 times each candidate architecture. The Levenberg-Marquardt algorithm is adopted as training method and early stopping is applied in order to prevent overfitting; hence, the training algorithm is stopped once the squared error on the validation dataset begins to increase. The architecture showing the lowest squared error on the validation set, which is assumed to have the best generalization ability, is finally selected as optimal and then simulated on the testing set, in order to quantitatively evaluate its performances.

With reference to data deseasonalization, we tried different solutions. The most effective one proved to be the fitting of a periodic regressor of type $f(\omega) = \sum_{k=1}^{k=N} a_k sin(k\omega t) + b_k cos(k\omega t)$, i.e.a linear combination of sine and cosine terms up to a harmonic N. Its parameters a_k and b_k have been estimated, on each pollutant time series, via a standard least square technique.The regressor including both the yearly angular frequency ($\omega_1 = 2\pi/365days^{-1}$) and the weekly angular frequency ($\omega_2 = 2\pi/7days^{-1}$) is obtained as:

$$R(t) = c + f(\omega_1) + f(\omega_2) \quad (1)$$

where c is a constant term. Models are then trained -with reference to the pollutants variables- on the standardized residuals time series, while meteorological data are just standardized (i.e. converted to a zero average and a unit standard deviation).

	Net_R	Net_S
Average goodness		
ρ	0.94	0.91
MAE	7.70	8.71
Threshold detection		
n_0	127	127
CPO	0.81	0.87
CPP	0.91	0.83
FA	0.09	0.17

Table 1: 1-days ahead prediction performances.

The most suitable configuration of the regressor for each polluttant (i.e., the number of harmonics to be considered) is investigated up to $N = 2$, thus avoiding higher parametrization of the regressors. Indeed, a similar attempt using periodic regressors (Kolehmainen et al. [2001]) reported a prediction worsening with respect to the use of simply standardized data; this was probably due to the too high number of harmonics they considered, which returned a too poor signal.

The testing set performances for the networks identified using the regressors (Net_R) or the usually standardized data (Net_S) are given in Table 1, and show a satisfactory prediction accuracy for both models. Performances are assessed in terms of average prediction ability by means of the true/predicted correlation ρ and by the mean absolute error MAE. Moreover, the problem of predicting the exceedances of the attention threshold is evaluated through of a series of indicators:

- CPO, the ratio of *C*orrectly *P*redicted exceedances N_{CP} to the number of total *O*bserved exceedances N_O:

$$CPO = \frac{N_{CP}}{N_O} \quad (2)$$

- *CPP*, the ratio between the number N_{CP} of *C*orrectly *P*redicted exceedances to the number N_P of totally *P*redicted exceedances:

$$CPP = \frac{N_{CP}}{N_P} \tag{3}$$

- *FA*, the ratio between the number of *F*alse *A*larms and the number N_P of totally *P*redicted exceedances:

$$FA = \frac{N_{FA}}{N_P} = 1 - CPP \tag{4}$$

The model trained on the residual time series shows an appreciable improvement on the average indicators. With regard to threshold exceedances prediction however, Net_S shows higher percentage (*CPP*: 87% vs 81%) of detection. On the other hand however, an above threshold prediction given by Net_R is far more reliable than by Net_S, with significantly better values of both CPP and FA.

For a comparison of these results with modelling approaches other than neural networks, one can refer to the work by Corani and Barazzetta [2004], who addressed the same case study using a traditional linear predictor (ARX model). They report, for instance, a true/predicted correlation of about $.89$ and a MAE of about $11\mu g/m^3$. The non-linearity of ANN appears therefore to allow a significant improvement of the prediction quality.

Given the satisfactory results on the 1-day prediction, we tried the more ambitious target of 2-days ahead forecast. To this end, we add further meteorological input variables to the model: temperature, rainfall, humidity pressure and wind speed measured over day (*t*) and (*t+1*). Such data constitute a set of improper inputs, since their values are not available at prediction time (9 a.m. of day *t*). However, they provide interesting indications, if one considers that the prediction performances obtained in this way constitute an upper bound of what can be achieved by inserting in the model the actual forecasts of such variables, obtained by means of a meterological model. Polluttant time series have been again deseasonalized by means of periodic regressors, given the valuable contribution of such a technique on the 1-days prediction.

Since the prediction target is much more difficult in this case, we try to improve the neural networks identification procedure performing a set of experiments for the identification of the architecture via the Optimal Brain Surgeon pruning algorithm (Hassibi and Stork [1993]), besides the classical trial and error approach. The basic idea of pruning algorithms is to start from a fully connected network, considered large enough to capture the desired input-output relationship. Then, they compute some measure of the contribution of each parameter to the problem solution, and consequently prune the less influential one from the network, to generate a new partially connected model, containing one parameter less. In this way, weights and neurons considered redundant are eliminated, significantly reducing the amount of guesswork needed for model selection. The network showing the lowest validation error between the many pruned architectures generated is finally chosen as predictor, consistently with the model selection criterion adopted for fully connected networks. The selected pruned network architectures may contain one order of magnitude less parameters than the fully connected ones, and are hence very parsimonious, thus providing a greater generalization power.

	Net_{2d}	Net_{2dP}
Average goodness		
ρ	0.76	0.76
MAE	13.08	12.89
Threshold detection		
n_0	127	127
CPO	0.73	0.72
CPP	0.75	0.76
FA	0.25	0.24

Table 2: 2-days ahead prediction performances.

The performances of the network designed by trial and error Net_{2d} and by pruning Net_{2dP} on the testing set are given in Table 2. One can easily notice that the performances of the two networks are very close to each other, and that a strong decrease takes place with respect to the 1-days prediction case. The true-predicted correlation decreases from $.93$ to $.76$, and the threshold exceedance detection indexes becomes significantly worse.

Since a quantitative meteorological forecast was not available, it is impossible to verify the performances which may be achieved with this additional information. However, the results will certainly lie between the upper bound already found and a lower bound constituted by a model trained without any improper meteorological information. Such a model shows a true/predicted correlation of about 0.74, with CPO and FA being respectively 0.72 and 0.29. The gap between the two approaches is not substantial and indicates that it is worth investigating more advanced descriptions of the meteorology (e.g. at a synoptic level) and more effective ways of exploiting such information.

4 CONCLUSIONS

The results on the prediction computed at 9a.m. for the current day are clearly satisfactory, with a true/predicted correlation of about 0.94 and a index of correctly predicted exceedances higher than 0.80. Data deseasonalization seems a valuable approach to increase of some points the average performances indicators; improvements are however less clear on the threshold exceedances prediction indicators. The 2-days prediction appears as an open problem, and the extension of air quality forecast horizons is likely to require a great research effort. In our opinion, dramatical performances improvements are not to be expected by studying new prediction algorithms: indeed, neural networks constitute a flexible non-linear modelling approach, able to learn very complex relationship from data. On the other hand, the availability of more advanced meteorological data, able to describe the air masses motion in the atmosphere (e.g. vertical profiles of wind speed and temperature, mixing height), can greatly increase the informative content of the input variables set, and may thus allow more significant improvements of air quality predictions.

We remark that, although presently no clear trend is detected on the PM10 time series, the situation may evolve over time, thus requiring a retraining of the predictor. Neural networks cannot be easily updated, and in fact it will be necessary to identify ex novo both the structure and parameters of the network, in order to have an up-to-date predictor. From this point of view, it is worth to mention that lazy learning, a local linear modelling approach, can constitute a viable alternative to neural networks; in fact, according to (Birattari et al. [1999]), this method may provide comparable prediction performances, allowing at the same time a quicker design and an easier update of the predictor.

REFERENCES

AMA (Municipal Environment and Mobility Agency). *Relazione sullo stato dell'ambiente del comune di Milano*. Milan, 2003.

Birattari, M., G. Bontempi, and H. Bersini. Lazy learning for modeling and control design. *Int. J. of Control*, 72:643–658, 1999.

Bishop, C. *Neural networks for pattern recognition*. Oxford University Press, 1995.

Corani, G. and S. Barazzetta. First results in the prediction of particulate matter in the Milan area. In *9th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Garmisch-Partenkirchen, 2004.

Gardner, M. and S. Dorling. Artificial neural network (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, 6(32):2627–2636, 1998.

Hassibi, B. and D. Stork. Second order derivatives for network pruning: Optimal brain surgeon. In S.J. Hanson, J.D. Cowan, C. G., editor, *Proceedings of Advances in Neural Information Processing System*, pages 164–171, 1993.

Kolehmainen, M., H. Martikainen, and J. Ruuskanen. Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment*, 35:815–825, 2001.

Ostro, B., L. Chestnut, N. Vichit-Vadakan, and A. Laixuthai. The impact of particulate matter on daily mortality in Bangkok, Thailand. *Journal of Air and Waste Management Association*, 49: 100–107, 1999.

Hazard Assessment of Debris Flows by Credal Networks*

A. Antonucci,[a] A. Salvetti[b] and M. Zaffalon[a]

[a]Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA)
Galleria 2, CH-6928 Manno (Lugano), Switzerland
{alessandro,zaffalon}@idsia.ch

[b]Istituto Scienze della Terra (IST)
Via Trevano, CH-6952 Canobbio, Switzerland
andrea.salvetti@supsi.ch

Abstract: Debris flows are destructive natural hazards that affect human life, buildings, and infrastructures. Despite their importance, debris flows are only partially understood, and human expertise still plays a key role for hazard identification. This paper proposes filling the modelling gap by using *credal networks*, an imprecise-probability model. The model uses a directed graph to capture the causal relationships between the triggering factors of debris flows. Quantitative influences are represented by probability intervals, determined from historical data, expert knowledge, and theoretical models. Most importantly, the model joins the empirical and the quantitative modelling levels, in the direction of more credible inferences. The model is evaluated on real case studies related to dangerous areas of the Ticino Canton, southern Switzerland. The case studies highlight the good capabilities of the model: for all the areas the model produces significant probabilities of hazard.

Keywords: Debris flows; credal networks; imprecise Dirichlet model; probability intervals; updating.

1 Introduction

Debris flows are among the most dangerous and destructive natural hazards that affect human life, buildings, and infrastructures. Starting from the '70s, significant scientific and engineering advances in the understanding of the processes have been achieved (see Costa and Wiekzorek [1987]; Iverson et al. [1997]). Yet, human expertise is still fundamental for hazard identification as many aspects of the whole process are still poorly understood.

This paper presents a *credal network* model of debris flow hazard for the *Ticino canton*, southern Switzerland. *Credal networks* (Cozman [2000]) are imprecise-probability models based on the extension of *Bayesian networks* (Pearl [1988]) to sets of probability mass functions (see Sec. 2.2). *Imprecise probability* is a very general theory of uncertainty developed by Walley [1991] that measures chance and uncertainty without sharp probabilities.[2]

The model represents expert's causal knowledge by a directed graph, connecting the triggering factors for debris flows (Sec. 3.1). Probability intervals are used to quantify uncertainty (Sec. 3.2) on the basis of historical data, expert knowledge, and physical theories. It is worth emphasizing that the credal network model joins human expertise and quantitative knowledge. This seems to be a necessary step for drawing credible conclusions. We are not aware of other approaches with this characteristic.

The model presented here aims at supporting experts in the prediction of dangerous events of debris flow. We have made preliminary experiments in this respect by testing the model on historical cases of debris flows happened in the Ticino canton. The case studies highlight the good capabilities of the model: for all the areas the model produces signifi-

*Thanks to the Swiss Federal Office of Topography for providing the digital elevation model, and to the Swiss Federal Statistical Office for the landuse, soil suitability, and geotechnical maps. Bayesian network updating has been computed by the software *SMILE*, developed at the Decision Systems Laboratory of the University of Pittsburgh. Extreme mass functions have been obtained by D. Avis' vertex enumeration software *lrs*. The authors of these public software tools are gratefully acknowledged. This research was partially supported by the Swiss NSF grant 2100-067961.

[2]See Walley [1996b] for a thorough comparison of imprecise probability with other measures of uncertainty popular in artificial intelligence, such as belief functions and possibility measures.

cant probabilities of hazard. We make a critical discussion of the results in Sec. 4, showing how the results are largely acceptable by a domain expert.

2 BACKGROUND

2.1 Debris Flows

Debris flows are composed of a mixture of water and sediment.
Three types of debris flow initiation are relevant: erosion of a channel bed due to intense rainfall, landslide, or destruction of a previously formed natural dams. According to Costa [1984] prerequisite conditions for most debris flows include an abundant source of unconsolidated fine-grained rock and soil debris, steep slopes, a large but intermittent source of moisture, and sparse vegetation. Several hypotheses have been formulated to explain mobilization of debris flows. Takahashi [1991] modelled the process as a water-saturated inertial grain flows governed by the *dispersive stress* concept of Bagnold. In this study we adopt Takahashi's theory as the most appropriate to describe the types of event observed in Switzerland.
The mechanism to disperse the materials in flow depends on the properties of the materials (*grain size*, *friction angle*), channel slope, flow rate and water depth, particle concentration, etc., and, consequently, the behavior of flow is also various.

2.2 Methods

Credal Sets and Probability Intervals. We restrict the attention to random variables which assume finitely many values (also called *discrete* or *categorical* variables). Denote by $\mathcal{X}$ the possibility space for a discrete variable X, with x a generic element of $\mathcal{X}$. Denote by $P(X)$ the mass function for X and by $P(x)$ the probability of $x \in \mathcal{X}$. Let a *credal set* be a closed convex set of probability mass functions. $\mathcal{P}_X$ denotes a generic credal set for X. For any event $\mathcal{X}' \subseteq \mathcal{X}$, let $\underline{P}(\mathcal{X}')$ and $\overline{P}(\mathcal{X}')$ be the *lower and upper probability* of $\mathcal{X}'$, respectively, defined by $\underline{P}(\mathcal{X}') = \min_{P \in \mathcal{P}_X} P(\mathcal{X}')$ and $\overline{P}(\mathcal{X}') = \max_{P \in \mathcal{P}_X} P(\mathcal{X}')$. Lower and upper (conditional) expectations are defined similarly. Note that a set of mass functions, its convex hull and its set of *vertices* (also called *extreme mass functions*) produce the same lower and upper expectations and probabilities.
Conditioning with credal sets is done by element-wise application of Bayes rule. The posterior credal set is the union of all posterior mass functions. Denote by $\mathcal{P}_X^y$ the set of mass functions $P(X|Y=y)$, for generic variables X and Y. We say that two variables are *strongly independent* when every vertex in $\mathcal{P}_{(X,Y)}$ satisfies stochastic independence of X and Y.
Let $\mathbb{I}_X = \{\mathbb{I}_x : \mathbb{I}_x = [l_x, u_x], 0 \leq l_x \leq u_x \leq 1, x \in \mathcal{X}\}$ be a set of probability intervals for X. The credal set originated by $\mathbb{I}_X$ is $\{P(X) : P(x) \in \mathbb{I}_x, x \in \mathcal{X}, \sum_{x \in \mathcal{X}} P(x) = 1\}$. $\mathbb{I}_X$ is said *reachable* or *coherent* if $u_{x'} + \sum_{x \in \mathcal{X}, x \neq x'} l_x \leq 1 \leq l_{x'} + \sum_{x \in \mathcal{X}, x \neq x'} u_x$, for all $x' \in \mathcal{X}$. $\mathbb{I}_X$ is coherent if and only if the related credal set is not empty and the intervals are tight, i.e. for each lower or upper bound in $\mathbb{I}_X$ there is a mass function in the credal set at which the bound is attained (see Campos et al. [1994]).

The Imprecise Dirichlet Model. We infer probability intervals from data by the *imprecise Dirichlet model*, a generalization of Bayesian learning from multinomial data based on soft modelling of prior ignorance. The interval estimate for value x of variable X is given by $[\#(x)/(N+s), (\#(x)+s)/(N+s)]$, where $\#(x)$ counts the number of units in the sample in which $X = x$, N is the total number of units, and s is a hyperparameter that expresses the degree of caution of inferences, usually chosen in the interval $[1, 2]$ (see Walley [1996a] for details). Note that sets of probability intervals obtained using the imprecise Dirichlet model are reachable.

Credal Networks. A credal network is a pair composed of a directed acyclic graph and a collection of conditional credal sets. A node in the graph is identified with a random variable X_i (we use the same symbol to denote them and we also use "node" and "variable" interchangeably). The graph codes strong dependencies by the so-called *strong Markov condition*: every variable is strongly independent of its nondescendant non-parents given its parents. A generic variable, or node of the graph, X_i holds the collection of credal sets $\mathcal{P}_{X_i}^{pa(X_i)}$, one for each possible joint state $pa(X_i)$ of its parents $Pa(X_i)$. We assume that the credal sets of the net are *separately specified* (Walley [1991]): this implies that selecting a mass function from a credal set does not influence the possible choices in others.
Denote by $\mathcal{P}$ the *strong extension* of a credal network. This is the convex hull of the set of joint mass functions $P(\boldsymbol{X}) = P(X_1, \ldots, X_t)$, over the t variables of the net, that factorize according to $P(x_1, \ldots, x_t) = \prod_{i=1}^{t} P(x_i | pa(X_i)) \quad \forall (x_1, \ldots, x_t) \in \times_{i=1}^{t} \mathcal{X}_i$. Here $pa(X_i)$ is the assignment to the parents of X_i

consistent with $(x_1, \ldots, x_t)$; and the conditional mass functions $P(X_i \,|pa\,(X_i))$ are chosen in all the possible ways from the respective credal sets. The strong Markov condition implies that a credal network is equivalent to its strong extension. Observe that the vertices of $\mathcal{P}$ are joint mass functions. Each of them can be identified with a Bayesian network (Pearl [1988]), which is a precise graphical model. In other words, a credal networks is equivalent to a set of Bayesian networks. This makes credal networks inherit some of the advantages of Bayesian nets, such as compactness of uncertainty representation and easy visualization, while presenting the additional characteristic to permit modelling based on weaker, and hence often more realistic, assumptions.

Computing with Credal Networks. We focus on the task called *updating*, i.e. the computation of $\underline{P}(X|E=e)$ and $\overline{P}(X|E=e)$. Here E is a vector of *evidence variables* of the network, in state e (the *evidence*), and X is any other node. The updating is intended to update prior to posterior beliefs about X. The updating can be computed by (i) exhaustively enumerating the vertices P_k of the strong extension; and by (ii) minimizing and maximizing $P_k(X|E=e)$ over k, where $P_k(X|E=e)$ can be computed by any updating algorithm for Bayesian networks (recall that each vertex of the strong extension is a Bayesian network).
The exhaustive approach can be adopted when the vertices of the strong extension are not too many. In general, non-exhaustive approaches must be applied as the updating problem is *NP-hard* with credal nets (Ferreira da Rocha and Cozman [2002]) also when the graph is a polytree. A polytree is a directed graph with the characteristic that forgetting the direction of arcs, the resulting graph has no cycles. In the present work the type of network, jointly with the way evidence is collected, make the exhaustive approach viable in reasonable times. Note that the exhaustive algorithm needs credal sets be specified via sets of vertices. We used the software tool *lrs* (http://cgm.cs.mcgill.ca/~avis/C/lrs.html) to produce extreme mass functions from probability intervals.

3 THE CREDAL NETWORK

3.1 Causal Structure

The network in Fig. 1 expresses the causal relationships between the topographic and geological characteristics, and hydrological preconditions, already sketched in Sec. 2.1. The leaf node is the depth of debris likely to be transported downstream during a flood event. Such node represents an integral indicator of the hazard level.

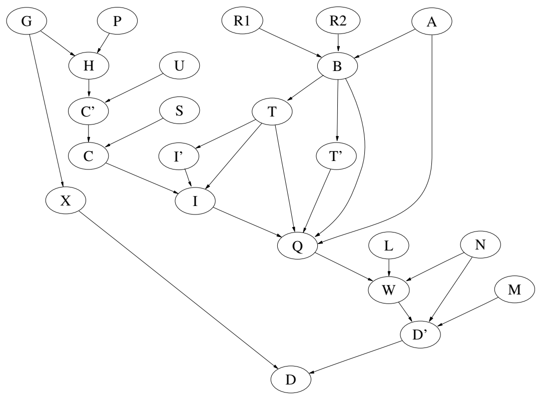


Figure 1: The causal structure.

In the following we describe the considerations that led to the network in Fig. 1. Node G represents the characteristics of the bedrock (*geology*) in a qualitative way. Debris flows require a minimum thickness of *colluvium* (loose, incoherent deposits at the foot of steep slope) for initiation, produced from a variety of bedrock. This is embedded in the graph with the connection to node X (*actual available debris thickness*) and expresses the propensity of different rock types to produce sediment. Additionally, bedrock properties influence the rate of infiltration and deep percolation, so affecting the generation of surface runoff and the concentration in the drainage network. This is accounted for by the connection of the geology to the *hydrologic soil type* (H), which influences the *maximum soil water capacity* (C').
The *soil permeability* (P), i.e. the rate at which fluid can flow through the pores of the soil, has to be further considered. If permeability is low, the rainfall will tend to accumulate on the surface or flow along the surface if it is not horizontal. The causal relation among geology and permeability determining the different hydrologic soil types was adopted according to Kuntner [2002]. The basic assumption is that soils with high permeability and extreme thickness show a high infiltration capacity, whereas shallow soils with extremely low permeability have a low infiltration capacity.
The *land use cover of the watershed* (U) is another significant cause of debris movement. It characterizes the uppermost layer of the soil system and has a definite bearing on infiltration.
We adopted the *curve number method* (USDA [1993]) to define the infiltration amount of the precipitation, i.e. the maximum soil water capacity. This method distinguishes hydrologic soil types which are supposed to show a particular hydrologic

behavior. For each land use type there is a corresponding curve number for each hydrologic soil type.
The amount of rainfall which cannot infiltrate is considered to accumulate into the drainage network (*surface runoff*), increasing the water depth and eventually triggering a debris flow in the river bed.
These processes are described by the deterministic part of the graph, related to runoff generation and Takahashi's theory, which takes into account topographic and morphologic parameters, such as *slope* (N) of the source area, *watershed morphology* (R_1 and R_2), *area* (A), *channel width* (L), and *precipitation intensity* (I').
The channel width is obviously decisive to determine the *water depth* (W), given the runoff generated within the watershed according to the standard hydraulic assumptions. Field experience in the study region indicates that debris flows often start in very steep and narrow creeks, with reduced accumulation area upstream.
The complexity and the organization of the channel geometry is therefore usually low and almost similar in the debris flow prone watersheds. For this reasons it was decided to adopt only three categories of channel width.
The climate of the regions in which debris flows are observed is as varied as geology and this was accounted for by defining several climatological regions, with different parameters of the depth-duration-frequency curve. In addition to the *duration* (T) and *effective rainfall intensity* (I) of a storm that ultimately produces a debris flow, the antecedent soil moisture conditions is recognized as an important characteristic. The significant period of antecedent rainfall varies from days to months, depending on local soil characteristics. According to the curve number theory, the transformation law to the *effective maximum soil water capacity* (C) depends only on the five-days antecedent rainfall amount corresponding to different *soil moisture conditions* (S).
We used the *linear theory of the hydrologic response* to calculate the *peak flow* (Q) values produced by constant-intensity hyetographs. We used the *multiscaling framework for intensity duration frequency curve* (Burlando and Rosso [1996]) coupled with the *instantaneous unit hydrograph* theory, proposed by Rigon et al. [2004]. Accordingly, the time to peak is greater than the rainfall duration and the *critical storm duration* (T') is independent of rainfall return period. The instantaneous unit hydrograph was obtained through the *geomorphological theory* (Rodriguez Iturbe and Valdes [1979]) and the *Nash cascade model of catchment's response*, where the required parameters (B_1 and B_2) were estimated from *Horton's order ratios* (R_1 and R_2), according to Rosso [1984].
By using the classical river hydraulics theory, the water depth in a channel with uniform flow and given discharge, water slope and roughness coefficient can be determined with the *Manning-Strickler formula* (see Maidment [1993]).
The *granulometry* (M), represented by the average particle diameter of the sediment layer, is required to apply Takahashi's theory. The *friction angle* was derived from the granulometry with an empirical one-to-one relationship. Takahashi's theory can finally be applied to determine the *theoretical thickness of debris* (D') that could be destabilized by intense rainfall events. The resulting value is compared with the *actual available debris thickness* (X) in the river bed. The minimum of these two values is the leaf node of the graph (D).

3.2 Quantification

Quantifying uncertainty means to specify the conditional mass functions $P(X_i|pa(X_i))$ for all the nodes X_i and the possible instances of the parents $pa(X_i)$. The specification is imprecise, in the sense that each value $P(x_i|pa(X_i))$ can lie in an interval. Intervals were inferred for the nodes G, P, U, N, H, and C', from the GEOSTAT database (Kilchenmann et al. [2001]) by the imprecise Dirichlet model (with s=2). The expert provided intervals for nodes L, M, R_1, R_2, and X. Functional relations between a node and its parents were available for the remaining nodes; in this case the intervals degenerate to a single 0-1 valued mass function. We detail the functional part in the rest of the section.
As mentioned in Sec. 2.1, the antecedent soil moisture conditions were accounted for by using the curve number method. The parametrization of the instantaneous unit hydrograph was obtained by using the number of theoretical linear reservoirs by which the basin is represented, $b_1 = 3.29 \cdot r_1^{0.78} \cdot r_2^{0.07}$; and by the time constant of each reservoir, $b_2 = .7 \cdot 0.251 \cdot (r_1 \cdot r_2)^{-.48} \cdot a^{0.38}$. Here b_1 depends on Horton's ratios, and b_2 is also function of the average travel time within the basin. For this we assumed the empirical expression reported by D'Odorico and Rigon [2003].
Given b_1 and b_2, following Rigon et al. [2004], we calculate the two characteristic durations, t and t', by solving the following system of two equations: $\alpha = [\frac{t}{b_2} \cdot (\frac{t'}{b_2})^{b_1-1} e^{-t'/b_2}]/[\gamma(b_1, \frac{t'}{b_2}) - \gamma(b_1, \frac{t'-t}{b_2})]$, and $\frac{t}{t'} = 1 - e^{-\frac{t}{b_2} \cdot \frac{1}{b_1-1}}$, where γ is the *incomplete lower gamma function* and α is a parameter, corresponding to the exponent of the *multiscaling*

intensity duration frequency curve.

We assume that these are in the form $i' = a(\tau_r) \cdot t^{-\alpha}$, where a is function of the return period τ_r of the event. To evaluate the effective intensity of rainfall, we have to impose the following transformation, taking account of the (effective) curve number, the corresponding dispersion term, and of the rainfall duration: $i = (i' \cdot t - \lambda(c)/10)^2/(i' \cdot t - \lambda(c)/10) + \lambda(c) \cdot 1/t$, where $\lambda(c) = 254 \cdot (100/c - 1)$ is the water depth absorbed by the soil of given curve number. The peakflow (Rigon et al. [2004]) can then be expressed as $q = a \cdot i/\alpha \cdot t/b_2 \cdot t'^{b_1-1}e^{-t'/b_2}$, and the corresponding waterdepth is $w = q/25\, l^{5/3}\, \sqrt{\tan n}$.

According to Takahashi [1991], we evaluate the debris thickness as $d' = w[k(\tan m'/\tan n - 1) - 1]^{-1}$. The relation is linear, with a coefficient taking into account the local slope n and the internal friction angle m' (which can be obtained from the granulometry m). $k = C_g(\delta_g - 1)$, with $\delta_g = 2.65$ the relative density of the grains, and $C_g \simeq 0.7$ the volumetric concentration of the sediments. The variables involved in the expression for d' must satisfy the constraints $1 + 1/k_2 \leq \tan m/\tan n \leq 1 + 1/k \cdot (1 + w/m)$. If the inequality on the left-hand side is violated, shallow landslides can occur also in absence of water depth, but technically speaking these are not debris flows. If the remaining inequality fails, the movable quantity is thinner than the granulometry and no flow can be observed.

d' is a theoretical value for the movable quantity, which does not take into account how much material is physically available. As the actual movable quantity cannot exceed the available material x, the final relation is given by $d = \min\{x, d'\}$.

4 Case Studies

We validate the model in preliminary way by an empirical study involving six areas of the Ticino canton. The network was initially fed with the information about the areas reported in Tab. 1, the estimated rainfall intensity on them for a return period of 10 years, and the geomorphological characteristics of the watershed. The estimated rainfall intensity is the expected frequency level of precipitations in a certain region during a future period. Using the estimated rainfall intensity allowed us to re-create the state of information existing 10 years ago[3] about precipitations in the areas under consideration. This is a way to check whether the network would have been a valuable tool to prevent the debris flows that actually happened in the six areas. The results of the analysis are in Tab. 2. We use the probabilities of defined debris thickness to be transported downstream as an integral indicator of the hazard level.

Table 1: Details about the case studies. (Note that P is not available. This is the common case as evaluation of permeability presents technical difficulties.)

Node	Cases					
	1	2	3	4	5	6
G	Gneiss	Porphyry	Limestone	Gneiss	Gneiss	Gneiss
A	0.26	0.32	0.06	0.11	0.38	2.81
M	10–100	≤10	≤10	100–150	≤10	150–250
U	Forest	Forest	Forest	Vegetation	Forest	Bare soil
N	20.8	19.3	19.3	21.8	16.7	16.7
L	4	6	4	8	4	8
R_1	0.9	0.6	0.7	0.9	0.9	0.8
R_2	1.5	3.5	3.5	3.5	2.3	2.1

Table 2: Posterior probabilities of node D, i.e. of the movable debris thickness (in centimeters). The probabilities are displayed by intervals in case 2.

Thickness	Cases					
	1	2	3	4	5	6
<10	0.011	[0.084,0.087]	0.083	0.196	0.087	0.005
10–50	0.048	[0.263,0.273]	0.275	0.388	0.139	0.013
>50	0.941	[0.639,0.652]	0.642	0.416	0.774	0.982

In cases 1 and 6 the evidences are the most extreme out of the six cases and indicate a high debris flow hazard level, corresponding to an instable debris thickness greater than 50 cm. In case 6 the relatively high upstream area (2.81 km^2), large channel depth, and the land cover (bare soil, low infiltration capacity) explain the results. In case 1 the slope of the source area (20.8°) plays probably the key role. In cases 2 and 3 the model presents a non-negligible probability of medium movable debris thickness. Intermediate results were obtained for case 5 due to the gentler bed slope (16.7°) as compared with the other cases. In case 4 the hazard probability is more uniformly distributed, and can plausibly be explained with the very small watershed area and the regional climate, which is characterized by low small rainfall intensity as compared with other regions.

We simulated also the historical events, by instantiating (as opposed to using the estimated rainfall intensity) the actual measured rainfall depth, its duration and the antecedent soil moisture conditions. Also in this setup the network produced high probabilities of significant movable thickness. (The probabilities are not reported for lack of space.)

As more general comment, it is interesting to observe that in almost all cases the posterior probabilities are nearly precise. This depends on the strength of the evidence given as input to the network about the cases, and by the fact that the flow process can partially be (and actually is) modelled functionally.

[3] The number of years is an arbitrary choice.

Now we want to model the evidence in even more realistic way with respect to the grain size of debris material. Indeed, granulometry is typically known only partially, and this limits the real application of physical theories, also considered that granulometry is very important to determine the hazard.

We model the fact than the observer may not be able to distinguish different granulometries. To this extent we add a new node to the net, say O_M, that becomes parent of M. O_M represents the observation of M. There are five possible granulometries, m_1 to m_5. We define the possibility space for O_M as the power set of $\mathcal{M} = \{m_1, \ldots, m_5\}$, with elements $o_{\mathcal{M}'}$, $\mathcal{M}' \subseteq \mathcal{M}$. The observation of granulometry is set to $o_{\mathcal{M}'}$ when the elements of $\mathcal{M}'$ cannot be distinguished. $P(m|o_{\mathcal{M}'})$ is defined as follows: it is set to zero for all states $m \in \mathcal{M}$ so that $m \notin \mathcal{M}'$; and for all the others it is *vacuous*, i.e. the interval $[0, 1]$ (the intervals defined this way must then be made reachable). This expresses the fact that we know that $m \in \mathcal{M}'$, and nothing else.

Let us focus on case 6 for which the observation of grain size is actually uncertain. From the historical event report, we can exclude that node M was in state m_1 or m_2. We cannot exclude that m_4 was the actual state (m_4 is the evidence used in the preceding experiments), but this cannot definitely be established. We take the conservative position of letting the states m_3, m_4 and m_5 be all plausible evidences by setting $O_M = o_{\{m_3,m_4,m_5\}}$. The interval probabilities become $[0.002, 0.008]$, $[0.010, 0.043]$, and $[0.949, 0.988]$, for debris thicknesses less than 10, in the range 10–50, and greater than 50, respectively. We conclude that the probability of the latter event is very high, in robust way with respect to the partial observation of grain size.

5 CONCLUSIONS

We have presented a model for determining the hazard of debris flows based on credal networks. The model unifies human expertise and quantitative knowledge in a coherent framework. This overcomes a major limitation of preceding approaches, and is a basis to obtain credible predictions, as shown by the experiments. Credible predictions are also favored by the soft-modelling made available by imprecise probability through credal nets.

The model was developed for the Ticino canton, in Switzerland. Extension to other areas is possible by re-estimating the probabilistic information inferred from data, which has local nature.

Debris flows are a serious problem, and developing formal models can greatly help us avoiding their serious consequences. The encouraging evidence provided in this paper makes credal networks be models of debris flows worthy of further investigation.

REFERENCES

Burlando, P. and R. Rosso. Scaling and multiscaling depth duration frequency curves of storm precipitation. *J. Hydrol.*, 187(1/2):45–64, 1996.

Campos, L., J. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2(2):167–196, 1994.

Costa, J. E. *Physical geomorphology of debris flows*, chapter 9, pages 268–317. Costa, J. E. and Fleisher, P. J. - Springer-Verlag, Berlin, 1984.

Costa, J. H. and G. F. Wiekzorek. *Debris Flows/Avalanches: Process, Recognition and Mitigation*, volume 7. Geol. Soc. Am. Reviews in Engineering Geology, Boulder, CO, 1987.

Cozman, F. G. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

D'Odorico, P. and R. Rigon. Hillslope and channel contributions to the hydrologic response. *Water Resources Research*, 39(5):1113–1121, 2003.

Ferreira da Rocha, J. C. and F. G. Cozman. Inference with separately specified sets of probabilities in credal networks. In Darwiche, A. and lastname Friedman, N., editors, *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI-2002)*, pages 430–437. Morgan Kaufmann, 2002.

Iverson, R. M., M. E. Reid, and R. G. LaHusen. Debris-flow mobilization from landslides. *Annual Review of Earth and Planetary Sciences*, 25:85–138, 1997.

Kilchenmann, U., G. Kyburz, and S. Winter. *GEOSTAT user handbook*. Swiss Federal Statistical Office, Neuchâtel, 2001. In german.

Kuntner, R. *A methodological framework towards the formulation of flood runoff generation models suitable in alpine and prealpine regions*. PhD thesis, Swiss Federal Institute of Technology, Zürich, 2002.

Maidment, D. R. *Handbook of Hydrology*. McGraw-Hill, 1993.

Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, 1988.

Rigon, R., P. D'Odorico, and G. Bertoldi. The peak-flow and its geomorphic structure. *Water Resources Research*, 2004. In press.

Rodriguez Iturbe, I. and J. B. Valdes. The geomorphologic structure of hydrologic response. *Water Resources Research*, 15(6):1409–1420, 1979.

Rosso, R. Nash model relation to Horton order ratios. *Water Resources Research*, 20(7), 1984. 914–920.

Takahashi, T. *Debris Flow*. A.A. Balkama, Rotterdam, 1991. IAHR Monograph.

USDA. *Soil Conservation Service*. United States Department of Agriculture, Washington D. C., 1993. Hydrology, National Engineering Handbook, Supplement A.

Walley, P. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.

Walley, P. Inferences from multinomial data: learning about a bag of marbles. *J. R. Statist. Soc. B*, 58(1): 3–57, 1996a.

Walley, P. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996b.

Sustainable Marine Resource Management: Lessons from Viability Analysis

J. Kropp [a], K. Eisenack[a], and J. Scheffran[a]

[a]Potsdam Institute for Climate Impact Research
P.O. Box 601203
14412 Potsdam, Germany

Abstract: Marine natural resources are under pressure worldwide. Management and surveillance systems are often inappropriate to guarantee a sustainable resource utilization since the knowledge on fisheries and/or stocks is limited. Additionally, institutional failures, e.g. unsuitable regulatory policies, have accelerated resource exploitation in several cases. Modelling is often considered as a very effective tool for studying the behaviour of complex systems, but a variety of difficulties arise if one has to deal with uncertain knowledge or inhomogeneous data of different quality. In this paper we present a method that is capable both for (i) integration of sparse or limited knowledge from different disciplines and (ii) provides a test-bed for an assessement of different management regimes.

Keywords: fisheries management; imprecise knowledge; qualitative modeling; viability theory

1 INTRODUCTION

Following sustained interest from policy makers, recent years have seen a number of modelling efforts examining the effects of commercial fisheries. The issue is recognized as highly important, as stated also by the FAO [2001]. In their report on the state of world fisheries the FAO mentioned that 50% of the world's fish resources are fully exploited, 20% overexploited and 10% depleted. Even though overfishing has been a fact since historical times [Jackson et al., 2001] the problem has gained a new quality due to the industrialization of fisheries. The activities of highly capitalized fishing companies have reduced community biomass by 80% within 15 years of exploitation [Myers and Worm, 2003]. In many cases the fishing industry can only be sustained at an economic level by paying high amounts of subsidies, while at the same time increased capitalization and efficiency put additional pressure on the stocks [Banks, 1999; Gréboval and Munro, 1999; Munro, 1999; Pauly et al., 2002; Pauly, 2003]. As a consequence of this intricate situation an ongoing debate on adequate control and management instruments is taking place. Especially the following questions are discussed frequently:

1. Do fishery management strategies focus too much on the ecological system part, an approach sometimes coined "ichthyocentrism" [Lane and Stephenson, 2000]?

2. How to deal with inherent uncertainties in the dynamics of the fishing industry as well as those of the fish stocks [Whitmarsh et al., 2000]?

3. What are potential benefits and risks of so-called co-management strategies [Potter, 2002]?

In our paper we address the issues mentioned above by applying a technique which considers inhomogeneous and uncertain knowlege from the biological, economic and political domain. The framework of viability theory [Aubin, 1991] can be used to assess management strategies and highlights the role of fish stock estimates in the recommendations for catch quota allocations. It allows normative sustainability criteria, e.g. for employment or environmental protection, to be included into systems analysis. All possible trajectories can be computed and checked concerning their properties violating these normative settings or not. Analytical frameworks like this are used to an increasing extent in sustainability sciences [Bene et al., 2001; Eisenack and Kropp, 2001; Aubin and Saint-Pierre, 2004].

Contact: kropp@pik-potsdam.de, Phone: +49-3312882526

2 MANAGEMENT REGIMES IN FISHERIES AND MODELLING METHODS

In this section we briefly introduce some concepts from fishery management and viability theory.

Co-management is implemented to increase participation of fishermen in the management of marine resources [Jentoft et al., 1998; Charles, 2001]. By introducing a degree of responsibility for the resource, it is expected that the compliance with regulations, e.g. catch quota, gear type, etc., will be higher [Mahon et al., 2003]. This is in contrast to the prevalent type of management, where a governmental authority imposes restrictions on the fishery. We refer to this strategy as top-down management, since fishing firms have no direct influence on the regulatory measures.

The problem of ichthyocentrism focuses on scientific institutions which play an essential part in both management regimes. Typically, these institutions provide biomass estimates which are an important basis for catch restrictions or quotas. It is often claimed that such an approach – in comparison to research efforts on the behaviour of resource users – puts too much emphasis on the resource itself. The question is whether we need a deeper understanding of the fisheries as a whole, or if it is sufficient to have better knowledge on the behaviour of fish stocks.

Viability theory [Aubin, 1991] investigates whether trajectories of a controlled dynamical system will stay within a prescribed region of the phase space given by the sustainability criteria (the so-called constrained set). This allows management targets to be formalized and takes into account the limitations of knowledge. A trajectory respecting the criteria is called viable or sustainable. However, not just a single trajectory is tested for viability, but the set of all trajectories which result from possible control paths (management decisions). The set of all initial conditions for which there exists at least one control path keeping the resulting trajectory in the constrained set forever is called the viablity kernel. The shape of the kernel depends on the management regime, and thus the approach can be used to identify sustainable management regimes.

3 THE MODEL

The change of biomass x of a fish stock depends on the recruitment $R(x)$ and the total catch h, given by $\dot{x} = R(x) - h$. We study the fish stock above a threshold $\underline{x} > 0$ assuming that its density is sufficiently low to allow an accelerating growth if there are no activities of fishing firms, i.e. $\forall x > \underline{x}$: $R(x) > 0$ and $D_x R(x) > 0$, where D_x denotes the partial differential operator with respect to x. Below $\underline{x}$ our knowledge about the recruitment behaviour is uncertain, and therefore we ignore this part of the phase space.

Co-management is exercised by a fishery council, where representatives from different groups (fishing firms, authorities, and scientific institutions) participate and negotiate about the allocation of catch quotas q_i to the groups $i = 1, \ldots, n$. The resulting total harvest is $h = \sum_i q_i$. When the members agree on their quota, the decision of the council is approved by a governmental authority. The resulting restrictions are executed by a management organization which works in close collaboration with local fishermen. The negotiations are opened by the scientific institution, which provides a preliminary recommendation h_0 for the total catch. Subsequently each group of the fishing industry tries both (i) to get a large share of the total harvest h and (ii) to increase h above the catch recommendation h_0 in order to improve profits. The profit π_i of group i depends on the quota q_i and on the available amount of fish x, which is the same for all fishing firms. It also depends on the efficiency of boats, fishing gear and technological equipment which varies between the groups [Scheffran, 2000]. Furthermore deviation costs d_i have to be considered which result from an exceeding of the scientific recommendation. They are due to public perception, the risk of being excluded from the co-management framework and a stronger need for good public relations. We define $\pi_i(q_i, x) = p\, q_i - c_i(q_i, x) - d_i\left(\sum_{j=1}^{n} q_j - h_0 \right)$. Here, the market price p is assumed to be exogeneous. The cost function c_i depends on a realized harvest q_i and the biomass of fish. It increases in q_i since more effort is needed for a larger catch, and decreases in x due to higher densities of the exploited stocks. If $\sum_j q_j - h_0$ becomes negative, d_i vanishes because deviation costs do not come into play if the sum of all quotas is below the scientific recommendation. It is reasonable to assume that d_i is a monotonicly increasing function. Each group of fishing firms can propose a quota allocation to obtain an optimal profit for a given price and fish stock. At Nash equilibrium the quota allocation assigns an individual quota q_i to each group i that maximizes π_i with respect to q_i for given p, x and quotas of the other participants. Assuming that all π_i are concave and continuously differentiable with respect to q_i, the equilibrium is provided by solving

$\forall i = 1, \ldots, n : D_{q_i}\pi_i = 0$. In our analysis we study the case of two fishery groups and specify c_i and d_i as

$$c_i(q_i, x) := \frac{{}_iq_i + \beta_i q_i^2}{x};$$

$$d_i(q_1{+}q_2{-}h_0) := \begin{cases} 0 \quad \text{if} \quad q_1 + q_2 < h_0 \\ \kappa_i(q_1 + q_2 - h_0)^2 \quad \text{otherwise.} \end{cases}$$

The parameters ${}_i, \beta_i$ $(i = 1, 2)$ are positive and represent the technical efficiency. The political power of each group i is expressed by the positive parameter κ_i. Both functions are continuously differentiable (x is positive), and the resulting profit functions π_i are concave with respect to q_1, q_2. The Nash equilibrium for given a p, x, h_0 is obtained by solving
$D_{q_1}\pi_1 = p - \frac{{}_1+2\beta_1 q_1}{x} - 2\kappa_1(q_1 + q_2 - h_0) = 0$,
$D_{q_2}\pi_2 = p - \frac{{}_2+2\beta_2 q_2}{x} - 2\kappa_2(q_1 + q_2 - h_0) = 0$,
for q_1, q_2.
From this, the total harvest resulting from the negotiations evaluates to

$$h(x, h_0) = q_1 + q_2 = \frac{upx + wxh_0 - v}{\beta_1\beta_2 + wx}, \quad (1)$$

where $u := \frac{1}{2}(\beta_2 + \beta_1) > 0$, $v := \frac{1}{2}({}_1\beta_2 + {}_2\beta_1) > 0$, and $w := \beta_1\kappa_2 + \beta_2\kappa_1 > 0$. It is below the profit optimum in the case of absent deviation costs and above the initial scientific recommendation. The impact of h_0 on h is

$$D_{h_0}h(x, h_0) = \frac{wx}{\beta_1\beta_2 + wx} > 0, \quad (2)$$

and

$$D_x h(x, h_0) = \frac{vw + (up + wh_0)\beta_1\beta_2}{(wx + \beta_1\beta_2)^2} > 0, \quad (3)$$

i.e. the negotiation equilibrium increases with the abundance of fish. The difference between the initial recommendation and total harvest increases with abundance too, because $D_x\big(h(x, h_0) - h_0\big) = D_x h(x, h_0)$.

4 VIABILITY CONSTRAINTS

For the assessment of recommendation strategies, i.e. for controls $h_0(t)$, we will specifiy sustainability criteria in the form of a constrained set (cf. Sect. 2). We will emphasize that these criteria are normative, i.e. they are not solely defined by scientists, but they are outcomes of the public debate which shall be analysed in this contribution. Here, we exemplarily investigate the following constraints:

(i) Prevent a decline of the fish stock, i.e. the condition $\dot{x} \geq 0$ has to be fulfilled at any time.

(ii) Ensure that $\forall t : x(t) \geq \underline{x}$, i.e. that situation exists with a relatively certain recruitment estimate.

(iii) Require a minimum harvest $\underline{h}$ covering fixed costs and that employment and/or food safety is sustained.

We rewrite $\dot{x} = R(x) - h(x, h_0)$, where R is monotonicly increasing for $x \geq \underline{x}$, and the function h has the form given in eq. (1). Then, constraint (i) implies

$$h_0 \leq \frac{R(x)(wx + \beta_1\beta_2) + v}{wx} - \frac{up}{w} =: \bar{h}_0(x). \quad (4)$$

Its differential is $D_x\bar{h}_0(x) =$

$$D_x R(x) + \frac{\beta_1\beta_2(xD_xR(x) - R(x)) - v}{wx^2}. \quad (5)$$

Observe that the expression may become negative. By definition, $R(x) \equiv h(x, \bar{h}_0(x))$, and therefore $D_xR(x) \equiv D_xh(x, \bar{h}_0(x))$. This implies that $D_x\bar{h}_0(x) = \frac{D_xR(x) - D_xh(x,h_0)}{D_{h_0}h(x,h_0)}$, where $D_{h_0}h(x, h_0) > 0$ (cf. eq. 2). If the negotiation result $h(x, h_0)$ increases in x faster than the recruitment $R(x)$, e.g due to high prices p (cf. eq. 3), the maximal viable harvest recommendation $\bar{h}_0(x)$ decreases in x.

For viability constraint (ii) it is sufficient that the derivative $\dot{x}$ is non-negative only for $x = \underline{x}$, i.e. $h_0 \leq \bar{h}_0(\underline{x})$. Constraint (iii) requires a minimum harvest $\underline{h}$ at each time, yielding

$$h_0 \geq \frac{\underline{h}(wx + \beta_1\beta_2) + v}{wx} - \frac{up}{w} =: \underline{h}_0(x). \quad (6)$$

The partial derivative of $\underline{h}_0(x)$ with respect to x simplifies to $D_x\underline{h}_0(x) = -\frac{\underline{h}\beta_1\beta_2 + v}{wx^2} < 0$. Thus, for increasing fish stocks lower initial recommendations can guarantee economic viability. Constraints (ii) and (iii) hold at the same time if $x > \underline{x}$ and $h_0 \geq \underline{h}_0(x)$, or if $x = \underline{x}$ and $\underline{h}_0(\underline{x}) \leq h_0 \leq \bar{h}_0(\underline{x})$, which implies $\underline{h} \leq R(\underline{x})$. Hence, the compatibility of the viability constraints (ii) and (iii) depends only on the recruitment function and the desired harvest. When recruitment at the minimum stock size $\underline{x}$ is higher than the required harvest, there is no contradiction between economic and ecological targets. Otherwise, when the stock attains $\underline{x}$, it must be decided whether to sacrifice conservation or harvest objectives. In the first case the viability kernel, i.e. the set of initial conditions allowing sustainable control, is the space $\{x \mid x \geq \underline{x}\}$, while

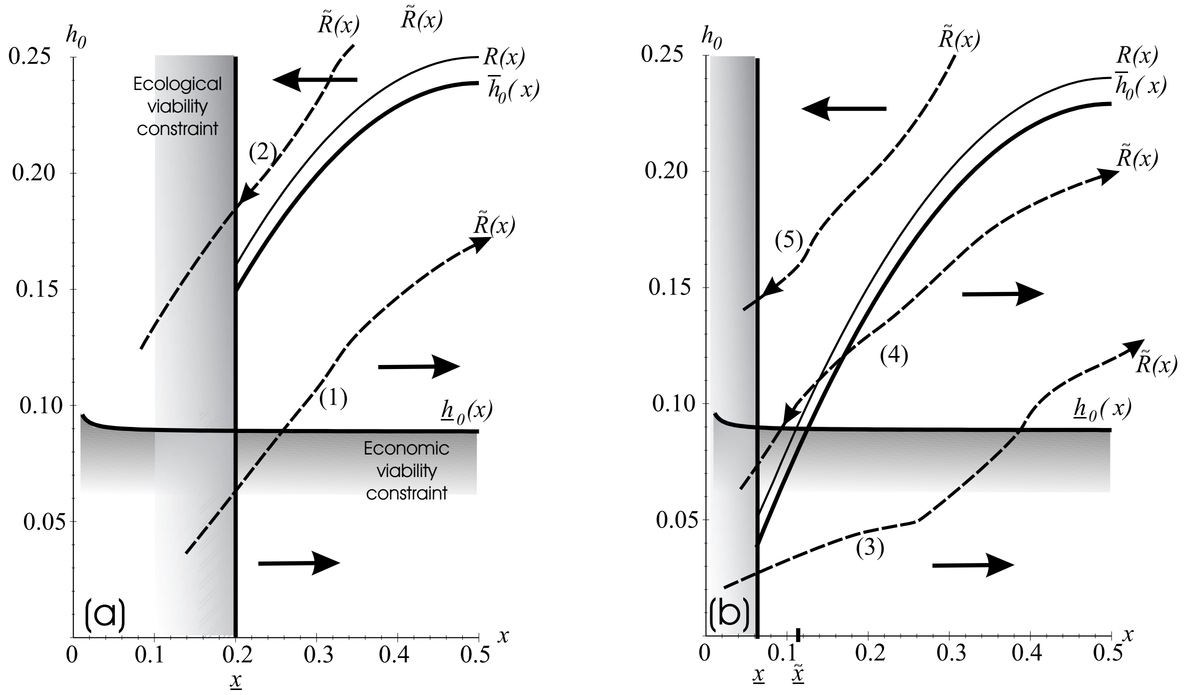


Figure 1: Phase space representation for five possible outcomes of ichthyocentric control. The shaded areas indicate the boundaries of the constrained set, the dashed lines represent possible "real world" realizations of estimated recruitment $\tilde{R}$. The bold arrows indicate how the stock evolves. Note that the dashed lines provides an additional information: the arrows signify the trajectories resulting from the corresponding recruitment estimate. (a) corresponds to the case where $R(\underline{x}) > \underline{h}$, and (b) to $R(\underline{x}) < \underline{h}$.

in the second case the viability kernel is the subset $\{x \mid x \geq \tilde{\underline{x}}\}$, where $\tilde{\underline{x}}$ is defined by the unique solution of $R(\tilde{\underline{x}}) = \underline{h}$.

5 Assessment of Recommendation Strategies

Whilst the system remains in the viability kernel it is *possible* to find harvest recommendations that guarantee viability forever. However, this does not imply that such a strategy is necessarily chosen. We assess whether this is the case for an ichthyocentric or for a qualitative recommendation strategy (see below). For the ichthyocentric strategy, the harvest recommendation h_0 is assumed to be identical to the estimated recruitment $\tilde{R}(x)$, assumed to be increasing with x (for possible realizations of $\tilde{R}(x)$, see Figure 1). We observe that under the presumption that recruitment is estimated exactly ($R = \tilde{R}$), the fishery is exposed to a high risk:

The permanent endeavour of fishing companies during the negotiations to accomplish higher catch quotas has the consequence that the stock will *necessarily* decrease below $\underline{x}$, since $h(x, R(x)) > \bar{h}_0(x)$, and thus $\dot{x} < 0$.

Only in the cases of strongly underestimated recruitment (cases (1) and (3)) is the resource within safe limits. For the cases (2) and (5) the recruitment is overestimated leading to a decrease of stocks and a violation of the ecological constraint. Finally, case (4) partly overestimates and partly underestimates the recruitment. Only if $R(\underline{x}) < \underline{h}$, a rapidly decreasing recruitment estimate leads to an economically non-sustainable state before the ecological criterion is violated.

The qualitative control strategy takes into account uncertainties and is performed in the sense of qualitative reasoning concepts, where systems dynamics is only characterized in terms of thresholds and trends [for details see, Kuipers, 1994; Kropp et al., 2002]. Assume that the values of $x, R, \underline{h}_0$ and $\bar{h}_0$ are not exactly known, but that we can observe whether x is decreasing or increasing and whether the total catch h exceeds $\underline{h}$ or not. In such a situation a control rule can only contain qualitative directives implying an increase or decrease of h_0 (see Table 1). A fast change (rule #1) means that h_0 is set to a level where x increases instantaneously. If the change is slow (rules #3 and #4), we only increase the harvest recommendation if x has been increasing for a longer time. If rule #2 or #3 come into

rule #	observed situation	response
(1)	x decreases and $h > \underline{h}$	decrease h_0 fast
(2)	x decreases and $h < \underline{h}$	decrease h_0 fast
(3)	x increases and $h < \underline{h}$	increase h_0 slow
(4)	x increases and $h > \underline{h}$	increase or decrease h_0 slow

Table 1: Qualitative control strategy consisting of four rules. If the surveillance authority, e.g. the scientific institution, observes the situation shown in the second column it can provide policy advice as indicated in the third column.

play, viability is already violated. Thus they have to be interpreted as crisis management. In Figure 2 the regions corresponding to the different rules are shown. They are separated by the graph of the function $\bar{h}_0(x)$ where $\dot{x} = 0$, and $\underline{h}_0(x)$ where $h = \underline{h}$. Horizontal arrows denote changes in the stock size resulting from the recommendation h_0. Vertical arrows represent the reaction of the scientific institution. Trajectories evolve in the directions given by the arrows. If rule #3 is applied the viability constraint (iii) is not fulfilled. Economic viability can be achieved by increasing h_0 and as a result region (2) or (3) can be entered. Indeed, if stocks are at a very low level and h_0 increases too fast, region (2) is more likely. Thus, an increase of h_0 should be slow, although this keeps the system economically unprofitable for a longer time. Otherwise a sharp cut in harvests is necessary in order to avoid the risk of crossing the threshold $\underline{x}$. When region (4) is approached, the associated rule prevents the system from leaving it. In this region of the phase space additional steering options are possible which not only meet economic and ecological constraints, but also allow a adjusted extension of the harvest in order to improve the profits.

If catches decrease below $\underline{h}$, rule #3 forces the system back. The same holds if catches increase too strongly: when x starts to decrease, rule #1 forces the system back to region (4). If h_0 is far above $\bar{h}_0(x)$ the decrease has to be fast to avoid a transition to region (2). For the same reasons h_0 must be changed slowly in region (4) to prevent it from exceeding $\bar{h}_0$ too much. The qualitative strategy is still risky in region (1) if we do not reduce harvest recommendations fast enough. However, once region (4) is entered, the qualitative strategy guarantees viable recommendations.

6 CONCLUSIONS

Management objectives in fisheries are rarely achieved in practice and the debate on adequate

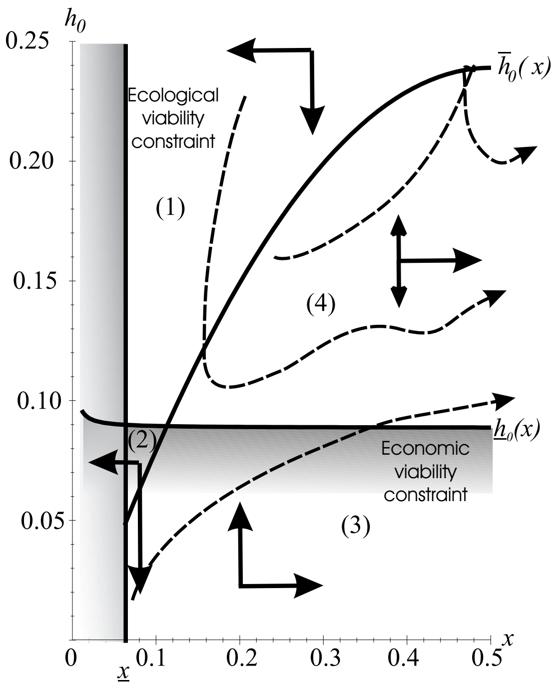


Figure 2: Phase space portrait for qualitative control: The shaded lines represent the viability constraints, the dashed arrows indicate trajectories in accordance with the control strategy. The numbers correspond to the control rules (cf. Table 1, right panel).

strategies is ongoing. Although it is not expected that a unique solution exists, the need for cross-disciplinary analyses taking into account uncertainties still remains. Our analysis shows that the outcome of co-management regimes depend on biological, *as well as* on economic and political factors. In particular, participatory management strategies, such as co-management, are not *per se* sustainable Recommendations purely based on the observation of fish stocks expose the fishery to a high economic and the resource to a high ecological risk. More flexible strategies, e.g. based simply on qualitative information on the state of the fishery, can guarantee minimum harvest levels and an increasing fish stock.

In summary, much more research is needed, but we hold that the presented methodology opens a promising road towards a better understanding of the intrinsic processes – including ecological, economic, and social issues – in fisheries. The viability concept supply a valuable tool for risk assessments in fisheries prone to non-sustainable developments. Therefore, future work will be directed to the introduction of additional viability constraints focussing on the profits realized by fishing firms and different capital stocks.

References

Aubin, J.-P. *Viability Theory.* Birkhäuser, Boston, 1991.

Aubin, J.-P. and P. Saint-Pierre. An introduction to viability theory and management of renewable resources. In Kropp, J. and Scheffran, J. (eds.), *Decision Making and Risk Management in Sustainability Science*, New York, 2004. Nova Sci. Publ. forthcoming.

Banks, R. Subsidising EU fleets: capacity reduction or capital subsidisation. In Hatcher, A. and lastname Robinson, K. (eds.), *Overcapacity, Overcapitalisation and Subsidies in European Fisheries*, pages 200–209, Porthsmouth, 1999. University of Portsmouth, cemare.

Bene, C., L. Doyen, and D. Gabay. A viability analysis for a bio-economic model. *Ecological Economics*, 36: 385–396, 2001.

Charles, A. *Sustainable Fishery Systems*. Blackwell, Oxford, 2001.

Eisenack, K. and J. Kropp. Assessment of management options in marine fisheries by qualitative modelling techniques. *Marine Pollution Bulletin*, 43(7-12): 215–224, 2001.

FAO (ed.). *The state of world fisheries and aquaculture 2000*, Rome, 2001. Food and Agriculture Organization of the United Nations.

Gréboval, D. and G. Munro. Overcapitalization and excess capacity in world fisheries: underlying economics and methods of control. In Gréboval, D. (ed.), *Managing Fishing Capacity: Selected Papers on Underlying Concepts and Issues*, FAO Fisheries Technical Paper 386, Rome, 1999. Food and Agriculture Organization of the United Nations.

Jackson, J. B. C., M. X. Kirby, W. H. Berger, K. A. Bjørndal, L. W. Botsford, B. J. Bourque, R. H. Bradbury, R. Cooke, J. Erlandson, J. A. Estes, T. P. Hughes, S. Kidwell, C. B. Lange, H. S. Lenihan, J. M. Pandolfi, C. H. Peterson, R. S. Steneck, M. J. Tegner, and R. R. Warner. Historical overfishing and the recent collapse of coastal ecosystems. *Science*, 293(5530): 629–638, 2001.

Jentoft, S., B. J. McCay, and D. Wilson. Social theory and fisheries co-management. *Marine Policy*, 22: 423–436, 1998.

Kropp, J., K. Zickfeld, and K. Eisenack. Assessment and management of critical events: The breakdown of marine fisheries and the north atlantic thermohaline circulation. In Bunde, A., Kropp, J., and Schellnhuber, H. J. (eds.), *The Science of Disasters: Climate Disruptions, Heart Attacks, and Market Crashes*, pages 192–216, Berlin, 2002. Springer Verlag.

Kuipers, B. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press, Cambridge, 1994.

Lane, D. E. and R. L. Stephenson. Institutional arrangements for fisheries: alternate structures and impediments to change. *Marine Policy*, (24): 385–393, 2000.

Mahon, R., S. Almerigi, P. McConney, C. Parker, and L. Brewster. Participatory methodology used for sea urchin co-management in barbados. *Ocean & Coastal Management*, 46: 1–25, 2003.

Munro, G. R. The economics of overcapitalisation and fishery resource management: A review. In Hatcher, A. and Robinson, K., ed.s, *Overcapacity, Overcapitalisation and Subsidies in European Fisheries*, pages 7–23, Portsmouth, 1999. University of Portsmouth, cemare.

Myers, R. A. and B. Worm. Rapid worldwide depletion of predatory fish communities. *Nature*, 423: 280–283, 2003.

Pauly, D. Ecosystem impacts of the world's marine fisheries. *Global Change Newsletter*, (55): 21–23, 2003.

Pauly, D., V. Christensen, S. Guénette, T. J. Pitcher, R. Sumaila, C. J. Walters, R. Watson, and D. Zeller. Towards sustainability in world fisheries. *Nature*, 418: 689–695, 2002.

Potter, B. Predatory politics: group interests and managment of the commons. *Environmental Politics*, 11(2): 73–94, 2002.

Scheffran, J. The dynamic interaction between economy and ecology - cooperation, stability and sustainability for a dynamic-game model of resource conflicts. *Mathematics and Computers in Simulation*, 53: 371–380, 2000.

Whitmarsh, D., C. James, H. Pickering, and A. Neiland. The profitability of marine commercial fisheries: a review of economic informations needs with particular reference to the UK. *Marine Policy*, 24(3): 257–263, 2000.

Cellular Automata Modelling of Seagrass in the Orbetello Lagoon

Stefano Marsili-Libelli, Elisabetta Giusti

Dept. of Systems and Computer Engineering, University of Florence

Via S. Marta, 3 - 50139 Florence, ITALY

Email:marsili@ingfi1.ing.unifi.it

Abstract: This paper describes the evolution in time and space of wigeongrass *(Ruppia maritima)* meadows in the Orbetello lagoon, in central Italy, where the control of the submerged vegetation, with a critical coexistence between macroalgae and macrophytes, is the key management problem. While macroalgae are liable to cause dystrophic crises, macrophytes oxygenate and stabilise the sediment and thus control the nutrient flux into the water. This model was developed for the Orbetello Lagoon Managerial Office to predict the development of both groups and test the actions to favour macrophytes over macroalgae, in the context of a decision support system. A previous model was developed to account for the interactions between nutrients and the submerged vegetation in a 2-D spatial context, and this paper presents a further refinement, with the dynamics of wigeongrass *(Ruppia maritima)* including a hydrodynamic model for the water movements and an ecological model describing the interactions between nutrients and the submerged vegetation.

***Keywords**: Cellular automata, Evolutionary computing, Water quality models, Ecological modelling.*

1. INTRODUCTION

The Orbetello lagoon, schematically shown in Figure1, is located along Italy's west coast. It consists of two shallow coastal reservoirs with a combined surface of approximately 27 km^2, an average depth of 1 m and is connected to the Tyrrhenian sea through one port at each end of the western lagoon and one at the south end of the eastern lagoon. The two lagoons are linked by a narrow passage under a bridge of the road connecting the little town of Orbetello with Mount Argentario. Two water-quality monitoring stations, indicated by the two circles in Figure 1, transmit hourly physicochemical data to the Orbetello Lagoon Managerial Office headquarters. The five squares indicate the major antropogenic point-source pollution discharges. The main problem in the Orbetello lagoon is the control of the submerged vegetation, both in biomass and inventory, given the problematical coexistence between macroalgae and macrophytes, mainly wigeongrass.

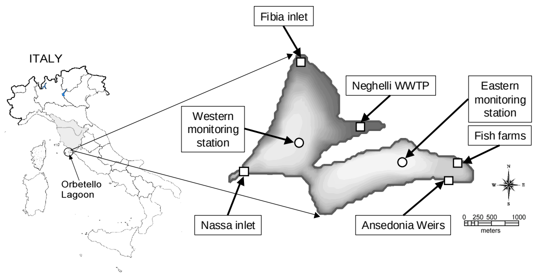


Fig. 1. General view of the Orbetello lagoon, showing the locations of the monitoring stations and the main discharge sites.

While macroalgae may cause dystrophic crises, macrophytes oxygenate and stabilise the sediment and thus control the nutrient flux into the water. A mathematical model has been developed to describe the development of both groups and test the actions to favour macrophytes over macro-algae. This model takes into account the interactions between nutrients and both groups of the submerged vegetation in a 2-D spatial context including a hydrodynamic model for the water movements. A preliminary version of this model has already been developed (Marsili-Libelli and Giusti, 2004). Using a regular grid of cells measuring 200 x 240 m the model operates with two interlocked modules: each cell implements the kinetics of nutrients, vegetation and their interactions, running on an hourly basis to keep track of the circadian cycles, whereas the advection/diffusion model runs on a daily basis to account for mass transfer among neighbouring cells. The model output consists of daily variations in nutrient concentrations and vegetation biomass showing the relative abundance of either group as a function of environmental conditions.

In the original model (Marsili-Libelli and Giusti, 2004) the wigeongrass dynamics consists of a growth and decay balance modulated by several environmental factors. This new contribution attempts at improving this dynamics by modelling the spread mechanism through seed dispersal and germination. The resulting model improves the accuracy of long-term predictions describing the propagation of wigeongrass meadows over several successive years.

The paper is organised as follows: after recalling the main features of the basic model, the ecology of wigeongrass is recalled with an emphasis on the elements which constitute the propagation mechanism. In the following section the mathematical details of the new wigeongrass dynamics are introduced and discussed, and in the last section some simulation results are presented.

2. THE SUBMERGED VEGETATION IN THE ORBETELLO LAGOON

The submerged aquatic vegetation in the lagoon is composed of macroalgae and macrophytes. Their differing physiological requirements tend to produce a mutual exclusion with either group colonising separate areas, depending on the local hydrology, water chemistry and sediment composition. Macroalgae, though of epiphytic origin, float in dense mats and absorb a large quantity of nutrients, eventually producing sudden blooms followed by dystrophic crises. On the other hand macrophytes, being rooted to the bottom play a key role in determining the oxidised or reduced state of the sediments, which is the primary factor controlling nutrient cycling. Selective harvesting is therefore the key problem in the lagoon management and a mathematical model is required to describe the development of either group in time and space, indicating the location and extent of "hot spots" where pre-emptive harvesting may be beneficial.

The macroalgae species commonly found in the Orbetello lagoon are *Chaetomorpha linum, Cladophora vagabunda, Gracilaria verrucosa*, and *Ulva rigida*. They are well documented in the literature (Coffaro and Sfriso, 1997; Duarte and Ferreira, 1997; McGinty and Wazniak, 2002; Naldi and Viaroli, 2002). Though their biological characteristics may be slightly different, in the sequel they will be globally referred to as macroalgae and modelled as a single state variable.

Macrophytes are represented in the Orbetello lagoon by *Ruppia maritima*, commonly known as wigeongrass. In the last few years wigeongrass has been expanding at the expenses of macroalgae in the form of rather compact meadows (Di Biasi et al., 2003). The ecology of the most common *Ruppia* species is well described in the literature (Calodo and Duarte, 2000; Touchette and Burkholder, 2000; da Silva and Asmus, 2001) and the characteristics described in those papers form the knowledge basis for the present model.

3. WIGEONGRASS PHYSIOLOGY AND ECOLOGY

Wigeongrass (Ruppia maritima L.) is a very common and widespread submersed macrophyte (a very thorough description can be found in Kantrud, 1991). In habitats such as the Orbetello lagoon the plant behaves as a perennial. It reproduces by releasing seeds (drupelets) at the top of emerged stalks. The dry weight of below-ground parts averages about 30-45% of maximum seasonal biomass. The below-ground biomass develops best at well-oxygenated sites in coarse-textured bottom sediments, which are low in free H_2S. In fact, complete degeneration of the root system can occur in very highly reduced organic sediments. The root system is delicate and unable to penetrate deeply into sediments. This makes the species susceptible to water turbulence.

In sexual reproduction *R. maritima* produces a large number of seeds about two weeks after flowering. Ripe drupelets are transported short distances in floating vegetation, but may travel longer carried by winds, fish and waterfowl. In temperate climates, drupelets usually lie dormant underwater until the following spring. Most drupelets are found in the upper 5 cm of bottom sediment.

In asexual propagation, *R. maritima* expands by rhizomes. Rapid growth of rhizomes on overwintering plants begins about the same time as drupelet germination and, like germination, is probably temperature controlled. The peak development is controlled by light and temperature; however, times of maximum light and temperature may not be in phase and can pose some problems for wigeongrass. If light exceeds photosynthetic saturation levels, the plants may be temperature-stressed and attain higher biomass later in the growing season when water temperatures are lower.

The interactions included in the basic ecological model (Marsili-Libelli and Giusti, 2004) are shown in Figure 2.

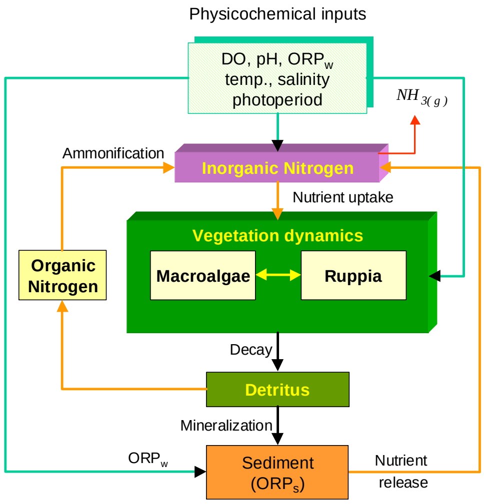


Fig. 2. Basic interaction in the ecological model of the Orbetello lagoon.

4. CELLULAR AUTOMATA FOR WIGEONGRASS DEVELOPMENT

The basic growth and death mechanism underlying the vegetative cycle of wigeongrass, already included in the previous model, is here enhanced to take the form of a ***cellular automata***, including seed production, transport, burial and germination, as described in Sect. 3. In addition to the basic dynamics, the automata consists of a number of rules controlling the growth of *Ruppia* within the cell and seed dispersal if favourable conditions occur in adjacent cells. These conditions are determined by other model variables and are updated at each daily time-step.

A detailed ecological model for *R. chirrosa* was developed by Calodo and Duarte (2000) and for *R. maritima by* da Silva and Asmus, (2001). The sediment pH and ORP appear to have a primary influence on the establishment and development of *Ruppia* meadows, which have increased their cover in the last ten years, whereas there has been a constant decline of macroalgae (de Biasi et al., 2003; Lenzi et al., 2003). The correlation between nitrogen and submersed vegetation was also considered, following the reported link between total nitrogen and macroalgal volume (McGinty and Wazniak, 2002). For space reasons only the *Ruppia* section of the model is reported here, whereas the dynamics for nitrogen and macroalgae can be found in the previous paper (Marsili-Libelli and Giusti, 2004).

4.1 Basic wigeongrass growth dynamics

The wigeongrass dynamics is a balance between growth ρ and decay Ω_R, mediated by the nutrient cell quota N_{int}^R

$$\frac{dR}{dt} = (\rho - \Omega_R) \cdot R \tag{1}$$

$$\frac{dN_{int}^R}{dt} = v_{NH_4}^R \cdot \frac{NH_4}{NH_4 + K_{NH_4}} \cdot + v_{NO_3}^R \cdot \frac{NO_3}{NO_3 + K_{NO_3}} - \rho_N \cdot N_{int}^R \tag{2}$$

The specific growth rate in Eq. (1) is the product of four limiting factors

$$\rho = \rho_{max} \cdot g(d) \cdot f_R(T) \cdot f_R(N_{int}^R) \cdot f_R(R), \tag{3}$$

depending on photoperiod *d*

$$g(d) = 1 - \frac{1}{1 + b \cdot e^{a(d - f_0)}}, \tag{4}$$

temperature *T*

$$f_R(T) = \frac{1}{1 + \left(\left(\frac{T - T_o}{c} \right)^2 \right)^d}, \tag{5}$$

density function *R*

$$f(R) = 1 - e^{-\left(\frac{R - R_{max}}{SL} \right)}, \tag{6}$$

and internal nitrogen quota N_{int}^R

$$f(N_{int}^R) = \frac{N_{int}^R - N_{min}}{N_{cri} - N_{min}}, \tag{7}$$

whereas the decay is a function of temperature.

$$\Omega_R = SR \times (0.098 + e^{-6.59 + 0.2217 \cdot T}). \tag{8}$$

Equations (1 - 8) describe the dynamics of the adult wigeongrass population in each cell of the lagoon.

4.2 Cellular automata for seed dynamics

An improvement to this basic model is presented here, consisting of:

1. Seed production and propagation;
2. A set of rules describing seeds dispersal, burial and germination, as a function of environmental conditions, to produce new plants on the following year.

The cellular automata approach is particularly suited to model the year-to-year evolution, rather than just the seasonal bloom which is less important in the case of seagrass which tend to form long-term, rather compact colonies. To describe the seed propagation the scheme of Figure 3 defines the reference directions of water movements along the horizontal (W-E) direction *u*, and the vertical (N-S) direction *v*.

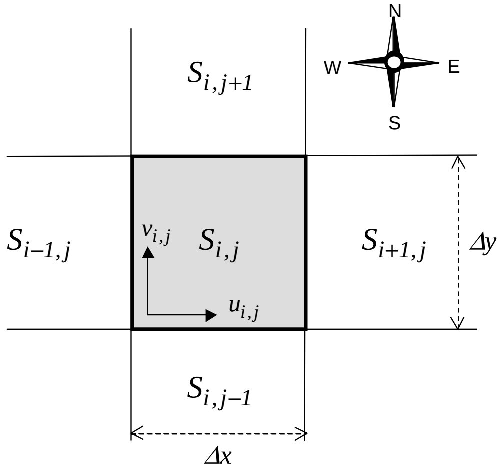


Fig. 3. Cell interactions through advection and diffusion in the spatial discretization. *S* represents the seed concentration in each cell.

It is supposed that as far as sexual reproduction is concerned, the scheme of Figure 4 represents the relationship between adult plants and seeds. The adult plants generate seeds with rate Ps, the seeds lie dormant until the next season when a fraction G generates new plants. Given the perennial nature of *Ruppia* in mild climates, the new plants increase the standing population according to the scheme of Figure 4. This, however, does not take into account the spatial dimension of seed propagation nor the influence of the environmental conditions in seed burial and germination. To link seed production to adult biomass and take into account seed dispersal, the following cellular automata is introduced, which considers seed transport by water movements as defined in Figure 3.

Equation (9) describing the seeds dynamics was derived directly in discrete-time form to fit into the daily module governing the cell evolution, with daily Δt time steps. In Eq. (9), $P_s R_{i,j}^t$ represents the seed production rate of the standing crop at time *t* in the (i,j) cell. The other terms model water transport, where the boolean parentheses account for the flow direction and the fact that seed deposition is possible only if the flow does not exceed a maximum shear velocity Φ, which is assumed here 0.4 m s^{-1} (McGinty and Wazniak, 2002).

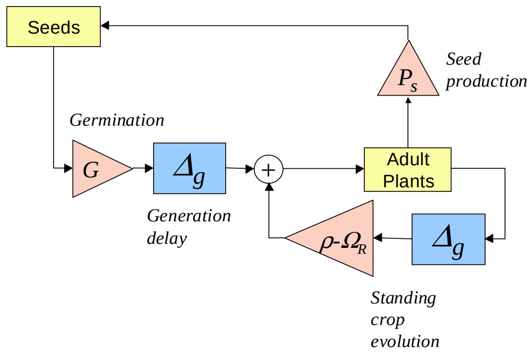


Fig. 4. Interaction between seeds and adult plants, including the generation delay Δ_g.

$$
\begin{aligned}
V_{i,j} S_{i,j}^{t+1} = & V_{i,j} S_{i,j}^{t} + \Delta t P_s R_{i,j}^{t} \\
& - \Delta t S_{i,j}^{t} A_x \left| u_{i,j} \right| \\
& - \Delta t S_{i,j}^{t} A_y \left| v_{i,j} \right| \\
& + \Delta t S_{i-1,j}^{t} A_x \left| u_{i-1,j} \right| \times \left(\Phi > u_{i-1,j} > 0 \right) \\
& + \Delta t S_{i+1,j}^{t} A_x \left| u_{i+1,j} \right| \times \left(-\Phi < u_{i+1,j} < 0 \right) \\
& + \Delta t S_{i,j-1}^{t} A_y \left| v_{i,j-1} \right| \times \left(0 > v_{i,j-1} > \Phi \right) \\
& + \Delta t S_{i,j+1}^{t} A_y \left| v_{i,j+1} \right| \times \left(-\Phi < v_{i,j+1} < 0 \right)
\end{aligned} \tag{9}
$$

Equation (9) is active only during the seed production period, which according to the physiology outlined in Sect. 3, lasts for about two weeks after the first flowering. From the modelling viewpoint, it is therefore assumed that Eq. (9) is activate when the total biomass is below 10% of the maximum vegetative peak and runs only for a limited number of days. At the end of this period the seed distribution is considered final and recorded into a matrix which is introduced as additional initial condition at the start of next year simulation, together with the existing standing crop at the end of the previous year.

4.3 Environmental conditions for seed germination

Other growth conditions regards the state of the sediment. Though no definitive results are available, from literature information and direct observations it can assumed that sediment composition in terms of organic content, Oxido-Reduction Potential (ORP) and pH represent important factors for the germination of seeds and the expansion of the meadow. These facts have been coded into a fuzzy inferential engine based on the following four rules (out of a possible 3x3=9 set of rules) of Table 1, which yield the actual germination percentage of sowed seeds, using the Fuzzy Logic Toolbox (Gulley and Jang, 1995) in the Matlab programming environment.

Table 1
Fuzzy rules for seed germination.

R_1: if pH is medium and ORP is high then G is high
R_2: if pH is high and ORP is high then G is low
R_3: if pH is low and ORP is low then G is low
R_4: if pH is medium and ORP is medium then G is medium

The fuzzy qualifiers (low, medium, high) for the two variables *(pH, ORP)* were defined as shown in Figure 5.

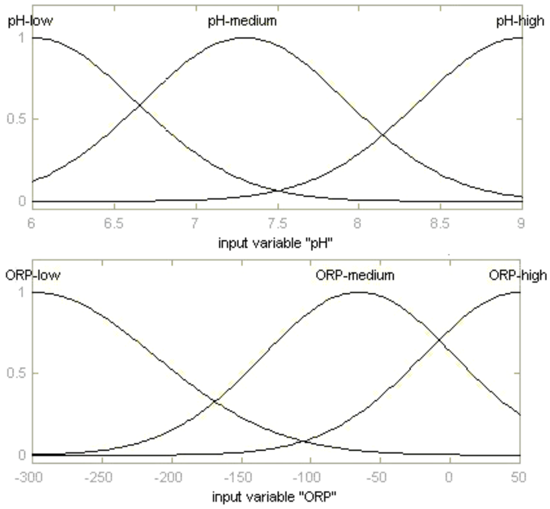


Fig. 5. Fuzzy membership for pH and ORP used in the fuzzy rules of Table 1.

The implication surface resulting from the fuzzy rules of Table 1 as a function of pH and ORP has the form shown in Figure 6. The seed dynamics Eq. (9) and the fuzzy rules of Table 1 form the cellular automata which, together with Eqs. (1 -8) form the wigeongrass dynamics.

Three successive years have been simulated using the weather data of 2002. The results are shown in Figure 7, showing the maximum and minimum extent of the wigeongrass meadow, with a clear expanding trend superimposed to seasonal fluctuations. The meadow tends to spread in the direction of the prevailing winds from NW and germinate where currents are slower.

The second summer is the time when the highest growth is observed, with the greatest in-meadow heterogeneity. After that, more regular growth development ensues and produces a further expansion in the third growing season.

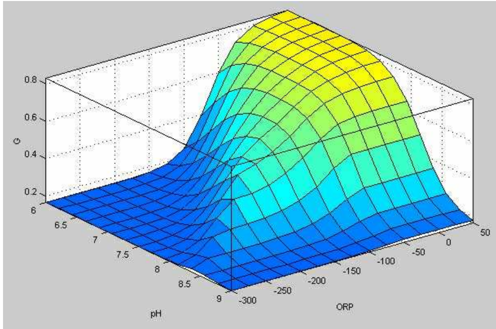


Fig. 6. Implication surface of seed germination rules of Table 1 as a function of pH and ORP.

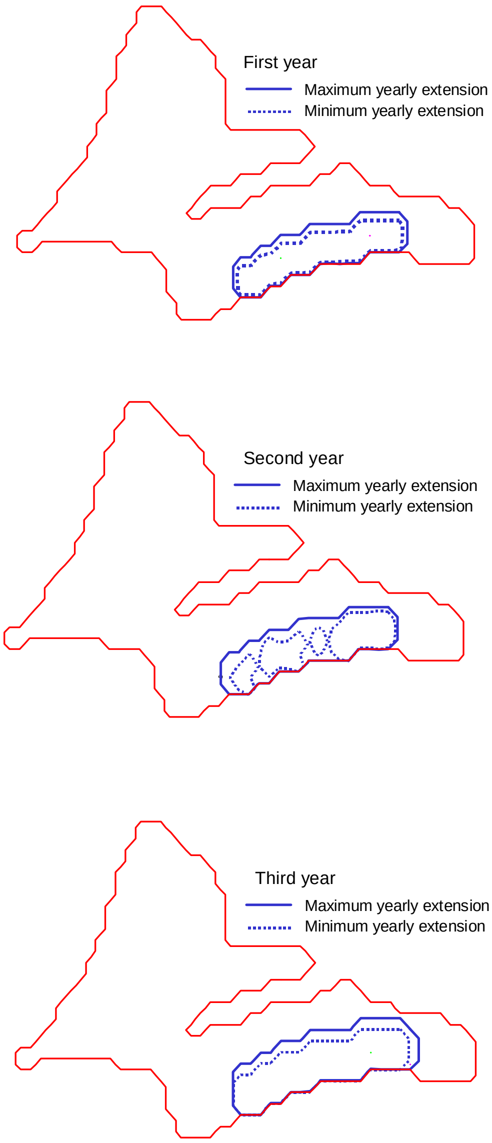


Fig. 7. Yearly fluctuations and long term expansion of the wigeongrass meadow during three consecutive years.

As a concluding remark, Figure 8 shows the general layout of the model, from the software engineering point of view.

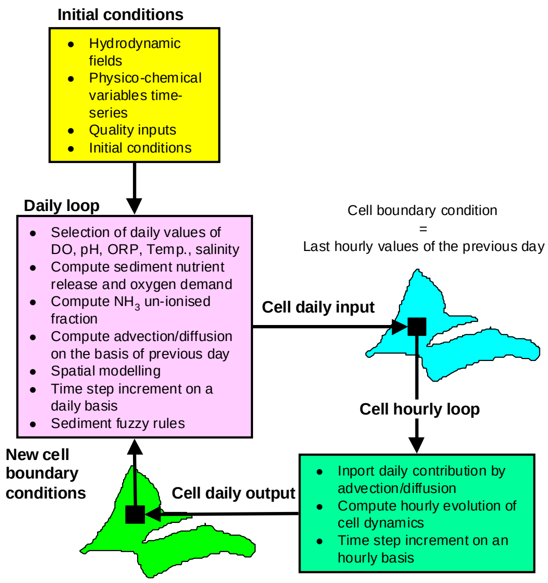


Fig. 8. Software engineering of the ecological model for the Orbetello lagoon.

5. CONCLUSION

This paper has presented an improvement to a previous model describing the evolution of the wigeongrass (*Ruppia maritima*) in the Orbetello lagoon in the context of an ecological model including the dynamics of nutrients and competing groups of submerged vegetation. The previous growth dynamics of wigeongrass has been augmented with the seed model, which consists of a dynamical equation, in the form of a cellular automata and a set of knowledge-based rules defining the favourable conditions for seed spreading and germination. These rules have been implemented as a fuzzy inferential engine. The augmented dynamical model has been integrated into the existing model, which is now capable of describing the observed spread of the *Ruppia* meadows in the Orbetello lagoon over a time horizon of several successive years. The model results are in agreement with the observed behaviour of the existing wigeongrass meadows, which expand slowly but steadily wherever favourable conditions are met.

6. ACKNOWLEDGEMENT

This research was supported by the Orbetello Lagoon Managerial Office under contract n. 96/988 of 24.07.2003.

7. REFERENCES

Calodo, G., and P., Duarte, Modelling growth of *Ruppia cirrhosa*. *Aquat. Bot.,* 68, 29 - 44, 2000.

Coffaro, G., and A., Sfriso, Simulation model of *Ulva rigida* growth in shallow water of the lagoon of Venice. *Ecol. Model.* 102, 55 - 66, 1997.

da Silva, E.T., and M.L., Asmus, A dynamic simulation model of the widgeon grass *Ruppia maritima* and its epiphytes in the estuary of the Patos Lagoon, RS, Brazil. *Ecol. Model.* 137, 161 - 179, 2001.

de Biasi, A.M., Benedetti-Cecchi, L., Pacciardi, L., Maggi, E., Vaselli, S., and I., Bertocci, Spatial heterogeneity in the distribution of plants and benthic invertebrates in the lagoon of Orbetello (Italy). *Oceanol. Acta* 26, 39 - 46, 2003.

Duarte, P., and J.G., Ferreira, A model for the simulation of macroalgal population dynamics and productivity. *Ecol. Model.* 98, 199 - 214, 1997.

Gulley, N and J.S.R., Jang, *Fuzzy Logic Toolbox User's guide*, The Mathworks Inc. Natick, MA, 1995.

Kantrud, H. A., *Wigeongrass* (*Ruppia maritima* L.*): A literature Review.* U.S. Fish and Wildlife Service, Fish and Wildlife Research 10, 58 pp., 1991.

Marsili-Libelli, S., and E., Giusti, Modelling the interactions between nutrients and the submersed vegetation in the Orbetello Lagoon, to appear in *Ecol. Model.*, 2004.

McGinty, M. and C. Wazniak, Understanding the role of macroalgae in shallow estuaries. Maritime Institute, Linthicum, Maryland, pp. 36, 2002.

Naldi, M. and P., Viaroli, Nitrate uptake and storage in the seeweed *Ulva rigida* C. Agardh in relation to nitrate availability and thallus nitrate content in a eutrophic coastal lagoon (Sacca di Goro, Po River Delta , Italy). *J. Exp. Marine Biology and Ecology* 269, 65 - 83, 2002.

Touchette, B.W. and J.M., Burkholder, Review of nitrogen and phosphorus metabolism in seagrasses. *J. of Exp. Mar. Biol. Ecol.* 250, 133 - 167, 2000.

Decision-Support System Workbench for Sustainable Water Management Problems

M.S.Morley[a], **C.K.Makropoulos**[b], **D.A.Savic**[a] **& D.Butler**[b]

[a] *Centre for Water Systems, University of Exeter, Devon, United Kingdom*
email: M.S.Morley@exeter.ac.uk

[b] *Urban Water Research Group,*
Department of Civil and Environmental Engineering,
Imperial College London, United Kingdom

Abstract: Decision Support Systems (DSS) comprise a wide-range of computer-enabled applications that are based on some form of analytical model, commonly linked to a database. Coupled with the visualization and spatial analysis facilities provided by a Geographic Information System (GIS), a unifying framework can be developed to promote the uptake of advanced decision support technology across a wide range of stakeholders. This paper describes an architecture for a Decision Support Workbench to facilitate the rapid development of DSS applications. Custom DSS implementations supporting decision-making for a wide range of problems may be generated through an extensible, interactive environment, featuring "drag and drop" object-oriented components and dynamic connections between them. Also presented are a number of components for the workbench derived from existing tools for integrated modelling, spatial visualization and advanced decision support. The DSS Workbench prototype is demonstrated through an example development of a DSS for water distribution systems modelling and optimization. The developed DSS includes: (1) simulation modelling tools (e.g. EPANET), (2) optimization algorithms based on the Evolutionary Computing principles, (3) database tools for data storage and manipulation, and (4) spatial analysis tools based on GIS. The new DSS is tested on a case study of water distribution network design. Evolutionary Computing, which uses a computer model of the principles of Darwinian evolution to "evolve" good designs, is used here to design a pipe network. This is a highly complex problem for which classical solution techniques such as linear programming or gradient-based methods are often inappropriate or sometimes hopelessly inadequate. The solutions obtained demonstrate the feasibility of developing a DSS Workbench and its useful implementation for water distribution network modelling and optimisation.

Keywords: Evolutionary Computing, Decision Support Systems, Network Optimisation, Workbench

1. INTRODUCTION

Decision Support Systems (DSS) in general and Spatial Decision Support Systems in particular, represent an attempt to assist the decision-making process by a set of intelligent, knowledge-based techniques. These techniques are of particular importance when the problems are complex and semi-structured [Densham, 1991], as is the case in the urban water management environment [Makropoulos et al., 2003]. A major UK EPSRC Research initiative, the Sustainable Urban Environment Programme (SUE) is currently investigating, through its various projects, issues of spatial decision support for sustainable decision-making (CoDES Project), including the sustainable management of the urban water cycle (WaND Project). Due to the complexity and extensive scope of this aim, the development of an extensible environment facilitating the generation of diverse DSS Tools for multiple end-users was considered *sine qua non*. The development and a first application of this environment (Workbench) will be discussed.

2. DECISION SUPPORT FRAMEWORK

The Decision Support System Framework (Framework) is a software engine that underpins a Workbench application and coordinates interaction between a number of modules that can be combined to constitute a specific Decision Support Tool. The Framework may be used either in the context of the Workbench, facilitating the interactive generation of Decision Support Tools or as a library suite that can be used with conventional programming languages for tightly integrated solutions.

2.1.1 Modules

The basic element of the Framework is the Module. The basis of the Module system is the Model – Control – View paradigm popularised by SmallTalk [Goldberg and Robson, 1989] in which software is decomposed into three distinct "layers" for presentation, operation and data analysis/simulation. These layers are also well suited for dividing the application across client/server boundaries. Modules could include analytical engines, input/output interfaces, display libraries, generic optimization routines (e.g. a GA library) etc. Within each "layer" of the Framework, there are prototypes for the different types of modules, which define certain core behaviours for that type – for example, the Data Provider type has to implement a mechanism for discovering the type and structure of the data therein. Figure 1 shows the available module types and their use in constructing a Decision Support Tool.

2.2 Model

The model layer comprises the information upon which the Decision Support System Tool acts. These data may be in many forms: external database connections, spatial data, temporally variable data and software models.

2.2.1 Data Provider

The Data Provider Module type incorporates raw data sources such as database tables, GIS spatial data etc. along with a number of core data components that provide generic access to relational and hierarchical database forms and manipulation of spatial data. At present, access to relational databases in the Framework is achieved through an abstraction layer, which "wraps" a Microsoft ODBC connection to a database. The abstraction layer provides a standard interface to the Framework that allows it to query the relational database as to its data structures and to forward SQL queries to the database. Use of a standard interface allows the abstraction layer, in future, to present an unchanged interface to other database technologies, not limited to ODBC, which would be more appropriate for other platforms. Hierarchical databases – which are, naturally, more appropriate for representing some types of data - are supported in the form of XML-based databases. XML documents are commonly parsed into tree-like structures for manipulation using the Document Object Model (DOM). DOM was initially a proprietary interface specification developed by Microsoft but, unusually, the interface has been adopted by the wider community as a standard. The Framework exploits the hierarchical nature of DOM documents to manipulate databases with heterogeneous structures.

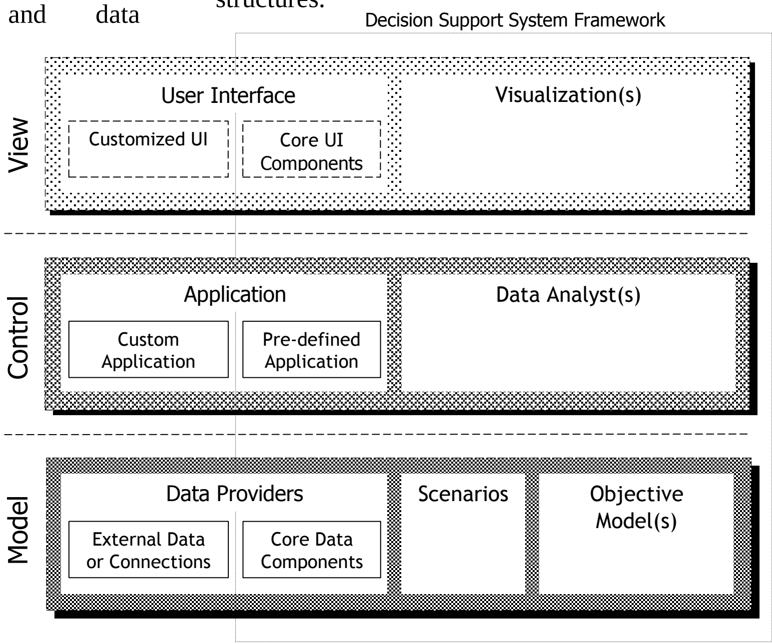


Figure 1: Anatomy of a Decision Support Tool

Access to GIS data from multiple sources is facilitated through the exposure of the ODBC connection if an ODBC aware spatial middleware is being employed. For local GIS tables, the Framework provides core data components for the inclusion of tables from MapInfo and Arc/Info. There is a presumption, however, that the GIS routines employed for visualization are able to accept the data from the data provider – as no translation facility is provided.

2.2.2 Objective Model

The Objective Model Modules represent the real-world environment on which the Decision Support System acts – a necessarily "fuzzy" constituent. For clean and wastewater systems, for example, the key Objective Model component is that of the *pipeline* network model.

2.2.3 Scenario

In order to simplify the structure of the Objective Model, the Scenario concept has been introduced. Scenarios represent instance-specific or temporally variable data for use in "what-if" analysis to record changes in state of the Objective Models or to act as a repository for results from Data Analyst components. Scenarios operate on the metadata exposed by the Objective Model and can be generated automatically as a snapshot of an objective model at a particular time or semi-automatically on a subset of the Model at a given stage. Once created, a Scenario can be used to revert the objective model to the state that it describes. Returning to the pipeline network example, Scenarios could be used to describe the control settings of valves, reservoirs and tanks across the network as well as for storing the results of a 24-hour hydraulic simulation.

2.3 Control

The control layer is responsible for the functional operation of the Decision Support Tool. Despite appearances in the View layer, the Control layer implements all but the most basic functionality of the application. This distinction is vital as it allows for a high-degree of independence between the View and Control layers allowing for the efficient replacement and variation of user interfaces. Whilst the View layer might be responsible, for example, for the processing of mouse clicks, the Control layer will ultimately determine what action that click performs.

2.3.1 Application

The Application module type is responsible for creating and managing all the components in a Decision Support Tool. At present, the Framework provides a standard Application type for undertaking optimization of an Objective Model. Other standard types will be added in future. Custom Application types can be created and added as with any other Module.

2.3.2 Data Analyst

Data Analysts are the actors in the Framework that perform some analysis or act on the data in Objective Model in some fashion. Examples of these components include Genetic Algorithm optimization, GIS spatial analysis and a hydraulic solver etc. Data analysts commonly store their results in Scenarios and present their results to Visualization components.

2.4 View

The topmost layer of the application is responsible for the interaction with the end user and the presentation of results. At the time of writing, the view layer is sparsely populated with key Visualization components – most notably a GIS routine for viewing a network. All user interface processing takes place within the View layer. The key characteristic of View layer Modules is that they are inherently passive – in that they do not respond to actions, as such, rather they utilize data exposed by the Control layer.

2.4.1 User Interface

The User Interface Modules can be neatly broken down into those constituents provided as core components by the Framework and those added to extend the core components in order to add support required by custom Application types. The core components envisaged for the system include common visual GUI elements such as forms, buttons and the common graphical controls associated with dialog boxes. Because of the variation in implementation of GUI elements across operating systems it is anticipated that only a sub-set of controls will ultimately be included as part of the Framework provided UI. Additional elements can be added to the system programmatically provided they conform to the Framework's interface requirements.

2.4.2 Visualization

Visualizations encompass facilities such as GIS maps, charts, 3D rendering etc. and access data exposed, particularly by Data Analysts, for display. Currently, the visualizations implemented include a GIS window, which is used to provide a graphical display of a network model with thematic mapping capabilities as well a basic charting Module, which can be use for displaying result comparisons.

2.5 Intra-Module communication

Behind the scenes of the Framework is a sophisticated software component, which has been termed an Object Interoperation Manager (OIM). This component is responsible for marshalling the interactions between the Modules in a Decision Support Tool. The OIM implements a standard interface for components to interact. In doing so it obviates the need for the components to be implemented in the same technology. The DSSF provides template wrappers for Windows DLLs, Windows COM objects, Windows .NET assemblies and CORBA objects. This interface can be developed to include support for web services accessed over SOAP (Simple Object Access Protocol) – this being functionally similar to CORBA's IIOP. Modules need to be registered with the Framework before they can be used in a Decision Support Tool. Registration takes place by passing an XML document to the Framework containing basic information on the purpose and functionality of the Module and, crucially, where the Framework can find the Module in question to instantiate it. Once registered with the Framework any Tool can have access to the Module.

2.5.1 Metadata

In order to meet the extensibility design criterion, it became apparent that the component system needed the ability to "publish" information about itself in such a fashion that Modules could query the capabilities of their peers and the data contained therein. At a more fundamental level, the attributes of individual elements within the Modules – particularly those associated with the model implementation – should also be able to be queried by user-interface Modules.

The absence of a built-in metadata facility in standard C++ and concerns about extending such functionality into other languages prompted the adoption of a simple metadata engine that is used for the representation of public data within Modules as well as facilitating the transfer of data between them. The desire to support diverse object implementations such as COM and CORBA also militated against adopting anything other than a proprietary metadata engine.

2.5.2 Thread safety

Multithreading enables an application to run many processes simultaneously and to take advantage of the multi-processor support provided by the Microsoft Windows NT and Linux operating systems. The use of this symmetric multiprocessing (SMP) allows operations to run concurrently on different processors in the same machine, or to parallelise the execution of object functions. The extensible nature of the Framework militates against adopting a standard threading model, as it is difficult to predict the needs of the potential modules that might be incorporated into a Decision Support Tool. Consequently, the OIM implements a centralized messaging system that facilitates communication between the Modules in a thread-independent fashion. Each Module is required to poll periodically this queue to discover if there are any messages waiting for it. The mechanics of this system are hidden from Module developers in the stubs of the Module code.

2.6 Workbench

The workbench application provides a "front-end" to the Framework – allowing the "drag-and-dropping" of the Modules registered with the Framework to create bespoke Decision Support Tools. In future, it will further allow rudimentary visual development of interfaces for these tools. Object-oriented "inspectors" allow the end-user to define the relationships between the Modules in terms of their attributes and events that can be .

One of the virtues of developing such Modules for use in the Workbench is that will implicitly have a full, coherent API. In turn, this means that it will be relatively straightforward to reuse the components within a more conventional programming context. This allows for the development of more complex applications than would otherwise be possible using the interactive Workbench technique.

3. DECISION SUPPORT TOOL: PIPE NETWORK DESIGN

Solving optimisation problems related to water distribution networks is recognized as an NP-hard analysis that has conventionally been approached using a number of techniques including hill climbing, linear and dynamic programming. Evolution algorithms represent an alternative, proven strategy for approaching these problems. The prototype application (Figure 2) seeks to demonstrate that the optimisation approach used in later versions of GAnet (Morley et al. [2001]) can be replicated in a user-extensible form within the Framework using a number of Modules. This application allows for the optimal design, on a least-cost basis, of a hydraulic network based on user-defined performance criteria.

3.1 Genetic Algorithm Library

<u>Data Analyst</u>. Genetic Algorithms (GAs) are stochastic algorithms whose search methods model mimic genetic inheritance and Darwinian mechanisms of survival [Michalewicz, 1996]. Individuals in nature evolve by developing characteristics allowing them to adapt in a particular environment. These characteristics are incorporated in their genes and the best ones survive through the process of natural selection, from one generation to the next. The individuals in the case of GAs are solutions to a given problem. The more robust ones survive and evolve. The metaphor is conceptually a powerful one: Darwinian natural selection and survival can be considered as one of the most complex and demanding "optimisation" problems ever encountered. Evolution is a natural process and is only "natural" that a mathematical equivalent would be used for some of our own complex optimisation problems [Walters & Savic, 1994, Makropoulos, 2003].

3.2 Network Model

<u>Objective Model</u>. Previous commercial work on optimisation software for water networks, (including Atkinson et al. [1998]), has highlighted the absence of an agreed standard for representing water network infrastructure and operating conditions. It has proved necessary to translate the representation of network infrastructure from the clients' network-modelling software into a form that can be understood both by GIS applications and for by a hydraulic network solver that can be automated for the purposes of performing the optimization. This earlier version used a commercial hydraulic network solver - directly manipulating the network structure within the solver. Extending this software to accommodate the differing modelling packages used by water companies was hampered by differing conventions

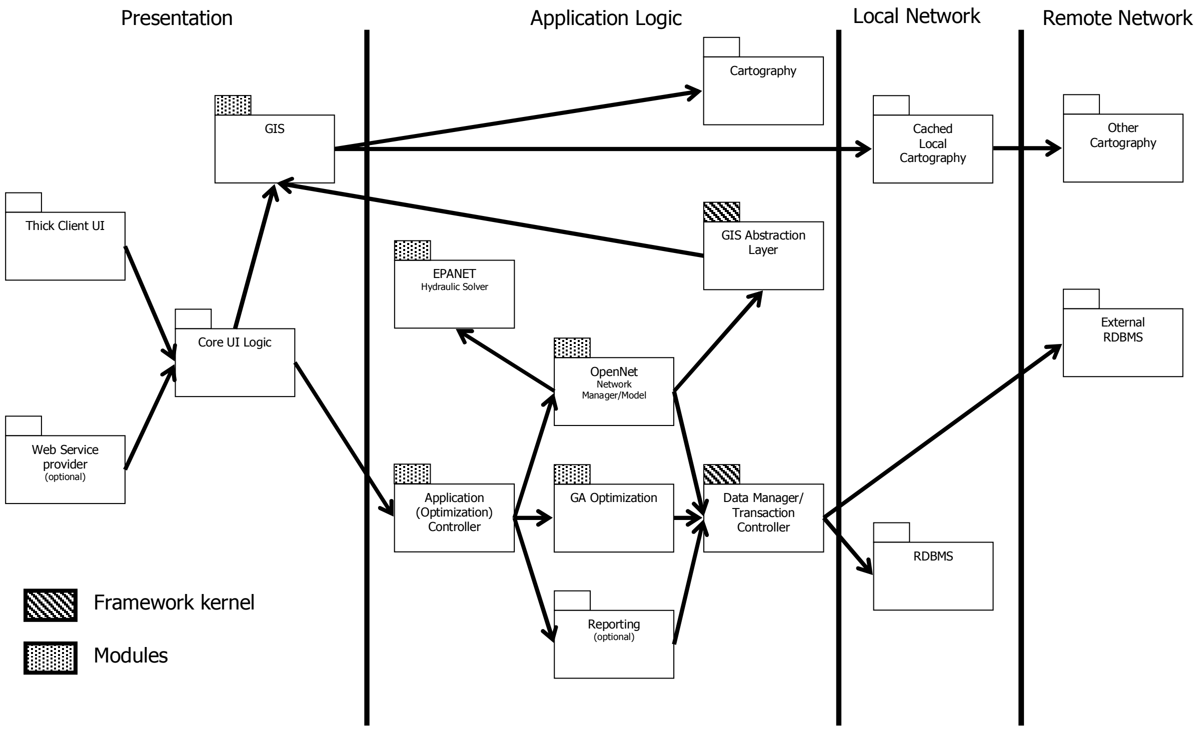


Figure 2: UML Deployment diagram for DSS Framework prototype application

and, in particular, different representations of similar hydraulic elements such as valves. To this end, an object-oriented class-library, known as OpenNet [Morley et al., 2000], has been developed as an abstraction to hide the inner workings of the network solver from the optimisation software. This class library is allied to an XML metafile representation of the network, which is designed to facilitate easier dissemination of network infrastructure data as well as storing the data in a human readable form. For the purposes of integration with the Decision Support System Framework, OpenNet has been completely revised and rewritten to accommodate the meta-data engine used by the Framework that ensures that all properties of the network may be enumerated and accessed by other Modules.

3.3 Hydraulic Network Solver

Data Analyst. A customized version of the EPANET 2 hydraulic solver [Rossman, 1993] has been abstracted and encapsulated in a Windows Dynamic Link Library (DLL). In order that the solver may continue to be used with network model Modules other than OpenNet, a small helper class – not shown in Figure 2 – facilitates the seamless hydraulic solution of an OpenNet pressurised network by automatically reflecting within EPANET, changes made in the OpenNet network.

3.4 Relational Database

Data Provider. In this prototype, the provision of a relational database connection is undertaken to demonstrate the feasibility of such a connection. The data table used by the Decision Support Tool is a simple look up table of available pipe diameters.

3.5 Thematic map

Visualization. A GIS interface component is provided, based on a third party COM object, which allows the visualization of the hydraulic network configuration. As with the hydraulic network solver, a helper class facilitates the seamless data transfer between network and GIS representation. In future the interface between such modules will be further abstracted removing the need for the helper classes.

4. CASE STUDY: NEW YORK TUNNELS (BENCHMARK) PROBLEM

The objective of the New York Tunnel (NYT) problem is to determine the most economically effective design for addition to the existing system of tunnels that constituted the primary water distribution system of the city of New York [Schaake and Lai, 1969]. The twenty-one existing pipes in the network are each considered for duplication to reinforce supply capacity. For each pipe 15 discrete pipe diameter choices available: 36, 48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 180, 192 & 204 inches – along with a “do nothing” option that leaves the existing pipe unduplicated. This gives a solution space of $16^{21} = 1.93\times10^{25}$ possible network design combinations. The minimum available head requirement at all nodes is fixed at 255ft except for nodes 1, 16, 17 that are 300, 260 and 272.8ft respectively. The objective function for the optimization is non-linear: $C_{ij} = 1.1 D_{ij}^{1.24} L_{ij}$ where: C is cost in dollars, D is diameter in inches and L is length in feet. In operation, the Decision Support Tool constructed within the framework, determined the minimum cost option to be $38.80m in line with the result reported by Murphy et al. [1993] and Savic and Walters [1997]. Work to extend the validation by implementing a multiobjective GA analysis of the same problem [Savic, 2002] is underway.

5. CONCLUSIONS

The development of a software Decision Support Framework provides a means of creating Decision Support Tools from a suite of modular building-blocks as part of either an interactive “Generator” application or for use as a more conventional software library. Within the framework additional modules have been developed for innovative optimization and modelling strategies as well as providing the platform for the development of the sustainable urban water tool deliverable from Work Package 6 of the WaND project. This Work Package envisages the development of a GIS-based toolbox to support the development-specific planning and preliminary design of sustainable urban water management practices under different new development scenarios and will serve as one of the main innovative deliverables of this project.

6. ACKNOWLEDGEMENTS

This work is jointly developed by the ‘Water Cycle Management for New Developments’ (WaND) project [*www.wand.uk.net*] and the “Consortium for Decision Support” (CoDeS) project - both funded under the Engineering & Physical Science Research Council’s “Sustainable Urban Environment” Programme by EPSRC, UK government and industrial collaborators.

7. REFERENCES

Atkinson, R.M., M.S. Morley, D.A. Savic and G.A. Walters, GAnet: The Integration of GIS, Network Analysis and Genetic Algorithm Optimization Software for Water Network Analysis, in *Hydroinformatics 98*, V. Babovic and Larsen, L.C. (eds.), Balkema, Rotterdam, 357-362. 1998.

Densham, P. J. Spatial Decision Support Systems. In: *Geographic Information Systems: principles and application*, Maguire, D.J., Goodchild, M.S. and D.W. Rhind (eds.), 403-412, Longman, Harlow, U.K., 1991.

Goldberg, A. and D. Robson, *Smalltalk-80: The Language*, Addison-Wesley, Reading, U.S.A., 1989.

Makropoulos C., D. Butler and C. Maksimovic, A Fuzzy Logic Spatial Decision Support System for Urban Water Management. *Journal of Water Resources Planning and Management - ASCE*, *129*(1), 69-78, 2003.

Makropoulos, C. *Spatial Decision Support for Urban Water Management*. PhD Thesis, Imperial College London, UK, 2003.

Michalewicz, Z. *Genetic Algorithms + Data Structures = Evolution Programs. 3rd edition*, Springer-Verlag, 387pp, New York, 1996.

Morley, M.S., R.M. Atkinson, D.A. Savic and G.A. Walters, GAnet: Genetic algorithm platform for pipe network optimisation, *Advances in Engineering Software*, *32*(6), 467-475, 2001.

Morley, M.S., R.M. Atkinson, D.A. Savic and G.A. Walters, OpenNet: An application-independent framework for hydraulic network representation, manipulation and dissemination, 4th Hydroinformatics Conference of the IAHR, Iowa City, U.S.A., 2000.

Murphy, L.J., A.R.Simpson and G.C.Dandy, *Pipe Network Optimization using an Improved Genetic Algorithm*, Research Report R109, Department of Civil and Environmental Engineering, University of Adelaide, 51pp, 1993.

Rossman, L.A. *EPANET User’s Manual*. United States Environmental Protection Agency, Cincinnati, U.S.A., 1993.

Savic, D.A., Single-objective vs. Multiobjective Optimisation for Integrated Decision Support, In: *Integrated Assessment and Decision Support*, Rizzoli, A.E. and A.J. Jakeman (eds.), Proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society, Lugano, Switzerland, *1*,7-12. 2002.

Savic, D.A. and G.A. Walters, Genetic Algorithms for the Least-cost Design of Water Distribution Networks, *Journal of Water Resources Planning and Management - ASCE*, *123*(2), 67-77, 1997.

Schaake, J. and D. Lai, *Linear Programming and Dynamic Programming Applications to Water Distribution Network Design,* Report 116, Department of Civil Engineering, MIT, 66pp, Cambridge, U.S.A. 1969.

Walters, G.A. and D.A. Savic, Optimal Design of Water Systems Using Genetic Algorithms and Other Evolution Programs, keynote paper in *Hydraulic Engineering Software V Vol. 1: Water Resources and Distribution*, W.R. Blain and K.L. Katsifarakis (eds.), Computational Mechanics Publications, Southampton, UK, 19-26, 1994.

Application of Genetic Algorithms for the Optimisation of Multi-Pollutant Multi-Effect Problems

Stefan Reis, **Steffen Nitter and Rainer Friedrich**

Institute of Energy Economics and the Rational Use of Energy,
University of Stuttgart, Germany
Stefan.Reis@ier.uni-stuttgart.de

Abstract: In this paper, crucial aspects of the implications and the complexity of interconnected multi-pollutant multi-effect assessments of both air pollution control strategies and the closely related reduction of Greenhouse Gas (GHG) emissions will be discussed. The main aims of the work described here are to identify the core problems which occur when trying to apply current state-of-the-art methodology to conduct integrated assessments – in this context, cost-benefit assessment (CBA) as well as cost-effectiveness assessment (CEA) – using sophisticated computer models and propose solutions to the problems identified. The approaches described will display the integrated use of databases, efficient Genetic Algorithms (GA) and already existing software tools and models in a unified model framework. The first part of the paper discusses the need for new developments in one particular field of Integrated Assessment Models (IAMs), the use of (typically) country-specific single pollutant abatement cost curves, which have been applied in a large number of modelling approaches with the aim to find cost effective solutions for given air quality targets. However, research conducted to find such cost effective solutions for the non-linear problem of tropospheric ozone abatement (dealing with two primary pollutants and their rather complex relationship to form tropospheric ozone) identified basic problems of cost-curve based approaches even in this two-pollutant case. The approach discussed here solves the key problems identified, making extensive use of databases in order to provide fast and high quality model input for CEA and CBA. In addition to that, the application of Genetic Algorithms will be discussed as a means to address extremely complex, vast solution spaces which are typical for the tasks IAMs are set to solve nowadays. In the final part of the paper, diversity increasing operators and methods to increase the performance of the GA to find optima are described and first results of extensive model runs are discussed.

Keywords*: genetic algorithms; optimisation; environment; air quality; climate change*

1. SCOPE

As air quality limit values have become more stringent during the last 20 years, and with the need to reduce emissions of greenhouse gases beyond the scope of no-regret technologies to achieve the Kyoto commitments, the costs of emission control has constantly gained importance. In international negotiations for instance the protocols to the UNECE Convention on Long-Range Transboundary Air Pollution (CLRTAP), the analysis of cost-effective ways to reduce emissions played a major role since the 1970s. Binding emission control targets for Sulphur Dioxide (SO_2) and Nitrogen Oxides (NO_x) were agreed upon with the backing of model calculations of the related costs of control strategies (cf. Amann et. al 1996, 1999; Ap-Simon 1994a,b, and 1996; Bailey 1996 as well as Gough et al. 1995, 1998), and even the most recent protocol to the CLRTAP, the Gothenburg Protocol was designed on the basis of IAM calculations using the RAINS model developed by IIASA.

2. PROBLEM FORMULATION

Integrated Assessment Models (IAMs) applied in this context mostly use(d) single abatement cost curves as input to their optimisation tools, in order to identify the least-cost ways to achieve given reduction targets, and to assess the overall costs of strategies. Typically, the analysis focussed on a single pollutant (e.g. SO_2, NO_x) with a (usually) linear relationship between emissions and concentrations, respectively emissions and effects. The case of acid rain and acidification in general (Gough et. al. 1995) is one of the most prominent examples, where reductions of emissions of SO_2 and/or NO_x would usually lead to reduced deposition in the same order of magnitude. The assessment models had to take into account transport of pollutants through the air to some extent, in order to map the regional distribution of deposition changes, while chemical transformation of pollutants did not play a major role yet.

When air pollution by tropospheric ozone became the focus, the modelling task turned more difficult, as the relationship between the emissions of ozone precursor substances NO_x and Non-Methane Volatile Organic Compounds (NMVOCs), and to some extent Carbon Monoxide (CO) as well as the formation of ground level ozone is not linear. Thus, the assessment models needed to include more complex mechanisms to account for these non-linearities. At the same time, the results not only needed to address the question, which measures to take in order to achieve a cost-effective solution, but instead had to tell, which pollutant to be controlled, and where – thus making the optimisation task more complex as the location and distribution of emission sources matters (cf. Friedrich and Reis, 2000). But even apart from the modelling aspects, the case of two pollutants to be controlled introduced an issue that has not been properly solved up to date. While in a single pollutant case, the costs per unit of pollutant controlled by a specific measure are usually easy to pinpoint, given sufficient information about activity rates, the technology and the economics of the respective installations, the situation becomes more difficult as soon as two pollutants were to be controlled, and measures existed, which would reduce the emissions of both pollutants when installed, usually with differing efficiency, thus creating the need for allocating cost proportions to single pollutant cost curves. Even though there are some possible ways to split the costs into different proportions and allocate these to different single abatement cost curves, most of them are more or less arbitrary and reflect more the preference of the model developer than anything. Furthermore, this approach implies a source for inconsistencies, as model solutions may lead to results where the same measure is applied due to its position in one cost curve and excluded because of a later position in a different cost curve.

In this paper, these particular problems shall be discussed, with a focus on the current development towards multi-pollutant multi-effect assessment models, where a robust and transparent methodology to solve this problem could prove to be vital. In the second part, a new methodology to model the application of technical and non-technical emission control measures and the respective costs of abatement is introduced and discussed in detail. Finally, an outlook will be given with respect to the application of this new methodology in the European research project MERLIN (see http://www.merlin-project.info).

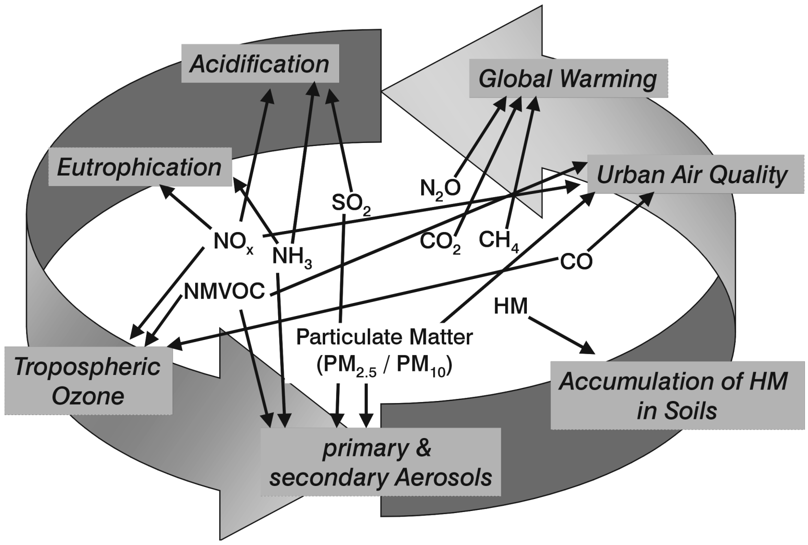


Figure 1. Illustrating the Multi-Pollutant Multi-Effect Environment for IAMs

3. IAMS IN A MULTI-POLLUTANT MULTI-EFFECT ENVIRONMENT

To the same extent that the knowledge about air pollution and its impacts increased, the development of more and more complex models to analyse various aspects of air pollution in an integrated way took place. *Figure 1* illustrates the level of complexity which is characteristic for current assessment tasks. And as it was indicated above, model concepts that have been suitable to address comparatively simple modelling and optimisation tasks are far less fit to cope with this level of complexity.

3.1 Limitations in the Use of Single Pollutant Abatement Cost Curves

As it was indicated in the problem formulation, single abatement cost curves for one specific pollutant and a source sector or country are widely used in IAMs. They usually serve the purpose to provide a function of costs and related emission (abatement) levels in a computable way, for instance as input to optimisation algorithms. However, the limitations of such single abatement cost curves are obvious, in particular in the view of the correlations between different pollutants and effects as indicated in *Figure 1*. Furthermore, generating abatement cost curves as input to optimisation leads to an artificial constraint of the models that are applied to find optimal solutions for a given task. As abatement options have to be ranked e.g. according to their unit costs (€/t), vital issues, such as a different abatement efficiency of a specific option depending on at what stage it is taken, cannot be accounted for (abatement measures applied to the same source sector often mutually influence their abatement efficiency, for instance, a measure that is applied first reducing x% of emissions from a specific source reduces the absolute efficiency – in terms of *tonnes of pollutant abated* – of a second measure applied to the same sector, and vice versa).

Hence, a new approach has to be taken that is able to reproduce the complex interconnections between pollutants and effects, but at the same time has to be transparent and simple enough to keep uncertainties to a minimum. Here the extremely fast increase in both computer speed and data storage and handling capacities provides the basis for innovative solutions. Basically, the same data as would be needed to generate abatement costs curves is collected, with more level of detail even to improve the reproduction of sector-specific characteristics. This comprises the following main data types:

- data on stock and activities (e.g. number of vehicles and annual mileage)
- data on measures (e.g. applicability, efficiency, implementation degree, costs)
- 'meta-data' (information on relationships between measures)

Instead of trying to process and split this data into single abatement cost curves, the optimisation model is given full access to the databases, thus being able to select, apply and evaluate abatement options with a considerable degree of freedom. And as an additional benefit, this approach permits the inclusion of structural changes due to the implementation of abatement options, for instance increasing an activity of one sector in order to reduce that of another.

This 'measure-matrix-approach' creates a number of additional modelling opportunities, e.g. by making it possible to assess single measures, individual sectors or whole countries/regions with simple presets, as no pre-processing of data is needed. Moreover, it does reflect the real-world characteristics of abatement options to a far greater extent than before, as in most cases, costs of abatement options are expressed relative to its application on stock or by activities. In addition to that, abatement options usually address not only one single pollutant, but a portfolio of different pollutants, either reducing or increasing emissions. This is of particular importance for the assessment of multi-effect problems, as such analyses usually have to achieve conflicting targets. Finally, this approach is not limited to mere technical abatement options, as it can include structural measures (e.g. changes in the sectoral structure of electricity generation etc.) and non-technical abatement options in the same way.

3.2 Intelligent Algorithms to Solve Complex Problems

A second critical issue for the assessment of complex multi-pollutant multi-effect problems is that of optimisation. While IAMs to date usually apply either linear optimisation algorithms (Amann et al., 1999), or simple iterative approaches (Friedrich and Reis, 2000), finding optimal solutions in a solution space as complex and vast as it is characteristic for multi-effect problems needs faster approaches. Here, evolutionary (as well known as 'genetic') algorithms (EA) can be the ideal tool, even though they have not been widely applied in the field of air pollution

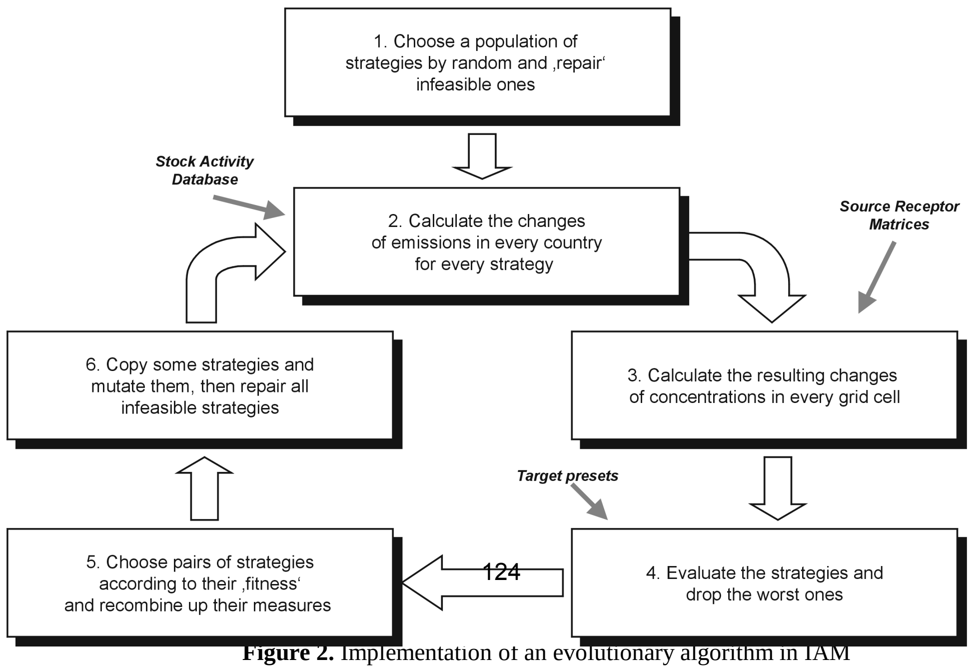


Figure 2. Implementation of an evolutionary algorithm in IAM

modelling yet (cf. Loughlin et al., 2000). As their name suggests, EAs optimise in a way similar to that of nature, using concepts such as recombination, mutation and fitness for survival to induce a process of evolution towards an optimal solution. An exemplary implementation is described here.

The optimisation algorithm as it is applied in the MERLIN project forms the core of an IAM to conduct cost-effectiveness (CEA) and cost-benefit analysis (CBA) of combined strategies to reduce air pollutant and greenhouse gas emissions simultaneously. This IAM is termed Optimisation Model for Environmental Integrated Assessment (OMEGA-2, the first OMEGA model was developed for the optimisation of Ozone abatement strategies, see Friedrich and Reis, 2000). The implementation of an evolutionary algorithm to identify optimal bundles of abatement measures is illustrated in *Figure 2*.

The decision to apply EA emerged, as it became clear that the problem to be solved was characterised by a vast solution space, as hundreds of different abatement options could be combined. For this particular situation, other approaches that were investigated, for instance global or local random choice, gradient based algorithms or divide and conquer strategies could not offer satisfactory performance. On the other hand, a 'black-box' situation had to be avoided, as for this particular task, the pathway to an optimal solution can provide as vital information as the solution itself.

In principle, the problem to be solved can be formulated as follows: from all possible abatement options ('measures'), the set of measures has to be identified, which fulfils all criteria (in this case air quality limit values and GHG emission limits) simultaneously at least costs.

Steps 1 and *2* form the initialisation to start the optimisation and enter the loop, where – in our case – abatement options are selected to reduce a variety of emissions to air. In step 3 the resulting changes of concentrations of pollutants are calculated, using so-called source-receptor matrices (SR-M). To reduce computational effort in this step, the resolution of the matrices will first be reduced and then gradually increase every generation run, until the finest grid resolution of 50x50 km will be achieved.

Thus using the total costs of the abatement measures and the preset thresholds that are still exceeded, the strategies can be evaluated in step 5. This approach allows different weights for limit values that are not achieved, introducing a so-called "fitness value" which will then be used to discard the worst performing strategies.

Pairs of strategies (parent-generation) are selected according to their degree of fitness, which will pass their measures on to two newly formed strategies (child-generation). In a first step, an n-point crossover mutation approach will be implemented in the algorithm, as it is illustrated in *Figure 3* (in this case, a two-point cross-over,) where the parent measures are cut in a number of pieces, which are then recombined to form the offspring.

The position of the measures within the strategies will play an important role as well. If, for instance, two strategies with sufficient fitness are selected in step 5, it would be harmful, to place measures of the first strategy which, for instance, focus mainly on reduction of one particular pollutant in one country at the beginning and those of the second strategy at the end. In this case their offspring (the next generation) would probably consist of one strategy, that has no such measures at all and one that has twice as much as needed, thus resulting in unbalanced individual strategies in the next generation.

To overcome this, groups of measures are formed that have more or less similar effects, where some measures may be members of several groups. This will be done automatically, so new measures can easily be added to the measure database. The strategies will consist of several sections, and every section can only include measures of one single group. So the mixing up of measures in step 5 will either be done by copying the whole measure group of one parent strategy or by n-point-crossover. Aside from solving the problem mentioned above, the measure groups also allow small variations of the strategies, as follows.

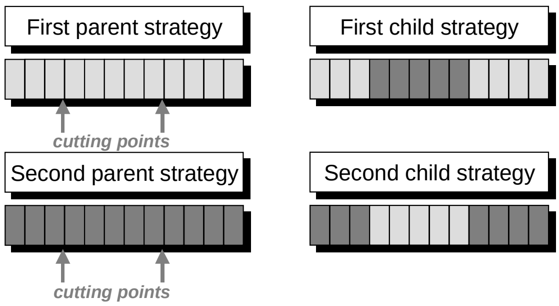


Figure 3. Illustration of a 2-point-crossover mutation

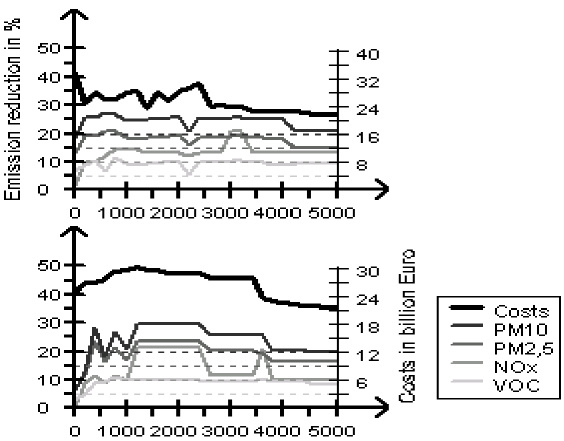
Most evolutionary algorithms simulate mutation of the individuals. In step 6, some strategies are chosen by random, and one or more of its measures will be replaced by other ones, that roughly, but not exactly, have the same effect. Because this is the case for measures of the same group, each one which fits into the same position of the strategy (and thus is a member of the same section) can be chosen. To make sure, that the fitness cannot decrease from generation to generation, the chosen strategy shall be duplicated, and only the copy will be allowed to mutate. This combination of a global search method (the crossover of strategies, done in step 5) and a local one (mutation as done in step 6) is often considered to be the key to the power of EAs in optimisation problems.

The following improvements of the EA-approach are in the process of being implemented and tested:

- Inclusion of special strategies as subsets of the starting population, to direct the search to regions of the solution space, which indicate potential for local/global optima.
- Preference for pairs consisting of solutions of the same neighbourhood to support local search.
- Enhancement of the fitness of young solutions, i.e. leaving the mutation operator enough time to improve them locally, so they are not prematurely suppressed by older ones.
- Use of diversity increasing operators, preventing the search to ignore promising regions too early in favour of few strategies with high fitness.
- Simulation of SINEs (short interspersed elements) to provide points to the crossover operator where cutting is done with increased probability.

As first result of OMEGA tested on some subsets of the final measure and stock/activity databases indicate, the diversity increasing operators seem to be most promising, since the whole population soon converges to some local optimum (see *Figure 4*), hence the mating operator did not have strong effects anymore.

Probably the most prominent of diversity



increasing operators are the “Messy Genetic Algorithms”, proposed by Goldberg in 1989 and generalised by van Veldhuizen in 1999 to multi-objective problems. Unfortunately this approach is not suited well for interacting measures which in some cases might in-/exclude others or modify their effects. Hence, to force the algorithms to cover at least the major part of the search space, the target presets were altered by some vanishing trigonometric function, which takes the generation number as an argument. This can best be recognised by *Figure 5*, showing the optimisation progress restricted to the solvent use sector, where more or less only one single pollutant (non-methane volatile organic compounds, short: NMVOC) is of importance.

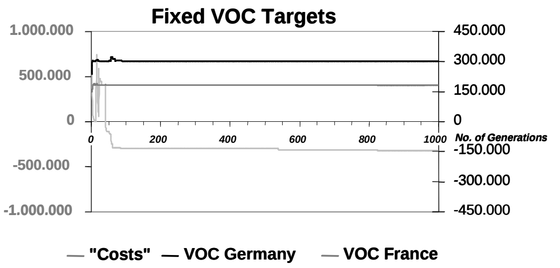


Figure 4. Premature convergence to local optimum

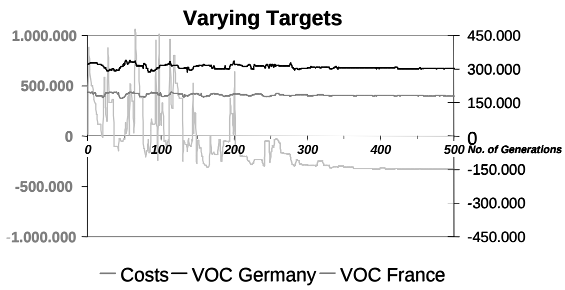


Figure 5. Changing VOC-Targets

Figure 6. : Twofold evaluation

Of course the algorithm computing time increases until it converges, but the optimum found in most runs was slightly better. Further improvement was achieved, when the best strategy of every generation was also evaluated by the fixed reduction targets and stored if it was the best one found so far, to be included in the population from time to time. Looking at this best abatement strategy of every generation the costs can only increase if it came closer to some reduction target (in any country), which was not met before.

For comparison *Figure 6* shows two typical optimisation runs, at the top with oscillating emission reduction targets and only a simple evaluation with regard to the distance to oscillating reduction targets, and below an optimisation run where every strategy is evaluated twofold (using the distance to the fixed reduction targets as well as to the oscillating targets). The upper figure indicates that when the optimisation approaches a threshold, which was not met so far, costs can increase in the early stage of the optimisation, but in contrast to the lower figure, shows a monotonic decrease of costs afterwards.

4. SUMMARY AND CONCLUSIONS

The previous sections have given an insight in both the most prominent problems for IAM in a multi-pollutant multi-effect environment, and innovative solutions to tackle them. To some extent, it is astonishing, that only little development in the field of IAMs for CEA and CBA could be witnessed during the last decade, while increasing computing power and growing knowledge about causes and effects of air emissions improved the situation for modelling to quite some extent.

However, as the problems to be addressed are marked by increasing levels of complexity, new approaches need to be taken. This is even more true, as the next development steps for IAMs are quite predictable. On the one hand, cross-media approaches need to be established, as research has already identified the importance of, for instance, deposition of air pollutants into surface water and soils. In a similar way, carbon sequestration in soils or oceans, or emissions of specific pollutants from soils to air are of importance. On the other hand, economic evaluation of environmental costs, both costs of abatement and external effects of environmental pollution, gains more and more importance for policy implementation. Thus, the full integration of models and tools for macroeconomic assessment of key indicators (e.g. GDP, employment effects, distributional effects, burden sharing etc.) has to be realised.

The benefits of this approach are obvious: No allocation of costs to single abatement cost curves is needed. Furthermore, measures are either applied, or not, reflecting a real-world choice of options. As the model selects measures from a database, maximum flexibility is achieved, new measures can easily be introduced, or others removed. Finally, the order in which measures are applied is taken into account. Hence, the problems described in Sect. 3.1. cannot occur.

5. ACKNOWLEDGEMENTS

The authors wish to acknowledge the input and contributions from the project team working on the EU funded RTD project MERLIN[1]

6 REFERENCES

Amann M et al.: Cost-effective Control of Acidification and Ground-Level Ozone. First Interim Report to the European Commission, DG-XI, Laxenburg, 1996

Amann M et al.: Final report on work done under EU acidification and ozone strategy. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, 1999

ApSimon H M, Warren R F, Wilson J J N: The abatement strategies assessment model (ASAM): applications to reductions of sulphur dioxide emissions. *Atmospheric environment* 28 (4), 649–664, 1994

ApSimon H M and Warren R F: Transboundary Air Pollution in Europe. *Energy Policy 24 (7)*, 631–640, 1996

Bailey P D: Emissions and Abatement Costs of Nitrogen Oxides for Use in Integrated Assessment Models. Report to the Working Group on Strategies and the Task Force on integrated Assessment Modelling, LRTAP Convention, UNECE, Stockholm Environment Institute at York, York, 1996

Faucheux S, G Froger: Decision-making under environmental uncertainty. *Ecological Economics 15 (1)*, 29–42, 1995

Friedrich R, Reis S (eds): Tropospheric Ozone Abatement – Developing Efficient Strategies for the Reduction of Ozone Precursor Emissions in Europe. Springer Publishers, 2000

Goldberg D. E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA, 1989

Gough C A, Chadwick M J, Biewald B, Kylenstierna J, Bailey P D, Cinderby S: Developing optimal abatement strategies for the effects of sulfur and nitrogen deposition on European scale. *Water, Air and Soil Pollution 85*, 2601–2606, 1995

Gough C, Castells N, Funtowicz S: Integrated Assessment: an emerging methodology for complex issues. *Environmental Modelling and Assessment 3, 19–29*, 1998

Loughlin D H, Ranjithan S R, Baugh J W, Downey-Brill E: Application of Genetic Algorithms for the Design of Ozone Control Strategies. *J. Air & Waste Management Assoc. 50*, 1050–1063, 2000

Simpson D, Olendrzynski K, Semb A, Støren E and Unger S: Photochemical oxidant modelling in Europe; multi-annual modelling and source-receptor relation-ships. EMEP/MSC-W Report 3/97, July 1997

van Velduizen, D.A.: Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations, 1999

Warren R F, ApSimon H M: Uncertainties in integrated assessment modelling of abatement strategies: illustrations with the ASAM model. *Environmental Science and Policy 2*, 439–456, 1999

[1] http://www.merlin-project.info

Using Interactive Archives in Evolutionary Multiobjective Optimization: Case Studies for Long-Term Groundwater Monitoring Design

P.M.Reed and V.K. Devireddy

Department of Civil and Environmental Engineering, The Pennsylvania State University,

212 Sackett Building, University Park, PA 16802-1408,

Abstract: Monitoring design is a problem of paramount importance to the environmental engineering field because environmental observation data provide the sole means of assessing if engineered systems are successfully protecting human and ecologic health. The monitoring design problem is extremely challenging because it requires environmental engineers to capture an impacted system's governing processes, elucidate human and ecologic risks, limit monitoring costs, and satisfy the interests of multiple stakeholders (e.g., site owners, regulators, and public advocates). Evolutionary multiobjective optimization (EMO) has tremendous potential to help resolve these issues by providing environmental stakeholders with a direct understanding of their monitoring tradeoffs. This paper demonstrates how **ε**-dominance archiving and automatic parameterization techniques for the NSGA-II can be used to significantly improve the algorithm's ease-of-use and efficiency for computational intensive applications. Results are presented for a 4-objective groundwater monitoring problem in which the archiving and parameterization techniques for the NSGA-II combined to reduce computational demands by more than 90-percent relative to prior published results. The methods of this paper can be easily generalized to other multiobjective applications to minimize computational times as well as trail-and-error parameter analysis.

Keywords: Groundwater; Monitoring; Design; Water resources management; Multiobjective optimization; Genetic algorithms

1. INTRODUCTION

This paper demonstrates the use of evolutionary multiobjective optimization (EMO) to design groundwater monitoring networks for conflicting objectives. Long-term groundwater monitoring (LTM) can be defined as the sampling of groundwater quality over long time-scales to provide "sufficient and appropriate information" to assess if current mitigation or contaminant control measures are performing adequately to be protective of human and ecological health Task Committee on Long-Term Groundwater Monitoring Design [2003]. The LTM problem is ideal for demonstrating how EMO can aid environmental engineers because of the tremendous expense and complexity of characterizing groundwater contamination sites over long time periods.

The 4-objective monitoring design problem presented in this paper is solved using a modified version of the Nondominated Sorted Genetic Algorithm-II (NSGA-II) Deb et al. [2002], which will be termed the **ε**-dominance NSGA-II in this paper using the abbreviated notation, **ε**-NSGA-II. The **ε**-NSGA-II demonstrates how **ε**-dominance archiving (Laumanns et al. [2002], Deb et al. [2003]) can be combined with a parameterization strategy for the NSGA-II Reed et al. [2003] to accomplish the following goals: (1) ensure the algorithm will maintain diverse solutions, (2) eliminate the need for trial-and-error analysis for parameter settings (*i.e.*, population size, crossover and mutation probabilities), and (3) allow users to *sufficiently* capture tradeoffs using a minimum number of design evaluations. A sufficiently quantified trade-off can be defined as a subset of Pareto optimal solutions that provide an adequate representation of the Pareto frontier that can be used to inform decision making.

In this paper, section 2 overviews prior **ε**-dominance archiving and parameterization studies used in the development of the **ε**-NSGA-II. Section 3 discusses the 4-objective groundwater monitoring test case used to demonstrate the **ε**-NSGA-II. Sections 4 and 5 provide a more detailed description of the **ε**-NSGA-II and its performance for the groundwater monitoring test case, respectively.

2. PRIOR WORK

The **ε**-NSGA-II combines the external archiving techniques recommended by Laumanns et al. [2002] and Deb et al. [2003] with automatic parameterization techniques (Reed et al. [2003]) developed to eliminate trial-and-error parameter analysis for the NSGA-II. A primary drawback of using EMO methods for environmental applications lies in the large costs associated with assessing performance (*i.e.*, algorithmic reliability and solution quality). The common practice of assessing performance for a distribution of random seeds employed in the EMO literature is often prohibitively expensive in terms of computational costs and in terms of the time that must be invested by users. The goal of the automated parameterization approaches developed by Reed et al. [2003] is to eliminate the need to assess algorithmic performance for a distribution of initial random number seeds and instead focus on the NSGA-II's reliability and efficiency for a single random seed. Reliability is addressed in the approach by adaptively increasing the size of the population. The method uses multiple runs in which the nondominated solutions are accumulated from the results of the successively doubled population sizes. The runs (and successive doubling of population sizes) continue until either the user-defined maximum run-time is reached or sufficient solution accuracy has been attained.

3. MONITORING TEST CASE PROBLEM

3.1 Test Case Data

The test case developed for this study uses data drawn from a 50 million-node flow-and-transport simulation performed by Maxwell et al. [2000]. The simulation provided realistic historical data for the migration of a hypothetical perchloroethylene (PCE) plume in a highly heterogeneous alluvial aquifer. The hydrogeology of the test case is based on an actual site located at the Lawrence Livermore National Laboratory (LLNL) in Livermore, California. Data were provided for a total of 58 hypothetical sampling locations within a 29-well multi-level monitoring network. If the i^{th} monitoring well was selected for sampling then PCE is sampled at all the possible sampling locations along its vertical axis.

3.2 Problem Formulation

The 4-objective monitoring test case used in this paper combines both the spatial redundancy and geostatistical approaches to monitoring design. The **ε**-NSGA-II and quantile kriging are combined to quantify the tradeoffs among the following four performance criteria: (1) cost, (2) squared relative estimation error (SREE), (3) the relative global mass error (MAE), and (4) local uncertainty as measured by kriging estimation variances. Cost is a linear function of the number of PCE samples that are used in a given monitoring design. SREE measures how the interpolated picture of the plume using data only from wells included in the κ^{th} sampling plan compares to the result attained using data from all available sampling locations. Likewise, the global mass objective error in the total mass of PCE in the subsurface. Lastly, local uncertainty is estimated using the sum of the estimation standard deviations (i.e., the square root of estimation variances) from kriging (for more details see Reed and Minsker [2004]).

3.3 Plume Interpolation using Quantile Kriging

Quantile kriging was selected for plume interpolation in this study based on the findings of Reed et al. [2004], who present a comprehensive performance analysis of 6 interpolation methods for scatter-point concentration data, ranging in complexity from intrinsic kriging based on intrinsic random function theory to a traditional implementation of inverse-distance weighting. Quantile kriging was shown to be the most robust and least biased of the interpolation methods they studied. Additionally, the method's non-parametric uncertainty estimates successfully predicted zones of high estimation error for each test case.

4. OVERVIEW OF THE ε-NSGA-II APPROACH

The proposed algorithm consists of three steps. The first step utilizes the NSGA-II with a starting population of 5 individuals to initiate EMO search. The initial population size is set arbitrarily small (i.e., 5 in this paper) to ensure the algorithm's initial search is done using a minimum number of function evaluations. Subsequent increases in the population size adjust the population size commensurate with problem difficulty. In the second step, the **ε**-NSGA-II uses a fixed sized archive to store the nondominated solutions generated in every generation of the NSGA-II runs. The archive is updated using the concept of **ε**-dominance, which has the benefit of ensuring that the archive maintains a diverse set of solutions. **ε**-dominance requires the user to define the precision with which they want to evaluate each objective by specifying an appropriate ε value for each objective.

The third step checks if the user-specified performance and termination criteria are satisfied and the Pareto optimal set has been sufficiently quantified. If the criteria are not satisfied, the population size is doubled and the search is continued. When increasing the population, the initial population of the new run has solutions injected from the archive at the end of the previous run. The algorithm terminates if either a maximum user time is reached or if doubling the population size fails to significantly increase the number of nondominated solutions found across two runs. The following sections discuss the **ε**-NSGA-II in greater detail.

4.1 Searching with the NSGA-II

Development of the **ε**-NSGA-II was motivated by the authors' goal of minimizing the total number of function evaluations required to solve computationally intensive environmental applications, eliminate trial-and-error analysis for setting the NSGA-II's parameters, and avoid the need for random seed analysis. The dynamic population sizing and injection approach applied in the **ε**-NSGA-II exploits computationally inexpensive small populations to expedite search while increasing population size commensurate with problem difficulty to ensure the Pareto optimal set can be reliably approximated.

The initial population size, N_0 is set to some arbitrary small value (*e.g,* 5), as it is expected that the adaptive population sizing scheme will adjust for an undersized population. A randomly selected subset of the solutions obtained using the small population sizes are injected into subsequent larger populations, aiding faster convergence to the Pareto front. This can be viewed as using series of "connected" NSGA-II runs that share results so that the Pareto optimal set can be reliably approximated. Computational savings should be viewed in two contexts: (1) the use of minimal population sizes and (2) elimination of random seed analysis. Note that the number of times the population size needs to be doubled varies with different random seeds, though exploiting search with small populations will on average dramatically reduce computational times. Moreover, our approach eliminates the need to repeatedly solve an application for a distribution of random seeds.

The NSGA-II's remaining parameters are set automatically based on whether an application is being solved using a real or binary coding. The 4-objective problem solved in this paper is solved using binary coded variables, uniform crossover with a probability of 0.5, and a probability of mutation equal to 1/population size [for more details see Reed et al. [2003]].

4.2 Archive Update

The **ε**-dominance archiving approach is particularly attractive for environmental applications because it allows the user to define the precision with which they want to quantify their tradeoffs while bounding the size of the archive and maintaining a diverse set of solutions. The concept of **ε**-dominance requires the user to define the precision they want to use to evaluate each objective. The user specified precision or tolerance vector **ε** defines a grid for a problem's objective space, which biases search towards the portions of a problem's decision space that have the highest precision requirements. The **ε**-dominance archive improves the NSGA-II's ability to maintain a diverse set of nondominated solutions by only allowing 1 archive member per grid cell. In the case when multiple nondominated points reside in a single grid cell, only the point closest to the lower left corner of the cell (assuming minimization) will be added to the on-line archive thereby ensuring convergence to the true Pareto optimal set Laumanns et al. [2002], Deb et al. [2002].

4.3 Injection and Termination

The **ε**-NSGA-II also seeks to speed convergence by pre-conditioning search with larger population runs with the prior search results attained using small populations. In prior efforts, any attempts to inject solutions found using small population into subsequent runs made the NSGA-II prematurely converge to poor representations of the Pareto optimal set, especially for problems with greater than 2 objectives. The **ε**-domination archive's ability to preserve diversity plays a crucial role in overcoming this limitation. As described previously in Section 3.1, the **ε**-NSGA-II begins search with an initial population of 5 individuals from which the **ε**-nondominated solutions identified in this initial run are stored in the archive.

The archive at the end of each run contains **ε**-nondominated solutions that can be used to guide search in future runs and speed up convergence to the Pareto front. This is achieved by injecting **ε**-nondominated solutions from the archive at the end of the run with population size *N* into the initial population of the next run which has a population size *2N*. Two scenarios arise when the **ε**-NSGA-II injects solutions from the archive generated with a population size *N* into the initial generation of a run with a population size *2N.*

In scenario 1, the archive size *A* is smaller than the subsequent population size *2N.* In this case, 100-percent of the **ε**-nondominated archive solutions are injected into the first generation of the subsequent run with *2N* individuals. We have found that the number of injected solutions should be maximized to aid rapid convergence. The **ε**-dominance archive in combination with successive doubling of population size guarantees the **ε**-NSGA-II will maintain sufficient solution diversity. In scenario 2, the archive size *A* is greater than the next population size *2N.* In this case, *2N* **ε**-nondominated archive solutions are selected randomly and injected into the first generation of the next run, again maximizing the impact of injected solutions.

The termination of search across all runs (i.e., across all populations used) compares the rate increase in the archive size at the end of two successive runs of the **ε**-NSGA-II in which the first run uses a population of N sampling designs to evolve a **ε**-nondominated set composed of A individuals, while the second run uses a population of 2N designs to evolve a **ε**-nondominated set of K individuals. The results of these runs are used in equation (1), to define which of the two following courses of action will be taken: (1) population size is again doubled, resulting in 4N individuals to be used in an additional run of the **ε**-NSGA-II or (2) the algorithm stops to allow the user to assess if the **ε**-nondominated set has been quantified to sufficient accuracy.

$$\begin{aligned} &if \quad \Delta < \left(\frac{|K - A|}{A}\right)100 \quad then \;, \; double \quad N \\ &and \quad continue \quad search \quad else \quad stop \quad search \end{aligned} \tag{1}$$

The solutions obtained in the archive at the end of the final run represent the Pareto front with the user defined accuracy. In this study, Δ was set equal to 10-percent as recommended by Reed et al. [2003]. Section 4 demonstrates the efficiency of the ε-NSGA-II in solving high order environmental problems.

5. RESULTS AND DISCUSSIONS

The efficiency of the ε-NSGA-II in solving high order multiobjective optimisation problems is demonstrated in this section by solving the 4-objective ground water monitoring problem described in Section 3. The experiments designed in this section are aimed at highlighting the algorithms efficiency in capturing an approximate trade off using a minimum number of functional evaluations.

The problem was solved using various precision settings as described in Table 1. The resolution of the solutions obtained decreases from setting 1 to 5. Setting 1 is of the highest resolution and hence each of the ε's are set to arbitrarily small value values. This would be an ideal setting for a user who would like the entire trade off and does not mind compromising on the computational complexity. To highlight the efficiency of the algorithm in reducing the computational time with a decrease in resolution, four other settings are chosen. In setting 2, the values of ε_{SREE} and ε_{MAE} are set to 0.01 and 10^{-4} respectively based on the range of objective values obtained from setting 1. Setting 3 is obtained by increasing the value of $\varepsilon_{Uncertainty}$ by a factor of 100 while at the same time keeping the other resolutions' constant. Similarly settings 4 and 5 are obtained by increasing ε_{SREE} and ε_{MAE}. ε_{Cost} is not varied in the experiments because it is a discrete integer function of the number of sampling points used.

Table 1. The resolution settings used in solving the ground water problem.

Setting No.	ε_{Cost}	ε_{SREE}	$\varepsilon_{Uncertainty}$	ε_{MAE}
1	1	10^{-5}	0.01	10^{-6}
2	1	0.01	0.01	10^{-4}
3	1	0.01	1	10^{-4}
4	1	0.1	1	10^{-4}
5	1	0.1	1	0.01

The problem was solved using binary coded variables and the parameter settings described in section 4. Table 2 summarizes the number of solutions obtained and the number of function evaluations required for each of the parameter settings. The effects of varying the resolution on the number of solutions found by the ε-NSGA-II as well as the number of function evaluations required are demonstrated in Figure (1). The number of nondominated solutions found by the ε-NSGA-II does not significantly decrease when the resolution is reduced from the values used in setting 1 to those in setting 2. Setting 1 required the highest resolution, which as expected had the highest number of ε-nondominated solutions and reduced the number of function evaluations required to solve the problem by more than 80% over the prior published results Reed and Minsker [2004]. Figure (1) demonstrates that as user-specified resolution requirements decrease, the number of function evaluations reduces by an order of magnitude relative to the 450,000 evaluations utilized by Reed and Minsker [2004].

Table 2. The number of function evaluations and the number of solutions obtained for each of the parameter settings.

Setting No.	No. of Function Evaluations	No. of Solutions
1	82740	880
2	76500	862
3	69900	631
4	51985	515
5	37260	463

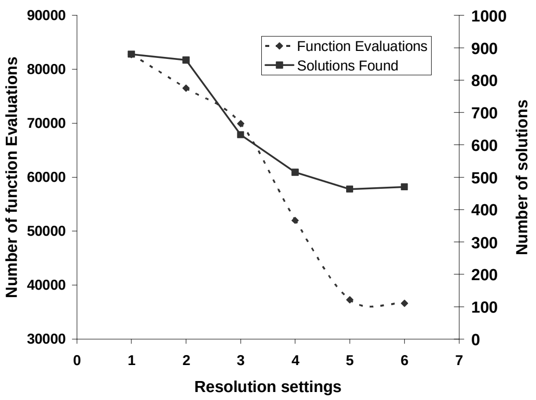


Figure 1. Variation of the number of solutions found and the number of function evaluations required with different resolution settings.

In the LTM application a single constituent is being monitored at 29 monitoring wells, which results in a decision space of more than 500 million possible sampling designs (i.e., 2^{29} sampling designs). Using the ε-NSGA-II to identify the subset of sampling designs that are optimal with respect to the 4 objectives used in this application, reduces the set of designs that must be considered from 500 million to less than a 1000 designs identified on the Pareto surface. Although the 4-dimensional Pareto surface cannot be visualized, the set of solutions designs can inform decision making as follows.

This process begins by analyzing pairs of the objectives that are known to conflict. These 2-dimensional tradeoffs are subsets of the overall nondominated set. These tradeoffs are found by identifying only those solutions that are nondominated in terms of cost and one other objective, independent of the remaining objectives' values [e.g., the Cost—Uncertainty tradeoff in Figure (2)].

Figures (2) shows the Cost—Uncertainty tradeoff generated by Reed and Minsker [2004] designated as the "prior method" as well as the results attained by the ε-NSGA-II using ε settings 1 and 5. The ε-NSGA-II closely approximates the results of the prior method using up to 90-percent fewer function evaluations. These results highlight that a tremendous amount of the computation time originally used to solve this application was spent seeking unnecessarily high-precision results. Non-domination sorting at 6-digits of precision and beyond is computationally expensive and does not significantly improve the representation of the tradeoffs used for decision making.

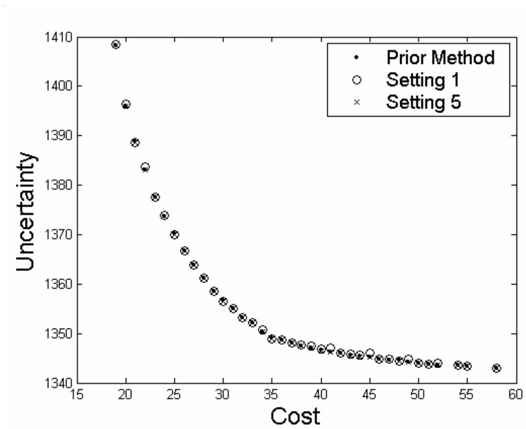


Figure 2. Cost—Uncertainty tradeoff.

6. CONCLUSIONS AND FUTURE WORK

The **ε**-NSGA2 demonstrates how **ε**-dominance archiving can be combined with a parameterization strategy for the NSGA-II to accomplish the following goals: (1) ensure the algorithm will maintain diverse solutions, (2) eliminate the need for trial-and-error analysis for parameter settings (*i.e.*, population size, crossover and mutation probabilities), and (3) allow users to *sufficiently* capture tradeoffs using a minimum number of design evaluations. A sufficiently quantified tradeoff can be defined as a subset of nondominated solutions that provide an adequate representation of the Pareto frontier that can be used to inform decision making. Results are presented for a 4-objective groundwater monitoring case study in which the archiving and parameterization techniques for the NSGA-II combined to reduce computational demands by greater than 90-percent relative to prior published results. The methods of this paper can be easily generalized to other multiobjective applications to minimize computational times as well as trail-and-error parameter analysis.

7. REFERENCES

Deb, K., M. Mohan and S. Mishra, A Fast Multi-objective Evolutionary Algorithm for Finding Well-Spread Pareto-Optimal Solutions. Kanpur, India, Indian Institute of Technology, 2003.

Deb, K., A. Pratap, S. Agarwal and T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II." IEEE Trans. Evol. Computation **6**(2): 182-197, 2002.

Laumanns, M., L. Thiele, K. Deb and E. Zitzler, "Combining Convergence and Diversity in Evolutionary Multiobjective Optimization." Evolutionary Computation **10**(3): 263-282, 2002.

Maxwell, R., F. S. Carle and F. B. Tompson, Contamination, Risk, and Heterogeneity: On the Effectiveness of Aquifer Remediation. UCRL-JC-139664. Livermore, CA, 2000.

Reed, P., T. Ellsworth and B. S. Minsker, "Spatial Interpolation Methods for Nonstationary Plume Data." Ground Water **42**(2): 190-202, 2004.

Reed, P. and B. S. Minsker, "Striking the Balance: Long-Term Groundwater Monitoring Design for Conflicting Objectives." Journal of Water Resources Planning and Management **130**(2): 140-149, 2004.

Reed, P., B. S. Minsker and D. E. Goldberg, "Simplifying Multiobjective Optimization: An Automated Design Methodology for the Nondominated Sorted Genetic Algorithm-II." Water Resources Research **39**(7): 1196, doi:1110.1029/2002WR001483, 2003.

Task Committee on Long-Term Groundwater Monitoring Design, Long-Term Groundwater Monitoring: The State of the Art. Reston, VA, American Society of Civil Engineers, 2003.

A Multi-Model Approach to Analysis of Environmental Phenomena

O. Giustolisi[a], A. Doglioni[a], D.A. Savic[b] and B.W. Webb[c]

[a] *Civil and Environmental Engineering Department, Engineering Faculty of Taranto, Technical University of Bari, Italy - o.giustolisi@poliba.it*

[b] *Centre for Water Systems, Department of Engineering, School of Engineering, Computer Science and Mathematics, University of Exeter, United Kingdom*

[c] *Department of Geography, School of Geography, Archaeology & Earth Resources, University of Exeter, United Kingdom*

Abstract: This paper introduces a novel data-driven methodology named Evolutionary Polynomial Regression (EPR), which permits the multi-purpose modelling of physical phenomena, through the simultaneous solution of a number of models. Multipurpose modelling or "multi-modelling", enables the user to make a more robust choice of those models aimed at (a) the knowledge based on data modelling, (b) on-line and off-line forecasting, and (c) data augmentation (i.e. infilling of missing data in time series). This methodology is particularly useful in modelling environmental phenomena, for which it is usually impossible to obtain physical data at a laboratory scale. In particular, the non-linearity of phenomena and non Gaussian nature of background noise make on-line forecasting complex, and where data are available, they often contain discontinuities (i.e. missing data). The use of EPR in modelling and analysis is illustrated by application to a case study containing all these limitations. The application of EPR to thermal behaviour of a stream gives not only a good physical insight of the phenomenon, but also allows infilling of missing data, resulting in good models that forecast the water temperature.

Keywords: Data reconstruction; Knowledge discovery in data; Environmental modelling; Evolutionary computing; Evolutionary Polynomial Regression.

1. INTRODUCTION

Modelling of environmental phenomena usually relies on sampled data, which are often incomplete. Ideally, analysis should provide new insights into the phenomena, give accurate forecasting of the output for a range of inputs and outputs and fill in gaps in data records. This can be achieved by creating a range of specific models, i.e. models chosen for well-defined purposes, although the construction and choice of the models is often challenging. Environmental phenomena are typically non-linear in their dynamics and affected by non Gaussian background noise. In the models, these effects must be reproduced as accurately as possible. The temptation is to use complex non-linear modelling strategies, to better describe the phenomena. Unfortunately, the answers from these are very difficult to interpret from a physical aspect.

An additional problem relates to discontinuities, i.e. gaps, often present in data records. On the one hand, we are interested in "reconstructing" that information contained in missing data, without losing the physics of the phenomenon. On the other hand, we do not know how to choose the best model for this purpose, because we have no data to define a traditional performance indicator.

This paper presents the Evolutionary Polynomial Regression (EPR) technique a novel, model-based reconstruction technique capable of reconstructing data series containing information about the physical phenomena [Giustolisi et al., 2004a]. It also provides simple well defined effective models useful both for on-line forecasting and simulation.

Such models usually are simple polynomial structures where each monomial can contain user-defined functions. These structures can improve physical interpretation of the studied phenomenon too [Giustolisi et al., 2004b]. EPR has the advantage of combining evolutionary algorithms with traditional numerical regression [Giustolisi and Savic, 2004a]. EPR is an incremental development of a hybrid methodology [Davidson et al. 1999; 2003] which incorporated least squares optimization within symbolic regression.

Thus, EPR is a hybrid system capable of producing a series of polynomial models, from which one can choose those considered best for a particular purpose. It is unlikely that the same model would be selected for short gap reconstruction, for forecasting the phenomenon (with a particular time horizon), or for gaining physical insight. This approach is possible with EPR because it does not have a rigid structure, but allows a multi-structure strategy with multiple performances where each different structure has its own advantages for a specific modelling goal.

EPR is tested and demonstrated on a UK environmental case study analysing thermal behaviour of a river. Air temperature (input) and water temperature (output) data were available, but the data series had several gaps of different duration. Therefore, several models were constructed to reconstruct (infill) data [Bennis et al., 1997], obtain a model for on-line forecasting; and discover some new knowledge about the dynamics of the heat transfer process over a short time scale. In summary, the case study contains all the features that typify the analysis of an environmental phenomenon.

2. THE EVOLUTIONARY POLYNOMIAL REGRESSION

2.1 A general portrait

EPR is a data-driven hybrid method for a multi-model approach based on evolutionary computing. A general EPR expression may be given as

$$\hat{y} = \sum_{j=1}^{m} f\left(\mathbf{X}, a_j\right) + a_o \qquad (1)$$

where $\hat{y}$ is the estimated output of the system/process; a_j is a constant value; f is a function constructed by the process; $\mathbf{X}$ is the matrix of input variables; and m is the length (number of terms) of the polynomial expression (bias excluded) [Giustolisi and Savic, 2004a].

The general functional structure represented by $f(\mathbf{X},a_j)$ is constructed from elementary functions by EPR which uses a Genetic Algorithm (GA) strategy [Goldberg, 1989]. The GA is employed to select the useful input vectors from $\mathbf{X}$ to be combined. The building blocks (elements) of the $f(\mathbf{X},a_j)$ structure are defined by the user based on physical process understanding. While the selection of feasible structures to be combined is done through an evolutionary process, the parameters a_j are estimated by the Least Square (LS) method.

The process starts with a GA searching through the space of user defined exponents, which must include the value of zero, thus allowing a simple expression to be generated by discarding unnecessary components of $\mathbf{X}$. The next step consists of determining the a_j values by simple LS.

The LS is performed in an original way, by searching for only positive values. This is because negative terms usually have a poor physical meaning, as they simply balance positive terms returning a better description of noise. Neglecting negative terms constrains the search space thus gaining computational efficiency without losing accuracy. Moreover, we can hypothesize that sometimes the pressure to find parsimonious expressions could improve the search of physically based equations. In this way, EPR returns models that are probably less appropriate to on-line forecasting, but have the advantage of giving physical insight, consistent with the multi-model concept.

2.2 Some properties of EPR

EPR is a technique for data-driven modelling, successfully tested on environmental problems [Giustolisi et al., 2004a; Giustolisi and Savic, 2004b]. The combination of the GA for finding feasible structures and the LS for training a few positive constants of those structures implies some advantages. In particular, the GA allows a global exploration of the error surface thanks to specifically defined objective (cost) functions. Through such functions we can set criteria for the search: (a) avoiding the overfitting of models to training data; (b) pushing the methods towards simple structures; and (c) avoiding superfluous terms representative of the noise in data. EPR shows robustness and in every situation can get a model truly representative of data.

The use of LS for evaluation of positive constant values a_j is not compromised in working with series containing insufficient data. Indeed, LS performed on short-length expressions shows that long time series are not necessary for proper evaluation of those constants.

In this scenario, the interesting feature of EPR is in the possibility of getting more than one model. Each of these models can be used for a specific purpose. For instance, we can get a model for a short time forecasting, another one for long time forecasting, another one for simulation, etc. Each different model can be trained according to specific cost functions.

A further feature of EPR is the high level of interactivity between the user and the methodology. The former can use physical insight to make hypotheses on the elements in the function $f(\mathbf{X},a_j)$ and on its structure, see Eq. (1). Choosing the proper objective (cost) function and assuming prechosen elements in Eq. (1) (external information), and working with dimensional information enables refinement of final models [Giustolisi et al., 2004].

Finally, the best models are chosen on the basis of their performances on a test set of unseen data. For this purpose, the data set is split in two subsets: (1) the subset used for building models, named training set, and (2) the subset used for testing the model, named test set. It is important to emphasise that the test set is never used in the phase of model construction, thereby allowing us to evaluate the generalisation capacities of each model. Thus, an unbiased performance indicator is obtained on real capability of the models. Nevertheless, a bootstrap technique can also be applied to increase the robustness of model evaluations.

3. THE CASE STUDY

3.1 The River Barle

The River Barle is the main tributary of the upper River Exe. It is located in a rural zone of Southwest England [Webb et al., 2003]. Our data collection consists of two years of hourly air (input) and water temperature samples (output). Each sample is referred to a window of 6 hours of a solar day covering the periods: 1-6; 7-12; 13-18; 19-24. We reasonably assume that the chosen windows are representative of the thermal dynamics at a day scale. Both air and water temperatures show two main periodic components: the annual and the daily cycles [Webb et al. 2003].

Further details about data and sampling location can be found on Webb at al. [2003].

3.2 Background to data

Before starting the modelling phase, we divided data into two subsets (training and test) each made up of 1460 samples, covering a solar year and affected by missing samples [Table 1].

Gaps in data are randomly distributed. While the longest gap is located in the test set, the 124-sample gap corresponds to 31 missing days. It should also be noted that no pre-processing was executed on gaps prior to passing the data to EPR.

A comprehensive examination of the data confirmed that the quality of samples is sufficiently good. There are neither occasional nor systematic errors which could affect modelling.

Table 1. Features of gaps contained in data.

Length of gaps in *6-hour samples*	Length of gaps in *hourly samples*	Number of gaps
1	6	14
2	12	2
3	18	1
28	168	2
29	174	1
63	378	1
65	390	1
124	744	1

4. THE MODELLING PHASE

4.1 The strategy

The modelling phase was done as follows:

- The structure of Eq. (1) is assumed polynomial.
- Each monomial term of Eq. (1) consists of elements from **X** raised to pre-specified power values.
- No hypotheses are made about a_0, besides its positive sign.
- The assumed range of possible exponents of terms from **X** is (0; 0.5; 1; 2).
- The maximum length of polynomial structures was assumed to be 5 terms plus bias.
- 7 objective (cost) functions have been used.
- The LS search is performed for positive coefficients only (negative ones are *a-priori* discarded).

- Data were scaled between 0 and 1; the outputs were rescaled before being listed.

Each objective (cost) function is based on the use of the Sum of Squared Error (SSE); the differences among them relate to the way the SSE is computed [Giustolisi and Savic, 2004a]. In summary the following cost functions were used:

- Soft Cross Validation, SCV from SSE evaluated on the whole training set.
- Rigid Cross Validation, RCV from SSE evaluated on 50% of samples of the training set.
- Control of Variance, CVP of each Parameter a_j.
- Penalization of Complex Structure, PCS.
- Penalization of Variance, PV.
- Traditional SSE evaluation, SSE.
- Control on Variance CVT of each monomial term, contained in the polynomial expression.

Details about cost functions can be found in Giustolisi and Savic [2004a]. The method of modelling ensured that the complex models with large monomial terms focussed on describing the physical process, rather than the background noise.

The presence of seven objective (cost) functions enabled a more robust multi-modelling approach, in which models can be selected according to different, appropriate objective (cost) functions to represent the most robust choices. Thus, each model has its specific utility according to the purposes previously described in the multi-model scenario. In our case study, when the search was constrained to polynomial expressions only made by 1 term plus bias, the same equation was always obtained for every cost function. By constraining the search to an expression of 2 terms plus bias, similar models were found and in some cases (e.g. PCS, PV, CVT) they were the same. Furthermore, for the same cost function, we can observe that EPR does not select more terms than it actually needs. For example, if the maximum polynomial length was set to 5, EPR could return an expression of 2 terms, because it does not consider longer expressions better than 2. Therefore, assuming a cost function, it is not unusual that after a well defined polynomial length, EPR goes on selecting the same model as optimum. On these bases, we can make a robust choice of the model. A model selected by different cost functions, or preferred to a longer expression by the same cost function, is likely to be a good model. Among those models assumed as robust choices, the best is selected to suit the purpose. To infill gaps in data, models would be selected according to the gap length. If physical insight was required, selection would be based on those models that were easily interpretable, i.e. with a clear physical meaning.

The choice of the best models for our purpose is made on the basis of their performances on the test set. We use the Coefficient of Determination, CoD, as the main performance indicator,

$$\mathrm{CoD} = 1 - \frac{N-1}{N} \frac{\sum_N \left(\hat{W} - W_{exp}\right)^2}{\sum_N \left(W_{exp} - \mathrm{mean}\left(W_{exp}\right)\right)^2} \qquad (2)$$

where N represents the number of samples, W_{exp} represents the measured water temperature, $\hat{W}$ represents the value of water temperature returned by the model.

Furthermore, a bootstrap procedure [Efron, 1979] was used on the test set data, rather than taking a simple value of CoD. Thus, the bootstrap CoD is an average value of the CoDs evaluated by re-sampling data 1000 times from the test set. To ensure improved evaluation of the models, the standard deviation of the CoD values, reported as percentage of the average value are used. The bootstrap is particularly helpful for infilling missing data, since there are no data for comparison with model results.

4.2 EPR results

EPR returned 13 different models: we selected 3 models among them as optimal. The three selected models are,

$$\begin{aligned} W_t = {} & 0.30574 \cdot A_t^{0.5} \cdot A_{t\text{-}1} + 0.50436 \cdot W_{t\text{-}1} + \\ & + 0.31731 \cdot W_{t\text{-}4} \cdot A_t^{0.5} + 0.013418 \end{aligned} \qquad (3)$$

$$\begin{aligned} W_t = {} & 0.078319 \cdot W_{t\text{-}4} + 0.23946 \cdot W_{t\text{-}3}^{0.5} \cdot W_{t\text{-}4}^{0.5} \cdot A_{t\text{-}1}^{0.5} + \\ & + 0.49433 \cdot W_{t\text{-}1} + 0.31486 \cdot A_t \cdot A_{t\text{-}1}^{0.5} + 0.0047945 \end{aligned} \qquad (4)$$

$$W_t = 1.0073 \cdot W_{t\text{-}1}^{0.5} \cdot A_t^{0.5} \qquad (5)$$

where the subscript t stands for time, in terms of 6-hour sampling rate and A refers to air temperature.

Eq. (3) is the best performing for 1-step-ahead prediction. Eq. (4) is the best working in 2-step-ahead, 4-step-ahead and 6-step-ahead prediction (one step corresponds to 6 hours). Eq. (5) was chosen as the best working in simulation because of its more likely physical behaviour, than the best CoD-best-working model for simulation, which generates unlikely overestimated values for peak zones, because it does not contain W terms. Indeed, it does not take into consideration the effects related to the thermal inertia of the stream, through the thermal capacity of water.

Table 2 shows the performances of the resulting models, in term of average CoD and percentage standard deviation.

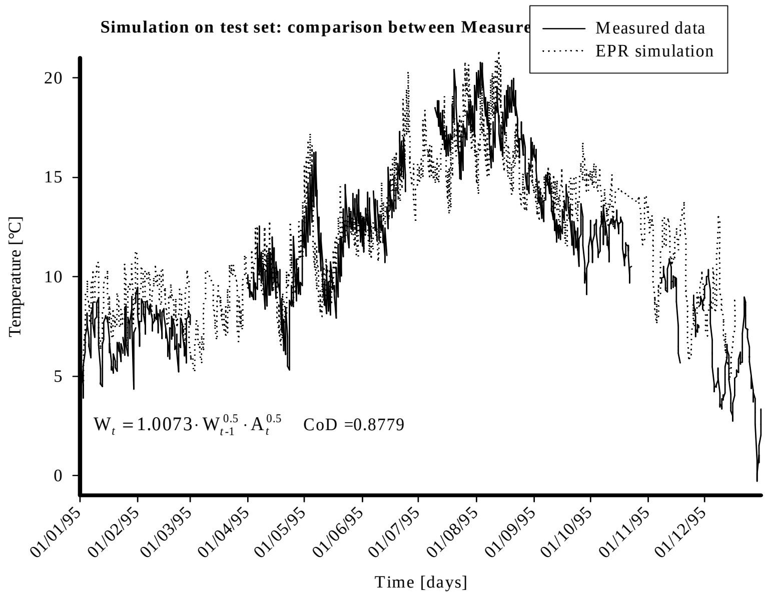


Figure 1. Comparison between measured data and EPR simulated data.

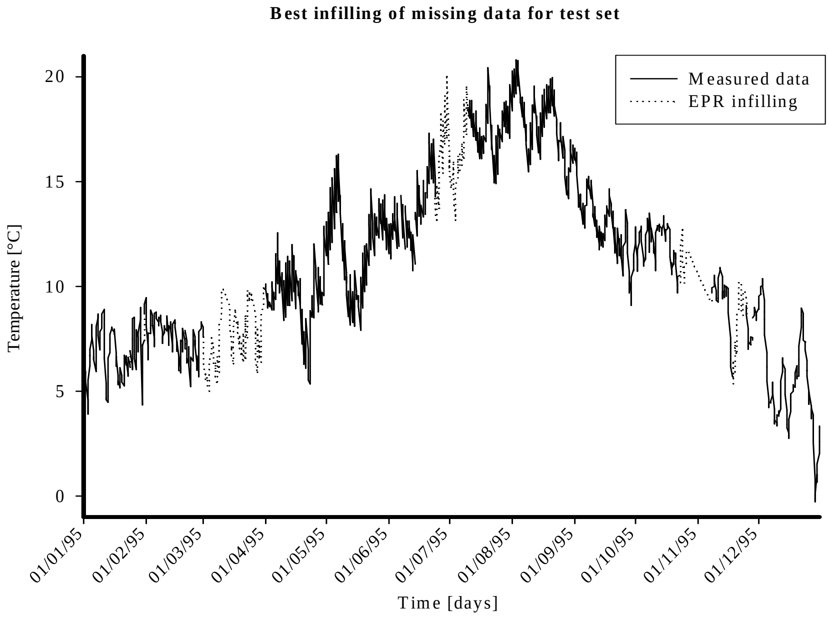


Figure 2. Best infilling of missing data.

Table 2. Statistics on model performances.

EPR model	Evaluation on	Average CoD	Standard deviation %
(3)	1-step-ahead	0.984	0.101
(4)	2-step-ahead	0.971	0.173
	4-step-ahead	0.958	0.256
	6-step-ahead	0.934	0.400
(5)	Simulation	0.878	0.412

The three models were obtained by using the following cost functions:

- CVT for Eq. (3).
- SCV for Eq. (4).
- RCV for Eq. (5).

Note that all the cost functions produced models with similar performances. We consider similar performances as strong indicators of the robustness of EPR methodology.

4.3 Comments on EPR results

All models found have a simple structure, enable the gaps in data to be reconstructed and are good at on-line forecasting and simulation. Such structures can allow a physical interpretation. In particular, Eq. (3) and Eq. (4) suggest a strong link between the output water temperature at time t and the water temperature at the time t-1. This interpretation is confirmed by the presence in both equations of the term W_{t-1}, multiplied by the higher coefficient in the expression. This occurred very frequently in all models. Another frequent term is the product between A_t and A_{t-1}, indicating a likely effect of the air temperature at time t and t-1 on water temperature. Webb confirmed this by physical observations and with a different approach to data analysis [Webb et al., 2003]. Further terms contained by models are considered of uncertain origin, and more likely associated with the background noise in data. The simulation model, whose time plot is represented in Figure 1, has a very compact unbiased expression. This is due to the similar shape, on average, of the curves representing the time plot of air temperature and water temperature. Previous studies [Webb et al., 2003; Mohseni and Stefan, 1999] underline the quasi-linear relationship between water and air temperatures, which is confirmed by our simulation model. Neglecting the stochastic information from measured data, the simulation emphasizes the quasi-linear relationship, and the W_{t-1} component seems to explain the non-linear behaviour that occurs in particular ranges of temperature [Webb et al., 2003]. Finally, in Figure 2, we can see the best infilling of data in the test set. We infilled using Eq. (4) for short and medium-size gaps, and Eq. (5) for long-size gaps. Maximum care was taken to ensure that the reconstructed values were physically possible. However, the missing samples ranging between the 23/10/95 and the 08/11/95 (right side of Figure 2), are not well reconstructed, because of the same size gap in the input data series. Therefore we linearly approximated the missing air temperatures, thereby obtaining reproduction of water temperatures.

5. CONCLUSIONS

EPR results for the case study show the effectiveness of the multi-model approach in dealing with environmental problems. We proved the ability of EPR to get parsimonious and efficient models, which can be flexibly adapted to an accurate online forecasting and simulation. The case study confirmed the real capabilities of the multi-model approach enabled by EPR. Additionally, the multi-model EPR strategy not only gave a good physical insight of the phenomenon, but also helped fill missing data, resulting in models that forecast the water temperature. The analysis of similar models returned by different objective (cost) functions ensured a robust choice of the best models. The cost functions were of general type, and not designed specifically for this case study, suggesting that EPR can be used without much customising for a particular problem.

6. REFERENCES

Bennis, S., Berrada, F. and Kang, N., Improving single-variable and multivariable techniques for estimating missing hydrological data, *J. of Hydrology*, 191, 87–105, 1997.

Davidson, J.W., Savic, D.A., & Walters, G.A., Symbolic and Numerical Regression: A Hybrid Technique for Polynomial Approximators, In Advances in Soft Computing: Soft Computing Techniques and Applications, R. John and R. Birkenhead (eds.), Physica-Verlag, Heidelberg, pp.111-116. 1999.

Davidson, J.W., D.A. Savic, and G.A. Walters, Symbolic and Numerical Regression: Experiments and Applications, *Information Sciences*, 150 (1/2), pp. 95-117, 2003.

Efron, B., Bootstrap Methods. Another Look at the Jackknife, *The Ann. of Statist.*, 7:1 – 26, 1979.

Giustolisi, O., and Savic, D.A., A Symbolic Data-driven Technique: Evolutionary Polynomial Regression, *J. of Comp. in Civil Eng.*, ASCE, in preparation, 2004a.

Giustolisi, O., Savic, D.A., Decision Support for Water Distribution System Rehabilitation Using Evolutionary Computing, ACTUI seminar, Exeter, 2004b.

Giustolisi, O., Savic, D.A., and Doglioni A., Data Reconstruction and Forecasting by Evolutionary Polynomial Regression, 6th Int. Conf. on Hydroinformatics, Singapore, 2004a.

Giustolisi, O., Savic, D.A., Doglioni A., and Laucelli, D., Knowledge discovery by Evolutionary Polynomial Regression, 6th Int. Conf. on Hydroinformatics, Singapore,, 2004b.

Goldberg, D.E., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison and Wesley, Reading, Mass., USA, 1989.

Mohseni, O., and Stefan H. G., Stream temperature/air temperature relationship: a physical interpretation, *Journal of Hydrology*, Elsevier, 218, 128 – 141, 1999.

Webb, B.W., Clack, P.D., and Walling, D.E., Water-air temperature relationships in a Devon river system and the role of flow, *Hydrological Processes*, No.17, pp.21-37, 2003.

An Evolutionary-based Real-time Updating Technique for an Operational Rainfall-Runoff Forecasting Model

Soon-Thiam Khu[a], **Edward Keedwell**[a] and **Oliver Pollard**[b]

[a] *Centre for Centre for Water Systems, School of Engineering, Computer Science and Mathematics, University of Exeter, UK EX4 4QF. Email: {s.t.khu; e.d.keedwell}@ex.ac.uk*

[b] *Environment Agency (Southwest Region), UK.*

Abstract: Error-correction is widely known to be one of the effective methods of real-time updating and tends to be the easiest method to implement and couple with existing simulation models. Methods such as autoregressive (AR) or autoregressive integrated moving average (ARIMA) have been widely used but the main disadvantage of such approaches is the prior assumption of the form of error correlation. Genetic programming (GP), a relatively new evolutionary-based technique, can be used to generate a suitable expression linking the observations, simulation model results and the error in the simulation for the purpose of error correction. In this study, GP functions as an error correction scheme to complement runoff forecasting model used by the UK Environment Agency (Southwest region) known as WRIP. WRIP is a transfer function-based operational forecasting software which uses radar rainfall as input. The proposed method is tested on a flashy catchment in Devon, UK. Hourly runoff forecasts of different updating intervals are performed for forecast horizons of up to six hours. The results show that the proposed updating scheme is able to forecast the runoff quite accurately for all updating intervals considered and particularly for those updating intervals not exceeding the time of concentration of the catchment. These results formed part of an ongoing feasibility studies by the UK Environment Agency and the proposed method will be tested on other catchments in the future.

Keywords: genetic programming; real-time updating; rainfall-runoff; artificial neural network; forecasting

1. INTRODUCTION

A catchment flood forecasting system is a system that takes information on the past and current states of meteorological conditions and those of the catchment, as inputs to it, and forecasts the catchment's response into the future. In real-time forecasting, however, the originally forecast values may be updated or modified as measured data become available and, thus, prediction errors can be determined and used for forecasting. In real-time runoff forecasting with rainfall runoff simulation models, rainfall time series up to the desired runoff forecast horizon must be available. The required rainfall time series within the runoff forecast horizon may be estimated with, for example, a non-linear prediction method. In this study, the measured rainfall time series, at any runoff forecast horizon, is made available to evaluate the performance of the proposed evolutionary algorithm-based error updating scheme.

In the Environment Agency (especially for the Southwest region), a key element of the flood forecasting strategy is to use rainfall-runoff modelling in order to increase the lead time for key flood forecasting sites, and to allow the flood duty officer to assess the severity of a flood event. Historically, rainfall-runoff modelling has been done by predicting river flow from total rainfall. This approach was based on an assumption that the only significant floods occur when the catchment is saturated. In reality, how wet the catchment is determines what proportion of rainfall is effective, ie makes its way to the river. More recently, the concept of effective rainfall has been used for flood forecasting with some success.

A recent UK Environment Agency R&D report (Environment Agency, 2003) on real-time modelling of flood discussed several issues on real-time updating. They are:
(i) Updating often improves the accuracy of forecast and hence should be adopted unless there is a good reason not to (e.g. poor data quality);
(ii) Updating should not be used to account for or compensate the use of a poorly structured or inappropriate model;
(iii) Since most updating procedure cannot account for timing error, model calibration can play an important role in adjusting for time lag; and
(iv) In cases where backwater effects predominate the flow (hence rating), it may be necessary to apply updating over a limited flow or level ranges only.

All the above issues highlighted by the R&D report are relevant to MATH/ WRIP model used in this study and, hence, shall be taken into account and addressed in turn:

The aims of this study are to: (1) compare the forecasts of a calibrated rainfall-runoff model, WRIP, with and without the evolutionary-based real-time error updating scheme; (2) evaluate the effectiveness of using total and effective rainfall flood forecasting in conjunction with the proposed evolutionary-based error updating scheme; and (3) suggest how far in the future, i.e. the maximum forecast horizon, the proposed error updating scheme can be used with confidence.

2. AUTOMATIC ERROR UPDATING

Real-time updating algorithms can be grouped into four categories: (i) input updating; (ii) state updating; (iii) parameter updating, and (iv) error updating. Figure 1 shows a schematic diagram of a generic real-time updating system for a computer simulation model (WMO, 1992). It can be seen that all the above updating algorithms rely on or attempt to improve forecasts results by examining previous forecast performance and allowing for adjustments in either (i) input variables (such as precipitation or air temperature); (ii) state variables (state updating); or (iii) model parameters (parameter updating); or (iv) model output values (error correction).

2.1 Previous studies

In 1992, the World Meteorological Organisation conducted a real-time intercomparison study which compared 14 rainfall-runoff models with different updating schemes (WMO, 1992). They found that for small catchments, the Kalman filter approach seemed to perform better than other schemes.

Toth et al. (1999) proposed the used of ARMA and ARIMA models update the results of a conceptual rainfall-runoff model. Their scheme was to model the differences between the measured and simulated runoff from the R-R model, and using the forecasted differences to adjust the simulated runoff. They found that the ARMA(1,1) scheme seemed to be able to improve the forecasted runoff best. Brath et al. (2002) later examined the used of neural networks and non-parametric methods and found that neural networks provided great improvements in the discharge forecast.

The concept of error updating using evolutionary-based methods was first proposed by Khu et al. (2001). In their work, genetic symbolic regression was used to update the MIKE11/NAM model and the results were found to be better than the auto-regressive or Kalman filter method proposed in the WMO intercomparison report (WMO, 1992).

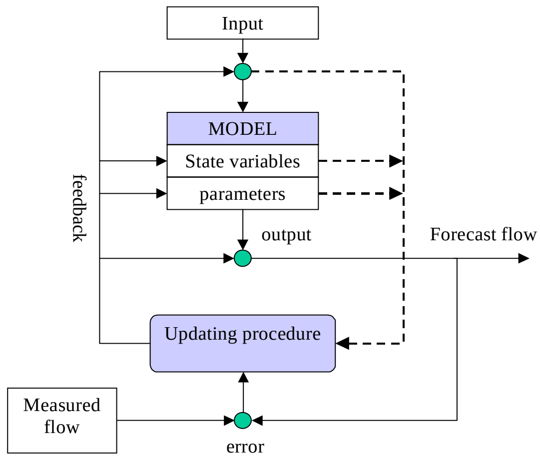


Figure 1: Schematic diagram of different updating modes (adapted from WMO, 1992)

2.2 Genetic Programming approach

Genetic Programming (GP) is a relatively new domain-independent method for evolving computer programs to solve, or approximately solve, problems (Koza, 1992). In engineering applications, GP is frequently applied to model structure identification problems. In such applications, GP is used to infer the underlying structure of either a natural or experimental process in order to model the process numerically. GP inferred models have the advantages of (1) generating simple parsimonious expressions and

(2) offering some possible interpretations to the underlying process.

GP belongs to a class of probabilistic search procedures known as Evolutionary Algorithms (EAs) which includes Genetic Algorithms (GA) (Holland, 1975), Evolutionary Programming (EP) (Fogel et al., 1966) and Evolutionary Strategy (ES) (Schwefel, 1981). These techniques use computational models of natural evolutionary process for the development of computer based problem-solving systems. All evolutionary algorithms function by simulating the evolution of individual structures via processes of reproductive variation and fitness based selection. The techniques have become extremely popular due to their success at searching complex non-linear spaces and their robustness in practical applications.

One successful application of GP in automatic program discovery is that of symbolic regression, instead of the traditional numerical regression. In traditional numerical regression, one pre-determines the functional form, either linear or higher order, and the task is to determine the coefficients. In symbolic regression, the task is to both find a suitable functional form and determine the coefficients. Hence, GP involves finding a mathematical expression, in symbolic form, relating a finite sampling of values of a set of independent variables (x_i) and a set of dependent variables (y_j).

GP can be viewed as an extension of Genetic Algorithm (GA) in terms of the basic principles of operations. Like GA, GP works with a number of solution sets, known collectively as a population, rather than a single solution at any one time. With a large number of solution sets, it gives both techniques the advantage of avoiding the possibility of getting trapped in the local optima.

There are, however, two major differences between GP and GA. They are:

(i) GP works with two sets of variables, instead of one set of variables as in GA. One set of variables, known as the terminal set, contains the independent variables and constants, $\{x_i\}$, similar to GA. The other set, known as the functional set, contains the basic operators used to form the function, *f( )*. For example, the function set may contain the following operators { +, -, *, /, ^, log, sin, tanh, exp, ….} depending on the perceived degree of complexity of the regression. Thus, the symbolic regression is performed using these two variable sets and it is possible to derive a large number of possible functional relationships to fit the data.

(ii) In most EAs, the length of the solution set is normally fixed. In GP, however, the length is allowed to vary from one solution set to another. This variation in length is due to the two genetic operators, crossover and mutation. The flexibility in the structure length increases the search space significantly.

The solution sets in each iteration are collectively known as a generation. In GPs, the size of a population does not have to be the same from one generation to the next. The solutions of the very first generation are usually generated through a random process. However, those of the subsequent generations are generated through genetic operations. Each possible solution set can be represented and visualized in either parse tree form or in Polish notation (Lukasiewicz, 1957). As the population evolves, new solution sets replace the older ones and are supposed to perform better. The solution sets in a population associated with the best fit individuals will, on average, be reproduced more often than the less-fit solution sets. This is known as the Darwinian principle of the "survival of the fittest".

The basic procedure of GP can be described as follows:

1. generate the set of initial population;
2. evaluate each parse tree and assign the fitness;
3. form the temporary population by selecting candidates according to their fitness. This temporary population is called the mating pool. Candidates with higher fitness are given greater probabilities to mate and hence, to produce children or offspring;
4. choose pairs of parse tree from the temporary mating pool randomly for mating and apply the genetic operator called crossover. Crossover is the exchange of genetic material (such as fitness, composition) between two selected candidates;
5. select a crossover site where the material will be exchanged randomly, thereby resulting in the creation of offspring;
6. apply another genetic operator known as mutation which randomly changes the genetic information of the candidate;
7. copy the resultant chromosomes into the new population;
8. evaluate the performance of the new population;
9. repeat steps 3-8 until a predetermined criterion is reached.

3. OPERATIONAL R-R MODEL

The operational rainfall-runoff model used in this study is the WRIP (Weather Radar Information Processor) system originally developed for the southwest regional office, UK Environment Agency. WRIP is a conceptual model that utilises the concept of transfer functions and physical realisable transfer function (PRTF) (Han, 1991) to model the relationship between rainfall and runoff from a catchment. The PRTF model is linear and time-invariant, and accommodates the non-linear time-variant nature of rainfall-runoff process by three parameters (shape, volume and timing). The transformation of these parameters into basic transfer function parameters guarantees stability. The simulated flow can then be merged with the telemetered flow data to produce the predicted hydrograph.

WRIP's main source of error (apart from rainfall forecast) is that the transfer function coefficients are determined based on calibrated events and hence, the model is essentially deterministic. The calibrated model should forecast well if an event similar to the calibration event occur, but not for "dissimilar" events. Hence, there is a need to adjust the model performance based on dynamic simulation.

4 PROPOSED SCHEME

4.1 Linking GP with WRIP

A GP toolkit was developed to facilitate the integration of the updating scheme with the operational flood forecasting model, WRIP. Since WRIP is used in real-time, the computational efficiency of the updating scheme is of primary importance. The standard forecasting interval for the Environment Agency is 1 hour, with lead times of up to 6 hours depending on catchments, hence, the forecasting period is not of a major constraint.

WRIP has a real-time link-up with radar information direct from the UK Meteorological Office which provides radar rainfall information at 5 minutes intervals. In other words, the flood duty officer can change their flood forecasting model once they receive the latest radar rainfall information. In practice, WRIP is usually configured to update the model forecast at 15 minutes intervals.

Two different forms of automatic error updating algorithm are investigated. They are genetic programming (GP) and artificial neural networks (ANN). A genetic programming toolkit was developed to facilitate the analysis and details of the genetic programming method used have been reported in Khu *et al.* (2001).

The procedure of application of GP for real-time updating can be expressed as follows and shown in Fig. 2. The WRIP model is first used to simulate the discharge, *QSIM*, for the whole period of interest based on the rainfall data, *R*. The proposed procedure is then used to compute the prediction error, ε, by comparing the simulated discharge, *QSIM*, with the observed discharge, *QOBS*, for time, *t*. The new simulated or improved discharge, $QIMP_t$, is computed by adjusting $QSIM_t$ for each forecast lead-time within the forecast horizon.

Mathematically, the measured discharge, $QOBS_t$, at time *t*, can be expressed as:

$$QOBS_t = QSIM_t + \varepsilon_t \qquad (1a)$$

or

$$\varepsilon_t = QOBS_t - QSIM_t \qquad (1b)$$

GP is used to infer the functional relationship, *F( )*, between the simulated discharges and the past simulation errors, and the present simulation error. For lead time of 1 hour, the functional relationship for GSR prediction error, $\hat{\varepsilon}_{t+1}$, may be expressed as follows:

$$\hat{\varepsilon}_{t+1} = F\{QSIM_{t+1},\ QSIM_t,\ .\ QSIM_{t-4},\ \varepsilon_t,\ \varepsilon_{t-1}\ ..\ \varepsilon_{t-4}\} \qquad (2a)$$

and the forecast improved discharge, $QIMP_{t+1}$, can be obtained from:

$$QIMP_{t+1} = QSIM_{t+1} + \hat{\varepsilon}_{t+1} \qquad (2b)$$

For lead time of 2, 3,…, α hours, the recursive form of Eq. (2) can be written as:

$$\hat{\varepsilon}_{t+\alpha} = F\{QSIM_{t+\alpha}\ ,.,\ QSIM_{t+\alpha-4}\ ,\ \hat{\varepsilon}_{t+\alpha-1}\ ,.,\ \hat{\varepsilon}_{t+\alpha-4}\} \qquad (3a)$$

$$QIMP_{t+\alpha} = QSIM_{t+\alpha} + \hat{\varepsilon}_{t+\alpha} \qquad (3b)$$

It should be noted that the values of $\hat{\varepsilon}_{t+\alpha}$ in Eq. (3a) may be either the actual errors at instances when measured data are available or GP derived errors.

4.2 Application

The proposed evolutionary-based updating method is applied to simulate flow for Bishops Hull, River Tone, a rural catchment upstream of Taunton. It is important to get reliable flood forecasts at Bishops Hull as Taunton has come close to flooding on several occasions in the past few years. Predictions of the timing, shape and scale of the flood response are superior using effective rainfall instead of total rainfall. Improved flood predictions are possible using total rainfall, but extra care is needed in adjusting the volume runoff to initial catchment wetness index. Use of total rainfall will not improve the prediction of the timing and shape of the flood hydrograph, and will cancel out any benefits due to real-time updating of predicted flows using observed flows.

A series of experiments was conduct to investigate the suitability of real-time error updating/ correction. Since error correction can be used in conjunction with either total rainfall predicted flow or effective rainfall predicted flow, the effectiveness on both types of flow was investigated. Table 1 show the results of applying GP and ANN error updating on both total and effective rainfall for the period December 1999 (calibration). It can be seen that the error updating results using GP are comparable to that using ANN. Similar good results can be obtained from validating the algorithm for an unseen rainfall-runoff period (Apr 2000).

From the results in Table 1, it can be seen that both GP and ANN can effectively improve the forecast runoff for up to 5 time-steps ahead, using total rainfall forecasting mode. If the effective rainfall mode is used, although the un-corrected WRIP's forecast was better than that of using total rainfall, the results of error correction can only be effective up to 4 time-steps ahead.

5 CONCLUSIONS

This paper discusses a novel technique of coupling a conceptual rainfall-runoff model with an evolutionary-based error correction/ updating technique. The results showed that with the addition of a suitable error updating/ correction such as genetic programming (GP) or artificial neural networks (ANN), the results of WRIP had improved significantly. This applies for both calibration and validation data set of the catchment under investigation.

It is shown that the proposed error correction scheme, when applied on WRIP, is more effective for forecast using total rainfall. Both GP and ANN were able to provide good forecast for up to 5 time-steps (i.e. 5 hours) ahead.

The resultant formula (not shown here due to space limitation) from the GP error updating is transparent to the flood warning engineers and can be interpreted as an advanced form of auto-regressive formulation (Khu et al., 2001). Any over or under prediction by the WRIP software will be captured via the simulated flow terms. Moreover, an error in the correction terms from the previous time-step will be automatically adjusted by the error terms. In this sense, the updating technique using GP and ANN are actually learning from the mistaks of previous WRIP prediction and GP error correction.

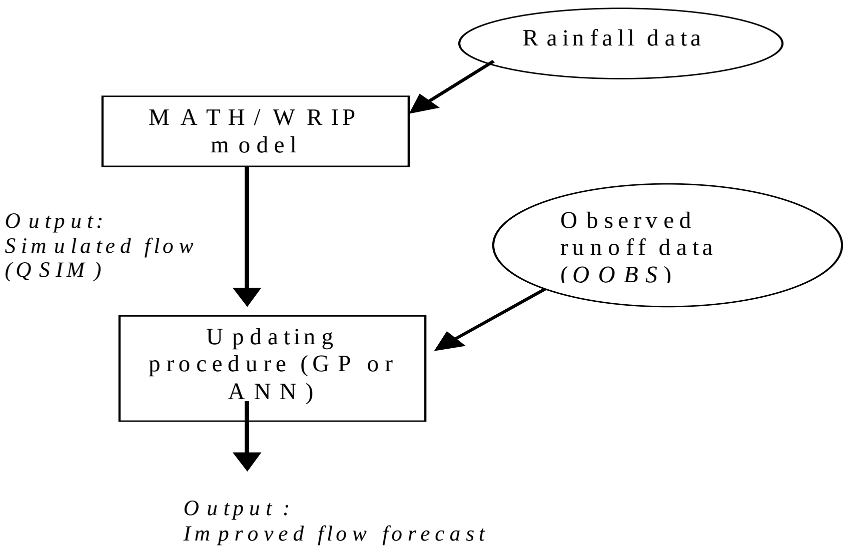


Figure 2: Flowchart of error correction in WRIP model

Table 1: Root Mean Square Error for Different Prediction Lead-times for GP and ANN

		Root Mean Square Error, RMSE (m^3/s)					
		Using Total rainfall			Using Effective Rainfall		
	Lead-time (hours)	WRIP	Calibration	Validation	WRIP	Calibration	Validation
GP	1	8.54* / 4.53**	0.90	0.63	6.34* / 2.59**	0.96	0.73
	2	----	1.25	1.15	----	1.49	0.67
	3	----	1.76	1.47	----	2.52	1.42
	4	----	2.68	1.90	----	3.68	3.21
	5	----	3.26	2.35	----	3.70	3.11
ANN	1	8.54* / 4.53**	0.61	0.63	6.34* / 2.59**	0.90	0.61
	2	----	2.00	0.96	----	1.25	1.24
	3	----	3.09	1.96	----	1.76	1.94
	4	----	5.70	3.34	----	2.68	2.49
	5	----	6.79	3.31	----	3.26	3.33

Note: Under WRIP column – "" applies to calibration data set; "**" applies to validation data set.*

6 REFERENCES:

Brath A., A. Montanari and E. Toth. Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. *Hydrology and Earth System Sciences, 6*, 627 – 640, 2002.

Environmental Agency. *Fluvial Flood Forecasting – Real-time Modelling*. R & D Technical Report W5C-013/5/TR. 2003.

Fogel, L. J., Owens, A. J. and Walsh, M. J., (1966). Artificial Intelligence through Simulated Evolution. John Wiley, New York.

Han, D. *Weather Radar Information Processing and Real-time Flood Forecasting*. PhD. Thesis. University of Salford, UK. 1991.

Holland, J. H., (1975). Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.

Khu, S.T., S.Y. Liong, V. Babovic, H. Madsen and N. Muttil. Genetic programming and its application in real-time flood forecasting. *J. American Water Resources Assoc., 36(2)*, 439- 452, 2001.

Koza, J.R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press, Cambridge, MA. 1992.

Lukasiewicz, J. Aristotle's Syllogistic from the Standpoint of Modern Formal Logic. Oxford , Clarendon Press. 1957.

Schwefel, H.P.. Numerical Optimization of Computer Models. John Wiley, Chichester. 1981.

Toth, E., A. Montanari and A. Brath. Real-time flood forecasting via combined use of conceptual and stochastic models. *Phys. Chem. Earth (B), 24(7)*, 793-798, 1999.

WMO. *Simulated Real-Time Inter-Comparison of Hydrological Models.* WMO Operational Hydrology Report no. 38, WMO no. 779. World Meteorological Organization, Geneva. 1992.

A fast Evolutionary-based Meta-Modelling Approach for the Calibration of a Rainfall-Runoff Model

Soon-Thiam Khu[a], Dragan Savic[a], Yang Liu[a] and Henrik Madsen[b]

[a] *Centre for Centre for Water Systems, School of Engineering, Computer Science and Mathematics, University of Exeter, UK EX4 4QF. Email: {s.t.khu; d.savic; yang.liu}@ex.ac.uk*

[b] *DHI Water and Environment*

Abstract: Population-based search methods such as evolutionary algorithm, shuffled complex algorithm, simulated annealing and ant colony search are increasing used as automatic calibration methods for a wide range of water and environmental simulation models. However, despite the advances in computer power, it may still be impractical to rely exclusively on computationally expensive (time consuming) simulation for many real world complex problems. This paper proposed the use of meta-models to replace numerical simulation models for the purpose of calibration. Meta-models are essentially "model of the model". The meta-model used in this study is the artificial neural network and, when coupled with genetic algorithm, forms a fast and effective hybridisation. The proposed evolutionary-based meta-model reduces the number of simulation runs required in the numerical model considerably thus making the automatic calibration of computationally intensive simulation models viable. The new approach was developed and tested in the calibration of a popular rainfall-runoff model, MIKE11/ NAM, applied to the Treggevaede catchment in Denmark. Both the calibration and verification results for single objective calibration indicate that the proposed method is able to achieve the same or better calibration performance compared to published studies using traditional population-based search methods and yet required only about 40% of the simulation runs on average.

Keywords: evolutionary algorithms; meta-models; rainfall-runoff; artificial neural network; calibration

1. INTRODUCTION

Population-based search methods such as evolutionary algorithm (EA) (which includes genetic algorithms, evolutionary strategies, evolutionary programming etc), shuffled complex algorithm, and simulated annealing are powerful search algorithms that can be used for optimisation. These algorithms are increasing used as automatic calibration methods for a wide range of water and environmental simulation models, especially when there are a large number of calibration parameters and some, or all, of them are interacting with one another. However, the main weakness in using population-based search methods for automatic calibration is that they require a large number of fitness evaluation, thereby render them not suitable to calibrate computational intensive simulation models. It is not uncommon for large simulation models for run for up to an hour or more and with typical EA run requiring thousand (if not tens of thousands) of model evaluations, automatic calibration using EA for large simulation models may not be totally feasible.

Currently, there are two main approaches to resolve the problem using EA for model evaluation and calibration. They are: (i) using faster EA algorithms; and (ii) using more computing power. The first approach exploits the flexibility of EA to develop more efficient techniques requiring less function evaluations and hence, less model evaluations. Typical methods of this approach are: hybridization of EA with some form of heuristics (Deb and Beyer, 2001; Keedwell and Khu, 2003); enhancement to EA operators (reproduction and selection) (Salami and Hendtlass, 2003; Liong et al., 2001). The second approach uses the inherent parallel computing capability of EA and allows simultaneous multiple model simulation on multiple processors. (Kohlmorgen et al., 1999; Rivera, 2001).

However, there is a third method that can be effectively and effortless coupled with EA to enable the calibration of large water and environmental simulation models. To reduce the computational cost of model evaluations/ simulations, surrogate evaluation tools, i.e. meta-models, are used in place of the time consuming simulations. Meta-models, otherwise known as surrogate or approximate models, are essentially "model of the model" which may be used to approximate the simulation model. According to Emmerich et al. (2002), "a metamodel approximates a multivariate function using points that have already been evaluated…. and is considered to be a fast surrogate model to the exact evaluation model." A variety of meta-models exist (e.g. design of experiments, response surface methodology, Taguchi design, kriging, neural networks, multivariate adaptive regression splines) and Simpson et al. (2001) provides a comprehensive review of the use of meta-models for engineering design.

This paper proposed the use of meta-models to strategically replace numerical simulation models (the "Simulator") for the purpose of calibration. The meta-model used in this study is the artificial neural network (ANN) and, when coupled with genetic algorithm (GA), forms a fast and effective hybridisation. This paper starts with an overview of meta-models and its applications in engineering, specifically water resources and environmental engineering. GA and ANN are subsequently described followed by two examples of effective usage of evolutionary-based meta-models for parameter estimation / calibration. Finally section 5 will give concluding remarks and some discussions on future directions.

2. META-MODELS

Meta-models have been in existence for a fairly long period of time (Kleijen, 1975) and are widely used by the engineering design community to reduce the time require for full simulation. An extreme example is the use of meta-model in place of motor vehicle crash test simulations, where Ford Motor Company reports that one crash simulation on a full passenger car takes 36-160 hours (Gu, 2001).

The basic approach of using meta-model for design is as follows:

- Select a multivariate mathematical function (meta-model) which can be used to approximate the "Simulator";
- Run the "Simulator" for a small number of runs;
- Construct the meta-model and adjust the variables within the model to fit the run results from the simulator;
- Once the adjustments are complete, the meta-model is used in place of the "Simulator" for future evaluation of the new designs.

The above approach requires certain modification if the "Simulator" is constantly changing, such as the case during calibration. The modifications are:

- Make necessary adjustments by usually running the "Simulator" more times; and
- A mechanism to update the meta-model.

Regardless of the usage of meta-models, three steps are involved (Simpson et al., 2001):

1. choosing an experimental design for generating data;
2. choosing a mathematical model to represent the data; and
3. fitting the model to the observed data.

Each of the steps may have many options and the choice of option in each step give rise to different meta-models. For example, generating data using fractional factorial design and fitting the data onto a second order polynomial function using method of least squares regression gave rise to the meta-model known as "response surface methodology", while measured data may be fitted onto a network of artificial neurons using least squares with back-propagation giving rise to "artificial neural network" as a meta-model.

Meta-models have also been successfully applied to model a variety of water and environmental problems. Some examples are: the response surface method has been applied to predict numerical geophysical models (Tatang et al., 1997), reconstruction and interpolation of effluent plume in an estuary (Riddle et al., 2004) and calibration of urban drainage model (Liong et al., 1995); Kriging has been used to model spatio-temporal pollutant deposit trend through the atmosphere (Haas, 1998), spatial distribution of heavy metals in a river basin (Ouyang et al., 2002) and shallow water wave in an estuary (Gorman and Neilson, 1999); Artificial neural networks have been used to model the input-output behaviour of wastewater treatment plants (Belanche et al., 1999), deforestation simulation (Mas et al., 2004), prediction of pollutant trends in urban areas (Lu, et al., 2004); and many others applications.

3. EA-BASED META-MODELS

3.1 Genetic Algorithm

One of the most common EA is the genetic Algorithms (GAs). GAs are computationally simple yet powerful search algorithms based on the mechanics of natural selection and natural genetics, which combines an artificial survival of the fittest with genetic operators from nature. GAs mimic the adaptation of natural species and genetically evolve to suit their environment over many generations. Using this analogy, a mechanism involving selection, crossover, and mutation can be used to evolve a population of potential solutions towards improved solutions.

GAs are especially useful for complex optimisation problems where the analytical solutions are difficult to obtain and it has been used for water and environmental model optimisation and calibration since early 90s. However, one of the main obstacles when trying to apply GA in practical optimisation problems is the large number of function evaluations required.

3.2 Artificial Neural Networks

Artificial neural network (ANN) is a computing paradigm designed to mimic natural neural networks in the biological brain. ANNs are commonly thought of as universal approximators for function mapping. Multilayer perceptrons (such as feed-forward backpropagation algorithms) and radial basis functions (RBF) are commonly used ANNs. In this study, the RBF neural network has been suggested because its simple structure enables learning in stages, gives a reduction in the training time. A standard RBF network has a feed-forward structure consisting of two layers, a nonlinear hidden layer and a linear output layer, and uses a Gaussian function as activation function to transforming inputs. Clustering identifies the centre point and radius (i.e., mean and standard deviation) of the Gaussian function in each unit of the RBF network.

3.3 Integrating Meta-models with GAs

As stated in the introduction, one possible way of overcoming the problem of time consuming simulation in EAs (including GA) is to use meta-models in place of the simulation model. Many researchers, especially in engineering design, have examined strategies to integrate different meta-models with GA (Giannakoglou et al., 2001; Poloni et al., 2000; Ong et al., 2003).

The most direct way of integrating meta-models with GA is to replace the "Simulator" with the meta-model completely during evaluation of objective function in GA. However, in order to construct the meta-model, a small number of run of the "Simulator" is required. This is the experimental design mentioned in Section 2 and can be performed either using Taguchi method, Design of Experiments, response surface methodology or even using GA. Liong et al. (2001) detailed one such method using fractional fractorial design with central composite design to provide initial population for GA. Emmerich et al. (2002) used kriging as the meta-model and they found that Kriging provided the local error estimation which enable them to assess the reliability of the solutions. Giannakoglou et al. (2001) used radial basis function network as meta-model coupled with GA to optimise an airfoil shape design. Poloni et al. (2000) used a hybridisation of GA, ANN and local search method to optimise the design of a sailing yacht fin keel. The ANN acted as a surrogate model for 3D Navier-Stokes simulation of the fin keel while cruising.

Another potential usage of evolutionary-based meta-model is the evaluation of risk and uncertainty. Currently, different sampling approaches have been devised to perform fast and effective sampling. Monte-Carlo sampling (MCS) is commonly regarded as the most accurate approach but it requires thousands, if not tens of thousands, of model evaluation. Importance sampling, Metropolis algorithms, Latin Hypercube method etc., are fast alternatives but they are approximating the statistical properties of the MC samples. Recently, Khu and Werner (2003) proposed the use of meta-model (GA – ANN) to select regions of interest for sampling. Their method required only about 10% of the MCS method.

Despite the extensive works in evolutionary-based meta-models, little effort is place on overcoming the problem of "changing landscape". During the process of optimisation, the region of GA search will constantly change, and it is reasonable to assume that the meta-model will have to be suitability modified to account for such changes. As the search progresses, more information on the objective function will be obtained, and suitable mechanism should be implemented to utilise this additional information and update the meta-model. The example in the next section demonstrates such a scheme where a GA-ANN meta-model is used to calibrate a rainfall-runoff model and the meta-model is constantly but strategically updated with the latest information.

4. CALIBRATION OF A RAINFALL-RUNOFF MODEL

A fast evolutionary-based meta-model using an innovative hybridisation of GA and RBF is proposed for the automatic calibration of numerical simulation models. The GA is used to search for the optimal objective function in much the same way as any optimisation routine. The ANN is used to map (and adapt) to the response surface of the objective function and used as a fast surrogate for the NAM model at regular intervals. The concept of using RBF as surrogate models is not new. However, as the GA search progresses, the response surface of the objective function tends to change and therefore, the meta-model needs to be self-adapting to the changing landscape.

The hybrid algorithm (GA-RBF) is presented below and illustrated in Figure 1:

(1) Run GA for *g* number of generations of population size, *p*;
(2) Train RBF to map the response surface using (*g* * *p*) points generated by GA;
(3) Perform selection and reproduction in GA;
(4) Evaluate the new GA population using the trained RBF instead of NAM;
(5) Select *m* best individuals in the new population and evaluate the true fitness using the NAM model;
(6) Update RBF using the true fitness from (5);
(7) Perform steps (3) to (6) until the stopping criterion is met.

4.1 Application Example

The proposed meta-model was adopted for calibration of the MIKE 11/NAM rainfall-runoff model applied to the Danish Tryggevaelde catchment. The calibration parameters used are the same as those in Madsen (2000). This catchment has an area of 130km^2, an average rainfall of 710 mm/year and an average discharge of 240 mm/year. The catchment is dominated by clayey soils, implying a relatively flashy flow regime. For the calibration, a 5-year period (1 Jan. 1984−31 Dec. 1988) was used where daily data of precipitation, potential evapo-transpiration, mean temperature, and catchment runoff are available. For comparing the calibrate models, validation data covering the periods 1 Jan. 1979−31 Dec. 1983 and 1 Jan. 1989−31 Dec. 1993 were used. The following two of objective functions are used in this study:

Average Root Mean Squared-Error (RMSE) of low flow events:

$$F_1(\theta) = \frac{1}{M_l}\sum_{j=1}^{M_l}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}[Q_{obs,i} - Q_{sim,i}(\theta)]^2\right]^{1/2} \qquad (1)$$

Average Root Mean Squared-Error (RMSE) of peak flow events:

$$F_2(\theta) = \frac{1}{M_p}\sum_{j=1}^{M_p}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}[Q_{obs,i} - Q_{sim,i}(\theta)]^2\right]^{1/2} \qquad (2)$$

In Eqs. (1)−(2), $Q_{obs,i}$ is the observed discharge at time i, $Q_{sim,i}$ is the simulated discharge, M_p is the number of peak flow events, M_l is the number of low flow events, n_j is the number of time steps in peak/low event no. *j*, and θ is the set of model parameters to be calibrated. Peak flow events were defined as periods with flow above a threshold value of 4.0 m^3/s, and low flow events were defined as periods with flow below 0.5 m^3/s.

Preliminary optimisation runs showed that entire population converged around the global optimum after about 2000 model evaluations. Thus, for each test, a maximum number of model evaluations equal to 2000 were employed as a stopping criterion with the population size p=50 and total number of generations G=40. In this work, floating point representation was used together with uniform crossover. The crossover rate used was 0.9 and mutation rate = 0.1.

Sensitivity analysis was performed to provide some idea of the values of *g* and *m* to use and the values of g = 2 and m=15 was found to provide good results. Hence, the number of simulation runs for NAM model calibration using the hybrid method can be calculated by: $p * g + m * (G - g)$ = 50 × 2 + 15 × (40 − 2) =670, which is 33.5% of the required GA run.

4.2 Results and Discussions

A total of 10 random calibration runs were perform for each objective function (Eqns. (1) and (2)) and the evaluated performance statistics of the GA and the hybrid method (GA-RBF) are shown in Tables 1. Tables 1 also shows the validation results of applying the calibrated parameter set to the two different validation periods. The hybrid method (GA-RBF) gave very close results to those from GA (in terms of best, worst and mean RMSE) but required only about 40% of the GA simulation runs. Even though some of the calibration results indicated that GA-RBF were be better than GA, mixed but comparable results were obtained for the validation data sets. The small standard deviation indicated the hybrid method is very stable and able to reproduce consistent results.

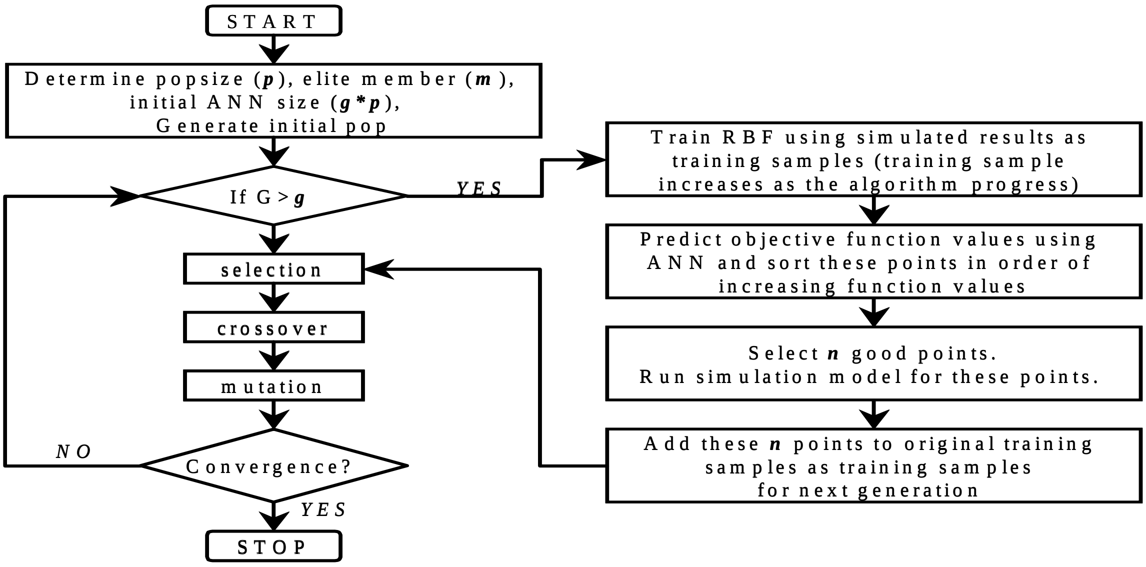


Figure 1: Flow chart of the evolutionary-based meta-model

Table 1: Calibration and validation results for low flow and peak flow RMSE

		Calibration data (1984−1988)		Validation data set 1 (1979−1983)		Validation data set 2 (1989−1993)	
		GA	GA−RBF	GA	GA−RBF	GA	GA−RBF
Low Flow RMSE (m^3/s)	Best RMSE	0.1345	0.1323	0.2065	0.1752	0.0986	0.1043
	Worst RMSE	0.1782	0.1697	0.2491	0.2516	0.1470	0.1519
	Mean	0.1543	0.1451	0.2215	0.2196	0.1217	0.1164
	STD	0.0119	0.0110	0.0140	0.0210	0.0166	0.0159
Peak Flow RMSE (m^3/s)	Best RMSE	1.1750	1.1687	1.1724	1.1836	1.0485	1.1325
	Worst RMSE	1.2378	1.2178	1.2564	1.2516	1.3945	1.3672
	Mean	1.2016	1.1966	1.2303	1.2165	1.2005	1.2386
	STD	0.0184	0.0175	0.0292	0.0227	0.0964	0.0798

5 CONCLUSIONS

This paper discusses the concept of meta-model and the integration between evolutionary algorithms and meta-models. It can be seen that there is significant advantage in using meta-models for water and environmental system simulation, design and calibration. However, one major problem for evolutionary-based meta-modelling is how to ensure that the meta-model is constantly relevant as the search progresses. To overcome this problem, a strategic and periodic scheme of updating the meta-model is proposed.

A novel evolutionary-based meta-model, using genetic algorithm and radial basis function neural network with dynamic updating, for hydrological model calibration has been proposed in this paper. It has been shown that the proposed method performed more efficiently when compared to GA. The results indicated that the proposed method was able to reduce the required simulation runs to 40% of GA while achieving comparable calibration and validation results. The results provided us with confidence that the proposed method is indeed a viable method to reduce the computation effort required in calibrating rainfall-runoff models while constantly updating the meta-model. The proposed methodology presents a viable option to calibration and optimise computationally intensive water and environmental simulation models.

Work is currently undergoing to (i) test the proposed algorithm on known different functions, (ii) extend the work to multiple objective functions; and (iii) extend the scheme to become even more adaptive with the changing landscape. Jin et al. (2002) addressed the dynamic landscape issue using "fuzzy" error rules to determine the frequency and timing of fitness approximation and it may be possible to incorporate such method with the proposed algorithm.

6 ACKNOWLEDGEMENTS

The financial assistance from ABPmer Ltd to Yang Liu is greatly appreciated.

7 REFERENCES:

Belanche, L.A., J.J. Valdes, J. Comas, I.R. Roda and M. Poch. Towards a model of input-output behaviour of wastewater treatment plants using soft computing techniques. *Env. Modelling & Software, 14*, 409-419, 1999.

Deb, K. and H.G. Beyer. Self-adaptive genetic algorithm with simulation binary crossover. Evolutionary Comp., 9(2), 197-221, 2001.

Emmerich, M., A. Giotis, M. Ozdemir, T. Back and K. Giannakoglou. Metamodel-Assisted Evolution Strategies. *Lecture Notes in Computer Sc. 2439*, 361-370, 2002.

Giannakoglou, K.C., A.P. Giotis, and M. Karakasis. Low cost genetic optimisation based on inexact pre-evaluations and the sensitivity analysis of design parameters. *Inverse Problems in Engrg., 9*, 389-412, 2001.

Gorman, R. M. and C. G. Neilson. Modelling shallow water wave generation and transformation in an intertidal estuary. *Coastal Engrg., 36(3)*, 197-217, 1999.

Haas, T.C. Statistical assessment of spatio-temporal pollutant trends and meteorological transport models. *Atm. Env., 32(11)*, 1865-1879. 1998.

Jin, Y., M. Olhofer and B. Sendhoff. A framework for evolutionary optimisation with approximate fitness functions. *IEEE Transactions on Evolutionary Computation, 6(5)*, 481-494, 2002.

Keedwell, E. and S.T. Khu. More choices in water system design through hybrid optimisation. *Proc. Int. Conf. on Comput. and Control for the Water Industry*, pp. 257-264, 2003.

Kleijnen, J.P.C. A comment on blanning's metamodel for sensitivity analysis: The regression metamodel in simulation. *Interfaces, 5(1)*: 21-23, 1975.

Khu, S.T. and M.G.F. Werner. Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. *Hydrol. Earth Sys. Sc., 7(5)*, 680-692, 2003.

Kohlmorgen, U., H. Schmeck and K. Haase. Experiences with fine-grained parallel genetic algorithms. *Annals of Op. Res., 90*, 203-219, 1999.

Lu, W.-Z., W.-J. Wang, X.-K. Wang, S.-H. Yan and J.C. Lam. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. *Env. Res., in press*, 2004.

Liong, S.Y., J. ShreeRam and W.T. Chan. Catchment calibration using fractional factorial and central composite designs-based response surface, *J. Hydr. Engrg., 121(8)*, 613-617, 1995.

Liong, S.Y., S.T. Khu and W.T. Chan. Derivation of Pareto front with genetic algorithm and neural networks. *J. Hydrol. Engrg., 6(1)*, 52-61, 2001.

Madsen, H. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, 235, 276-288, 2000.

Mas, J.F., H. Puig,, J.L. Palacio and A. Sosa-Lopez. Modelling deforestation using GIS and artificial neural networks. *Env. Modelling & Software, in press*, 2004.

Ong, Y.S., P.B. Nair and A.J. Keane. Evolutionary optimisation of computationally expensive problems via surrogate modelling. *AIAA J., 40(4)*, 687-696, 2003.

Ouyang, Y., J. Higman, J. Thompson, T. O'Toole and D. Campbell. Characterization and spatial distribution of heavy metals in sediment from Cedar and Ortega rivers sub-basin. *J. Contaminant Hydrol., 54(1-2)*, 19-35, 2002.

Poloni, C., A. Giurgevich, L. Onesti, and V. Pediroda. Hybridization of a multi-objective genetic algorithm, a neural network and a classical optimzer for a complex design problem in fluid dynamics. *Comput. Methods in Applied Mech. and Engrg., 186*, 403-420, 2000.

Riddle, A.M., R.E. Lewis and A.D. Sharpe. Environmental quality information processor (EQUIP). *Env. Modelling & Software, 19(1)*, 57-62. 2004.

Rivera, W. Scalable Parallel genetic algorithm. *Artificial Intel. Rev., 16*, 153-168, 2001.

Salami, M. and T. Hendtlass. A fast evaluation strategy for evolutionary algorithms. *Applied Soft Comput., 2*, 156-173, 2003.

Simpson, T.W., J. Peplinski, P.N. Koch and J.K. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engrg. With Computers, 17(2)*: 129-150, 2001.

Tatang, M.A., W. Pan, R.G. Prinn and G.J. McRae. An efficient method for parametric uncertainty analysis of numerical geophysical models. *J. Geophysical Res., 102*, 21925-21932, 1997.

Method for water distribution systems reliability evaluation

Tina Dasic and Branislav Djordjevic
Faculty of Civil Engineering, Belgrade, Yugoslavia
E-mail: mtina@grf.bg.ac.yu

Abstract: A new methodology for water resources systems reliability evaluation is presented. The proposed methodology considers both: mechanical reliability (probability of pipe failure) and reliability of hydraulic parameters in the nodes and links (pressure, velocity). On the basis of this methodology the model NetRel was developed. This model is useful for determining reliability of systems with different configurations and complexity. Also, methodology for optimal reliability allocation, based on genetic algorithms, is proposed. That methodology, coupled with the reliability evaluation method, is an efficient tool for solving problems of optimal allocation of water distribution network reliability.

Key words: reliability, water supply systems, networks, reliability of hydraulic parameters

1. INTRODUCTION

In common engineering practice water distribution systems are designed using only heuristic criteria. Determining the optimal configuration and network parameters that can meet required flow and pressure rate are the result of hydraulic and cost-benefit analyses. The probability of system failure and other reliability statistics are very rarely included in such analyses.

Unlike other technical systems (such as car, airplane and similar systems), where reliability is an important system characteristic rigorously determined and critically analyzed, the probability of failure in water resources systems relies on examining the possibility of meeting demand under some predefined "worst case" scenarios. As a result of such practice, certain system elements are over designed, but reliability of the entire system is usually inadequate.

In the phase of planning and design of the optimal system configuration, required reliability should be included as an important parameter. Existing practice, the mutual comparison of different systems without including reliability as criteria, can lead the designer to an unreliable solution that needs further repairs or remediation.

Due to the fact that the failure of water distribution systems causes serious consequences in the social and economical environment, these characteristics have become a field of examination in the last decade.

2. NETREL - MODEL FOR RELIABILITY EVALUATION

One of the reasons that reliability has not yet become a common phase in design practice is its complexity. While the reliability of other technical systems depend only on the network configuration and the failure rate of its elements (it is defined as mechanical reliability), the water distribution system has the additional request of meeting the network hydraulic parameters. So, the demand in some node will be satisfied (the node will fulfill its task) if it is physically connected to at least one source node and if pressure is in accordance with designed levels. This second probability can be defined as the probability of meeting hydraulic parameters, or hydraulic reliability for short.

In this model (NetRel) mechanical and hydraulic reliability are coupled in overall network reliability or mechanical-hydraulic reliability. It defines the probability that the network will meet specified hydraulic parameters (specified flow and pressure in nodes and/or velocity in pipes).

Two probability measures are defined:

1) network reliability - probability that established hydraulic parameters (usually the flow and pressure rates) will be satisfied in all demand nodes (links).
2) single node reliability - probability that the defined hydraulic parameters will be satisfied in a specified demand node. Single node reliability can be of great interest when it

denotes an important consumer (hospital, school, some dangerous industry,...).

All the nodes in the network are modeled as perfectly reliable. Each link is said to have a probability r_i of functioning at any point in time, and probability $q_i=1-r_i$ of being inoperative. Links are assumed to fail independently. These assumptions may be questioned and they will certainly be the topic of future investigations, but presently they are standard assumptions defined in almost all reliability calculations.

2.1 Mechanical reliability

Mechanical reliability of the water distribution network is calculated using the main theorem of binary function decomposition. The calculation process is divided into three models (subprograms): aggregation model, system decomposition and determination sub reliability.

2.1.1 Aggregation model

The aggregation model is used to aggregate a complex system into a less complex one (tree network or purely-looped network[1]), using series and parallel aggregations.

Series aggregation is performed by replacing two links incident to the same node, with probabilities r_1 and r_2, by one link. The probability of the new link operation can be determined as $r_1 \cdot r_2$ (Figure 1a). A parallel aggregation is performed by replacing two links connected to same nodes by one link. The probability of the new link can be calculated as $1-q_1 \cdot q_2$.

The described aggregation model cannot be used to establish overall network reliability, as the reduction of any demand node effects the overall network reliability. In that case the algorithm for the calculation of the "K-node reliability", developed by Satyanarayana and Wood (Wagner et al. [1988]) is used. K-nodes are defined as nodes of interest in the network (in water distribution systems K-nodes are demand nodes). Beside K-nodes there can be others, so called simple nodes (nodes with no water demand, used only to define system configuration) in the network.

According to this method, series aggregation can be performed if either of following two conditions is satisfied: node *v* is a simple node or all three nodes are K-nodes. In the first case, new link reliability is calculated as in the previously described series aggregation method (Figure 1a). If all three nodes are K-nodes (Figure 1b) the reliability of the new link can be calculated as $r_1 \cdot r_2/(1-q_1 \cdot q_2)$, and the system correction factor (denoted by Ω) is multiplied by $(1-q_1 \cdot q_2)$. The correction factor accounts for the necessity that the middle K-node be connected to the others, even when the middle K-node seems to "vanish". After the network has been reduced the reliability of the network is found by multiplying the probability by the correction factor.

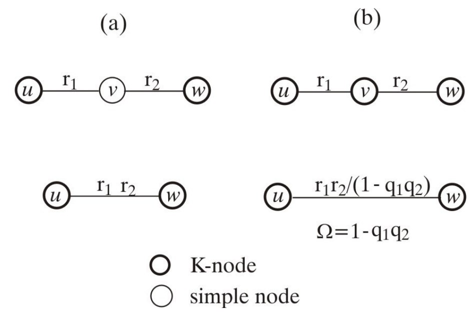


Figure 1. Series aggregations in K-nodes method

In parallel aggregations two or more links (connecting the same nodes) are replaced by new links, with greater reliability. So, there is no node reduction, and the aggregation in K-node method can be performed as previously described.

When the overall system reliability is calculated, all demand nodes are defined as K-nodes. When a single node reliability is calculated, only one node is a K-node, the node for which the reliability is calculated, while all other nodes are simple nodes (although they are defined as demand nodes).

2.1.2 System decomposition

If the result of series and parallel aggregations is a purely-looped network, the reliability of such a system can not be easily calculated. In that case the network has to be link reduced.

The method is based on main theorem of binary function decomposition. *i link decomposition* suppose the network is divided into two sub networks (Figure 2):

$(1_i, y)$ - sub network in which the reliability of the reduced link is equal to 1 ($r_i=1$). Physically, it means that nodes connected by link *i* can be coupled into one node;

$(0_i, y)$ - sub network with reduced link reliability equal to 0 ($r_i=0$), means that link *i* can be omitted in further reliability calculations.

[1] Purely-looped system denotes the looped system with no series or parallel connected elements.

The reliability of the new subsystem has to be multiplied by the probability that link *i* is going to operate correctly r_i and its probability of failure q_i, respectively. The overall system reliability can be calculated as:

$$R = r_i \, \psi(1_i, r) + (1 - r_i) \, \psi(0_i, r)$$

where:

$\psi(1_i, r)$ - reliability structure function of $(1_i, y)$ subsystem

$\psi(0_i, r)$ - reliability structure function of $(0_i, y)$ system

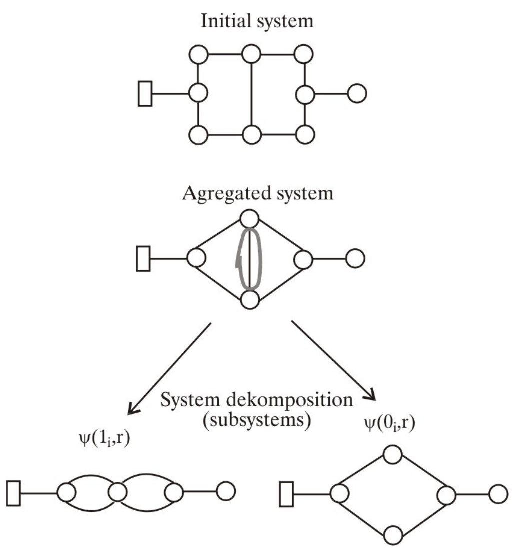


Figure 2. Principles of system aggregation and decomposition

If subsystems are purely-looped networks (i.e. the result of the aggregation model is not a tree network) then decomposition continues.

Generally, decomposition can be performed for any link in the network, but the reliability calculation will be faster if an optimal link is chosen. In this model the link for decomposition is found in two steps. Firstly, the search is performed on an aggregated network. The result of this search is the link that connects nodes in which maximal links joins. The chosen link is a path in the real (not aggregated) network. If there is at least one series connected link, it is chosen as a link for decomposition. If not, the first step is repeated. If there is more than one link with the same number of nodes repeated (in the first step), the one with the lowest link number in the path is chosen.

2.1.3 Reliability calculation

Network (or sub network) reliability can be calculated when it can be reduced to a tree of K-nodes by performing the aggregation model.

When single node reliability is calculated there is only one link connecting the source node and the demand node (specified as a K-node). Network reliability is equal to reduced link reliability. When overall network reliability is calculated the network can be aggregated to a tree network of K-nodes. Reliability can be calculated by multiplying link reliability, which is then multiplied by the reliability correction factor Ω.

Water distribution systems often rely on more than one source. That means there is more than one source node in the network. Such networks have to be modified when mechanical reliability is calculated. The problem is solved by adding an imaginary source node (Figure 3). That node is connected with real source nodes by imaginary links and has absolute reliability, its reliability is equal to 1. The new network is a network with one source node, whose reliability can be calculated as previously described.

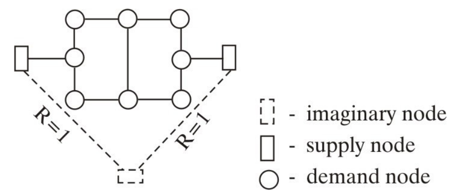


Figure 3. Imaginary node and imaginary links in network with more than one source node

2.2 Reliability of meeting hydraulic parameters

For water distribution systems connection to a source is not only necessary, but a sufficient condition to ensure that a given node is functional. That is why hydraulic calculation has to be included in determining mechanical-hydraulic reliability. Hydraulic calculation has to be performed for each subsystem for which the mechanical reliability is not equal to one (each node is connected to at least one source node).

Mechanical-hydraulic reliability can be calculated as:

$$R = \sum_{i=1}^{n} R_{meh}(S_i) \cdot R_{hyd}(S_i)$$

where n is the number of subsystems S.

If the chosen hydraulic parameters (there can be more than one parameter: pressure in nodes, velocity in links) are in specified boundary levels, the hydraulic reliability is equal to 1 (R_{hyd}=1), and mechanical-hydraulic reliability is equal to the mechanical reliability of the subsystem. If any parameter in any K-node (or link) of the system is not in desired levels, the probability of meeting the hydraulic parameters is not fulfilled, and the hydraulic reliability is equal to 0 (R_{hyd}=0).

The number of hydraulic calculations decreases if the hydraulic reliability of (0_i, y) subsystem is equal to 0. For such subsystems further decompositions will certainly form hydraulically unreliable subsystems.

3. RELIABILITY BASED OPTIMIZATION MODEL

Many authors have treated the problem of reliability allocation and optimization, but most of the attention to this issue has been given to the redundancy allocation problem. In this approach the minimum set of elements will be estimated, using genetic algorithms, in order to achieve specified system reliability with minimum cost.

The optimal design of water distribution network may be formulated as follows: for a given set of pipes and set of specified demand patterns at the nodes, find the combination of pipes which gives the minimum cost subjected to the constraint that the system reliability is better than the specified one.

The optimization model consists of several steps, usual for genetic algorithms:

1. An initial population of coded strings is generated randomly. Each string represents one network solution. The number of genes is equal to number of pipes included in optimization. Each gene is represented by one bit, which takes value 1 or 0, where 1 denotes that the pipe is included in network, and 0 denotes that the pipe is not included in the network.

2. The individuals in the current population are decoded. The result is a set of different networks. For each, network reliability is calculated using the previously described reliability evaluation model (including an additional presumption for very complex networks: only two pipes can fail at the same time). The fitness function is defined as the sum of investments in the system and a penalty function, included only for networks with reliability less than the defined one, and is calculated for each network.

3. A new population is generated using the selection operator. Individuals are selected according to their fitness, applying some of many selection procedures currently in use.

4. Crossover and mutation, genetically -inspired operators, introduce new individuals into the population.

5. Finally the replacement schemes are used to determine how the new individuals will be assimilated into the population.

Calculation procedure ends when value of fitness function is satisfied or some predefined number of iterations is achieved.

4. CASE STUDY

The proposed reliability evaluation and optimization methods are demonstrated on a hypothetical example of water distribution system. It is a system with single source node, one pumping station, 2 storage tanks, 16 demand nodes (K-nodes) and 37 links (Figure 4). This network is taken from Wagner et al, 1988, and is often used for hydraulic and reliability calculation, as it contains all the significant elements of a water distribution system.

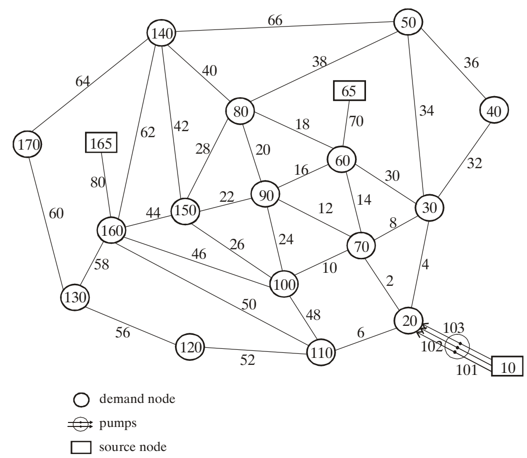


Figure 4. Case study network

The source (node 10) is at a low elevation, so water is pumped uphill from the river to downtown nodes. The pumping station is composed of three units in parallel (101, 102, 103). Pumps are assumed to fail 8 times a year, with mean repair time of 52 hours. It gives the reliability of each pump equal to 0.9543.

Demand nodes 30 - 110 are in the downtown zone, at relative elevation of 15.24 m. Nodes 120 - 170 are in a new part of town, at 36.58 m. Links are assumed to fail 0.62 times in a year on 1 km, and mean repair time is 3 days (72 hours). Specific reliability of 1 km long pipe is R_0=0.99492.

Besides the source node, there are two water tanks in the system (nodes 65 and 165). Water from these nodes can be released into the system, so they are defined as source nodes for reliability calculation. Characteristics of those nodes are same, with the same minimal and maximal water levels (Z_{min}=68.6 m, Z_{max}=76.2 m). Service head in the network is p=27.5 m, while the minimal required head in demand nodes is p=13.75 m.

Different reliability measures were calculated and analyzed for the described network. As a first step, the mechanical reliabilities for each demand node and for the overall system were calculated, using the NetRel model. Reliability of almost all nodes is close to 1. Such results were expected, as almost all nodes are connected with 4-5 links. Exceptions are nodes 40, 120 and 170, which are connected only with 2 links. That is why their reliability is slightly smaller comparing to the other nodes. The mechanical reliability of the entire network is naturally the smallest reliability, and presents the lower limit (R = 0.999611).

A result of the gathered reliability measures, and hydraulic calculations indicates that tanks are the critical elements for water supply in the system. When tanks are full, only one correctly working pump is required in the system for all nodes to be completely hydraulically satisfied (pressure in all nodes will be greater than the defined service head).

Results of mechanical-hydraulic reliability are used to calculate the average duration of water deficits in nodes and in entire the network (Figure 5). It is assumed that water tanks are full in analyzed time.

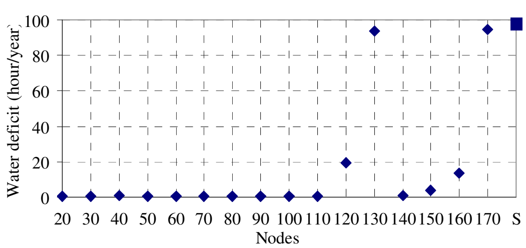


Figure 5. Nodes and network water deficits

In the case when water level in tank 65 is at a minimal level (a tank cannot be determined as a supply node and due to these circumstances cannot release water into a system), the pressure in nodes 130 and 170 will not be satisfied to service pressure level, even if all the pumps are operating. But, water deficit is not going to occur, as pressures in the aforementioned nodes will be greater than minimal network pressure (defined as 25 m).

Another scenario is when water in tank 165 is at a minimal level, tank 65 is full and all pumps are in service. Pressures in nodes of zone II would be at the minimal level (demanded water quantities would not be satisfied in those nodes). So, the critical element in the analyzed network is tank 165, on whose operation nodes of zone II depends.

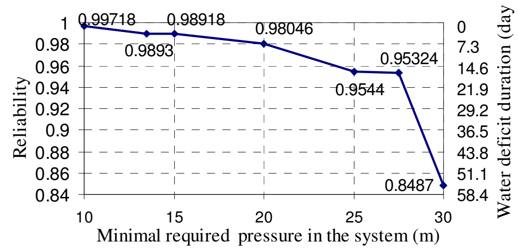


Figure 6. Network reliability and water deficit duration related to minimal required pressure in the system

Results of mechanical-hydraulic network reliability for different minimal pressures in the system are presented at Figure 6. It is assumed that tanks are full during the calculation. These reliabilities are lower than values of mechanical reliability, but they are relatively high, as a result of network connectivity and tank locations. The average duration of water deficit (expressed in days/year) is presented at Figure 6.

Node 170 of the analyzed network is the critical one, from the reliability point of view. This could be expected, as that node is in zone II and is connected to a network only through two links. In the cases when the connection to tank 165 fails, it is hydraulically not possible to supply this node from any other supply node. This is similar to node 130, with one difference - that node is connected to a network through three links. The water deficit for these two nodes is approximately 4 days/year, which is close to network reliability.

Reliability of node 120, although in zone II and connected to the network with two links (like node 170) is much higher. Its water deficit is nearly three times smaller (about one day in a year). The reason is the fact that node 120 is at the boundary between zone I and II. When the connection to tank 165 (source node) fails, node 120 can be supplied from other sources.

When tanks 65 and 165 are half full (water level is 72.9 m) network reliability is a bit smaller, with a value of 0.988362. Water deficit duration is approximately 102 hours/year. That is only a few hours in a year more than in the case when tanks are full. In both cases the system fails when connection to the tank 165 fails (link 80 fails), even if all other links in the network are operating. It is well known that the greatest reliability reductions are caused when a break in one link causes the network failure.

The optimization method is applied on presented water distribution system (for minimum allowable total head specified at each node as 27.5 m). Reliability of the system is 0.95324. The optimization problem is defined as follows:

- main objective: increase system reliability over 0.98: $R \geq R^{*} = 0.98$
- criteria: $f=C+C_K \rightarrow$ min, minimum of function defined as sum of additional investments in system (C) and penalty function (C_K), included only for networks with reliability les than the defined one $C_K=(R^{*}-R)\omega_K$
- variables: activities performed to increase system reliability. Two measures are considered: new pipes included in system (2 pipes are predicted connecting nodes 120-160 and 160-170) and pipes revitalization - increasing Hazen-Williams roughness coefficient from 70 to 125 (14 pipes)

On the base of the results of previously performed optimizations, methods based on steady state genetic algorithms were used. Analyzed methods differ in recombination type (1-point and 2-point crossover) and fitness function (according to rank and the real value of criteria function).

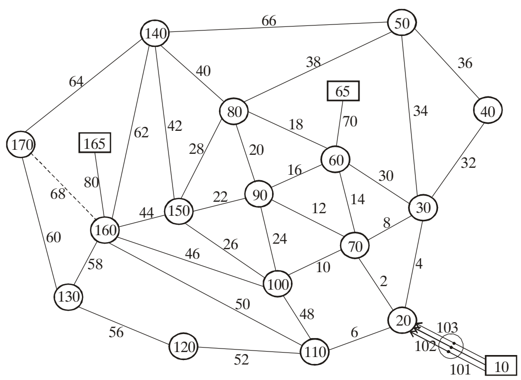


Figure 7. Optimal network solution

Obtained results indicate a fast convergence to an optimal solution: including one new pipe, pipe 68 at Figure 7. The solution, for all the examined models, is the global optimum (it was verified using enumeration method). The reliability of a new system is R = 0.98018.

5. CONCLUSIONS

The proposed model (NetRel) for calculating reliability of water distribution systems enables the determination of different reliability parameters for systems of different complexity. Mechanical and gathered mechanical-hydraulic reliability can be calculated for a single network node as well as for a whole network (probability of sufficient supply of all demand nodes in network). It can also be easily used for determining the reliability of other technical systems.

Furthermore, the reliability evaluation model has been incorporated in an optimization model (based on evolutionary programming). This model can be used for water distribution system optimization (for planning new systems or the reconstruction of an existing system). The optimization objective is to increase system reliability, with minimum investment in the system. The model was analyzed for different networks (different optimization problems). In a number of cases the results were very close or equal to the global optimums of the analyzed optimization problem.

6. REFERENCES

Beim, G. and B. Hobbs, Analytical simulation of water system capacity reliability, 2. A Markov chain approach and verification of the models, *Water Resources Research*, Vol. 24, No. 9, 1988.

Dandy, G., A. Simpson, L. Murphy, An improved genetic algorithm for pipe network optimization, *Water Resources Research*, Vol. 32, No. 2, 1996.

Dhillon, B. S. and C. Singh, Engineering reliability, new techniques and applications, A Wiley-Interscience, 1981.

Dasic, T and B. Djordjevic, Incorporation of reliability and maintenance proposals in hydrotechnical system design, *Vodoprivreda*, No. 169-170 (5-6/1997), Belgrade, 1997.

Gargano, R. and D. Pianese, Reliability as tool for hydraulic network planning, *Journal of Hydraulic Engineering*, Vol. 126, No. 5, 2000.

Goulter, I. C. and F. Bouchart, Reliability-constrained pipe network model, *Journal of Hydraulic Engineering*, Vol.116, No. 2, 1990.

Nahman, J. M., Dependability of engineering systems - modeling and evaluation, Springer, 2002.

Wagner, J., U. Shamir, D. Marks, Water distribution reliability: analytical methods, Journal of *Water Resources Planning and Management*, Vol. 114, 1988.

Whitely, D., A genetic algorithm tutorial, CSU, 1993.

Modelling the Dynamics of Public Attention towards Environmental Issues

Jens Newig, Julia Hesselmann,

Institute of Environmental Systems Research, University of Osnabrück, Germany. E-mail: jens.newig@usf.uni-osnabrueck.de

Abstract: Public attention, traced over time, often displays seemingly paradox behaviour: Contrary to what one might expect, public concern is seldom highest when the environmental conditions are the worst. Rather, concern most often rises when conditions have already become better. Furthermore, public attention dynamics frequently shows self-reinforcing behaviour resulting in distinct cycles, e.g. due to some trigger event. In this paper, public attention is viewed as a social macro-level phenomenon that is sought to be explained by the interaction of a multitude of single actors, namely individual citizens, the press, and politicians. First, a causal model based on rational choice theory is constructed in order to elucidate the mechanisms according to which public attention dynamics develop and to address the question of when and why public attention rises and falls. Key variables include the acuteness and visibility of the issue at stake as well as the ability to "solve" the underlying problem. Self-amplifying behaviour of agent interaction on different time scales adds to the complexity of the model. In a second step, an agent-based computer model is constructed from the conceptual rational choice model. It allows to reproduce the basic features of typical issue-attention cycles such as those analysed in empirical case studies. The model elucidates the causal mechanism and clearly displays the emergent structure, i.e. the typical, complex patterns of attention cycles. It thus serves to test and validate the conceptual model. In addition, it allows to incorporate additional conceptual refinements such as simultaneously tracking attention towards multiple issues. However, further research is needed to elucidate the mechanism(s) according to which public attention declines.

***Keywords**:* Agent-based modelling; rational choice, issue-attention, media coverage, public attention

1. INTRODUCTION

Different environmental problems receive different and varying degrees of public attention. Some issues are high on the public agenda when environmental conditions are most severe, while others start to be of high public concern only when the problem has already been halfway solved. Certain issues display both characteristics. Summer smog, for instance, is an issue that becomes "popular" each summer season when air concentrations of tropospherical ozone become high due to solar radiation. On a long-term track, however, quite a different picture could be observed (cf. figure 1): In the 1990s, when the issue of summer smog first entered the public arena and year by year became more important, ozone peak values had already begun to gradually decrease due to fewer emissions of ozone precursors by industry and transportation. What is more, public attention dynamics frequently shows self-reinforcing behaviour resulting in distinctive patterns. Thus, some issues might suddenly come up, often due to some trigger event, rise sharply – no longer in relation to the actual development of the environmental conditions –, and wane from the public agenda as quickly as they appeared.

Public attention being of chief relevance for political decision-making (cf. Mahon and Waddock 1992: 22), much research – albeit rather selective – has been conducted and many a theory has been put forward to describe and explain dynamics of public attention.

In this paper, a conceptual model based on previous research by the author will be presented, tested against empirical data and then implemented in the form of an agent-based computer model in order to verify the causal model and compare the model output with empirical data and thus test it on plau-

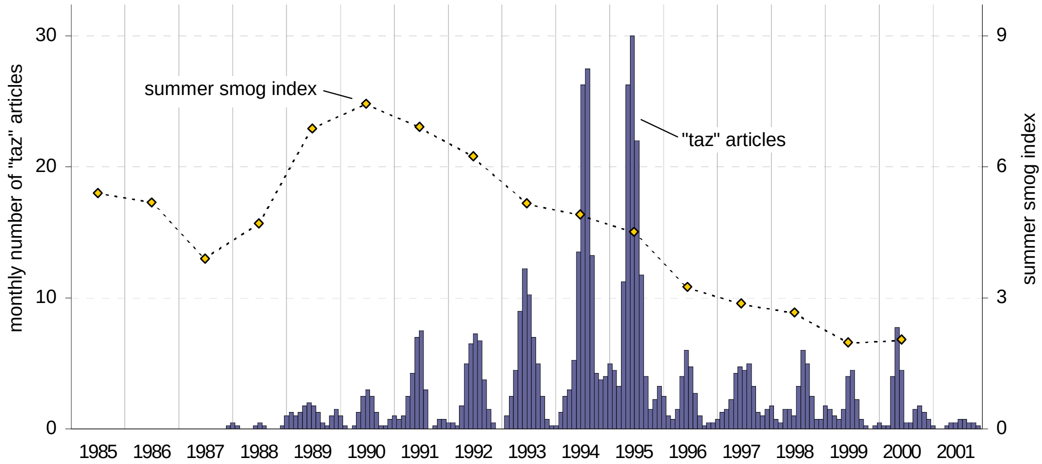


Figure 1: Public attention towards the "summer smog" issue, measured by monthly numbers of press articles[1], and an annual summer smog index. For both time series, 3-period centred moving averages have been used to even out random fluctuations.

sibility and to gain further insights into the proposed mechanisms. Central research questions to be tackled are: (1) When and why does public attention towards environmental issues rise and fall? (2) By what causal mechanism(s) can the dynamics of public attention be modelled?

2. CONCEPTS AND TERMINOLOGY

In order to avoid confusion with differing results in the literature, the fundamental concepts of "attention", "public", and "issue" shall be briefly discussed.

Attention denotes the resources (time and other) that people dedicate towards an issue and often signifies considerable political pressure. Regarded over time, attention can be conceived as an intensity (resource employment per time unit). Attention is also a scarce resource many issues are competing for (cf. Hilgartner and Bosk 1988; Zhu 1992). In contrast, "attitude" is commonly defined as "a learned predisposition to respond in a consistently favorable or unfavorable manner with respect to a given subject" (Fishbein and Ajzen 1975: 5), "opinion" denoting a verbalized attitude (cf. Zimbardo, Ebbesen and Maslach 1977: 20). Both do not necessarily involve resources.

In our model, we distinguish *public* from political, where the first refers to citizens (as potential voters), special-interest groups and the mass media, whereas the latter refers chiefly refers to politicians, parties and governmental officers. Furthermore, public attention as a sociological construct certainly involves an element of communication. If a multitude of private individuals devoted their attention towards an issue but there was no communication about it, this would not be public attention.

In today's democracies, public attention, as we conjecture, tends to form around *issues*. They focus on relatively small-scale problems like summer smog, BSE, ozone depletion, waste incineration and so forth that are sufficiently distinct to still be publicly perceived as units (cf. Dunlap and Jones 2002: 485–486). This is no longer the case with extensive problem areas such as air pollution or waste management, which merely serve as *categories* comprising and thereby classifying the multitude of different issues.

3. CONCEPTUAL MODEL

3.1 Theoretical assumptions

Some authors have sought to model public attention dynamics in a purely statistical manner by correlating aggregate measures of public attention with external variables or by regression analysis of attention time series (Henry and Gordon 2001; Soroka 2002). The disadvantage is that explanations remain on the aggregate level. Here, we opt for a different approach. Starting from individual actors, we seek to model the emergent phenomenon of public attention by the interaction of multiple individual actions. More abstractly speaking, our aim is to explain the social macro-level phenomenon by social micro-level processes (cf. Coleman 1990: 13–18). We base our theoretical model on the following general assumptions:

- Methodological individualism, i.e. actors (individual or collective) are the basic elements of analysis (not social systems).

[1] The German left-wing newspaper "taz" has been selected for good and early availability of electronic data. Although not representative in terms of its political content and audience, the "taz" follows basically the same attention cycles as all German newspapers (cf. Newig 2003).

- utility/interest-maximization under conditions of limited resources and cognitive capacity ("bounded rationality", cf. Simon 1972).

In order to keep our theory as simple as possible (and thus attain a high degree of generality, applicability and relevance), we distinguish only two different types of relevant actors whose interests and preferences, resources and restrictions will be portrayed in the following. In an earlier version of the conceptual model, we included four different types, adding politicians and interest groups to the ones presented here (cf. Newig forthcoming). For the sake of simplicity, we left out all interactions with political action in this first version of the computer model.

Citizens are generally interested in maintaining or improving the environmental conditions which affect them. As rational individuals, they dedicate their time and attention preferably towards those issues which they have an interest in. At the same time, however, citizens try to minimize the costs – monetary as well as cognitive – involved in finding out about how different political actors will deal with a specific environmental issue, which often requires understanding the environmental problem and the effect of proposed measures. Avoiding these (mental) costs, most citizens remain "rationally ignorant" (Downs 1957) towards most issues.

The *mass media*, considered as media enterprises free of any outward political influence, seek to maximize print runs or viewer levels. Therefore, they tend to cover the issues they believe the audience is interested in or concerned with. The publisher will discover whether or not a newspaper, for example, meets the expectations of its readers by monitoring print runs ("voting at the kiosk") and letters to the editor. Following this economic rationale, the mass media in our theory do not pursue any genuine *political* interests in the issue at stake.

3.2 Central causal mechanisms[2]

Possibly the most simple approach to issue-attention regarding environmental problems is the *problem-reaction model* employed in political science (cf. v. Prittwitz 1990: 103). According to this, public attention directly depends on the severity of the environmental condition (cf. figure 2): The higher the costs caused by the environmental problem (in terms of shortcomings in health or quality of life, or of the costs required to rectify these), the greater the stakeholders' discontent with the situation (cf. Opp 1996: 361–363, 368) and their interest in embarking on measures to improve the deficient situation and thus the greater the attention of all stakeholders towards the issue. This approach may be particularly applicable for threatening catastrophic events that are directly perceptible by the broad public and consequently lead to an immediate response in public attention. It particularly applies to environmental problems that have a direct effect on human health.

By contrast, the *capacity model* derived from social psychology tries to ascribe public attention to existing capacities for action, i.e. to resources for solving the problem at stake. The basic proposition is that deteriorations in environmental conditions remain unperceived unless they are or become technically solvable at economically viable costs. This is explained by the theory of cognitive dissonance (Festinger 1957). Accordingly, people generally strive to even out discrepancies between different perceptions – i.e. cognitive dissonances – or do not permit them to enter conscious reasoning in the first place (v. Prittwitz 1990). Environmental conditions that are significantly worse than the level of aspiration may constitute such discrepancies. Depending on the available resources, they can be resolved in different manners: When adequate resources to deal effectively with the problem are lacking, the aspirational level may be adjusted, or information regarding the actual state of the problem may be ignored or believed to be untrue in order to avoid cognitive dissonance and psychological stress. If, on the other hand, sufficient options for action are perceived, then actors will seek to implement measures to improve environmental quality, thus contributing to a rising public attention (cf. Opp 1996: 363).

In addition to these static mechanisms (both models hold independently of the temporal, historical development), public attention dynamics generally involves processes of self-organization as well (cf. Downs 1972: 38).

A completely new issue will at the beginning have difficulty to rise in public attention – even when the environmental conditions are quite severe and problem-solving resources are at hand –, because the mental costs of understanding the issue are often quite high, and the time needed to pay attention to an issue is scarce (cf. Zhu 1992; Neumann 1990). Every individual (or collective) actor can only receive a limited amount of information per time unit. In order to receive public attention, an environmental problem issue must be publicly perceptible and/or sufficiently simple to comprehend (visibility of the problem). Those already aware of the problem must both be willing and able to spread their knowledge and insight about the subject. In modelling issue-attention, this "stage" is the most sensitive. It is difficult to predict if and when the "critical mass" of theme-promoters is reached; "random" factors may be essential in determining whether or not and when an issue-

[2] For a detailed empirical test of basic model propositions cf. Newig forthcoming.

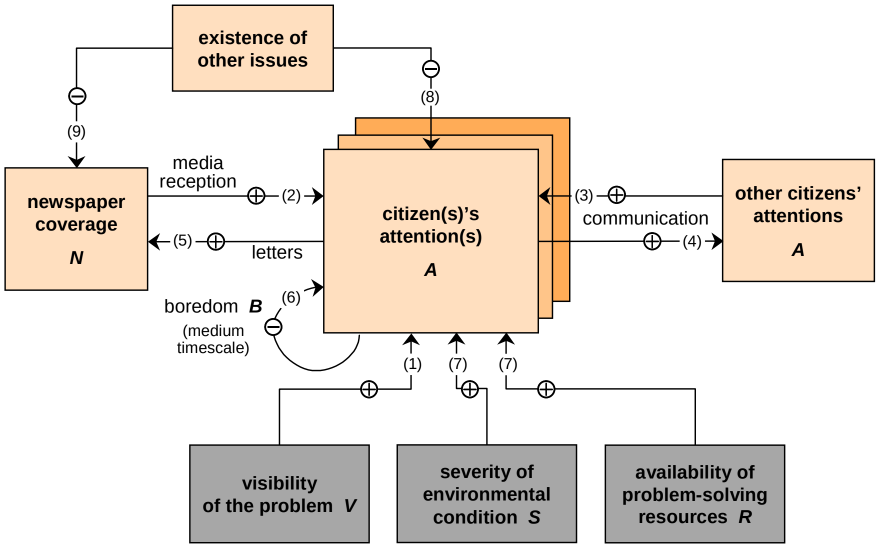


Figure 2: Elements of the causal model. Plus signs indicate positive, minus signs indicate negative feed-back. Numbers in brackets refer to the equations governing the computer model.

attention cycle is triggered. In such a non-linear dynamic system, small changes in initial conditions may lead to great changes in the resulting trend.

Once a critical mass is reached, the issue becomes a "fast-selling item" and attention towards it grows in an exponential manner, involving both interpersonal and mass media communication: The more individuals dedicate their attention to the issue and discuss it with others, the more individuals will be informed and possibly feel concerned about it. Thus, information costs decrease as more and more information becomes readily available, so it becomes rational for individuals to devote their attention to this issue rather than to others. An important factor in amplifying public attention are the mass media who – as rational agents – take up those issues they anticipate people to be interested in and in turn facilitate communication and information about the issue. The media thus act as *catalysts:* they increase the speed of news dissemination, but do not lead the discussion down a particular road[3]. Mathematically, this mechanism of self-amplification could be modelled in terms of diffusion theory (cf. Krampe 1989; Rogers 1995; Dearing and Rogers 1996).

Regarding the further "fate" of the issue-attention cycle, two different mechanisms exist: One is the "zero-sum game" of public attention (cf. Zhu 1992). Hence, when other issues gain in public attention, one issue may decrease simply due to increasing competition. The other, more important, mechanism acknowledges the fact that citizens get "fed up" and "bored" with an issue after having spent much attention on it for a long period of time (cf. Neumann 1990): Marginal benefits decrease, spending further attention on the issue provides less and less new information. This mechanism clearly introduces a negative feedback effect (cf. figure 2) on a medium time scale. The model thus becomes path-dependent, for it depends on the amount of time an issue has already received high public attention, whether it will continue to rise or start to decline. A third possible mechanism may consist in political action being taken (cf. Newig forthcoming) – this task will be left for further improvements of the model.

4. AGENT-BASED COMPUTER MODEL

4.1 Model structure

On the basis of the above considerations a computer model has been implemented in Java, using the social simulation environment Quicksilver[4]. The model represents a community of 20 citizens c and a newspaper company interacting with one another on a daily timescale. A number of n different issues i are modelled simultaneously. Every citizen has a fixed number of friends that are chosen randomly at the beginning of each model run.

For each issue, every citizen has a level of basic interest I_0 that depends on the issue's visibility V: Values of I_0 are randomly assigned between 0 and 1 according to a Gaussian distribution with mean V

[3] As much research as has been done to clarify whether the media influence the citizens or vice versa, it seems now that both are too closely intertwined to be able to decide on this: "In sum, there is considerable evidence that the direction of causality in the media-public relationship cannot be assumed" (Soroka 2002: 10). This may differ of course, from case to case.

[4] For more information, see http://java4u.sourceforge.net/ and http://www.usf.uos.de/projects/quicksilver/.

and standard deviation 0.1. They influence both the citizens' actual interests in the course of the model run $I_{i,c}$ and a level of tolerance towards the environmental condition regarding the same issue

$$T_{i,c} = 1 - I_{0,i,c} \,. \qquad (1)$$

The citizens' attention, defined on the interval [0,1], is determined in each time step t according to

$$A_{i,c,t+1} = A_{i,c,t} + A_{i,c,t}(\kappa C_{i,c} + \nu N_i - \beta B_{i,c})I_{i,c} \qquad (2)$$

with Greek letters indicating – throughout the paper – model parameters that weight the influencing factors according to their (relative) importance.

$C_{i,c}$ denotes a citizen's intensity of communication about an issue, given by

$$C_{i,c,t} = (\sum_{\tau=1}^{7} \frac{1}{\tau})^{-1} \sum_{\tau=1}^{7} \frac{1}{\tau} F_{i,c,t+1-\tau} \qquad (3)$$

with $F_{i,c,t}$ being the number of friends an actor talks to on day t, that depends on the actual $A_{i,c}$:

$$F_{i,c,t} = 0.6\, A_{i,c,t} + 0.4\, random[0..1] \,. \qquad (4)$$

Thus, the attention a citizen devotes to an issue depends on the intensity of communication not only of the present day but also – to a lessening extent – on the communication within the last week.

The intensity of newspaper coverage

$$N_{i,t+1} = N_{i,t} + \lambda N_{i,t}(\sum_{\tau=1}^{7} \frac{1}{\tau})^{-1} \sum_{\tau=1}^{7} \frac{1}{\tau} L_{i,c,t+1-\tau} \,. \qquad (5)$$

depends on the number of letters to the editor $L_{i,c,t}$ within the previous seven days. Whether or not a citizen writes to the editor is depends on whether his/her $A_{i,c,t}$ is greater than a randomly generated number between 0 and 1.

Boredom of an issue

$$B_{i,c,t} = (\sum_{\tau=1}^{365} \frac{1}{\tau})^{-1} \sum_{\tau=1}^{365} \frac{1}{\tau} A_{i,c,t+1-\tau} \qquad (6)$$

grows the more attention a citizen has already devoted to it, the "memory" lasting for one year with decreasing intensity.

The interest in an issue, which is initially given by the basic interest I_0, is altered according to the severity of the environmental problem S_i, the availability of problem-solving resources R_i and the individual level of tolerance versus a particular environmental problem T_i:

$$I_{i,c,t+1} = I_{i,c,t} + \sigma(S_{i,t} - T_{i,c}) + \rho S_{i,t} R_{i,t} \,. \qquad (7)$$

Both A and N are conceived as a "zero-sum" game. I.e., in each time step, after all other actions have been taken, both variables are normalized to the effect that the sums of all $A_{i,c}$ and all N_i, respectively, become equal to 1:

$$A_{i,c} = A_{i,c}(\sum_{j=1}^{n} A_{j,c})^{-1} \,; \qquad (8)$$

$$N_i = N_i(\sum_{j=1}^{n} N_j)^{-1} \,. \qquad (9)$$

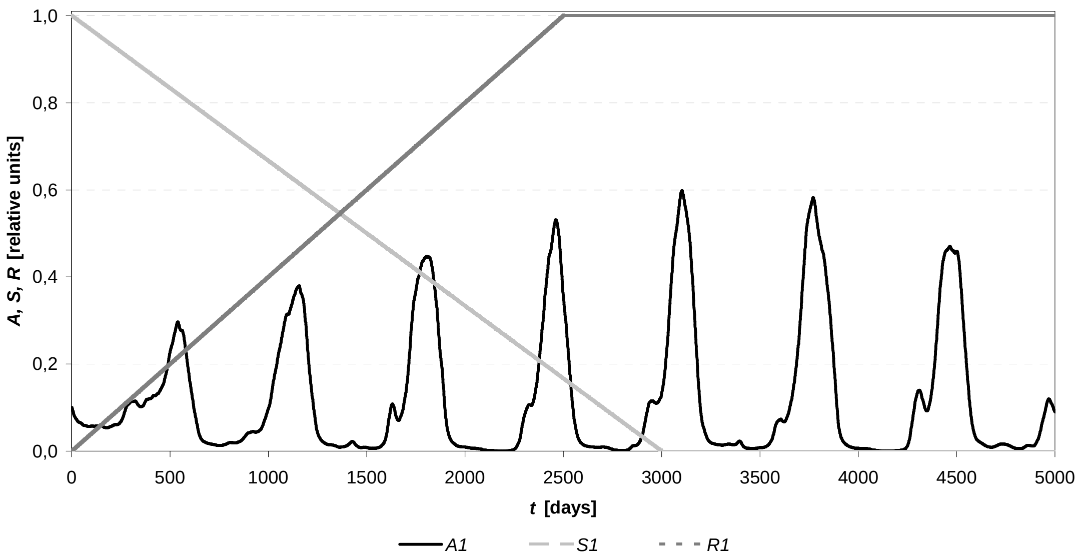


Figure 3: Example of model output for one of five issues, showing mean citizens' attention $\bar{A}_1$, severity S_1 and resources R_1 for this issue. κ = .35; ν = .2; β = .75; λ = .2; σ = .1; ρ = .1. Start values for all citizens' A_i: .1, .15, .15, .35 and .25, and N_i: .05, .1, .2, .4 and .25. V_i are .3 for issues 2 to 5, V_1 = .8. S_i and R_i are constant at .5 for issues 2 to 5.

4.2 Model results and discussion

The computer model being still in the test and validation phase, some preliminary results may already be shown. Figure 3 provides an example of a typical model output where n = 5 issues were tracked. It represents a scenario of rising problem-solving resources and a decreasing severity of the environmental situation, as often encountered in reality. The cyclical, yet not completely ordered time series of aggregated public attention[5] reflects well the non-linear, partly self-organising system of actors that is due to the positive feed-back mechanisms among citizens and the press, and the negative feed-back mechanisms of boredom and issue competition. However, public attention seems also susceptible to external variables: Both factors that positively influence attention – i.e., S_1 and R_1, – develop in an antagonistic way, with maximum positive influence expected at the intersection of both curves (cf. equation 7). Interestingly, the model output shows a time lag in the dynamics of the attention peaks, showing the largest attention cycle two to three periods after highest external influence. Only after some time, the lessening external stimulus becomes apparent in the attention dynamics.

5. CONCLUSIONS AND OUTLOOK FOR FURTHER RESEARCH

The proposed agent-based model, although still in its infancy, already reproduces basic features of empirical issue-attention time series, including (1) complex dynamics due to self-organisation of agents and agent communities, (2) the influence of external stimuli, i.e. the severity of the environmental condition and the availability of problem-solving resources, and (3) competition among issues. It thus supports the theoretical basis of the conceptual model.

It should be stressed, however, that the model in its present state of development does not yet allow to reproduce empirical time series of public (media or citizen) attention towards particular issues. Nor may the model parameters be attributed a specific meaning. Rather, it is expected that a further, systematic analysis of the model behaviour will yield relevant information regarding the (relative) importance of the model parameters and thus of the various influencing factors. Only then may the model be validated for specific historical issues. Moreover, an analysis of the actor network may result in deeper insights of the self-amplifying mechanism.

Possible improvements of the model structure will be to enlarge the number of issues in order to dampen the competition effect of each single issue, to allow new issues to appear in the course of a model run, and to refine the “boredom” mechanism which is still poorly understood on a conceptual level.

Thus, further research is needed to empirically test (and perhaps change) specific model assumptions, such as the boredom mechanism, the zero-sum game effect or the way in which external variables effect both media and citizens’ attention.

[5] The model output for aggregated citizens’ attention and for media coverage is typically very similar.

6. REFERENCES

Dearing, J.W. and E.M. Rogers. 1996. *Agenda-setting, Communication Concepts*. Thousand Oaks, Calif.: Sage.

Downs, A. 1957. *An Economic Theory of Democracy*.

Dunlap, R.E. and R.E. Jones. 2002. Environmental Concern: Conceptual and Measurement Issues. In *Handbook of Environmental Sociology*, eds. R.E. Dunlap and W. Michelson. Westport, Connecticut; London: Greenwood Press.

Festinger, L. 1957. *A theory of cognitive dissonance*.

Fishbein, M. and I. Ajzen. 1975. Belief, attitude, intention and behavior : an introduction to theory and research. Reading, Mass.: Addison-Wesley.

Henry, G.T. and C.S. Gordon. 2001. 'Tracking issue attention. Specifying the dynamics of the public agenda.' *Public Opinion Quarterly* 65:157-77.

Hilgartner, S. and C.L. Bosk. 1988. 'The rise and fall of social problems: A public arenas model.' *American Journal of Sociology* 94 (1):53-78.

Krampe, G. 1989. Ein Früherkennungssystem auf der Basis von Diffusionsfunktionen als Element des strategischen Marketing. In *Strategisches Marketing*, eds. H. Raffée and K.-P. Wiedmann. Stuttgart: Poeschel.

Mahon, J.F. and S.A. Waddock. 1992. 'Strategic issues management: An integration of issue life cycle perspectives.' *Business and society* 33 (1):98-101.

Neumann, W.R. 1990. 'The Threshold of Public Attention.' *Public Opinion Quarterly* 54:156-76.

Newig, J. 2003. *Symbolische Umweltgesetzgebung. Rechtssoziologische Untersuchungen am Beispiel des Ozongesetzes, des Kreislaufwirtschaft- und Abfallgesetzes sowie der Großfeuerungsanlagenverordnung*. Vol. 84, *Schriften zur Rechtssoziologie und Rechtstatsachenforschung*. Berlin: Duncker & Humblot.

Newig, J. forthcoming. 'Public Attention, Political Action - the Example of Environmental Regulation.' *accepted for publication in 'Rationality and Society'*.

Opp, K.-D. 1996. Aufstieg und Niedergang der Ökologiebewegung in der Bundesrepublik. In *Umweltsoziologie*, eds. A. Diekmann and C.C. Jaeger. Opladen: Westdeutscher Verlag.

Prittwitz, V.v. 1990. *Das Katastrophenparadox. Elemente einer Theorie der Umweltpolitik*.

Rogers, E.M. 1995. *Diffusion of innovations*. 4th ed. New York: Free Press.

Simon, H.A. 1972. Theories of Bounded Rationality. In *Decision and Organization: A Volume in Honor of Jakob Marschak*, eds. C.B. McGuire and R. Badner.

Soroka, S.N. 2002. *Agenda-setting dynamics in Canada*. Vancouver: UBC Press.

Zhu, J.-H. 1992. 'Issue competition and attention distraction: A zero-sum theory of agenda-setting.' *Journalism Quarterly* 69 (4):825-36.

Zimbardo, P.G., E.B. Ebbesen and C. Maslach. 1977. *Influencing attitudes and changing behavior: an introduction to method, theory and applications of social control and personal power*. 2nd ed, *Topics in social psychology*. Reading, Mass.: Addison-Wesley.

Modeling Flexible Plans for Agricultural Production Management

R. Martin-Clouaire and J.-P. Rellier

Unité de Biométrie et Intelligence Artificielle, INRA, BP27 Auzeville, 31326 Castanet Tolosan, France.

Abstract: Analyzing in what circumstance and why an agricultural production system performs acceptably well or fails requires models of management practices and work processes. A farmer's management task relies on planned and reactive behaviors that enable him to organize his work in function of known and exploitable regularities, and to adapt it to uncontrollable contingencies as they occur. Consequently, a farm production manager exhibits a decision-making behavior that seems to rely heavily on a kind of flexible plan. This paper aims at presenting a generic model of such plans that are similar to programs. Once represented in this framework a production management plan can be simulated in various exogenous conditions, which enables the study of the underlying production management behavior.

Keywords: Farm production management; Plan; Activity; Time; Event; Simulation

1. INTRODUCTION

Farm management consists in making and implementing the decisions involved in organizing and controlling a farm enterprise toward an objective that integrates socio-economic and environmental concerns. This paper focuses on production management that deals with the farmers' active role of manipulating the underlying biophysical system (crops and/or livestock) through technical operations. A farm manager must have beforehand an idea of how he intends to get where he wants his production system to move. The global management behavior can be seen as the result of the dynamic interpretation of his management strategy.

A strategy [Martin-Clouaire and Rellier, 2003] specifies in a flexible manner the plan that organizes the activities in time, the constraints that have to be satisfied in order to make them executable, the adjustments of the plan when particular events occur, and the preferences used to select the activities to be executed. Plan revision and execution must be interleaved because the external environment changes dynamically beyond the control of the farmers (due to weather influence in particular) and because relevant aspects are revealed incrementally. The commitment to particular actions must be delayed until run-time conditions are known. In particular, what can be executed is strongly constrained by the availability of resources and state-dependent requirements on the operations suggested by the plan. This paper focuses on the computational framework that supports the representation and simulation of such plans.

Because the nature of this mental object in the mind of managers is still largely unknown, we only aspire at providing the declarative means to specify the various types of constraints bearing on the activities involved in the production process, and making such plans dynamically interpretable by a mechanism that simulates the management behavior. Due to evolving and unpredictable circumstances the plans must be flexible with respect to the constraints that express what should be done, when or under what conditions in relation to the other activities and state of the production system. The commitment to executing particular activities is delayed until run-time conditions are known.

This paper is devoted to the presentation of:

- the modeling of flexible plans as a set of activities constrained by operators that play the role of control constructs used to specify sequential and concurrency composition, iteration, optional execution, choice between alternatives and wrapping of activities;
- the algorithm in charge of the runtime interpretation of the plan seen as a program.

2. MODELING PLANS

2. 1 Activities and primitive activities

The basic structure in a plan is the concept of activity. In its simplest form, an activity, which is then called a primitive activity, specifies something to be done on a particular biophysical object or location (e.g. a mob, a plant, a field or a set of these) by a performer (e.g. a worker, a robot or a set of these). Besides these three components, a primitive activity is characterized by local opening and closing conditions, defined by time windows and/or predicates referring to the biophysical state. These conditions are of use to determine at any time the activities that are eligible for execution consideration. For this purpose any activity has a status taking value in the set: *sleeping*, *waiting*, *open*, *closed* and *cancelled* (explained later).

The something-to-be-done component of a primitive activity is an intended transformation called an operation (e.g. the harvesting operation). The execution of an operation causes changes to the biophysical system, also called its effect, which are expressed in the form of values given to some variables of the biophysical systems. These changes are usually not instantaneous and take place progressively during the period of execution. In order to have the effect realized consistently with its definition the operation must satisfy some enabling conditions that refer to the current state of the biophysical system (e.g. the field to be processed should not be too muddy).

Primitive activities can be further constrained by adding temporal relations between them (sequencing, concurrency) and by using programming constructs enabling specification of choice of one activity among several, iteration, grouping and optional execution. To this end, a set of composition operators are used, the most important of which are: *before*, *meet*, *overlap*, *co-start*, *equal*, *or*, *and*, *iterate*, and *optional* (see next subsections). Any activity involving a composition operator is said to be non-primitive; a composition operator applied to an activity (primitive or not) defines another activity that may also be given local opening and closing conditions. A non-primitive activity is called the mother activity and the activities that are the arguments of the operator are called the child activities. The opening and closing of a non-primitive activity depends on its own local opening and closing conditions (if any), and of those of the underlying primitive activities that play a role through the composition operators. All the activities are connected; the only activity that does not have a mother is the plan. The plan is flexible in the sense that two different sequences of events are likely to yield two different realizations of the plan. The opening date of the same activity will not be the same in the two cases. Moreover some activities may be cancelled in one case and not in the other if they are optional or subject to context-dependent choices.

The passing of time and the evolution of the state of the production system may make true the conditions that govern the changing of status of the primitive activities. The change of status of activities is realized at particular times specified by the manager and when an operation is completed. Any change of status of an activity is propagated to the activities that are directly or indirectly connected to it via composition operators.

The meaning of the possible values of an activity status can now be explained. The value *sleeping* is given to all activities at creation time. It means that the opening and closing conditions do not have to be examined yet. The status turns to *waiting* as soon as the opening activities have to be examined. For instance, as soon as an activity finishes it becomes necessary to monitor those following it in a sequence specified with a *before* operator. The nominal plan is declared *waiting* at the starting time of a simulation. The status of an activity turns to *open* when its opening conditions are satisfied. The status changes from *open* to *closed* when the closing conditions are satisfied or, in case of primitive activity, when the underlying operation is completed. The status turns to *cancelled* when the activity becomes of no interest; this happens, for instance, once a choice among alternatives specified through the *or* operator has been made, making *cancelled* the non-selected alternatives.

The meaning of each operator used to construct à new activity by constraining other activities is defined by two sets of rules specifying:

- the preconditions that must be satisfied by the mother activity in order to enable the change of status of some of the child activity and vise versa;
- the post-conditions or effects of any change of status of one of the mother or child activities on the others.

Each type of activities constructed with such operators are visited in turn in the next subsections.

2.2 Sequencing constraints

To specify that two or more than two activities must be performed successively without any overlapping in the interval of time of their execution one can use the *before* operator and apply it to the child activities given in the order of the sequence desired. In other words, the activity *before(A B)* imposes that the activity B cannot have the status *open* before the status of A is *closed*. The order in time of the sequence is expressed by the order of the arguments of the operator. Any activity constructed using the *before* operator has two extra properties that enable specification of, if necessary, the delays between the opening of two consecutive activities, and between the closing of one of them and the opening of the next one.

The change of status of any of the involved activities is subject to the following preconditions.
In order for the mother activity status to become:
- *waiting* (resp. *open*), its first child must be allowed to turn to *waiting* (resp. *open*);
- *closed*, its last child must be allowed to turn to *closed*.

In order for the first child to become:
- *waiting* (resp. *open*), the mother must be allowed to turn to *waiting* (resp. *open*).

In order for any other child than the first one to become:
- *waiting*, the preceding activity must be *closed* or allowed to turn to *closed*.

In order for last child to become:
- *closed*, the mother must be allowed to turn to *closed*.

The effect of a change of status of any (mother or child) activity follows the following rules.
As soon as the mother turns to:
- *waiting* (resp. *open*), the first child turns to *waiting* (resp. *open*);
- *closed*, the last child turns to *closed*.

As soon as a child activity turns to:
- *waiting* and if it is the first child then, the mother turns to *waiting*. Otherwise, the preceding child tuns to *closed* (if not already so);
- *open* and if it is the first child then, the mother turns to *open*;
- *closed* and if it is the last child then, the mother turns to *closed*. Otherwise, the next child turns to *waiting* if possible.

Another operator used to specify a sequence is *meet*. It is very similar to *before* except that there should be no delay between the closing of a child and the opening of the next one. The only modifications on the preconditions and effects to a change of status of any of the involved activities are the following.
In order for any child but the first one to become:
- *waiting* or *open*, the preceding activity must be *open* and allowed to turn to *closed*.

In order for any child but the last one to become:
- *closed*, the next activity must be allowed to turn to *open*.

The effect of a change of status of any (mother or child) activity follows the following rules.
As soon as a child turns to:
- *open* and if it is the first child then, the mother turns to *open*. Otherwise, the preceding activity turns to *closed*.
- *closed* and if it is the last child then, the mother turns to *closed*. Otherwise, the next child turns to *open*.

Actually a *meet* activity behaves like a *before* activity involving no delay.

2. 3 Concurrency constraints

Several operators enable to specify that some activities have to remain open concurrently for some time. For instance, using the *overlap* operator, one can express that the interval of time in which the child activities (two or more) have the status *open* must intersect and the order of the opening of the children is also the order of their closing. Therefore this operator defines a ranking over the children; this is expressed through the order of its arguments. The mother activity constructed using the *overlap* operator has three extra properties that enable specification of, if necessary, the delays between the opening of two consecutive (wrt. the order of the arguments) activities, between the opening of a child and the closing of the preceding one, and between their closing.

The rules stating the preconditions are the following.
In order for the mother to become:
- *waiting* (resp. *open* or *closed*), its first child must be allowed to turn to *waiting* (resp. *open* or *closed*).

In order for the first child to become:
- *waiting* (resp. *open*), the mother must be allowed to turn to *waiting* (resp. *open*).

In order for any child other than the first one to become:
- *waiting*, the preceding activity must be *closed* or allowed to turn to *closed*;
- *closed*, the preceding child must be *closed*.

In order for the last child to become:
- *closed*, it must be allowed to turn to *closed*.

In order for any child but the last one to become:
- *closed*, the next activity must be *open*.

The effect of a change of status of any (mother or child) activity follows the following rules.
As soon as the mother turns to:
- *waiting*, the first child turns to *waiting*;
- *open*, the first child turns to *open* and the next turns to *waiting* if possible.
- *closed*, the last child turns to *closed*.

As soon as a child turns to:
- *waiting* and if it is the first child then, the mother turns to *waiting*. Otherwise, the preceding child turns to *open* if it is *waiting*;
- *open* and if it is the first child then, the mother turns to *open*. If it is not the last child, the next one turns to *waiting* if possible;
- *closed* and if it is the last child then, the mother turns to *closed*.

Other operators enable the specification of other kinds of concurrency. For instance, the operator *inclusion* can be used to constrain the intervals of time in which the children are open to be nested (each fitting within the one immediately larger).

The operator *co-start* (resp. *co-end*) imposes the simultaneous opening (resp. closing) of the children. The operator *equal* imposes that the children be open simultaneously and closed simultaneously.

2. 4 Iteration

The operator *iterate*, which has a single argument activity, specifies that the child activity be repeated within the time in which the mother activity is *open*. The mother must be given opening and closing conditions and the child or descendant activities should not appear elsewhere in the plan. The mother constructed using the *iterate* operator has two extra properties that enable specification of, if necessary, the delays between the opening of two consecutive iteration of the child, and between the closing of the child and the opening of its next iteration. The minimum and maximum numbers of iterations may also be specified. The only preconditions to a change of status of the child are that the mother be *waiting* or *open* in order for the child to turn to *waiting*, and that the mother be *open* in order for the child to turn to *open* or *closed*.

Concerning the effects, as soon as the mother activity turns to:
- *open*, the child turns to *waiting* if possible;
- *closed*, the child turns to *open* and the next turns to *waiting* if possible.

As soon as the child turns to *closed*, it turns to *waiting* unless the closing conditions of the mother are satisfied at that time.

The iteration process, which is controlled by a specific procedure, duplicates (instantiates in fact) the child activity as needed in agreement with the constraints of delay between repetitions and of limitations of the number of iterations if provided. These copies have a status changing from *sleeping*, to *waiting*, from *waiting* to *open*, from *open* to *closed*, and, exclusively for this case, from *closed* to *waiting*. These transitions continue as long as the mother is *open*.

2. 5 Optional activity

The *optional* operator applied to an activity expresses that if this one cannot be realized (i.e. it is too late with respect to the closing interval or the closing predicate cannot be satisfied) then, it is not a sufficient circumstance to declare the plan invalid. In other words, this operator enables specification of the child activity that should be realized if possible. The child or descendant activities should not appear elsewhere in the plan if not declared optional there too. The status of the mother can change to *waiting* if the child can turn to *waiting*. Analogous preconditions hold when substituting *waiting* by *open* or by *closed* and by permuting child and mother. The effects rules follow from the precondition rules (e.g. the child becomes *open* as soon as the mother becomes *open*).

When a mother activity made with the *optional* operator cannot be realized it status is forced to turn to *closed*.

2. 6 Disjunction and conjunction

The *or* operator enables specification of a possibility of choice between the child activities. When one of them is chosen (this can only be done after resource allocation which is not addressed in this paper) the others are turned to *cancelled* and therefore can no longer be considered for execution. The rules stating the preconditions are the following.

In order for the mother to become:
- *waiting* (resp. *open*), there must be at least one child that can be turned to *waiting* (resp. *open*);
- *closed*, all the children must be *closed* or *cancelled* or allowed to turn to *closed*.

In order for any child to become:
- *waiting* (resp. *open* or *closed)*, this child must be allowed to turn to *waiting* (resp. *open* or *closed)*.

The effect of a change of status follows the following rules.

As soon as the mother turns to:
- *waiting* (resp. *open*), all the children that are allowed to turn to *waiting* (resp. *open*) do so;
- *closed*, the only child that is still *open* turns to *closed*.

As soon as any child turns to:
- *waiting* (resp. *open*), the mother turns to *waiting* (resp. *open*).

As soon as the child turns to:
- *closed* (the other children being *cancelled* at this moment), the mother turns to *closed*.

The *and* operator enables to specify that the set of activities constituting the child activities should be realized so that the mother activity can end up *closed*. This operator plays the role of a wrapper. The rules stating the preconditions to a change of status are the following.

In order for the mother activity status to become:
- *waiting* (resp. *open*), there must be at least one child that can be turned to *waiting* (resp. *open*);
- *closed*, all the children must be *closed* or allowed to turn to *closed*.

In order for any child to become:
- *waiting* (resp. *open)*, this child must be allowed to turn to *waiting* (resp. *open)*.

The effect of a change of status follows the following rules.

As soon as the mother turns to:
- *waiting* (resp. *open*), all the children that are allowed to turn to *waiting* (resp. *open*) do so;

- *closed*, the children that are still *open* turn to *closed*.

As soon as any child turns to:

- *waiting* (resp. *open*), the mother turns to *waiting* (resp. *open*);
- *closed*, the mother turns to *closed* if all children are *closed* or can turn to *clos*ed.

3. UPDATING THE ACTIVITIES

3. 1 Algorithm

The advance of time and the evolution of the production system (the biophysical system in particular) may make true the opening and closing conditions of the activities. The updating of the status of the activities occurs at either examination times specified by the manager (typically at discontinuity points induced by new day or new week) or when an operation is terminated. The change of status is realized by a procedure that essentially checks that the opening and/or closing conditions can be satisfied and that the constraints linking this activity to others would be satisfied if the change proceeded. This procedure, applied to the plan, causes a recursive examination of all the activities that are not *sleeping*, *closed* or *cancelled*. Any activity whose change of status is validated is updated and the change is propagated immediately to the connected activities.

Normally the status updating process is repeatedly invoked until the plan is closed. In some cases, the plan cannot be closed, which reveals a plan failure. Such an inconsistency situation occurs when some preconditions to change cannot be satisfied (e.g. a *meet* activity in which the second child cannot be open although the first has just been closed). In other words, this happens when an activity that is not optional can no longer be open or when it cannot be closed without violating constraints that link them to other activities by composition operators. A more formal presentation of this updating process is given through the pseudo-code of the main procedures.

```
procedure: Update(activity)
    if      activity.situation not waiting and
            activity.situation not open
    then    return
    if      {activity.situation = waiting and
            it is no longer possible to open} or
            { activity.type = primitive and
            activity.situation = open and
            opening time is over and
            operation is not yet executing}
    then    if  activity.type = optional
            then TurnToClosed(activity); return
            else  exit("Plan failure")
    if      activity.situation = open and
            it is no longer possible to close
    then exit("Plan failure")
    switch activity.type
            case primitive
                if      ?OpeningValid(activity)
                then    TurnToOpen(activity)
            case iteration
                if      ?OpeningValid(activity)
                then    TurnToOpen(activity)
                if      situation = open
                then switch child.situation
                        case sleeping
                            TurnToClosed(activity)
                        case waiting
                            if  ?ClosingValid(activity)
                            then TurnToClosed(activity)
            case others
                for     each child do Update(child)
```

Two important predicates are used in Update: ?OpeningValid, ?ClosingValid. They return true if it is legal to open or close the argument activity. They call the two activity-dependent predicates ?CheckSonsIfOpen and ?CheckIfSonOpen. The latter, together with ?CheckSonsIfWaiting, ?CheckIfSonWaiting, ?CheckSonsIfClosed, and ?CheckIfSonClosed, implement the preconditions to changes defined for each composition operator. They themselves call ?OpeningValid, ?ClosingValid and ?WaitingValid. These three predicates are very similar in principle. The pseudo-code of ?OpeningValid is given below. For clarity, this code does not include all the bookkeeping structures and tests necessary to avoid loops.

```
predicate: ?OpeningValid(activity)
  if  activity.situation = open  then  return  true
  else
     if  {activity.situation = waiting or ?WaitingValid(activity)}
         and local opening conditions satisfied
    then
       if ?CheckSonsIfOpen(activity)
       then   for each mother do
              if  not ?CkeckIfSonOpen(activity, mother)
              then      return false
       else  return false
     else return false
```

Note that the predicates ?OpeningValid, ?ClosingValid and ?WaitingValid are also used in the operator-dependent procedures that implement the effect of a change of status of an activity.

Update calls the procedures TurnToOpen and TurnToClosed. Together with TurnToWaiting each of these procedures realizes the due changes of status of the argument activity and propagates the effect to the connected activities. Once they are called (either by Update or at the beginning of the simulation when the plan status is forced to change from *sleeping* to *waiting*) they perform all the required changes in the plan according to the operator-dependent rules.

3. 2 Example

An application of the concepts and mechanisms defined in the above sections has been used to describe glasshouse production system for tomatoes by Jeannequin et al. [2003]. For illustration, we consider here a highly simplified management plan that is actually only a part of a real one in this domain; this part should normally be considered with the other parts at the same time because they are likely to interact. The plan is the following:

before(iterate(PRUNING1), iterate(optional(PRUNING2)))

It expresses that two series of pruning activities have to be done successively and the pruning activities in the second series are optional. Both PRUNING1 and PRUNING2 are primitive activities that consist in applying a Prune operation to the plants of a particular glasshouse compartment. This operation removes young fruits from the most recent truss so as to leave only a limited number of them and prevent small sized fruit. The above two activities differ only by the resources that they require: the first one needs one worker of a particular type (e.g. highly qualified) whereas the second one needs one too but of another type (e.g. temporal labor). We assume that w1and w2 are workers of the first and second type respectively. w1 is available from day 0 to day 30 whereas w2, is hired from day 30 to the end of the season and might nevertheless be unavailable from time to time at random due to other duties. We assume he might be off for 6 consecutive days every 2 weeks (15 days) but he must stay at least five days when he comes back to his glasshouse job. The area of the glasshouse compartment is equal to 10 units and the pruning speed of a worker is 2 units per day.

The temporal specifications in the various activities are expressed on a daily scale. It is assumed that the plan itself (i.e. the before activity) has opening and closing windows equal to [0, 60] and [60,60] respectively. The opening window of the first pruning activity in the first series is [0, 5]. When a pruning activity is open at time t the opening window of the potential next iteration in the series is set to [t+10, t+15]. Any pruning activity has a closing predicate that forbids its closing later than 10 days after the execution of the underlying operation has started. The two arguments of the before activity have [0, ∞] as opening windows; their closing window are [30, 60] and [60, ∞] respectively. Finally the before activity is specified that the opening and closing windows of the potential first iteration of the second series is set to [t+10, t+15] where t is the opening date of the last iteration in the first series. Since the availability of w2 is stochastic the outcome of running the plan is stochastic too. One of the possible realisations is considered next.

The first series involves three pruning activities that are opened as soon as possible with respect to the delay constraints (at days 0, 10 and 20 respectively). They are never interrupted by resource unavailability so the execution of the operation always extend over 5 consecutive days. The first pruning activity in the second series behaves similarly for the same reason. At day 40 another pruning activity is opened but the operation cannot be performed because worker w2 is not available. Since w2 comes back only at day 46 and a prune operation cannot start executing later than 15 days after the opening of the pruning activity, this optional activity cannot be performed and is simply closed. The following candidate activity is opened at day 50 (i.e. 10 days after the previous opening). The prune operation is executed at days 50 and 51 when w2 is available. This is not enough to complete the activity, which resumes as soon as w2 is back at day 58. The operation ends at day 60, which complies with the delay requirement that the activity ends within 10 days after its beginning. As specified, the execution of the plan stops at the end of day 60.

4. CONCLUSION

The conceptual and computational model of plans presented in this paper has been developed for and inspired by production management problems in agriculture. It seems nonetheless relevant in any production process that involves a single manager and that highly depends on uncontrollable factors, thus requiring flexible management plans.

Besides the runtime interpretation of plan other important tasks such as resource allocation or strategy adjustment are involved in plan execution and have not been addressed in this paper although already implemented.

Future research effort will be devoted to the issue of preference processing including consideration of subsidiary goals and anticipation of likely future.

5. REFERENCES

Jeannequin B., R. Martin-Clouaire, M. Navarrete, and J.-P. Rellier, Modeling management strategies for greenhouse tomato production. CIOSTA-CIGRV Congress, Turin, Italy, Sept. 22-24, 506-513. 2003.

Martin-Clouaire R., and J.-P. Rellier, A conceptualization of farm management strategies. Proc. of EFITA-03 conference, Debrecen, Hungary, July 5-9, 719-726. 2003.

Fairness Principles In Allocating Water: Integrating Views Of Different Agents

B. E. Nancarrow[a] **and G. J. Syme**[b]

[a]*Australian Research Centre for Water in Society, CSIRO Land and Water, Australia*
[b]*CSIRO Land and Water, Australia*

ABSTRACT

The allocation, or re-allocation, of water to achieve environmental sustainability in farming communities can be a source of considerable conflict. "What's fair" when sharing water between different agents and the environment becomes paramount in people's decision making. Issues such as self interest; efficient uses of water; business investments; viable communities and prior rights to water all play a part in the ways people view the fairness of the allocation decision. The different agents will form their fairness rules in different ways. For example, the small family farmer will not want to be disadvantaged by the greater economic capacity of the corporate farmer. Those who have not used their allocations in the past may not consider it fair if they stand to lose their allocations, in the interests of achieving environmental sustainability, to those who have previously used their allocations and been a part of the over-allocation problem. What role can economic instruments play in achieving fair water allocation? When does the environment have greater rights over individual and societal needs? This paper draws on a number of case studies in rural Australia to see if allocation rules can be framed for the different community groups in deciding what's fair in water allocation. More importantly, we explore whether investigation of these rules can contribute to gaining community consensus and suggest that Agent Based Modelling may be a useful tool for assisting in this.

KEY WORDS: social justice; fairness; water allocation; environmental sustainability; consensus.

1.0 INTRODUCTION

Currently, in Australia, negotiations between community and government are continuing on issues related to ensuring environmental sustainability while allowing for economically and socially resilient rural communities. "Fair" ways of balancing the self-interests of different groups and agents and accommodating those of the environment have yet to be incorporated in policy formulation and decision making despite escalating community conflict. It is not enough for the governments to espouse the policy of "equitable allocation" if the determinants of equity are unclear. There is a need to understand how people interpret equity, justice and other principles when the outcomes of decision making affect them personally. This will enable better prediction of judgements about the "fairness" or otherwise of government decision making, as well as addressing the likely impacts on the affected communities.

In deciding which measure of justice most clearly reflected the ways in which people judged justice, equity or fairness, a number of preliminary studies were conducted [Syme, Nancarrow and McCreddin, 1999]. We (and other authors) found that there seemed to be a large degree of overlap and correlation between people's views on procedural and distributive justice and the role of equity considerations within these judgements [e.g. Folger, 1996]. For this reason the concept of the fairness heuristic [e.g. van den Bos et al, 1997], which allowed for this interaction, was adopted. We found as Peterson [1994, p99] suggests "once an impression of fairness has been produced it becomes extremely resistant to change ...... because it provides a cognitively available summary judgement. People use their summary fairness judgement in lieu of a more complicated analysis of policy each time they are asked".

Fairness in the Syme *et. al.* [1999] studies was measured at two levels *"universal"* and *"situational"*. In the *universal* sense, we developed lay philosophies, or the principles and values that people wanted to see articulated in general terms in water allocation policy. S*ituational fairness* judgments were made when respondents were asked to assess fairness of specific contexts, such as the development re-allocation policies in a particular river basin.

It was found that *universal* fairness judgements were remarkably similar over a variety of surface and groundwater allocation problems in a variety of states and with a diverse range of respondents over a number of years. On the other hand, specific *situational* fairness judgements varied between case studies, but were still able to be categorized under broad universal fairness principles. Furthermore, one latter study tended to suggest that consensus could be reached in a community on a fair way to proceed which catered not only for self-interest but also for the social well-being of the community. Given the significance of this finding it was important to replicate it across further studies in current and “real” water allocation conflicts.

In this paper we provide a summary of the findings of the developmental studies as well as the application of the theory developed to case studies where decision makers and water users were faced with the requirement to re-allocate water to provide for environmental sustainability. We further suggest where Agent Based Modelling (ABM) may have assisted in the decision making processes.

2.0 THE STUDIES

Australia is one of the driest continents on earth. As a result of a century of poor environmental practices; the regulation of rivers that reverse natural systems to provide water to irrigate agriculture in summer; increased populations and better scientific knowledge, it has become evident that the rivers and the groundwater have been over-allocated and in many cases over-used and current practices are not sustainable. Case studies abound in Australia where there are arguments as to how the water will be reallocated between competing consumptive (eg. irrigation; urban water supply; industry) and non-consumptive (eg. environment; recreation) uses. These have provided the opportunity for both the developmental and applied studies mentioned previously. Figure 1 shows the locations of these studies that have been conducted since the early 1990s.

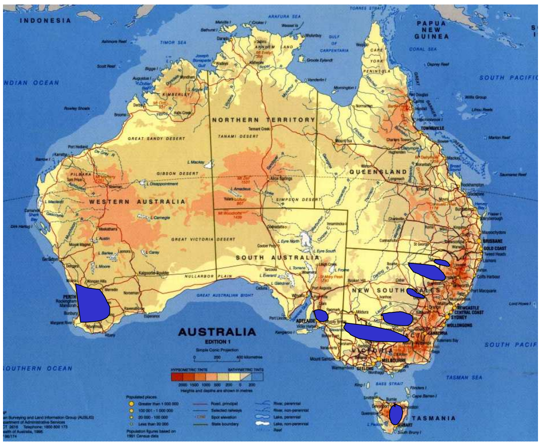


Figure 1: Map of Australia depicting the locations of the developmental and applied studies.

3.0 RESULTS OF UNIVERSAL FAIRNESS JUDGMENTS

Over a number of years and through a variety of qualitative and quantitative methods, a series of about thirty-five statements was developed that represented communities’ over-arching, lay philosophies of fairness when allocating water to multiple uses. These statements had been derived from Wenz’s [1988] review of the philosophy of environmental justice [see Syme and Nancarrow, 1996]. These included attitudes towards water as a common good, environmental rights, efficiency of use considerations, moral obligations between groups of water users, and economic rationalism through to virtue theory (those who already have the resource are inherently deserving) and different formulations of benefit/cost analysis. In addition, items relating to short and long term planning and procedural and distributive justice were embedded in them.

Measurement of these universal fairness principles through a questionnaire format on a five point agree/disagree Likert scale showed consistent results over more than 10 years and in a variety of water allocation decisions across Australia. It was shown that there is consistent strong (dis)agreement (>80% of study samples) on a number of universal principles such as:

- environmental management for future generations;
- the right of all sections of the community to have a say on how water will be allocated;
- the rights of the environment;
- the need for efficient water use;

and consistent disagreement on:

- prior rights or history of use determining future allocations; and
- the use of water trading markets to determine how water should be allocated.

However, it was also shown that these principles, although held at the universal level, may change given specific situations. For example, while prior rights to water may not be supported at the universal level of fairness, they can be supported in certain situations when the circumstances of those holding the rights are known.

A cluster analysis of the ratings of these universal fairness principles was conducted to determine whether there were identifiable, general philosophical stances taken by differing groups of people. This too resulted in consistent findings across multiple studies. Two identifiable groups constantly emerged from this analysis and a discriminant analysis was conducted to better understand the them. With minor queries, the groups in all studies could be labelled "public good" allocators and "private good" allocators. That is, one group was more inclined to a "social" emphasis on fair water allocation while the other was more "individually" focused on fairness. This is somewhat supportive of Rasinki's [1987] two factor model of equity. These two groups each accounted for approximately half the respondents in each study and their identification provided a mechanism for demonstrating whether the specific solutions to water allocation challenges at a local level were considered to be fair by people with alternative universal fairness judgements. Working towards this end would provide the fair solutions at a society level that were being advocated by planners.

4.0 RESULTS OF SITUATIONAL FAIRNESS JUDGMENTS

Discussions with the range of agents associated with the applied studies, revealed many and wide-ranging criteria that were being promoted to resolve their specific re-allocation problems. Some examples included:

- irrigators using approved water efficient irrigation and re-use systems should not have their allocations reduced;
- only those who have caused the problem should have their allocations cut;
- *all* licence holders should have allocations cut proportionally, according to the amount of water actually used (ie. small cut for small users, and larger cuts for larger users).
- reduce the annual irrigation season;
- irrigators using approved water efficient irrigation and re-use systems should not have their allocations reduced.

While these criteria at first seemed to be highly diverse, closer examination revealed that they could be categorised under five broad universal fairness themes. These were:

- Equality of opportunity
- Reward for hard work and investment
- Allocation through historical water use
- Allocation through water trading markets
- Promotion of water efficient management.

Of particular interest was the consistency across the studies in people's general acceptance of some of these themes as approaches for solving their specific water allocation challenges. Similarly there was consistency in the perceptions of the unacceptability of the other approaches. Tables 1 and 2 show the acceptability of some approaches as measured in the surveys and the percentages of respondents who considered other approaches to be totally unacceptable across four studies.

Table 1: Mean standardised ranking scores for acceptability of the five general approaches to solving the groundwater problems in four studies

Possible Approach	Mean Standardised Acceptability Rank			
	#1 N=287	#2 N=257	#3 N=65	#4 N=46
Efficiency & management	25.0	23.3	25.3	22.2
Equality of opportunity	24.9	25.2	25.7	23.0
Reward for hard work & investment	20.2	22.2	24.7	22.5
Historical use	16.2	14.5	16.4	17.8
Water Markets	13.7	14.8	13.8	14.5

Table 2: Percentage of respondents who considered the general approaches to problem solving to be *totally* unacceptable in four studies

Possible Approach	% of Total Sample			
	#1 N=287	#2 N=257	#3 N=65	#4 N=46
Historical use	35	19	14	22
Water markets	31	18	31	28
Reward for hard work & investment	17	9	12	2
Efficiency and Management	11	7	6	20
Equality of opportunity	9	7	11	15

Respondents were also asked to rate the acceptability of the individual situational criteria, grouped under the themes, for re-allocating the irrigation water. Again, the difference between universally held beliefs and situational circumstances was shown. While "history of use" was not supported as an acceptable approach to resolving the allocation conflict, in one of the studies, situational criteria under this approach were supported. Respondents' understanding of the particular circumstances allowed them to support the criteria.

Analyses were then carried out to ascertain the least number of situational criteria that could be used to develop a re-allocation solution that would ensure that the needs of everyone in the sample of respondents were included in at least one of the criterion. For example, one of the studies resulted in a "four item solution" that provided for 96% of the sample, as shown in Figure 2.

- *Any reductions in allocations should provide protection for the "family farm".*
- *A viability base should be set, and license holders with allocations less than that base should not receive cuts.*
- *Irrigators who are fully active, and have invested in their properties, should not have their allocations reduced by more than those who have little or no investment in their properties.*
- *Allocations should be reduced and the amount of water allowed to carry-over to the next year should be increased for a longer period of time.*

Only six percent of the survey sample did not have at least one acceptable criterion incorporated in the solution. On inspection, it was found that these six respondents had not answered the particular question in the questionnaire. As they could not be attributed to any particular demographic group, it was assumed that the solution was not marginalizing any specific group of agents..

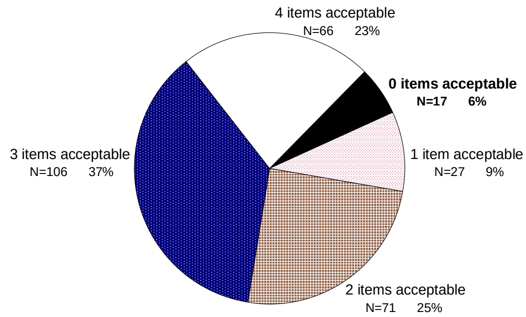


Figure 2: Percentages of respondents who nominated different numbers of situational criteria as acceptable in developing a solution to the water over-allocation problem.

As the four items show above, and in all of the studies, examination of the criteria that made up the situational solutions revealed that most respondents were promoting criteria that not only catered for their own self-interest, but also catered for other groups in the community. That is, people were making their decisions based on not only their own personal needs, but also the requirements for ensuring the ongoing viability of their communities.

5.0 WHAT NEXT?

The above results illustrate the complexity of people's fairness judgements when it comes to deciding how to allocate or re-allocate water to ensure environmental sustainability. People hold particular values at the universal level and these can carry through to specific situations in that people are very clear about acceptable and unacceptable management approaches. We identified two groups in society whose values in this regard are generally opposing.

We have also shown that people can provide a range of situational fairness criteria for developing solutions to the allocation problem, and these can be reduced to an operable number which still includes everyone in the solution.

We have also shown that rational choice plays little place in people's decision making in that they are as much concerned about the viability of the whole community's future as they are about their own.

In one of the studies, this point was reached and a way forward was recommended to the government department dealing with the problem. This entailed taking the four situational fairness criteria above and working with the community to develop a range of futures that incorporated these criteria and to come to a negotiated final decision on the fairest solution. This, we advised, would take about eighteen months, given the difficulty of the problem, the extent of conflict, and the potential hardship for many of the irrigators.

We were advised that, given the political agenda and the looming election for the state government, a solution was required in about four months and that they would have to find another way. We were dismissed.

It now seems to us that ABM may have provided the tool whereby this fairness process could have been completed and a negotiated solution achieved in the required time frame. Being social scientists, our understanding of the intricacies of ABM is limited, however, reading Kurz and Snowden [2003] reinforced our thoughts in this regard.

This article challenges the three basic assumptions in organizational decision support: those of order; rational choice and intent. Kurz and Snowden's (2003) discussion on the realities of contextual complexity and un-order epitomises much of what we have described above. A goal of their contextualized framework "…. to enable clients to achieve self-awareness rather than to provide "expert" advice" would be a highly valuable tool to use with Australian farmers who constantly provide derisive comments about the value of "experts" and the usefulness of their advice. In using their justice criteria as "rules" for evaluating different water sharing scenarios, people would be able to see their values influencing the decision making process. ABM could provide the vehicle to do this.

We assume there are other ABM and network frameworks that may be useful in this case, but this demonstrates that there may be some potential for a decision support system to finally be useful in the complex social arena. Achieving social justice in natural resources management is something that many government decision makers grapple with and the research described in this paper shows that it will not be easily understood by other than a social scientist. If ABM can provide an easy tool to assist lay people to achieve social justice, one could begin to expect its routine inclusion in NRM decision making and hence more cooperation by farmers in achieving environmental sustainability that is becoming more and more urgent worldwide.

6.0 REFERENCES

Folger, R., Distributive and procedural justice: multi-faceted meanings and interrelationships, *Social Justice Research*, *9*, 371-394, 1996.

Peterson, R.S., The role of values in predicting fairness judgements and the support of affirmative action, *Journal of Social Issues. 50* (4), 95-115, 1994.

Rasinski, K.A.., "What's fair or is it? Values differences underlying public views about social justice, *Journal of Personality and Social Psychology, 53,* 201-211, 1987.

Syme, G. J. and Nancarrow, B.E., Planning attitudes, lay philosophies and water allocation: A preliminary research agenda, *Water Resources Research, 32,* 1843- 1850, 1996.

Syme, G.J., Nancarrow, B.E. and McCreddin, J.A., Defining the components of fairness in the allocation of water to environmental and human uses, *Journal of Environmental Management*, 57, 51-70, 1999.

van den Bos, K., Lind, E.A., Vermunt R. and Wilke, H.A. M.. How do I judge my outcome when I do not know the outcome of others? The psychology of the fair process effect, *Journal of Personality and Social Psychology, 72,* 1034-1046, 1997.

Wenz, P.S., *Environmental Justice*. Albany: State University of New York Press, 1988.

Kurtz, C.F. and Snowden, D.J., The dynamics of strategy: Sense-making in a complex and complicated world, *IM Systems Journal, 42* (3) 462-483, 2003.

Heuristics to characterise human behaviour in agent based models

Claudia Pahl-Wostl and **Eva Ebenhöh**

Institute of Environmental Systems Research, University of Osnabrück, Germany

Abstract: Human behaviour is one of the key factors to understand the causes for common pool resource problems and to develop policies to promote more sustainable resource management regimes. Agent based models can help to investigate the role of important processes in this respect such as factors determining the degree of trust and cooperation in a group. We have chosen a pragmatic approach to represent human behaviour by assuming that agents can be characterised by a set of attributes and their behaviour can be described by a set of simple decision heuristics. Individual agents differ in their importance of attributes (e.g. fairness, cooperativeness, trust), in their rules how to choose a heuristic, and in their responses to social interactions. The assumptions are tested by using data from experimental economics describing the behaviour of players in simple games dealing with resource allocation. A set of specific attributes and heuristics was derived by analysing data from different games. The plausibility and generality of the behavioural model is tested by applying it to different data sets from different games. We expect from these simulations insights into behavioural patterns that determine processes of social learning and negotiation. The modelling approach will be applied and tested with data from case studies where actors make decisions in a real world context of dealing with a resource management problem.

Keywords: Heuristics, modelling human behaviour, experimental economics, cooperation, fairness

1. INTRODUCTION

As early as 1968 Hardin in his famous article entitled “The Tragedy of the Commons” described a situation where villagers were using a common field to graze their cattle. The commons tended to be overgrazed since each villager would graze to a point where the private costs equalled the benefits, and social costs were neglected. In general, such a situation applies to the problem of ‘common pool resources’. And the tragedy of the commons is a typical case of a ‘social dilemma’ where the maximization of the short-term self-interest of the individual, leaves all participants worse off than feasible alternatives. Each individual faces a trade-off between what is in his or her own short term interest and what is in the broader interest of the community in which he or she lives. A collective version of social dilemmas occurs frequently in the provision and management of public goods and may account for many environmental problems such as the overexploitation and pollution of water resources, arable land and the atmosphere. Hardin’s analysis suggested that the only solution to preventing such social dilemmas would be regulation of the commons by a central entity. This would argue in favour of governmental regulation and control as the most promising strategy for dealing with environmental problems and managing public goods. However, in her influential book, Elinor Ostrom (1990) provided evidence that Hardin’s analysis did not apply in general and that local communities have efficient ways of self-organizing and self-governance and may also prevent the degradation of resources on the base of voluntary cooperation. Hence, an alternative strategy would explore ways in which governmental intervention and the self-organizing capacity of communities interact and subtly reinforce themselves so as to develop more efficient and enduring resource management regimes. Such a strategy suggests as well a stronger role of participatory approaches in resources management to facilitate collective learning and choice processes. However, approaches to manage common pool resources are still hampered by a lack of understanding the nature of human behaviour.

The prisoners’ dilemma and game theoretical approaches have mainly been used to analyse human behaviour in the context of common pool resource problems (Gintis, 2000). The basic underlying assumption has been the profit-maximizing, rational homo economicus. Such approaches provide little potential for analysing the possibilities of cooperation and self-governance. Empirical and experimental evidence show considerable deviations from theoretical assumptions based on homo economicus. In particular, the emerging field of experimental economics has developed a set of games and has collected a rich data base on human action in a number of experimental situations where fairness and equity issues in the allocation of rewards play an important role. The potential for innovation from these insights has not been fully exploited yet. Whereas most approaches to explain behaviour have been based on analytical mathematical approaches and extensions of game theory, more recently agent based modelling has been used to analyse and explain data from economic games (e.g. Duffy, 2001; Deadman, 2000; Ebenhöh and Pahl-Wostl, 2004). Agent based simulation offers a major methodological breakthrough in the ability to investigate the role of different processes determining human behaviour in more detail. The method is also very flexible since it is not linked to a specific disciplinary paradigm (Pahl-Wostl, 2002a).

We have decided to capture essential elements of human behaviour in resource allocation problems by assuming that agent behaviour is guided by heuristics (Gigerenzer and Selten 2001) and agents learn from experience.

2. MODELLING APPROACH

In the chosen modelling approach decisions are not necessarily based on elaborate calculations, but on heuristics, including simple rules humans may follow in making their decisions. Those heuristics depend on different attributes characterizing individual agents. Another very important aspect is agent diversity. Agents differ not only in their individual attribute values, but also in the heuristics they use.

We use data from economic experiments to support our heuristics approach (see section 2.2). In experimental economics various experiments are conducted with human subjects, placed in an artificial, laboratory environment in which they have to solve tasks or play games with or against each other. Usually those games are played anonymously, in order to prevent the formation of a social environment. In these situations, decision making of the human subjects is supposed to be abstracted from a specific context and social interactions. This may be seen as a breach with respect to actual human behaviour in day-to-day situations. However, this method's main advantage is that it produces comparable and reproducible data. The set of attributes presented in the next section were derived from a number of experiments from different games during the process of developing agent based models that can explain observed behaviour.

2. 1 Attributes

So far, our set of attributes includes nine different characteristic traits. All attributes are represented as real numbers in the interval between 0 and 1, where 0 implies that this trait is not important for the agent and 1 implies that it is very important. In our model, the numbers are random numbers, equally distributed, and the attributes are independent from each other.

Cooperativeness: How important is group utility for an agent? A high cooperativeness indicates the willingness to spend individual resources in order to further group resources. Mainly this is associated with an increase in efficiency.

Fairness (concerning others): How important is it for an agent that the other agents get roughly as much as it? This is a purely comparative attribute that does not refer to efficiency increases. Agents with a high fairness are willing to spend money in order to equalize the outcome. Note that this fairness does not yet consider equity considerations in reward allocations.

Conformity: How important is it for an agent to appear to be as others expect it to be? Agents with a high conformity may play fairly because they feel they are expected to, and not because of their own high fairness.

Fairness concerning me: How important is it that the agent's payoff is roughly as high as the other agents' payoffs? Agents with a high fairness concerning me are easily annoyed at being treated unfairly. However, whether or not this annoyance leads to retaliating actions is defined by negative reciprocity.

Positive reciprocity: How important is it to return behaviour that is perceived as fair and cooperative, with fair behaviour? An agent with a high positive reciprocity feels committed to play cooperatively in a second move when the other agents have played cooperatively before.

Negative reciprocity: How much is an agent willing to pay in order to make another agent pay (more)? An agent with a high negative reciprocity feels committed to punish in a second move when the other agents have defected before.

Risk aversion: How risk averse is an agent? An agent with a high risk aversion will not invest anything in a project that yields a high but uncertain return.

Commitment: How important are the decision to be made and previously made agreements for an agent? A player with a high commitment to a project will invest in group utility even when its cooperativeness is low.

Trustworthiness: To which degree does an agent respond to trust placed in him or her by other agents with expected behaviour instead of being opportunistic?

In addition to all these attributes, agents hold expectations about the attribute values of other agents. They learn from observed behaviour about the others' attribute values. These are referred to as expected cooperativeness, expected fairness etc. Note that trust is modelled as expected trustworthiness.

In literature on experimental economics attributes like the ones described above, often appear without being strictly defined. For example, fairness is rarely differentiated into fairness concerning me and fairness concerning others, although the distinction is quite apparent. Likewise, positive and negative reciprocity, are often considered as a single attribute reciprocity ("strong reciprocity" in Fehr and Rockenbach, 2003). Also, names can vary substantially. For example, what we refer to as "fairness concerning me" is essentially the same as "annoyance", expressed by subjects in post-experimental questionnaires in an experiment by Fehr and Gächter, and "negative reciprocity" corresponds to "willingness to punish" in the same experiment (Fehr and Gächter, 2002). Recently Cox especially designed and conducted an experiment to discriminate between trusting, positively reciprocating, and altruistic, other-regarding behaviour (Cox, 2004). Cox uses trust and positive reciprocity quite similar to the attributes presented in this paper. However, we differentiate other-regarding behaviour further into cooperative and fair.

In this attributes approach lies a major difference to other work combining experiments with agent-based models (Duffy 2001, Deadman 2000).

2.2 Heuristics

Heuristics usually make sense only in a concrete decision environment. This is why the heuristics presented here are given as examples for a specific game. For our modelling approach, however, it is also important that the heuristics chosen are more generic to be applicable to a large range of situations in experimental games and empirical case studies. Therefore, the heuristics are kept as simple as possible.

As example for a decision environment consider the following two-player game, taken from an experiment by Fehr and Rockenbach (2003). Both players receive 10 money units (MU). The first player is asked to give any number of his or her money units to the second player. This gift is tripled by the experimenter. Then, the second player may return any number of MU from 0 to the tripled gift to the first player. This is not tripled. When the first player gave the gift, he or she is also asked to indicate, how much he or she would like to receive.

A main result of this experiment is, that most first players do place trust in second players and most second players reciprocate trust with returns greater than 0. First players gave 6.5 MU on average to second players. Second players returned more than 40 percent of the tripled investment to first players. The higher the gift of the first player, the higher was the return by the second player (Fehr and Rockenbach 2003, 138f.). However, some first players gave less than 5 MU (26.5%) and some second players kept everything to themselves (16%).

The first player's first decision may depend on cooperativeness, because of efficiency considerations. Risk aversion may play a role, because a gift of 10 money units may be returned doubled or even tripled, but the return might also be 0. Finally, the expected trustworthiness is an indicator for perceived risk and thus may also be important. The first player's second decision, how much he or she would like to receive, is probably only influenced by fairness concerning others and fairness concerning me.

Table 1: Some heuristics for the 1st player's decision:

A1	gift = 0 MU
A2	gift = 5 MU
A3	gift = 10 MU
A4	gift = cooperativeness * 10 MU
A5	gift = (1-risk aversion) * 10 MU
A6	gift = expected trustworthiness * 10 MU
A7	if (exp. trustworthiness < low limit) gift = 0 MU else if (exp. trustworthiness > high limit) gift = 10 MU else gift = expected trustworthiness * 10 MU
A8	if (expected trustworthiness > some limit AND risk aversion < another limit) gift = cooperativeness * 10 MU
A9	calculate expected return with expected fairness, expected fairness concerning me and expected positive reciprocity for different gifts and take gift with highest expected return.

Although the second player's task is very similar to the first player's first decision (deciding on an amount of money to give to the other player), the attributes needed are different. The second player's decision is not influenced by risk aversion and expected trustworthiness. However, it may be influenced by fairness considerations, both concerning others and concerning me, as well as positive reciprocity and trustworthiness.

Table 2: Some heuristics for the 2nd player's decision:	
B1	return = 0 MU
B2	return = gift
B3	return = 2*gift
B4	return = fairness * 2*gift
B5	return = (1-fairness concerning me) * 2*gift
B6	if (fairness < low limit) return = 0 else if (fairness > high limit) return = 2*gift else return = fairness * 2*gift
B7	if (trustworthiness < low limit) return = 0 else if (trustworthiness > high limit) return = asked return else return = fairness * 2*gift

2.3 Choosing Heuristics

Heuristics should be as simple as possible. In order to do that, we give certain agents certain heuristics according to their attributes, thus moving the case differentiation out of the heuristic (as in A7, B6, and B7). A typical way of doing so would be:

```
if (agents cooperativeness > high limit)
        use cooperative heuristic (A3)
else if (agents cooperativeness < low limit)
        use maximizing heuristic (A9)
else use default heuristic (A7)
```

By this, the agents' attributes determine their decision making behaviour in two ways. The chosen heuristic as well as the actual decision made by the chosen heuristic both depend on the attribute values (see figure 1).

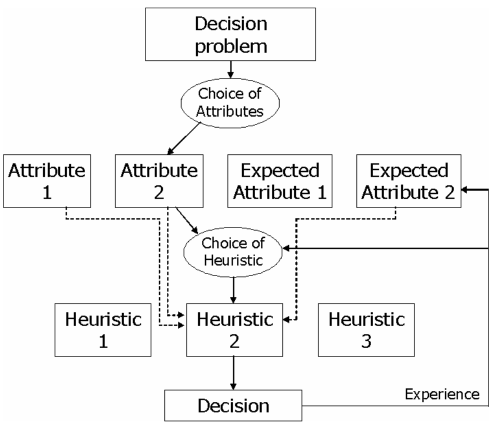


Figure 1: Role of attributes in the decision making process

2.4 Learning

Learning takes place in two different ways. The first and easier kind of learning is the adjustment of expected attributes to the experiences made by the agents. In our model, so far, the agents start with believing others to be exactly like themselves. If they make, for instance, cooperative experiences exceeding expectations the value of expected cooperativeness is increased. However, the adjustment is not necessarily exact, because the agents do not perceive the other agents' attributes directly, but only their decisions.

The second learning process affects heuristics. If an agent makes negative experiences using one heuristic, in some cases these experiences should lead to an exchange of that heuristic by another possible one. If, in the above case, an agent with a high cooperativeness uses *cooperative heuristic* and gets a return of 0, it might consider this heuristic to be inappropriate the next time a similar decision has to be made, although neither its own cooperativeness nor any expected attribute value has changed. In the example presented here, however, this kind of learning does not take place.

The attribute values themselves do not change over the time scale of the model simulations. Changes in attributes may occur over longer time scales, months or years.

3. TESTING ASSUMPTIONS WITH EXPERIMENTAL DATA

The example in the previous section shows that even with a simple decision task there may be a great number of possible heuristics that could explain human behaviour. In order to test model assumptions, we use data from experimental economics (Ebenhöh and Pahl-Wostl, 2004). Laboratory experiments provide us with a rich data base of individual human behaviour in simple controllable settings. By variations of the experimental settings, we can focus on different aspects of human behaviour. We base our choice of heuristics used in the model on the data and accompanying questionnaires. Furthermore, the process in which the agents choose between the heuristics is also derived from data. In order to be able to do this, we have to analyse individual and not only aggregated data. The representation of individual data is an important advantage of agent-based modelling compared to other modelling techniques (see also Duffy 2001, 309). However, it requires agent behaviour to be more psychologically plausible, than if only aggregated data were to be reproduced (Jager, Janssen, 2002, 99).

But even representation of individual data may not be sufficient to model the actual human behaviour. In the previous example, the game design does not allow us to distinguish trusting behaviour from genuinely cooperative behaviour. That is, we do not know if a first mover invested money, because he or she expected the other player to return a part of the tripled investment, or if the utility increase by tripling the investment was reason enough for him or her to give money away. Likewise we do not know, whether returned amounts are due to positive reciprocity or a sense of fairness (cf. Cox, 2004, p. 264 in a comment on a similar experiment by Berg, Dickhaut and McCabe, 1995). Reciprocity assumes that the player who trusts another player expects to trigger a social norm of fairness that is stronger than the possible desire to defect and maximize individual utility.

Such differences may be important to understand behaviour in real world settings. However, one question arises for all experimental approaches – how far can insights from such experiments be transferred to situations in real world settings?

4. PROSPECTS

The attribute approach presented in this paper is guided by empirical analysis and modelling practices rather than psychological theory. In psychology there are some "trait approaches" (Liebert and Spiegler, 1994) to explain human behaviour on the basis of dispositions, defined as "enduring, stable personality characteristic(s)" (Liebert and Spiegler, 1994, 156). For our modelling purpose, however, empirically tested traits, like the "big five" (neuroticism, extraversion, openness, agreeableness and conscientiousness) are too broad and general compared with the attributes chosen. In order to merge psychological theory with the model assumptions, one would have to conduct thorough empirical investigations on the correlation between the super traits and our attributes and how the super traits translate via attributes into observed behaviour.

To understand key behavioural phenomena as trust and cooperation requires an interdisciplinary approach in the social sciences combing insights from economics, psychology and sociology. Here data from experimental economics have their limitations. Social interactions are largely excluded. While experiments in economics often emphasize the generality of a situation and comprise monetary rewards and repeated trials, psychologists try to capture intrinsic motivations and the mental processes at work in a particular decision situation, what has been termed the *framing* of a decision problem. Hence it will be of major interest to test insights gained from experimental economics in real world settings where social interactions, context and framing play a major role.

The concepts developed in this modelling approach are currently tested in a number of case studies dealing with collective decision making processes and collaborative governance in the management of common pool resource problems. The typical methodological case study design includes participatory integrated assessment and modelling approaches. The decision context refers either to groups of stakeholders such as farmers engaging in collective action and cooperative governance of a common pool resource or to processes of social learning and negotiation in moderated group settings (Pahl-Wostl, 2002b).

Moderated group settings include an actors' platform with representatives from stakeholder groups who engage in processes of social learning and collective decision making over a period of 1-2 years. A typical platform is expected to involve the following sequence of steps (Pahl-Wostl, 2002b):

(1) Build up a shared problem perception in a group of actors, in particular when the problem is largely ill-defined (this does not imply consensus building).

(2) Build trust as base for a critical self-reflection, which implies recognition of individual mental frames and images and how they pertain to decision making.

(3) Recognize mutual dependencies and interactions in the actor network.

(4) Reflect on assumptions about the dynamics and cause-effect relationships in the system to be managed.

(5) Reflect on subjective valuation schemes.

(6) Engage in collective decision and learning processes (this may include the development of new management strategies and the introduction of new formal and informal rules and resource allocation schemes).

This sequence describes an idealized case. In reality quite a few obstacles may impede such cooperative learning and decision making processes. A crucial variable is the willingness to cooperate and trust between stakeholders participating in such a process (Panebianco and Pahl-Wostl, 2004).

In contrast to the experimental game settings the settings in case studies are determined by context and history. The actors know each other and hold expectations about the attributes of other individuals (in contrast to expectations about an average other agent in the game settings). They hold general expectations about expected degrees of cooperativeness and fairness that are determined by their prior experience and their cultural and social environment. At the same time the attributes of an individual actor are not assumed to be invariants but are assumed to be shaped in a long-term learning process determined by experience and the social and cultural environment. Hence one can conclude the social environment and the mutual expectations are partly socially constructed. And we can expect that attributes change over the time period of observation.

Role playing games have been used to generate trust in a stakeholder group and build a shared understanding of the problem (steps 1-3 in the above listed sequence). Such role playing games have been combined with agent based simulation models in an iterative fashion to elicit new insights about decision making strategies, and support the development of new strategies (e.g. Pahl-Wostl and Hare, in press, Barreteau et al 2001). The empirical investigations can be compared with the heuristics derived from experimental games.

Figure 2 shows an example of a bargaining routine that was elicited during a role playing game in a stakeholder platform, a situation where experimental results from dictator and ultimatum games can be quite useful (Ebenhöh and Pahl-Wostl, in review).

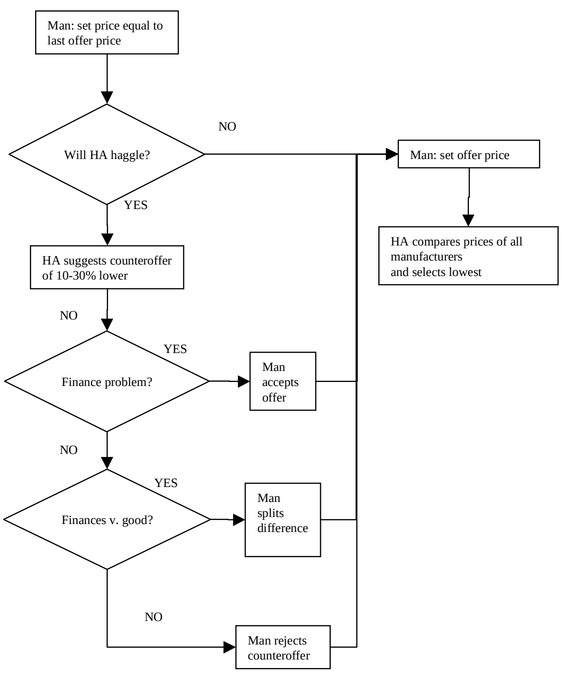


Figure 2 Bargaining algorithm between Housing Association and Manufacturer of Sanitary Technologies (Hare, Heeb and Pahl-Wostl, 2002)

Currently work is ongoing to test the applicability of the modelling approach described in section 3 in a number of case studies. Specific emphasis is given to:

- Applicability of attributes listed in section 3 to characterize actors in real world settings.
- Applicability of heuristics to characterize actor behaviour in real world settings.
- The dynamics of learning processes (individual attributes, expectations, and heuristics) over different time scales in real world settings.
- The importance of context and history in case studies for understanding human behaviour in contrast to the context free, mostly anonymous experimental game settings.

An improved understanding of the importance of trust and cooperation and the use of simple heuristics in learning and decision making processes will support the development of improved participatory approaches and decision support tools in the management of common pool resources.

5. REFERENCES

Barreteau, O., F. Bousquet, and J.-M. Attonaty, Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems. *Journal of Artificial Societies and Social Simulation* 4, 2001 <http://jasss.soc.surrey.ac.uk/4/2/5.html>

Berg, J., J. Dickhaut, and K. McCabe, Trust, reciprocity, and social history, *Games and Economic Behaviour*, Vol 10, 122-142, 1995

Cox, J. C. How to identify trust and reciprocity, *Games and Economic Behaviour*, Vol 46, 260-281, 2004

Deadman, P.J., E. Schlager, and R. Gimblett, Simulating Common Pool Resource Management Experi-ments with Adaptive Agents Employing Alternate Communication Routines. *Journal of Artificial Societies and Social Simulation* 3(2), 2000 <http://jasss.soc.surrey.ac.uk/3/2/2.html>

Duffy, J. Learning to speculate: Experiments with artificial and real agents. *Journal of Economic Dynamics & Control*, 25, 295-319, 2001

Ebenhöh, E. and C. Pahl-Wostl, An adaptive toolbox model: a Pluralistic Modelling Approach for Human Behaviour based on Observation. *Journal of Artificial Societies and Social Simulation*, 7(1), 2004 <http://jasss.soc.surrey.ac.uk/7/1/3.html>

Fehr, E. and S. Gächter, Altruistic punishment in humans, *Nature*, 415, 137-140, 2002.

Fehr, E. and B. Rockenbach, Detrimental Effects of sanctions on human altruism, *Nature*, Vol 422, 137-139, 2003

Gigerenzer, G. and R. Selten, *Bounded Rationality, The Adaptive Toolbox*. Cambridge, Massachusetts: The MIT Press. 2001.

Gintis, H. Game Theory Evolving. Princeton University Press, Princeton, NJ, 2000.

Hardin, R.G. The Tragedy of the Commons. Science 162: 1243-1248, 1968.

Hare,M., J. Heeb and C. Pahl-Wostl, The Symbiotic Relationship between Role Playing Games and Model Development: A case study in participatory model building and social learning for sustainable urban water management. Proceedings of the Meeting of the International Society for Ecological Economics, Sousse, Tunisia, 2002.

Jager W. and M. A. Janssen, Using artificial agents to understand laboratory experiments of common-pool resources with real agents. In: M. A. Janssen (editor), Complexity and Ecosystem Management. The Theory and Practice of Multi-Agent Systems, Bodmin, Cornwall, 2002

Liebert, R. M. and M. D. Spiegler, *Personality: strategies and issues*, Wadsworth, Belmont, California, 1994.

Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, New York, 1990.

Pahl-Wostl, C. Agent Based Simulation in Integrated Assessment and Resources Management. In: Rizzoli,A. and Jakeman,T. (eds). Integrated Assessment and Decision Support. Proceedings of the 1st biennial meeting of the International Environmental Modelling and Software Society, Vol 2, 239-250, 2002a.

Pahl-Wostl, C. Participative and Stakeholder-Based Policy Design, Evaluation and Modeling Processes, *Integrated Assessment*, 3(1), 3-14, 2002b.

Pahl-Wostl, C. and M. Hare, Processes of social learning in integrated resources management. *Journal of Community and Applied Social Psychology*, in press.

Panebianco, S. and C. Pahl-Wostl, Obstacles in launching a participatory modelling process. Submitted to iEMSs 2004.

Simulating stakeholder support for river management

P. Valkering [a], J. Krywkow [b], J. Rotmans [a], and A. van der Veen [b].

[a] *International Centre for Integrative Studies (ICIS),University of Maastricht, PO BOX 6200 MD, Maastricht, Netherlands*

[b] *University of Twente, Enschede, Netherlands*

Abstract: River management is a typical example of a complex problem involving a variety of stakeholder interests and fundamental environmental uncertainties. The Dutch government aims to take the different interests and views explicitly into account by allowing stakeholders to participate in the planning process. The aim of our research is to analyse this participatory process to investigate stakeholder support, their influence on the decision-making process, and the role of norms and co-operation. To this end, we developed an agent based model representing a negotiation among stakeholders. Stakeholder support for a river management strategy is modelled on the basis of the Theory of Reasoned Action and a theory of Social and Cognitive Action. For evaluating the different river engineering alternatives the Agent Based Model is coupled to an Integrated River Model that describes possible long-term impacts (e.g. flood risk, nature development) of river engineering options. We show how the coupled model framework can aid to analyse the participatory planning process of the ongoing Grensmaas project. Also, we assess how the policy outcome might change when the agents would take climate change into account.

Keywords: Participatory Agent Based Modelling, River modelling, Stakeholders, River management

1 INTRODUCTION

Agent Based Modelling (ABM) has been identified as a promising technique for the explicit representation of stakeholder perspectives in policy relevant research. Agent based models may be incorporated into Integrated Assessment modelling frameworks for a better representation of stakeholder behaviour in Integrated Assessment models allowing us to investigate stakeholder-environment interaction (Rotmans, 2002). Furthermore, agent based models can be used to structure participatory processes, stimulating social learning by sharing viewpoints among stakeholders (Pahl-Wostl, 2002).

In this paper we apply the approach of ABM to a case study of river management. We will focus on the river engineering project ‘Grensmaas’ which is currently ongoing in the Dutch province of Limburg. The Grensmaas project was initiated in 1997 to achieve three main goals (Maaswerken, 1998): 1) reduction of flood recurrence to 1:250 years, 2) the development of a minimum of 1000 ha of riparian nature, and 3) the extraction of a minimum of 35 million tons of gravel for national use. To this end, measures are planned to widen the Meuse to the north of the city of Maastricht over a length of some 40 km.

The Grensmaas project affects many stakeholders with a variety of interests. The main stakeholder groups of the Grensmaas project are the inhabitants of the region, farmers, nature organizations, and the gravel extracting companies. It is an explicit aim of the project organization to involve these stakeholders as much as possible in the decision-making process in order to develop an integrated strategy and a broad societal interest and support.

In this paper we analyse the participatory planning process of the Grensmaas project. We thereby use an agent based model to assess stakeholder support for a river management strategy on the basis of their goals and beliefs. In the first model prototype discussed here we assume the cognitive representation of goals and beliefs to be static. In future research we will further investigate and implement mechanisms of adaptive cognition.

With the current prototype we assess optimal strategies from the perspectives of individual stakeholders ('ideal' strategies), as well as plausible negotiation outcomes ('compromises'). Finally, we show how the negotiation process could change when all stakeholders would take climate change into account.

2 A MODEL OF STAKEHOLDER SUPPORT

2.1 Determinants of support

The conceptual model of stakeholder support is depicted in **figure 2.1.** It is based upon the well-known Theory of Reasoned Action (TRA) (Ajzen and Fishbein, 1980) and a theory of Social and Cognitive action (Conte and Castelfranchi, 1995). Stakeholder support is determined by three factors:

Attitude: In the TRA attitude towards a behaviour is described as 'the individuals positive or negative evaluation of performing that behaviour' (Ajzen and Fishbein, 1980). We interpret this notion as an evaluation of self-interest goals that arise from a so-called pre-cognitive objective relation of interest (Conte and Castelfranchi, 1995). In the case study of the Grensmaas typical self-interest goals are flood reduction for the inhabitants, profit for the gravel extracting companies and so on.

Norm compliance: In correspondence to the 'subjective norm' in the TRA we adopt the notion, norm compliance as a formal evaluation of normative goals. A normative goal is related to the agent's belief that he has a social obligation to perform (a) specific action(s). According to (Conte and Castelfranchi, 1995) normative goals can be issued by a so-called 'sovereign' agent. For our case study typical normative goals are the primary objectives of the Grensmaas project of flood recurrence, nature development, and gravel extraction issued by the 'sovereign' policymaker.

Social compliance: Goal adoption can be considered one of the main mechanisms of stakeholder interaction (Conte and Castelfranchi, 1995). Therefore we include the notion 'Social compliance' as an evaluation of the agents' adopted goals.

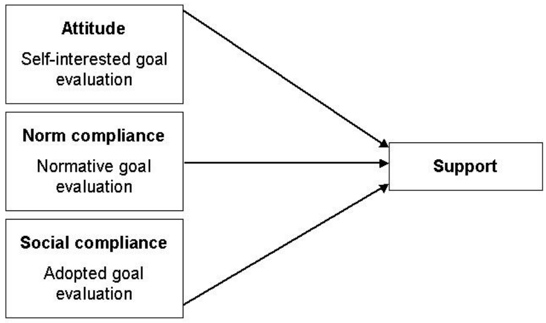


Figure 2.1: Determinants of stakeholder support

2.2 Calculating Support

The agents determine their support for a river management strategy by evaluating the expected impacts of the river management strategy in relation to their self-interest, normative, and adopted goals. For this evaluation we assume that the agents exhibit so-called lexicographic preferences. Lexicographic preferences imply that decision evaluations are made on the basis of a (ordered) set of minimal needs. When these needs are fulfilled, optimisation may occur on the basis of other criteria. The methodology has recently been discussed in the field of environmental economics (Stern, 1997) and applied, for example, for a stakeholder evaluation of wetland recreation (Spash, 2000).

The implementation of the lexicographic preference structure is shown in **figure 2.2**. Agents may express their goal satisfaction on a scale from –1 ('unacceptable') to 0 ('neutral') to 1 ('strongly supported'). Goal satisfaction is determined on the basis of two types of quantitative goal standards. Firstly, an agent can attach a minimal (maximal) requirement, referred to as a 'conditional standard' (*CS*). When its expected value for this goal is respectively below (above) this standard the agent will consider this goal value to be 'unacceptable', corresponding to lexicographic, non-compensatory behaviour. Additionally, agents may specify whether the goal should be optimised. They do so by expressing two optimisation standards: an optimisation zero point value OS_0 (the goal value for which their support is 'neutral') and a optimisation high value OS_H (the goal value for which their support is 'high').

The total support an agent attaches to a river management strategy is now simply calculated as the unweighted average of its goal evaluations. However, when one of its goals is evaluated as 'unacceptable' the river management strategy is considered 'unacceptable' (support is set to –1).

2.3 The negotiation outcome

The outcome of the negotiation process is calculated by maximizing support among

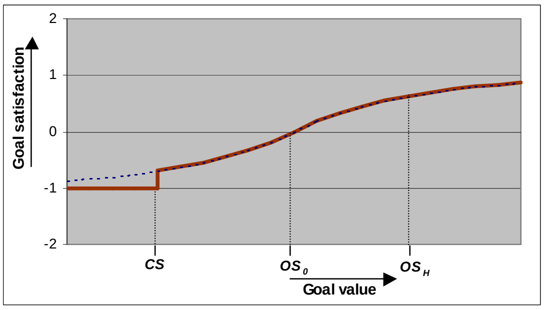


Figure 2.2: A typical lexicographic goal evaluation curve that specifies goal satisfaction as a function of the expected goal value. A goal evaluation curve is defined on the basis of a conditional standard *CS*, and/or optimisation standards OS_0 and OS_H.

stakeholders involved. This simple method allows us to assess 1) what would be 'ideal' river management strategies designed on the basis of one individual stakeholder perspective, and 2) a 'compromising' strategy designed on the basis of multiple perspectives. For compromising stakeholders are valued equally with the exception of the policy-maker and gravel extractor. These powerful parties must support the river management strategy (support > 0) for its overall approval.

3 THE INTEGRATED RIVER MODEL

For evaluating the different river engineering alternatives the agents use an Integrated River Model (IRM). The concept of the IRM is displayed in **figure 3.1**. The main input model variables are different river engineering measures - that together constitute a river management strategy. The main output variables are the long-term impacts with respect to flooding, nature, and agriculture as well as short-term costs and benefits (e.g. monetary costs, gravel extraction, and hindrance) associated with river engineering.

The model was implemented for a river cross-section representing the river Meuse at the location 'Borgharen'. The modules are based on basic principles of hydrology, hydraulics, groundwater dynamics, and nature development and are partially based upon existing expert modules (for example to assess flood and agricultural damage). The model was conceptually validated with experts from the Grensmaas project organisation, and partially numerically calibrated and validated with respect to their model results. For a detailed model description, see (Valkering *et al.*, 2004).

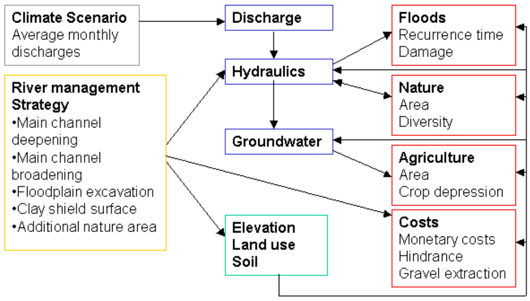


Figure 3.1: The Integrated River Model concept.

Estimating the impacts of river engineering involves numerous fundamental uncertainties in relation to climate change, nature development, and for estimating the monetary costs and benefits (Valkering *et al.*, 2004) According to (van Asselt, 2000) different legitimate interpretations of these uncertainties may exist. The agents may thus hold different perspectives on uncertainty as part of their subjective belief system. These perspectives are represented as value settings for uncertain IRM parameters related to climate change, hydraulic roughness, and costs and benefits. The agents feed these settings into the IRM, together with a set of river engineering measures, to calculate values for their goal criteria, see **figure 3.1**. These values form the basis of their outcome evaluation of the river management strategy described in the previous section.

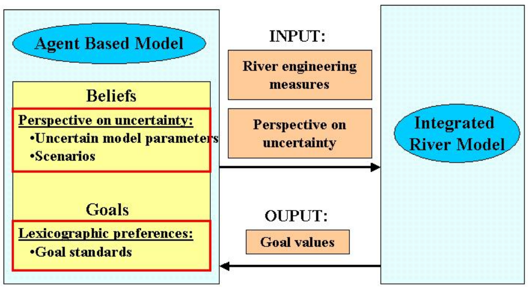


Figure 3.1: The interface between the Agent Based and Integrated River Model

4 STAKEHOLDER PERSPECTIVES

In the previous sections we have developed a framework for structuring stakeholder perspectives in terms of goals, quantitative goal standards, and perspectives on uncertainty. In this section we apply this framework for a preliminary assessment of stakeholder perspectives. The information was obtained from a series of stakeholder interviews, available governmental Environmental Assessment reports of proposed

Agent	Goal	Type	Direction	Conditional Standard	Optimization zero point value	Optimization high value
Policymaker	flood recurrence (yrs)	N	min	250	-	-
	nature area (ha)	N	min	1000	-	-
	ecosystem diversity (-)	N	min	0.7	0.7	1
	loss agricultural area (ha)	N	max	-	1000	2000
	Δ groundwater level (m)	N	min	-	0	-0.2
	hindrance (person*years)	N	max	-	20000	30000
	gravel extraction (*10^^{6} tons)	N	min	35	-	-
	profitability (%)	N	min	4	-	-
Citizen	flood recurrence (yrs)	SI	min	250	-	-
	hindrance (person*years)	SI	max	30000	20000	30000
	nature area (ha)	SI	min	1000	-	-
	ecosystem diversity (-)	SI	min	0.7	0.7	1
Nature org.	nature area (ha)	SI	min	1000	-	-
	ecosystem diversity (-)	SI	min	0.7	0.7	1
	Δ groundwater level (m)	SI	min	-0.2	0	-0.2
Farmer	flood recurrence (yrs)	SI	min	250	-	-
	loss agricultural area (ha)	SI	max	1000	0	1000
	Δ groundwater level (m)	SI	min	-0.2	0	-0.2
Gravel extractor	gravel extraction (*10^^{6} tons)	SI	min	-	35	70
	profitability (%)	SI	min	4	-	-

Table 4.1: Stakeholders' goals and conditional and optimisation goal standards. The goal standards determine the evaluation curve an agent uses to evaluate its goal, see **figure 2.2**. The goal 'type' indicates whether the goal is considered self interested (SI), normative (N) or adopted (A). The 'direction' indicates whether the standards refer to minimal or maximal requirements.

river management alternatives (Maaswerken, 1998), (Maaswerken, 2003), and on the official stakeholder commentaries to these reports. The description of the methodology for perspective elicitation is necessarily brief. A more elaborate description can be found in (Valkering *et al.*, 2004).

The main stakeholders of the Grensmaas project are represented by the corresponding agents 'policymaker', 'citizen', 'farmer', 'nature organization', and 'gravel extractor'. The stakeholders' goals and goal standards are presented in **table 4.1**. The policymaker is considered to be a sovereign agent issuing the norms of flood reduction, nature development, and gravel extraction in correspondence with the main objectives of the Grensmaas project. The negative side effects (loss of agricultural area, groundwater level decrease, and hindrance) are to be minimized (Maaswerken, 1998). The policymaker also issues the norm of profitability, since it requires the monetary costs of river engineering to be fully covered by the benefits of extracted gravel. The non-governmental stakeholders hold various self-interest goals as presented in **table 4.1**.

The adopted stakeholders' uncertainty perspectives are displayed in **table 4.2**. We assume that, in order to avoid risk, stakeholders adopt an uncertainty perspective that minimizes their goal fulfilment. The citizen, for example, claims that a high estimate of climate change has to be taken into account in the estimation of the safety level.

Uncertainty / Stakeholder	Climate change	Hydraulic roughness	Costs and benefits
Policymaker	No	Central	Central
Citizen	High	High	Central
Nature organization	No	Central	Central
Farmer	No	Low	Central
Gravel extractor	No	Central	Negative

Table 4.2: Assumed stakeholder perspectives on uncertainty

5 SIMULATION RESULTS

In this section we present an overview of some simulation results. In particular, we will construct optimal river engineering strategies for different boundary conditions. To this end, a river management strategy is represented as a set of river engineering parameters: main channel deepening (MC deepening), main channel broadening (MC broadening), floodplain excavation (FP excavation), surface elevation of the clay shield (CS level), and additional nature area (Add. nature). For simulating a river management strategy the parameters are varied within predefined ranges in order to maximize total agent support.

In **table 5.1** the following simulated river management strategies are displayed: 'Ideal' strategies maximizing support for each individual agent, and a 'compromising' strategy maximizing total support (see **section 2.3**). These optimal strategies are constructed on the basis of the stakeholder perspectives presented in **section 4**. The compromising strategy corresponds well with currently preferred river engineering alternative of (Maaswerken, 2003), see (Valkering *et al.*, 2004).

	cit	ge	pm	no	farm	comp
MC deepening (m)	1	2	-2	-2	-2	0
MC broadening (m)	100	200	50	50	0	100
FP excavation (m)	125	500	250	375	250	125
CS surface (m)	0	4	0	0	4	4
Add. nature (m)	250	0	375	125	0	250

Table 5.1: Calculated 'ideal' river management strategies for the agents citizen (cit), gravel extractor (ge), policy-maker (pm), nature organisation (no) and farmer (farm), and the 'compromising' strategy (comp).

5.1 A short analysis the Grensmaas project

The 'compromising' river engineering alternative can be characterized as, indeed, a compromise between the different stakeholders, but within the normative boundary conditions of safety, nature area and profitability set by the government. This is illustrated in **figure 5.1**, which shows calculated stakeholder support for the 'compromising' alternative. The goal evaluations for safety, nature area, gravel extraction, and profitability are equal to 1, indicating that the main project objectives are fulfilled. However, other criteria are valued negatively. In particular the farmer strongly objects to the expected loss of agricultural area and change in groundwater level.

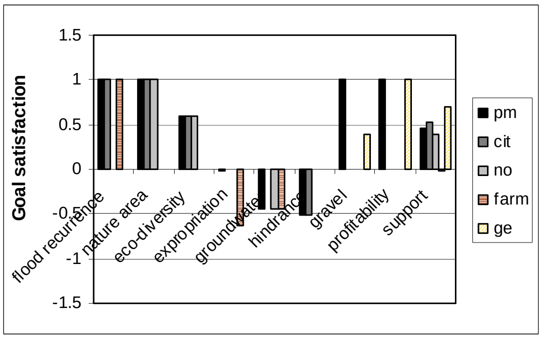
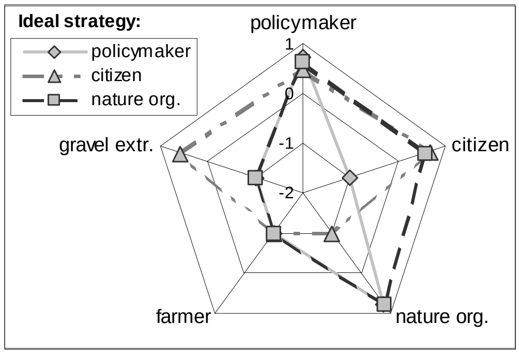


Figure 5.1: Calculated goal satisfactions and stakeholder support for the 'compromising' alternative.

The analysis of the 'ideal' strategies shows that the mutual agreement among the policy-maker, citizen and nature organization is large. Each one of their respective ideal strategies is valued relatively high among this group (see **figure 5.2 a)**) indicating common interests and win-win. The farmer and gravel extractor, on the other hand, generally disagree with the other parties, as illustrated in **figure 5.3 b)**. Their respective ideal strategies are valued high only by themselves and are considered unacceptable by the rest.

For the farmer the disagreement is related to conflicting goals. Its interest of agricultural land and proper groundwater conditions inherently conflict with the approach of river widening and nature development. For the gravel extractor, however, the disagreement results from a difference in uncertainty perspective with respect to the expected costs and benefits of the river engineering measures (see **table 4.2**).

a)



b)

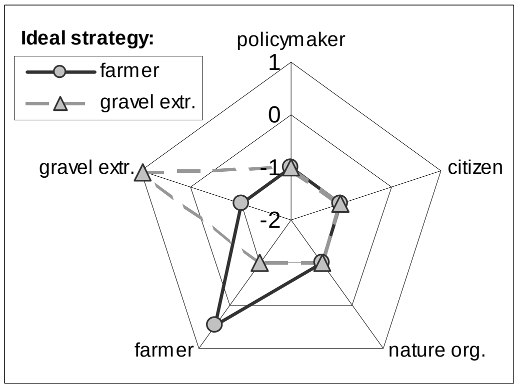


Figure 5.2: Calculated stakeholder support for the 'ideal' river management alternatives.

5.2 The case of climate change

The modelling framework allows us to assess how the negotiation process may change under changing boundary conditions. Consider, for example, the case of climate change. Climate change is currently considered an important issue for long-term water management in the Netherlands. It may lead to increasing peak discharges by some 20%.

The new 'compromising' strategy simulated for conditions of climate change contains large-scale riverbed broadening, in combination with raising the main channel riverbed. This would allow society to maintain current safety standards whilst mitigating decreases in the groundwater level, see **figure 5.3**. In this scenario the citizen would strongly object to the river engineering strategy, because of excessive hindrance levels, but he would be largely ignored.

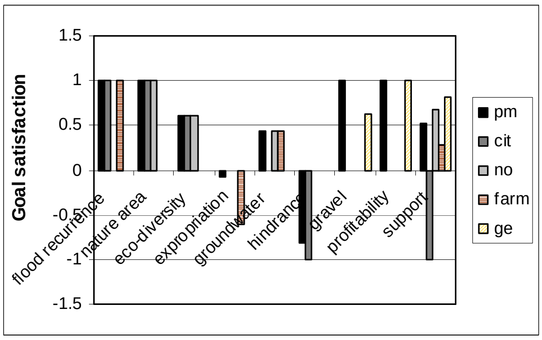


Figure 5.3: Calculated goal satisfactions and stakeholder support for the 'compromising' alternative under climate change conditions.

6 CONCLUSION

In this paper we have presented an Agent Based Model that describes stakeholder support and the outcome of a negotiation process among stakeholders of river management. The type of model that we developed must not be considered a 'truth machine' that predicts policymaking for river management. It rather provides a framework for a 'what-if' analysis. Given the goals and beliefs of stakeholders, the model calculates which river management strategy receives the maximal total support of the stakeholders.

We showed how the ABM can be used to analyse stakeholder perspectives and identify conflicting goals and mutual benefits among stakeholders. Also, the model can aid to reflect upon possible changes in the negotiation process (i.e. problems that are likely to emerge) as a result of uncertain future developments. The current model version is a 'stepping stone' for investigating

In future research we intend to implement mechanisms of goal adoption/rejection and belief change in the model framework, in correspondence with the notions of adaptive cognition and social learning, in order to study the interactions between the social system and river environment.

The current prototype is particularly suitable as a communication tool for application within participatory stakeholder processes. Using this model would encourage stakeholders to reflect upon their goals in a social context. When conflicting goals are revealed stakeholders may reconsider their goals and adopted standards. A participatory approach would thus be vital for a better understanding of the mechanism of cognition change. Eventually, we hope that the application of this model concept will induce social learning and the consideration of multiple perspectives in decision-making. This would be a small step further towards truly collaborative and sustainable river management.

7 ACKNOWLEDGEMENTS

We would like to thank all the participants of the FIRMA project for many stimulating discussions and fruitful collaborations during this research.

8 REFERENCES

Ajzen, I. and Fishbein, M. (1980) *Understanding attitudes and predicting social behavior*. London: Prentice-Hall.

Conte, R. and Castelfranchi, C. (1995) *Cognitive and social action*. London: UCL Press ltd.

Maaswerken (1998) Mer grensmaas (environmental assessment report), hoofdrapport a: Hoofdlijnen. Maaswerken.

Maaswerken (2003) Mer grensmaas 2003 (environmental assessment report), hoofdrapport. Maaswerken.

Pahl-Wostl, C. (2002) Participative and stakeholder-based policy design, evaluation and modeling processes. *Integrated assessment* 3(1), 3-14.

Rotmans, J. (2002). "Geintegreerde milieumodellen: Een retrospectie en prospectie.", Wageningen, Netherlands.

Spash, C. L. (2000) Ecosystems, contingent valuation and ethics: The case of wetland recreation. *Ecological Economics* 34(195-215.

Stern, D. I. (1997) Limits to substitution and irreversibility in production and consumption: A neoclassical interpretation of ecological economics. *Ecological Economics* 21(197-215.

Valkering, P., Krywkow, J., Rotmans, J. and Van der Veen, A. (2004) Simulating stakeholder support in a negotiation process: An application to river management. Submitted to: *Simulation: Transactions of The Society for Modeling and Simulation International*.

van Asselt, M. B. A. (2000) *Perspectives on uncertainty and risk : The prima approach to decision support*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Relating Choice of Agent Rationality to Agent Model Uncertainty - an experimental study

D. E. Reusser, dreusser@uos.de [a] , M. Hare[b] and C. Pahl-Wostl[a]

[a] Institute of Environmental Systems Research, University of Osnabrück, 49069 Osnabrück, Germany

[b]Seecon Deutschland GmbH, AVZ: 31/515, Albrechtstr. 28, 49069 Osnabrück, Germany

Abstract: The importance of model uncertainties arising from different assumptions about human behavior as opposed to parameter uncertainty is often neglected in integrated models for policy development. In this study, so-called agent model uncertainty is estimated in relation to the choice of agent rationality. A classification scheme is proposed which allows us organize decision models according to their deviation from full rationality. Five decision models covering the whole range from full rationality to maximum deviation from rationality (random decisions) are classified. They are then used in an existing integrated model simulating crop fertilizer usage and related threshold policies for groundwater protection. Using this model and the different decision models, two hypotheses are tested: 1) that agent model uncertainty increases with increasing deviation from rationality and 2) that agent model uncertainty increases for all decision models similarly and uniformly in response to an increase in noise in the model. Results are analyzed with respect to changes in policy and with respect to the level of weather influence on crop yield. Results show that agent model uncertainty varies with deviation from the purely rational in a non-linear way. Hypothesis two also does not hold. The degree of sensitivity of results with respect to uncertain parameters that the agent needs to consider is very much dependent on the decision model. Therefore it is suggested to test agent-based models for robustness and validity with respect to agent model uncertainty by using different categories of decision models that sample the range of possible rationalities.

Keywords: bounded rationality, agent-based modeling, agent model uncertainty

1 INTRODUCTION

Agent-based models are increasingly used to develop integrated models for developing policy scenarios in resources management. To do so one has to make assumptions on human behavior. The possibility on making choices are numerous [Pahl-Wostl, 2002], but only a few investigations [e.g. Hare and Pahl-Wostl, 2001] exist about the implications of different choices. Uncertainty exists about choice of a decision model from the wide variety of possible models. Agent model uncertainty is defined here as the uncertainty propagated through an agent-based model because of the uncertainty about the decision model to be used [Hare and Pahl-Wostl, 2001]. Since choice of decision model normally reflects an alteration in model structure of some kind, agent model uncertainty can be seen as a sub-class of model form uncertainty as defined by Morgan and Henrion [1990].

Perfect rational behavior (*PRB*) as defined by the Von Neumann - Morgenstern utility [Kreps, 1988] is convenient and successful for many problems, but it does not accurately describe human behavior in many real-world situations [McFadden, 1999]. Bounded rationality has been proposed to be a better concept to describe human decision strategies. However, the term bounded rationality has at least two different meanings [Sent, 1997], which leads to several dimensions along which decision models may deviate from rationality. It is of interest how choice of agent rationality relates to agent model uncertainty. Hypotheses about this relationship are proposed in section 2. An agent-based model is introduced in section 3 and five exemplary decision models are described in section 4 together with a classification of rationality. The decision models are applied to the agent-based model as described in section 5. Agent model uncertainties are then calculated and used to test the proposed hypotheses (Section 6). Conclusions are presented in section 7.

2 HYPOTHESES

Working hypotheses are formulated to support the development of a more systematic approach about how to account for different assumptions of agent rationality in agent-based models. In order to formulate hypotheses relating choice of agent rationality to agent model uncertainty, a classification scheme for decision models is proposed as shown

1. information availability
   - full information about environment, parameter uncertainties as *objective* probability distribution (OPD)
   - full information about environment, parameter uncertainties as *subjective* probability distribution
   - full information about environment, parameters game theoretic treatment
   - percepts as cause and effect
   - percepts without further information
   - percepts from search process with stopping heuristic
2. sampling of alternatives before the decision
   - all alternatives
   - some/one alternatives
   - no beforehand assessment
3. measure for assessment before the decision
   - utility function
   - prospect function (cognitive limitations)
   - probability function
   - other measure
   - no measure
4. selection of an alternative
   - best possibility
   - probability distribution (prob. dist.)
   - selection heuristics
     - ignorance-based decision making
     - One-reason decision making
     - Elimination heuristics
     - Satisficing heuristics for sequential search
   - Random selection.

Figure 1: Decision model classification scheme

in figure 1. The distinction into four dimensions (1. through 4. in figure 1) has its origin in artificial intelligence [Russell and Norvig, 1995, p. 419]. However sampling and measure for the assessment of alternatives before the decision are discussed as one step by Russell and Norvig [1995]. Classes within each dimension are ordered with increasing deviation from rationality in figure 1. Rationality and deviation from rationality are used in a strictly economic sense (PRB) for the classification and hypotheses. Note, that the economic sense can differ from personal intuition about what may or may not be described as rational behavior. See Reusser [2004] for more details on the suggested classification. Based on the classification scheme hypotheses are derived to relate agent model uncertainty to choice of agent rationality. Note that agent model uncertainty requires a reference rationality to allow comparison.

Hypothesis 1 *Agent model uncertainty with reference to a classical rational decision model* PRB *increases with increasing deviation from rationality.*

Hypothesis 2 *Noise in the decision problem is expected to increase uncertainty for all decision models uniformly which is expected to result in higher agent model uncertainty if a rational decision model (*PRB*) without noise is used as a reference.*

3 AGENT-BASED MODEL

The agent-based model in this work is based on an existing model by Hare and Pahl-Wostl [2001] investigating agent model uncertainty for nitrogen fertilization on crop fields. The model provides the environment for the decision models (*farmer* model). Low fertilizer application levels lead to low yield and low income, while high levels lead to high yield, high income but groundwater contamination and fines.
The *field* model determines the crop yield as a function of the fertilizer amount. The yield function (Figure 2) is parabolic and depends on the fertilizer nitrogen uptake effectiveness parameter values (ef). Saturation occurs for higher nitrogen fertilizer amounts. The yield function is also affected by the *climate* model, which reduces the uptake effectiveness randomly within a certain range (see Section 5) for each decision step. The varying uptake effectiveness due to the climate will be referred to as the model noise, about which the farmer decision models are ignorant. The income I is determined by the achieved yield Y together with parameters from the *market* model (price for crop, fertilizer and fertilizer application). The fertilizer that is not taken up by the crop enters the *groundwater* model. Groundwater pollution is not regarded as a problem by the win-oriented farmer A *regulator* model is in charge to decrease groundwater pollution in the presence of win-oriented farmers. If the applied fertilizer amount is higher than the *fertilizer threshold* level t_f, the *regulator* model may or may not $fine$ the farmer with probability p_{fined}.

4 DECISION MODELS

Five decision models differing in the choice of agent rationality will be compared in order to test the two hypotheses. The five decision models chosen for this study are meant to be an exemplary set for testing the hypotheses. Many more decision models are possible such as for example a stubborn agent, evolutionary learning agents, and a rational agent with parameter learning abilities. This would be subject to further research. Since agent model uncertainty is a sub-class of model form uncertainty, comparing different decision models means that different models are compared. Structural differences will be-

Table 1: Classification of farmer decision models

farmer model	information avail.	sampling	measure	selection
RationalFarmer PRB	full information, OPD	all alternatives	utility function	best alternative
ProspectFarmer	full information, OPD	all alternatives	prospect function	best alternative
ExperientialFarmer	percepts as cause or effect	some alternatives	probability function	prob. dist.
HeuristicFarmer	percepts as cause or effect	no assessment	no measure	selection heuristic
RandomFarmer	no information	no assessment	no measure	random selection

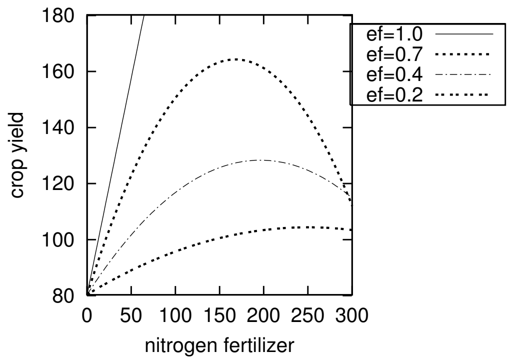


Figure 2: Crop yield as a parabolic function of the nitrogen fertilizer amount for different uptake effectiveness ef.

come clear in the following sections. Table 1 shows the differences between the models according to the classification scheme (Figure 1).

Using the *RationalFarmer* as a reference point for full rationality, we assume that deviation from rationality increases from the *RationalFarmer* to the *ProspectFarmer*, to the *ExperientialFarmer*, to the *HeuristicFarmer*, and to the *RandomFarmer*. The goal of each decision model is to select the desired possible yield goals the farmer sees to attain. This decision then directly influences the rate of fertilizer applied. For purposes of comparison, each agent shares a common decision space, selecting from possible yield goals of 110, 120, 130, 140, 150, 160, 170, or 180 bundles of crop [Hare and Pahl-Wostl, 2001].

4.1 Rational Decision Model

In the *RationalFarmer* decision model (*PRB*), a Von Neumann-Morgenstern expected utility EU [Kreps, 1988] for each yield goal Y is calculated as

$$EU(Y) = R_f(Y)*p(f|Y)+R_{nf}(Y)*(1-p(f|Y))$$

with $p(f|Y) = p_{fined}$ if $Y > t_f$ (fertilizer threshold) and $p(f|Y) = 0$ otherwise. The returns R_f and R_{nf} are defined as $R_f(Y) = I(Y) - fine$ and $R_{nf}(Y) = I(Y)$, respectively.

4.2 Decision Model Based on Prospect Theory

Prospect theory by Tversky and Kahneman [1992] reflects five important deviations from expected utility theory, that consistently occur if humans are making decisions. Because of these deviations and in contrast to rational decisions, monetary returns transform to utility in a nonlinear way Uncertain events such as the probability to be fined are weighted with a function $w()$, overestimating the probability for events with a low probability. In the *ProspectFarmer* decision model, the overall expected prospect $V(Y)$ for a specific yield goal is calculated as

$$V(Y) = -(fine)^{\beta} * w(p(f|Y)) + I^{\alpha}$$

with $\alpha \approx \beta \approx 0.88$, w as defined by Tversky and Kahneman [1992], and other symbols as defined before.

4.3 Experiential Decision Model

The *ExperientialFarmer* decision model is a more realistic decision model, which makes it possible to reproduce several stylized facts about human decision behavior as observed in psychological experiments [Arthur, 1993]. The basic idea is to make decisions based on accrued experience, which implies that learning occurs. Probabilities to choose a certain option change proportionally to the accumulated income from using this option. Therefore lock-in on suboptimal strategies may occur if differences between options are small or random processes are present in the decision environment, such as the noise due to the climate model.

4.4 Heuristic Based Decision Model

According to Bock, farmers often use the following heuristic to determine a yield goal: The average yield is calculated for a given field over a 4- or 5-year period and a safety margin $s = 5\%$ is added to that average [Bock and Hergert, 1991, therein Wiese et al., 1987]. Since this method lacks possible reactions to fines, the following extended version is used, where the *HeuristicFarmer* decision model does not add the safety margin if a fine was paid the year before:

$$Y_{Goal} = \frac{\sum_{y=0}^{t} Y_y}{t} * \left(1 + s - \sum_{y=0}^{t} b * s * r^y\right)$$

t is the time horizon used by the farmer to average the yield, Y_y is the yield achieved y years ago, and

r is a discount rate. b decides whether a past fine is considered or not: $b = 1$ if $\frac{fine_y}{I_y} > f$ and $b = 0$ otherwise. $fine_y$ is the fine payed y years ago, I_y is the income achieved y years ago, and f is a decision model parameter.

4.5 Random Decision Model

The *RandomFarmer* decision model uses a dice throw to choose the yield goal. It is used as reference for maximum deviation from rationality.

5 EXPERIMENTAL SETUP

The five decision models are applied to the agent-based model with four different fertilizer threshold levels $t_f \in (50, 75, 100,\text{unlimited})$. Additionally, the agent-based model is run on three noise levels (no noise, ef not reduced; medium noise, ef reduced between 0 and 0.2, and high noise, ef reduced between 0 and 0.4). This results in $5*4*3 = 60$ scenarios. Parameter uncertainty is accounted for with Monte Carlo simulations. For the resulting cumulative probability distribution, we want the 95 % confidence intervall for all percentiles to be about plus or minus 3 estimated percentiles. According to Morgan and Henrion [1990, p. 202] this requires 1000 Monte Carlo runs. Each simulation lasts 125 decision steps except for the *ExperientialFarmer* with 1250 decision steps in accordance to Hare and Pahl-Wostl [2001]. The higher number of decision steps for the *ExperientialFarmer* was chosen in order to reduce the effect of the learning phase.
Groundwater pollution level is used to estimate agent model uncertainties since it is a key variable for judging the efficiency of policy measures in this model. Pollution levels are averaged over all decision steps. Cumulative probabilities [Morgan and Henrion, 1990] are obtained by sorting average pollutions x from all Monte Carlo runs and assigning each result a probability of $\frac{1}{n}$.

5.1 Calculation of Agent Model Uncertainty

The mean square difference between two cumulative probability distributions obtained from different agent models is proposed as a measure for agent model uncertainty AMU.

$$AMU = \frac{1}{m}\sum_{i=1}^{m}(p(x_i) - p_{ref}(x_i))^2$$

with p_{ref} and p the cumulative probabilities for pollution level x_i for the reference scenario and the scenario of interest, respectively. Pollution levels were selected from the wide variety of possible model output parameters for calculation of AMU, because in the context of policy analysis, the main interest is whether the policy used by the regulator to reduce

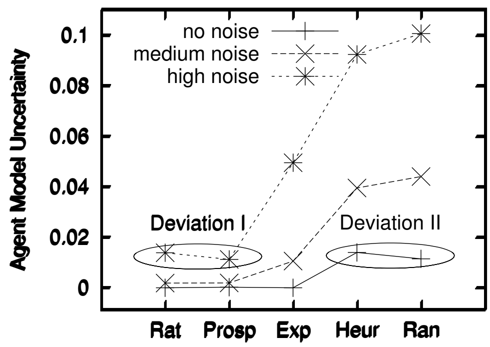


Figure 3: Agent model uncertainty for scenarios with a fertilizer threshold of $t_f = 50$ lbs.
Rat: *RationalFarmer*, Prosp: *ProspectFarmer*, Exp: *ExperientialFarmer*, Heur: *HeuristicFarmer*, Ran: *RandomFarmer*

groundwater pollution has the desired influence. The *RationalFarmer* scenarios with *no noise* are used as *reference on each fertilizer threshold level* (e.g. RationalFarmer scenario with $t_f = 50$ lbs and no noise is used for all scenarios with $t_f = 50$ lbs). x_i is set to $x_i \in (5, 10, 15, \ldots, 445, 450\ units)$ and therefore $m = 90$. $AMU = 0$ only for equal cumulative probability distributions. Therefore, for the reference runs (Rational-50-no-noise, Rational-75-no-noise, Rational-100-no-noise, and Rational-Baseline-no-noise) the agent model uncertainty is $AMU = 0$ by definition, since the same cumulative probability function is used as data and as reference.

6 RESULTS AND DISCUSSION

Agent model uncertainties AMU are discussed in the light of the need to test the two hypotheses (Section 2). However, for a more detailed discussion see Reusser [2004]. Figure 3 shows AMU on the y-axis for all scenarios with a fertilizer threshold of $t_f = 50$ lbs. The decision models are ordered along the x-axis with assumed increasing deviation from rationality. Different symbols indicate different noise levels. AMU for the other fertilizer threshold levels t_f are shown in figures 4 through 6. According to hypothesis 1 we expect increasing AMU for increasing deviation from rationality (left to right in the four figures). A first deviation from expectations (Deviation I) is observed in figures 3 and 4, with the AMU for the Prospect-50-high-noise scenario being lower than for the Rational-50-high-noise scenario. This is due to the fact that the *ProspectFarmer* decision model overestimates the probability to be fined. Because of this overestimation, the pollution level is slightly lower compared to the *RationalFarmer* decision model. Since the pollution for the reference scenario (Rational-50-no-noise) is lower as well (be-

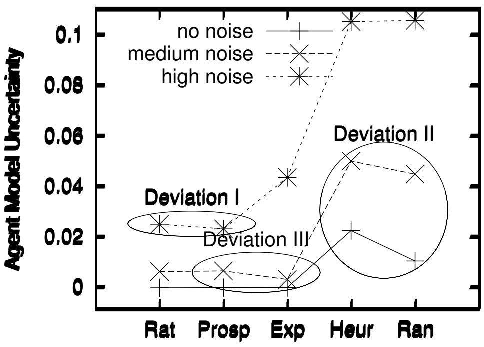


Figure 4: Agent model uncertainty for scenarios with a fertilizer threshold of $t_f = 75$ lbs.

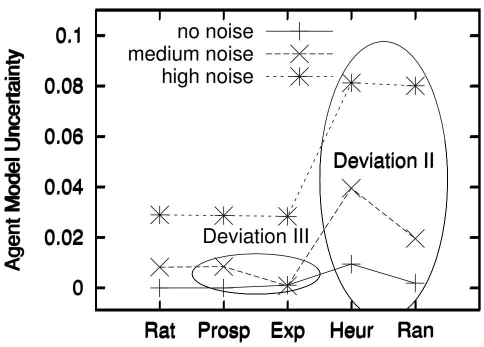


Figure 5: Agent model uncertainty for scenarios with a fertilizer threshold of $t_f = 100$ lbs.

cause of the lower noise level), a lower AMU is observed for the *ProspectFarmer* decision model at high noise levels.

The AMU for the the *HeuristicFarmer* decision model is higher than for the *RandomFarmer* decision model (Deviation II), which can be observed in all four figures. Note that the *RandomFarmer* decision model is the reference for maximum deviation from rationality. However, the *HeuristicFarmer* decision model results in a AMU that is even greater than this. The primary goal of the decision models is to optimize fertilizer usage for high income (it is not to avoid high pollution levels) and therefore such higher deviations are possible. It is noteworthy that the decision model which is supposed to imitate behavior of real farmers shows highest deviations (compared to the other decision models) from rational decisions.

A third deviation is observed in figures 4 through 6. AMU for the *ExperientialFarmer* is smaller than for the *ProspectFarmer* decision model. In fact, the *ExperientialFarmer* and *RationalFarmer* decision models are expected to result in equal decisions, „if alternatives are distinct, non-random and clearly different"[Arthur, 1993] (observed in figure 3). Equal results for the *Experiential*- and the

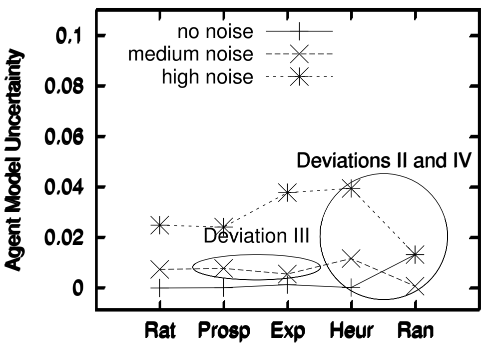


Figure 6: Agent model uncertainty for scenarios with no fertilizer threshold.

RationalFarmer decision model if noise is absent are not observed for higher fertilizer thresholds t_f because of the saturation in the parabolic yield function (Figure 2). Note, that the accrued experience based decision allows the *ExperientialFarmer* decision model to perform better than the *RationalFarmer* decision model on medium noise levels (lower AMU)

In figure 6 the increasing AMU for increasing deviation from rationality can not be observed as expected. In the case of the scenario with no fertilizer limitation, the maximum achievable yield is usually the most rational choice, since additional fertilizer costs are generally low compared to increase in income. All decision algorithms except for the *RandomFarmer* are able to choose a high yield goal if not restricted by a fertilizer threshold. Therefore, almost no difference is observed for the various decision models.

We expect higher AMU as the noise level increases (hypothesis 2). Higher agent model uncertainties are observed for higher noise levels in all four figures, as expected. The single exception is observed in figure 6, where the AMU for the medium noise level is lower than for the no noise level (Deviation IV). However, noise does not increase uniformly for all decision models in contrast to the expectations raised by hypothesis 2. Sensitivity of agent model uncertainty on noise (defined as the difference in AMU between the high-noise and no-noise scenario) is higher for the *Heuristic*- and *RandomFarmer* decision model compared to the other three decision models in figures 3 through 5. No explanation is available for the differing sensitivities toward noise.

7 CONCLUSIONS

In view of the simulation results, the two hypotheses fall. If our assumptions hold, agent model uncertainty does not increase linearly with increasing deviation away from full rationality. The agent model uncertainty of different decision models changes non-uniformly in response to increases in model noise. Even if the assumed ordering of decision models according to increasing deviation from full rationality is wrong, there is no possible order under which hypothesis 1 would stand given the results of our experimental model simulation. Under the assumption that AMU is a useful measure for agent model uncertainty, this means that it is necessary to consider testing decision models with different choice of rationality in agent based models, since uncertainty about which decision model to use can be a significant source of uncertainty in the outcome of such models [Hare and Pahl-Wostl, 2001, see also]. Special care is also necessary in interpreting the basic uncertainty of different decision models if the level of noise is high in the model, since different decision models have different sensitivities toward noise.

More research is necessary to test whether the uncertainty about which decision model to use is still a significant source of uncertainty with a non-linear influence if different model output parameters (such as income) for calculation of AMU are used. We expect this to be the case based on the findings of Hare and Pahl-Wostl [2001], where results for profit also showed increases in uncertainty level as a result of changing away from rational farmer towards a more boundedly rational one. While learning is not included as a separate topic in the classification scheme in figure 1, simulation results from the *ExperientialFarmer* decision model suggest that learning processes are of high importance for agent rationality and should be further investigated. This includes farmer agents with a heterogeneous mix of decision models.

In order to investigate the full range of agent rationality, it is suggested to test agent-based models with a set of at least four decision models, one from each of the following four categories: 1) A fully rational model can be used as a reference for a decision model that is based on optimization in a perfectly known world. As an alternative, a somewhat more realistic decision model according to prospect theory could be used. Differences in results are small for the two models and both models are based on *full information availability*. 2) A *learning* decision model such as the *ExperientialFarmer* should be included as well in order to investigate importance of non-linearities for learning processes. 3) A „likeliest case" empirical model (e.g. the *HeuristicFarmer*) imitating decision mechanisms of relevant stakeholders will allow to estimate the degree of deviation between results from the „likeliest case" model and results from the other decision models. Of course such a „likeliest case" model must be based on investigations to try to understand how people actually decide. 4) A random decision model should finally be the reference for maximum deviation from rationality.

References

Arthur, W. B. On designing economic agents that behave like human agents. *J. Evol. Econ.*, 3:1–22, 1993.

Bock, B. R. and G. W. Hergert. Fertilizer nitrogen management. In Follett, R., Keeney, D., and Cruse, R., editors, *Managing Nitrogen for Groundwater Quality and Farm Profitability*, Soil Science Society of America. Madison, WI, 1991.

Hare, M. and C. Pahl-Wostl. Model uncertainty derived from choice of agent rationality - a lesson for policy assessment modelling. In Giambiasi, N. and Frydman, C., editors, *Simulation in Industry: 13th European Simulation Symposium*, pages 854–859, 2001.

Kreps, D. M. *Notes on the Theory of Choice (Underground Classics in Economics)*. Westview Press, Colorado, 1988.

McFadden, D. Rationality for economists. *Journal of Risk and Uncertainty,*, 19(1):73–105, December 1999.

Morgan, M. G. and M. Henrion. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, Cambridge, 1990.

Pahl-Wostl, C. Agent based simulation in integrated assessment and resources management. In Rizzoli, A. and Jakeman, T., editors, *Integrated Assessment and Decision Support. Proceedings of the 1st biennial meeting of the International Environmental Modelling and Software Society*, volume 2, pages 239–250, 2002.

Reusser, D. E. Relating choice of agent rationality to agent model uncertainty - an experimental study. Diplomarbeit, Universität Osnabrück, 2004.

Russell, S. and P. Norvig. *Artificial Intelligence, A Modern Approach*. Prentice-Hall International (UK) Limited, London, 1995.

Sent, E.-M. Sargent versus simon: bounded rationality unbound. *Cambrdige Journal of Economics*, 21:323–338, 1997.

Tversky, A. and D. Kahneman. Advances in prospect theory, cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5: 297–323, 1992.

Coupled Human and Natural Systems: A Multi-Agent Based Approach

Monticino, M[a], Acevedo M., Callicott B., Cogdill, T, Ji M., and Lindquist, C.

University of North Texas, Denton, Texas, U.S.A. ([a]monticno@unt.edu)

Abstract. A major force affecting many forest ecosystems is the encroachment of residential, commercial and industrial development. Analysis of the complex interactions between development decisions and ecosystems, and how the environmental consequences of these decisions influence human values and subsequent decisions will lead to a better understanding of the environmental consequences of private choices and public policies. Determining conditions on the interactions between human decisions and natural systems that lead to long-term sustainability of forest ecosystems is one goal of this work. Interactions between human stakeholders are represented using multi-agent models that act on forest landscape models in the form of land-use change. Feedback on the effects of these actions is received through ecological habitat metrics and hydrological responses. Results are presented based on a study of a riparian area of the Dallas-Fort Worth (Texas, U.S.A.) region facing intense residential development.

***Keywords**:* Biocomplexity; Multi-agent models; Land-use change dynamics; Decision models.

1. INTRODUCTION

Few ecosystems are free of extensive human influence. A major force affecting many forest ecosystems is the encroachment of residential, commercial and industrial development. The complex interactions between development decisions and ecosystems, and how the consequences of these decisions may then influence human values and subsequent decisions is an important area of study. Analysis of these interactions will lead to a better understanding of the environmental consequences of private choices and public policies. This paper presents a coupled natural-human system model and analyzes the dynamics of land-use change under various scenarios for a rapidly urbanizing region of north Texas. The main focus here is on the human component, which uses multi-agent models to capture essential features of the decision processes and stakeholder values that lead to land-use changes. This work is part of an interdisciplinary Biocomplexity in the Environment project supported by the National Science Foundation with study sites in north and southeast Texas, and two study sites in Venezuela. Results from one of the Venezuela study sites are reported in another paper in these proceedings (Barros, et al [2004]).

The agents represent a variety of interacting human stakeholders, including municipal governments, land developers, landowners of large tracts of undeveloped land, and homeowners. For example, homeowner agents may decide to "protest" a proposed commercial development, thereby affecting the government agent's decision of whether to approve the development. A government agent's approval of a protested development may then lead to homeowner agents voting a new government agent into office. The decision models used by the stakeholder agents are based on decision-analysis utility functions derived from quantitative and qualitative surveys. As noted in Hoffmann, et al. [2001], the multi-agent approach accounts for the complex interactions between stakeholders that are an essential part of land-use change dynamics. The decision analysis framework provides a flexible structure for investigating likely outcomes of growth management strategies and the sensitivity of these outcomes to variations in stakeholder values. Barros, et al [2004] use a logic-based approach to model agent behavior as their work places less emphasis on encoding stakeholder values. The two approaches supplement each other and allow for future comparisons of coupled system dynamics

The natural systems portion of the coupled model includes a land-cover transition model, a hydrological model and a wildlife habitat model. The structure of each of these components is generic enough to accommodate the various study sites in the overall project, and yet allow the level of detail necessary to accurately represent specific systems. Thirteen land cover types, based on remote sensing studies (Newell, et al. [1997], CWRAM [2002]) are used for the north Texas study site. The types can be broadly categorized as vegetated-natural, vegetated-managed, and developed. Dynamics within the vegetated-natural category are dominated by succession from oldfield to wetland, upland or bottomland forest depending on topography. Succession is modeled with MOSAIC using parameters estimated from detailed gap-model simulations (see Acevedo, et al. [2001], and Monticino, et al. [2002]). Vegetated-managed dynamics and transitions to developed types are controlled by the human system model. All the natural systems models provide feedback to the human system. The land-cover transition model provides land-cover maps; the hydrological model outputs metrics derived from rainfall runoff, sediment yield, and nutrient concentration; and the wildlife habitat model gives metrics related to habitat quality.

2. DECISION FLOW

2.1 Study Area and Agent Classes

The study area represented by the model is a region of north central Texas (Denton County), U.S.A., experiencing rapid residential and commercial growth. Denton County grew from a population of 273,575 to 504,750 from 1990 to 2003. From 1995 to 2000, the percent of developed land doubled from 13% to 26.8%; and, in just the two-year period from 2000 to 2002, the number of housing units increased by over 10% (NCTCOG [2003]).

While this paper focuses on modeling the essential features of the decision processes that lead to land-use changes in this study area, an equally important objective of the work is developing a model framework flexible enough to be adapted to regions with other land-use dynamics and stakeholder interactions. In particular, the model was designed so that it would be straightforward to include other decision attributes, value systems and available actions.

A representation of the process of land-use change was developed for the study area based on formal focus group sessions and quantitative surveys of area residents, local developers, real estate agents, large landowners, and municipal government officials. Four main classes of stakeholders are defined. Landowner agents represent owners of large (undeveloped) parcels of land suitable for residential, commercial or industrial development. Developer agents model residential, commercial or industrial land developers. Homeowner agents represent collections of municipal residents within a particular tract of land. (Homeowner agents are assigned a weight representing the number of residents in the tract and their influence on land-use decisions – e.g., homeowner agents representing a large number of high-income residents are assigned a higher weight than agents representing sparsely populated low-income tracts.) Government agents characterize municipal governments that can approve, modify or reject a development proposal. Several types of agents are defined within each agent class. As discussed in section 3, agent types are characterized by value structures that influence the actions selected by the agent.

2.2 Agent Interactions and Decision Flow

The model is initialized by setting values for two sets of parameters. The natural system model uses the first parameter set. These parameters characterize the current land-use and cover type of each parcel of land in the study area. A parcel's description also includes physical metrics such as its percent of impervious surface and/or soil type, its slope and elevation. The natural system uses land-use information both to model the succession dynamics of undeveloped land and to provide feedback to the human system. For example, peak water flows from rainfall runoff at various points in the study area are passed to the human system to provide information to the stakeholder agents on how land-use changes have affected flooding patterns in the region. The second set of initialization parameters is used by the human system. These parameters involve ownership assignments to undeveloped parcels of land, assigning agent types to residential and undeveloped land parcels, and assigning the initial type of government agent. Once initialized, the decision/information flow between stakeholder agents and between the natural and human systems proceeds as follows. (A model schematic is available at www.geog.unt.edu/biocomplexity.)

- At the beginning of a time step (typically a one year increment), landowner agents decide whether to hold or to sell their land. If the decision is to sell, then the land becomes available to developer agents. Landowner agents that decide to sell their land become inactive in the model.

- Once land is made available for development, then a development category – residential, commercial or industrial – is selected probabilistically based on a development-potential map for the region. This map gives the likelihood of a development category based on factors such as proximity to roads, proximity to other developments, and inclusion in municipal jurisdictions.
- After the development category is chosen, a developer type is selected. Developer types are characterized by the development proposals they will make. The developer type is selected probabilistically as a function of the current type of government agent.
- The developer type selected submits a development proposal to the government agent. Homeowner agents affected by the proposal are also notified of the proposal.
- The homeowner agents then decide whether to protest the proposed development or not. The protest decision is based on the homeowner agent type, the development proposal, and the type of residential development in which the homeowner agent resides.
- The government agent decides whether to approve, approve with modifications, or reject the development proposal. The decision is based on the government agent type, proposal type, weights of the homeowner agents protesting the proposal, and environmental information provided by the natural systems model.
- Once government agent decisions are made for all pending proposals, any changes in land-use are passed to the natural system model. Any parcel that has become a residential development is assigned a homeowner agent. The agent type and weight is a function of the type of proposal approved.
- Before the next time increment, the human system model receives input (e.g., rainfall-runoff and landscape fragmentation information) from the natural systems model on the effects of the approved land-use changes. Based on this information and the government agent's decisions, homeowner agents may modify their values – i.e., change type.
- Homeowner agents then vote on the government agent type that will be in power for the next time iteration. Different homeowner agent types vote for the various government agent types with different probabilities. Election results are determined by the weights of the homeowner agents casting ballots. The new government agent is in place at the start of the next time increment.
- The next iteration begins again with the current set of landowner agents deciding whether to hold or sell their land.

3. AGENT DECISION MODELS

3.1 Decision Analysis Overview

Agents select their actions from a specified set of available actions. Intuitively, agents select the action that best conforms to their values. These values are quantified within a statistical decision analysis framework (see, for instance, Keeney and Raiffa [1993]). The decision analysis (DA) framework encodes the value tradeoffs and uncertainties inherent in stakeholder decisions. Mathematically, agents evaluate the worth of each available action according to a multi-attribute utility function and then select that action with the highest expected utility. Utility functions were developed from focus group sessions for the landowner, developer and government agent classes and from a formal conjoint analysis survey for the homeowner agents. The DA framework provides a consistent structure for adapting the model to other study areas where stakeholders may have different available actions and value structures. It is not uncommon to observe that elicited value models and the resulting decisions prescribed by a DA model may differ from the decisions actually observed – people are not always rational decision makers. However, the DA models used here provide important benchmarks for investigating the effect of growth management strategies on land-use dynamics, and for evaluating the sensitivity of these dynamics to variations and temporal changes in the elicited value structures.

3.2 Multi-attribute Utility Functions

Faced with making a decision, agents first define the set of possible consequences, $\{c_1(A), c_2(A), \ldots c_m(A)\}$, and their respective probabilities, $\{p_1(A), p_2(A), \ldots p_m(A)\}$, for each available action A. The value of consequence $c_i(A)$ is evaluated with respect to an additive multi-attribute function of the general form
$U(c_i(A)) = k_1 U_1(c_i(A)) + \ldots + k_n U_n(c_i(A))$.
The functions U_j represent the partial utilities of value attributes associated with the decision. The constants $k_1, k_2 \ldots k_n \geq 0$ indicate the relative value that the agent places on the respective

attributes. Following standard practice, the partial utilities functions take values between 0 and 1, and $k_1 + k_2 + \ldots + k_n = 1$. The expected utility of action A is $E[U, A] = \sum_{i=1}^{m} p_i U(c_i(A))$. Agents select the action with the maximum utility.

3.3 Landowner Agents

Each privately owned undeveloped parcel of land is assigned a landowner agent. Landowner agents (LAs) are assigned an initial wealth and a number of years that they have owned their parcel at initialization time. For many regions, the time that a landowner has owned a parcel is available from government records. If not, landowner agents are randomly assigned an ownership time. An agent's initial wealth is based on the assessed value of the land (from government records) and the current land-use. A landowner's value for wealth is assumed to follow a classic decreasing marginal utility model given by $U_W(m) = 1 - e^{-Rm}$. The value of the constant R characterizes the rate at which additional wealth is discounted (R can also be viewed as a measure of risk aversion). Each LA is assigned a value for R at initialization. Using a decreasing marginal utility model and assigning an initial wealth to each LA allows the model to represent landowners with different sensitivities to farming/ranching income and to changes in land prices.

Two actions are available to LAs – hold their land and maintain its current use, or sell it. Expected utility calculations are based on the possible consequences of each action with respect to three value attributes – wealth, tradition value and neighboring land-use. Wealth is the monetary return from an action – farming or ranching income if the land is held, or profits received from selling the land. Agents assess monetary return based on an economic trend model that provides nominal, high and low values (along with respective probabilities) for land prices and the present value over given time horizon for farm/ranch income. Land prices are also affected by government agent actions that tend to increase the cost of development. The partial utility for wealth is U_W. Tradition value represents the intrinsic worth of the land to the landowner. A farm that has been in a family for several generations may have a higher tradition value than a recently purchased "hobby" ranch. Accordingly, the partial utility for tradition, U_{Tr}, is a non-decreasing function of the time that the parcel has been owned by the LA. Neighboring land-use indicates the type of land-use surrounding the landowner's parcel. This attribute provides a way to measure the desirability of maintaining rural land-use when surrounded by residential or commercial development. The partial utility for neighboring land-use, U_{NL}, is a decreasing function of the percentage of developed land bordering the landowner. LAs project historical development trends to evaluate the potential value of U_{NL} for their current land and for a new location if they were to sell. The utility function for LAs is given by $U = k_W U_W + k_{Tr} U_{Tr} + k_{NL} U_{NL}$. The attribute weights, k_W, k_{Tr} and k_{NL}, indicate the relative value that a landowner places on wealth, tradition and neighboring land-use. Each LA agent is assigned a set of attribute weights. LA types are defined by their attribute weights along with their initial wealth and wealth discount rate. For example, taking $k_W = .6$, $k_{Tr} = .1$, and $k_{NL} = .3$ represents landowners primarily interested in wealth maximization, while taking $k_W = .3$, $k_{Tr} = .4$, and $k_{NL} = .3$ models landowners placing a higher value on the intrinsic worth of their land.

3.4 Developer Agents

There are three types of developer agents for each development category, labeled environmentally-sensitive, environmentally-moderate, and environmentally-insensitive. Developer agent types are characterized by the type of development that they are most likely to propose. For example, environmentally-sensitive residential developer agents are most likely to propose developments that preserve a high percentage of existing tree cover and leave more open space. Three development types are classified within each development category – environmentally-sensitive, environmentally-moderate, and environmentally-insensitive. Metrics defining the classification includes housing density, percent impervious surface, percent tree cover, and pollution emission. The likelihood of selecting a given developer agent type is a function of the government agent type and the development category. For example, if a progressive government agent is currently in office, then an environmentally-insensitive commercial developer is less likely to obtain a parcel than if an economic-growth government agent was in office. The likelihood of a developer agent type proposing a given development type is a function of the developer type and the government agent type.

3.5 Homeowner Agents

Two actions are available to homeowner agents (HAs) when faced with a neighboring development proposal – to protest the development, or not. An HA's utility function involves four attributes – economic property value, residential setting, neighboring land-use, and community effort – giving the utility function
$U = k_{EPV}U_{EPV} + k_{RS}U_{RS} + k_{NL}U_{NL} + k_{CE}U_{CE}$.
The partial utility for economic property value evaluates the consequence of a proposed development on the agent's home value. Residential setting represents the compatibility of residential development within the HAs immediate locality. Neighboring land-use corresponds to the suitability and perceived environmental effect of commercial and industrial land-use in a wider neighborhood around the agent. Community effort measures the perceived effort in taking a particular action. Four types of agents are defined – apathetic, property-value, neighborhood, and environmentalist. HA types are characterized by the form of the partial utility functions and the attribute weights. For example, an apathetic HA has a large value for k_{CE} and a partial utility U_{CE} that decreases rapidly as a function of perceived effort, making it unlikely that an apathetic HA will protest a development proposal. On the other hand, environmentalist HAs have high values for k_{RS} and k_{NL}, and are likely to protest most commercial and industrial development proposals. Property value HAs have a high k_{PV} value and the partial utility function U_{PV} is sensitive to decreases in property value. Neighborhood HAs place a high weight on residential setting.

The expected utility of an action is calculated by specifying the possible consequences of a development proposal with respect to each attribute and the respective probabilities of these consequences. Consequence probabilities are a function of the action, development proposal, HA type, and current type of government agent.

The probability of an HA changing to another type is a function of the development decisions made, the natural system feedback, and the current HA type. For example, if a property-value HA protested a commercial development eventually approved by the government agent and localized flooding increased because of parking lot runoff, then the agent is likely to change to an environmentalist agent. After possibly changing types, HAs vote for the type of government agent. The probability of voting for a particular government agent type depends on the HA type. Environmentalist HAs will vote for a progressive government agent with a high probability, while property-value HAs are more likely to vote for an economic-growth government.

3.6 Government Agents

Given a pending development proposal, the government agent (GA) selects one of three actions – approve, conditionally approve at a higher environmental sensitivity level, or reject. GAs select their action based on four attributes – business relations, citizen relations, environmental consequences, and tax base effect. Their utility function is $U = k_{BR}U_{BR} + k_{CR}U_{CR} + k_{EC}U_{EC} + k_{TB}U_{TB}$.
GA types are defined by the form of the partial utility functions and the attribute weights. Economic-growth GAs have attribute weights $k_{BR} = .4, k_{CR} = .1, k_{EC} = .1$ and $k_{TB} = .4$. Moderate and progressive GAs place more weight on community relations and environmental consequences.

The consequences of each action and their respective probabilities are evaluated with respect to the partial utility functions. For instance, the community-relations partial utility of approving an industrial development in spite of protesting HAs will be small; whereas, the business-relations partial utility approval will be high. Perceived environmental consequences of a potential action are a function of the GA type and feedback received from natural system model on environmental consequences of previous land-use decisions. As with the other agent classes, the GA evaluates the expected utility of each action and selects that action with the highest value.

4. SIMULATION RESULTS

Land-use change dynamics were simulated for several scenarios, varying by the initial distribution of landowner, homeowner and government types, and economic model assumptions. The model produced land-use change dynamics qualitatively similar to those observed in the study area. For instance, starting with an economic-growth GA, increasing land prices and stagnant farm/ranch income (as seen in the north Texas study site), LAs steadily sold their land for development. The first to sell were those with low to moderate personal wealth and who placed a high value on wealth. As more land was developed, LAs placing weight on neighboring land-use and tradition begin to sell. Eventually, only LAs placing a very high value on tradition and who were initially next to existing development were left. When government decisions on development proposals had only a moderate effect on land price trends, changes in land-use occurred fairly rapidly, before changes in homeowner and government types had an effect on land-use dynamics. On the other hand, when government development decisions had a more substantial effect on land prices, an interesting oscillatory effect was noticed. As initial landowners sold and development occurred, more homeowners began to protest and the government did not approve as many developments. This dampened the increase in land prices and slowed the rate that landowners sold and so slowed development. Homeowners and government then became less active, land prices started climbing again and another burst of development occurred with the subsequent increase in homeowner and government activism. Comparing these development cycles to empirical data is part of current model validation work. Simulations also investigated effectiveness of variations of proactive growth management strategies. One strategy that has been suggested is purchasing landowners' development rights in order to create open-space preserves. Landowners retain all land-use rights except development. Simulations were conducted to examine effective ways to selectively purchase development rights – in particular, investigating ways of purchasing rights so as to leverage the neighboring land-use values of other landowners to effectively take more land out of development. Two simple scenarios were compared. First a corridor of undeveloped land was set aside, and second the same amount of land was set aside but scattered across the study area. Both strategies generally resulted in land other than that set aside not being developed (in the absence of open-space preserves this land was developed). The scattered open-space scenario consistently resulted in a higher proportion of undeveloped land over a 25-year time horizon. Thus, neglecting any ecological disadvantages, scattering open-space preserves appears to offer a higher likelihood of limiting development.

5. CONCLUSIONS

The goal of this work was to develop both a specific model for the study area and a general framework that captures essential features of land-use change dynamics. Simulations produced qualitative patterns of land-use change similar to those observed in the north Texas study area. This helps validate the overall modeling approach as other sites are studied and more quantitative results are derived from the model. The simulations also illustrated key sensitivities of land-use dynamics to model assumptions. Principal drivers of land-use change are the land-price model and the sensitivity of the landowner agents' decision about whether to sell to changes in land prices. Accordingly, an important component of future research will be eliciting landowner values through quantitative surveys and developing a more comprehensive economic model. The model also indicates that decisions by resident agents to protest developments and subsequent government agent decisions to limit development may have effects in controlling land-use change over and above the specific properties targeted. Moreover, development management strategies may be augmented by geographically dispersing, when possible, open space preserves.

6. ACKNOWLEDGEMENTS

This work was supported by an NSF Biocomplexity in the Environment grant CNH BCS-0216722. We wish to thank the other grant participants for their valuable contributions.

7. REFERENCES

Acevedo M.F., S. Pamarti, M. Ablan, D.L. Urban and A. Mikler, Modeling forest landscapes: parameter estimation from gap models over heterogeneous terrain. *Simulation* 77, 53-68, 2001.

Barros, R., Davila, J., Moreno, N., Quintero, R., Ablan, M., Modeling of the biocomplexity at Caparo Forest Reserve, submitted to the Proceedings of the Second IEMSS Conference, Osnabruck Germany, 2004.

CWRAM, Watershed atlas for Ray Roberts, Lewisville, and Grapevine (Center for Watershed and Reservoir Assessment and Management), UNT, 2002.

Hoffman, M., Kelly, H., and Evans, T., Simulating land-cover change in Indiana: An agent-based model of de/reforestation, in *Complexity and Ecosystem Management: The Theory and Practice of Multi-agent Systems*, ed. M. A. Janssen, Northampton, Mass., Edward Elgar Publ., 218-247, 2002.

Keeney, R.L., Raiffa, H., *Decisions with Multiple Objectives*, Cambridge University Press, 569 pp., Cambridge, 1993.

Monticino M.G., Cogdill T. and M.F. Acevedo, Cell Interaction in Semi-Markov Forest Landscape Models. pp 227-232. In: Rizzoli A.E. and A.J. Jakeman (Eds.). *Integrated Assessment and Decision Support*, Proceedings of the First Biennial Meeting of the IEMSS. Lugano Switzerland, 2002.

Newell J., K.L. Dickson, J.W. Fitzgerald, S.F. Atkinson and M.F. Acevedo, A landscape ecological characterization of the Ray Roberts Lake study area, north-central Texas: temporal variations in landscape pattern 1987-1992, *Romanian Journal of Optoelectronics*, 5, 95-110, 1997.

NCTCOG, North Central Texas Council of Governments Demographic Estimates, www.dfwinfo.com, 2003.

Simulating Human Behaviour: the Invisible Choreography of Self–Referential Systems

David. Batten@csiro.au

CSIRO Manufacturing and Infrastructure Technology, Melbourne, Australia

Abstract: Challenges arise when it comes to capturing complex patterns of human behaviour in agent-based simulations. For example, human beings are not limited to one identity, to local levels of awareness or to acting on predetermined rules. Seemingly immune from these difficulties are some *self-referential* problems – situations where agents' forecasts act to create the world they are trying to forecast. The emergent complexity in these systems results more from ways in which agents interact and react rather than from their individual idiosyncrasies. A well-known example is the Bar problem, whose collective regularities are relatively insensitive to the vagaries of individuals. Some other socio-technical, socio-economic and socio-ecological systems of a self-referential nature are discussed. Many self-referential systems are intriguing because there is an air of inevitability about them. They seem to co-evolve in prearranged ways, as if under the spell of an invisible choreographer.

Keywords: Agent-based models; Bar problem; Self-referential behaviour; Sheep and explorers

1. INTRODUCTION

The pace of development in agent-based simulation models linking social, economic and ecological interactions has been rapid. Many simulation models representing facets of human behaviour have emerged [Arthur, 1994; Axelrod, 1997; Gilbert and Troitzsch, 1999, Barreteau and Bousquet, 2000]. Social scientists use simulation for several purposes, including discovery of collective regularities. Some have been built on strong behavioural foundations, where the mental models of the artificial agents are supported by empirical data or stakeholders' views. In others, agents' strategies have been chosen arbitrarily or stochastically and the outcomes tested in computational experiments. Both approaches have yielded interesting results.

In a recent paper to the IBM Systems Journal, Kurtz and Snowden [2003] assert that there are some important contextual differences between the behaviour of human and ant colonies, making it difficult to simulate humans using computer models. Unlike social insects, we are not genetically hardwired for cooperative behaviour. They list several challenges when it comes to capturing complex patterns of human behaviour in agent-based simulations:

(1) Humans are not limited to one identity or any common set of emotions;
(2) Humans are not limited to acting in accordance with predetermined rules;
(3) Humans are not limited to acting on local patterns.

On first sight, such challenges look daunting. Clearly it is difficult to consider all scales of human awareness simultaneously, instead of choosing one circle of influence when devising mental models to represent human behaviour in a specific context. On further reflection, however, these challenges may not be critical for simulating certain social collectives. One class of social systems that seems relatively immune is *self-referential* systems – situations where the forecasts made by agents serve to create the world they are trying to forecast. The emergent complexity in these systems arises more from ways in which the agents interact and affect each other and less from each agent's individual idiosyncrasies.

The purpose of this paper is to look at one self-referential system (the Bar Problem) and show that its collective outcomes are insensitive to the vagaries of individuals. Although it does not necessarily follow that all self-referential systems have this property, there is a similar air of inevitability about many of them. They appear to co-evolve in prearranged ways, as if under the spell of an invisible choreographer.

2. THE HUMAN IDENTITY PROBLEM

2.1 K&S Argument Number 1

K&S stress that "In a human complex system, an agent is anything that has identity, and we constantly flex our identities both individually and collectively." We play different roles at the same time – e.g. parent, spouse, employee or neighbour – and will behave differently depending on the context. Collectively, we might belong to a dissenting community group. When faced with a common threat, however, we might change this identity and throw our support behind the same government that prompted our dissent. Their key point is that it is often impossible to know in advance which unit of analysis we are working with. K&S are critical of several attempts to overcome this "unit of analysis" problem in the social simulation literature:

(1) They argue that identity goes deeper than *norms*, a concept often used to explain group behaviour [Axelrod, 1997];
(2) They argue that much internal diversity and patterning is suppressed if individuals are modelled as agents;
(3) They argue that use of an idea or "meme" as the unit of analysis is insufficient to capture the dynamics of multiple identities [Dawkins, 1976].

Despite the weight of their arguments, the identity problem may not always be as serious as K&S suggest. Who a person is may not be as vital for exploring the emergent properties of certain human collectives as representing the heterogeneous mix of agents' behavioural strategies and their interactions in the appropriate co-evolutionary context. When a mix of strategies co-evolves over time, any flexing of individual identities is unlikely to have much impact on collective outcomes. One setting where individual identities play a limited role is the bar problem, an early agent-based simulation of a complex adaptive system reported by Brian Arthur [1994].

2.2 The Bar Problem Defined

Consider a system of N = 100 agents deciding independently each week whether or not to go to their favourite bar next Thursday. Space is limited, so the evening is enjoyable only if the bar is not too crowded (say $N_{max} = 60$). There is no collusion or prior communication among agents. Knowing the bar attendance over the past few weeks, each bar-loving agent simply decides independently to go if he expects less than N_{max} to attend and to stay home if he expects more than N_{max} to go.

This problem is a metaphor for a broad class of social situations: e.g. urban traffic congestion, canteen crowding, queue lengths at big events, fishing strategies and many other commons or coordination problems. It has some interesting properties. First, if a decision model existed that agents could rely upon to forecast attendance, then a deductive solution would be possible. No such model exists. Irrespective of past attendance figures, many plausible hypotheses could be adopted to predict future attendance. Because agents' rationality is bounded, they are forced to reason *inductively*. Second, any shared expectations will be self-defeating and broken up. If all agents believe *most* will go, then *nobody* will go. By staying home, that common belief will be destroyed. If all agents believe *few* will go, then *all* will go, thus undermining that belief. The result is that agents' expectations must always differ.

Perplexed by the intractability of this problem, Arthur created a computer simulation in which his agents were given attendance figures over the past few months. Also, he created an "alphabetic soup" of several dozen predictors replicated many times. After randomly ladling out k of these to each agent, each kept track of his k different predictors and decided whether to go or not according to a preferred predictor in his set. This preferred predictor could be chosen in a variety of ways, although Arthur adopted the currently most accurate predictor for each agent in his simulations.

Each predictor is a means of deciding between two simple alternatives: GO or DON'T GO. It could be tied to a much richer set of decision criteria, including an agent's multiple identities or moods. When deciding whether to go to our favourite bar, for example, our simultaneous roles as a parent, a spouse or employee impact differently on our decision at different times. Provided the desire is to go to an uncongested bar each week, however, a richer soup of identity-flexing hypotheses is unlikely to alter the results of Arthur's simulations (see the next section) in any qualitative way.

3. THE INTENTIONALITY PROBLEM

3.1 K&S Argument Number 2

K&S argue how difficult it is to simulate free will and complex intentionality. Simulations have addressed cooperation, reputation, gossip, reciprocity, lying and trust, but are yet to address other aspects – like duplicity, rumour, self-deception, manipulation, stress, confusion, ambiguity and charisma. Although the list is challenging, not all of it is relevant to the aims of those engaged in agent-based simulation. Whether humans do act in accordance with certain predetermined rules or not is not the primary question, but rather can we identify certain patterns when agents behave *as if* governed by predetermined mechanisms (such as certain cognitive processes). In this respect, human being and ants are similar. We still do not know if ants behave according to certain predetermined mechanisms, but at least we can model and study them as if this is the case.

A key difference between the strong-AI school of the seventies (hoping to reproduce human intelligence) and those doing agent-based simulation is that the latter are not expecting to simulate how people will behave in every instance [Gilbert and Troitzsch, 1999]. Their main aims are: (1) to show how heterogeneous micro-worlds of individual behaviours interact to generate macroscopic regularities, and (2) to show how alternative collective outcomes may evolve under different conditions. Both are sensitive to *interactions between agents*, rather than to the agents' individual predilections.

3.2 The Bar Problem Simulated

Once decisions have been made in Arthur's simulated bar, agents learn the new attendance figure, updating the accuracy of their own set of predictors. Then decisions are made for the following week. In this kind of problem, the set of predictors acted upon by agents – called the set of *preferred* predictors – determines the attendance. But the attendance history also determines the set of preferred predictors. We can think of this set as forming a kind of *ecology* (John Holland's term). Of interest is how this ecology evolves over time.

The simulations show that weekly attendance fluctuates unpredictably, but mean attendance always converges to sixty in the long run. The predictors self-organize into an equilibrium pattern or ecology in which (on average) 40% of the preferred predictors forecast above sixty and 60% below sixty. This 40/60 split remains although the population of preferred predictors keeps changing in membership forever. The emergent ecology is rather like a forest whose contours do not change, but whose individual trees do. Similar results appeared throughout Arthur's experiments, robust to changes in the types of hypotheses (read identities or moods). There is another intriguing result. Although, the computer-generated attendance results look more like the outcome of a random process rather than a deterministic one (see Figure 1), there is no inherently random factor governing how many people attend. Weekly attendance is a deterministic function of the individual predictions, themselves being deterministic functions of the past attendance figures.

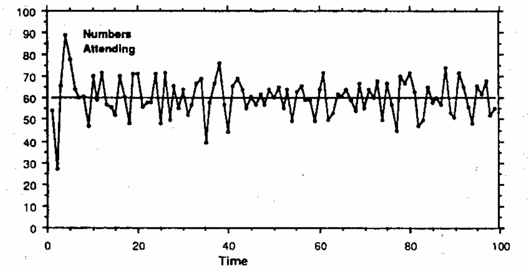


Fig. 1. A simulated 100-week record of attendance at El Farol.

4. SCALES OF AWARENESS PROBLEM

4.1 K&S Argument Number 3

The third challenge raised by K&S is the fact that it is difficult to consider all scales of human awareness simultaneously, instead of one circle of influence when devising a set of mental models to simulate human behaviour in a specific context. Although true, it evokes an earlier statement: people engaged in simulation are not expecting to represent what people may think, feel and do in every context. Instead they aim to explore the co-evolutionary space under "as if" scenarios – collective results that may emerge under particular conditions only. Such conditions are described in terms of the state of the agents, the environment in which they interact and the sets of rules that govern both agents and environment. Since the agents are interacting in or on this environment, the rules and state of the environment play a key part in focusing each agent's mind on the decision at hand and the choice of a pragmatic predictor or heuristic to apply to it.

Part of the art of agent-based simulation lies in identifying settings where the agents' decision criteria are focused and limited in number (e.g. fast and frugal), less emotive or theoretical. Heuristics allow agents to make smart choices quickly in the face of limited information, by exploiting the way that information is structured in some environments [Gigerenzer and Selten, 2001]. In many settings, there is little scope for sophisticated decision tools or emotive states because the decision itself is of a simple, binary kind – "GO or DON'T GO". In these situations, recollections of earlier experiences of a similar kind tend to combine with our current predisposition towards the event, leading to a final decision.

4.2 The Bar Problem Interpreted

The bar problem is a situation where a system of interacting agents can develop collective properties that are not at all obvious from our knowledge of the agents themselves. Even if we knew all agents' individual idiosyncrasies, we are no closer to anticipating the emergent regularities. In the absence of communication between the agents, individually subjective, boundedly rational expectations self-organize (under the influence of a strong aggregate *attractor*) to produce "collectively rational" behaviour [Arthur, 1994]. If we allow the agents to learn and communicate using an evolutionary process (a genetic programming algorithm), heterogeneity among the agents emerges in the form of role-playing and non-uniform social structures [Edmonds, 1999]. All these collective properties are features that emerge purely from the micro-dynamics.

Is the bar problem important? From several perspectives, it would seem to be. First, like the Prisoners' Dilemma, it is receiving more attention outside economics – as a metaphor for learning and bounded rationality. It has inspired a new literature in statistical physics on a closely related problem known as the Minority Game. In this game, each agent chooses one out of two alternatives every turn and those who end up in the minority are the winners. Like in the bar problem, numerical simulations of this game have displayed a remarkably rich set of emergent, collective behaviours [Challet and Zhang, 1998].

Second, the bar problem contains all the key elements of a *complex adaptive system*. It involves a *medium* number of agents, a number too large for hand-calculation or intuition but too small to use statistical methods applicable to very large populations. These agents are *intelligent* and *adaptive*, making decisions on the basis of rules of thumb or heuristics (like the bar predictors). Needing to modify these rules or come up with new ones if necessary, they reason *inductively*. Importantly, no single agent knows what all the others are (thinking of) doing, because each has access to limited information only.

Third, the bar problem is *self-referential* – a situation where forecasts made by individual agents act to create the world they are trying to forecast. Such systems have also been called *reflexive* or *co-evolutionary* [Batten, 2000]. In self-referential systems, the "best" thing to do (e.g. GO or DON'T GO) depends on what everyone else is doing. Since no single agent knows that, the best thing that they can do is to apply the predictor or heuristic that has worked best so far. The remainder of this paper will be devoted to a discussion of other self-referential systems and some examples of their nonlinear, dynamic properties.

5. SELF-REFERENTIAL SYSTEMS

What do these self-referential situations have in common? First, they are "GO or DON'T GO" decision problems at specific locations in space and time. Second, the best thing to do depends on what everyone else is doing at that time. Third, since no agent knows what all the others are doing, agents must decide using a predictor or heuristic that has worked well for them earlier. Finally, there is a risk that agents may be caught up in an undesirable collective outcome – such as congestion.

5.1 Socio-Technical Systems

Many examples of self-referential problems lie in the socio-technical arena. Socio-technical systems are ones in which human beings interact with each other in a physical environment built by humans. Situations of a qualitatively similar kind to the bar problem arise, such as canteen crowding, queue lengths at cinemas, crowds at sporting stadiums, and traffic congestion during peak periods on our roads and at our airports.

Like agent numbers turning up at the bar each Thursday night, the number of vehicles turning up on a specific road each day is unpredictable. If the traffic density is pushed beyond critical levels, however, it triggers unexpected phase changes in the traffic's collective behaviour. Simulation work using Cellular Automata has shown that the average speed drops rapidly once the density passes a critical value – corresponding to a jamming transition from free-flow to start-stop waves. Fluctuations in travel time from vehicle-to-vehicle go up very quickly, reaching a peak near the point of critical density. This emergent behaviour is quite striking.

Of interest are the adaptive strategies of drivers exposed to regular traffic jams. Downs [1962] identified two behavioural classes of driver: those with a low propensity to change their mode or route strategy, called *sheep*, and those with a propensity to change, called *explorers*. Explorers search for alternative options to save time. They are quick to learn and hold several heuristics in mind simultaneously. Sheep are more conservative and prone to following the same option. Empirical work in North America has confirmed the presence of sheep and explorer behaviour in real traffic [Conquest et al, 1993]. Sheepish drivers, who are unwilling to modify their commuting behaviour, made up about one-quarter of the sample.

5.2 Socio-Economic Systems

Socio-economic systems are ones in which human beings interact with one another in an economic environment designed by humans. Examples are stock and commodity exchanges, labour markets and trade networks. Adopting the "Santa Fe" complexity approach highlights their self-referential character. Because agents derive their expectations from an imagined future that is an aggregate result of other agents' expectations, there is a self-referencing of expectations that leads to deductive indeterminacy. As with the above-mentioned traffic example, agents' forecasts combine to create the very same world that they are trying to forecast.

Collections of beliefs and heuristics co-evolve in simulation experiments, revealing emergent features of socio-economic systems. Markets tend to mimic traffic systems. The beliefs and expectations of drivers are constantly being tested in a world that forms from their and others' beliefs and actions [Batten, 1998]. A confused investor is akin to a confused driver! Prediction for each means a beat-the-crowd anticipation of tomorrow's situation (stock prices or travel times). How individual agents decide what to do matters little. What happens depends more on the interaction structure through which they act, that is, who interacts with whom according to which rules.

In real and simulated stock markets, agents' expectations continually react and adapt to a market they create together. Observable states are often poised between the deterministic and the seemingly chaotic (i.e. between simple and complex). Given sufficient homogeneity of beliefs, for example, the standard equilibrium of the literature is upheld. As we turn the dial of heterogeneity of beliefs up, the market undergoes a phase transition and "comes to life" developing a rich psychology. It displays phenomena regarded as anomalies in the standard theory but observed in real markets – speculative bubbles, crashes, technical trading and persistent volatility [Arthur, Durlauf and Lane, 1997].

If we label these two regimes *simple* and *complex*, there is growing evidence that real markets live in the complex regime. There is also evidence that sheep and explorers co-exist in populations of trading agents. Such systems have been called *adaptive nonlinear networks* [Holland, 1988]. There are many such systems in nature and society, such as nervous systems, immune systems, ecologies and economies.

5.3 Socio-Ecological Systems

Socio-ecological systems are ones in which human beings interact with one another and other living systems in a natural environment. Self-referential problems of the "GO/DON'T GO" type arise in these systems, but are rarely recognized as such. Most are commons dilemmas in which agents are over-exploiting natural resources. Examples are degradation of national parks, overfishing of fisheries and destruction of coral reefs.

In fisheries, for example, the "best" thing to do in a fishing vessel definitely depends on what everyone else is doing. Allen and McGlade [1986] explored the implications of different fishing strategies and information flows among vessels. They found that such vessels exhibit one of two strategies – *Cartesian* or *Stochast*. Like sheepish drivers on our roads, Cartesians are risk-averse agents who choose well-known sites with the best possible return. Risk-taking Stochasts direct their search more randomly.

Using agent-based models, information flow among fishing vessels can be shown to have important effects on the dynamics and resource exploitation of a simulated fishery [Little et al, 2004]. Some vessels interact by obtaining information about where other vessels are fishing. Whether they share reliable catch information is unclear, but agent-based models can help to clarify the collective value of information-sharing. Also, they can be used to explore self-administered solutions that do not involve the market or the state.

Since commons problems are management or coordination problems, agent-based models can provide decision support on sustainable management strategies for them. For a review of such models, see Hare and Deadman [2003]. Representing all the essential components of socio-ecological systems is a vastly more challenging task than doing the same for socio-technical or socio-economic systems. In the latter, the slower dynamics (of a road network or an exchange system) can be ignored and the physical or economic environment treated as a constant. In many socio-ecological systems, however, the dynamics cannot be simplified in such a way. Slower and faster processes must be addressed together.

6. CONCLUSIONS

In this paper, we have discussed some *self-referential* systems, a class of social systems to which objections raised by Kurtz and Snowden [2003] may not always apply. Part of the explanation for this lies in the emergent complexity of these systems. At least some of their emergent regularities are less sensitive to individual behaviours and more to ways in which agents interact and react in totality.

The fundamental class of properties of the social world that agent-based simulation is opening to new understanding is that which occurs only in the dynamics produced by the interactions of the agents making up the system. Emergent novelty derives mostly from accumulated interactions between agents and the co-evolutionary learning that it engenders. This reflexive process may induce the traits of agents to change over time. If the collective outcomes of these self-referential systems turn out to be insensitive to the rich spectrum of idiosyncrasies that human beings possess individually, then it is quite reasonable to disregard these idiosyncrasies as inputs to the simulations because they are not central to the context under investigation.

7. ACKNOWLEDGMENTS

Helpful comments from Paul Atkins, George Grozev, Marco Janssen and David Snowden are gratefully acknowledged. The usual caveat regarding author's responsibility applies.

8. REFERENCES

Allen, P.M. and J.M. McGlade, Dynamics of discovery and exploitation: the case of Scotian Shelf fisheries, *Canadian Journal of Fisheries and Aquatic Sciences*, **43**: 1187-1200, 1986.

Arthur, W.B. Inductive behaviour and bounded rationality, *The American Economic Review*, **84**: 406-411, 1994.

Arthur, W.B., Durlauf, S.N. and D. Lane, *The Economy as an Evolving Complex System II*. Reading, MA: Addison-Wesley, 1997.

Axelrod, R. *The Complexity of Cooperation*. Princeton: Princeton University Press, 1997.

Barreteau, O. and F. Bousquet, SHADOC: A multi-agent model to tackle the viability of irrigated systems, *Annals of Operations Research,* **94**: 139-162, 2000.

Batten, D.F., Co-evolutionary learning on networks," in M. Beckmann et al, eds. *Knowledge and Networks in a Dynamic Economy*. Berlin: Springer-Verlag, 311-332, 1998.

Batten, D.F. *Discovering Artificial Economics: How Agents Learn and Economies Evolve*. New York: Westview, 2000.

Challet, D. and Y-C. Zhang, On the minority game: analytical and numerical studies, *Physica A*, **256**: 514, 1998.

Conquest, L.J. Effect of motorist information on commuter behaviour, *Transport Research B* **18**: 101-110, 1993.

Dawkins, R. *The Selfish Gene*. Oxford: Oxford University Press, 1989.

Downs, A., The law of peak-hour expressway congestion, *Traffic Quarterly,* **16**: 393-409, 1962.

Edmonds, B., Gossip, sexual recombination and the El Farol bar: modelling the emergence of heterogeneity, *Journal of Artificial Societies and Social Simulation*, **2**(3): 1-21, 1999.

Gigerenzer, G. and R. Selten, *Bounded Rationality: The Adaptive Toolbox*. Cambridge, Massachusetts: The MIT Press, 2001.

Gilbert, N. and K. Troitzsch, *Simulation for the Social Scientist*. Buckingham: Open University Press, 1999.

Hare, M. and P. Deadman, Further towards a taxonomy of agent-based simulation models in environmental management *Mathematics and Computers in Simulation*, **64**: 25-40, 2004.

Holland, J. The global economy as an adaptive process, in P.W. Anderson et al, eds. *The Economy as an Evolving Complex System*. Reading, MA: Addison-Wesley, 117-124, 1988.

Kurtz, C.F. and D.J. Snowden, The new dynamics of strategy: sense-making in a complex and complicated world." *IBM Systems Journal* **42**(3): 462-482, 2003.

Little, L.R. et al., Information flow among fishing vessels modelled using a Bayesian network," *Environmental Modelling and Software*, **19**: 27-34, 2004.

Strategic Environmental Assessment (SEA) as a Participatory Approach to Environmental Planning

Experiences from a case study with SEA in waste management

Stefan Salhofer, **Gudrun Wassermann, Erwin Binner**

Institute of Waste Management, BOKU – University of Natural Resources and Applied Life Sciences, Vienna, Austria (stefan.salhofer@boku.ac.at)

Abstract: A Strategic Environmental Assessment (SEA) is conducted to ensure that the environmental consequences of plans or programs (land use programs, traffic planning, waste management planning etc.) are identified and assessed. Participation is a mandatory element of the SEA process. All affected parties are required to be represented in the process. Until today, the SEA process has been applied to waste management in only a few cases. This paper describes the assessment of a waste management plan in the province of Salzburg, Austria. The process took place in 2003. As in other cases of SEA, several alternatives were considered and assessed. Participation involved the establishment of a project team including all the relevant authorities and institutions and two expert teams. All decisions throughout the process (e.g. selection of the framework, scenarios and indicators), had to be taken by a majority of the project team. The aim of the process was to pinpoint the pros and cons of the different scenarios rather than identify the "best solution." This paper describes the process and highlights the critical points.

Key words: Strategic Environmental Assessment; LCA; Participation; Waste Management

1 INTRODUCTION

Waste management has developed from the simple transport of waste out of the settlement areas to more complex systems including recycling and prevention of waste. As a result of the increasing complexity of waste management, adequate assessment tools have needed to be developed [Björklund, 2000]. From both a methodological and a practical point of view it is a complex task to compare alternatives with respect to environmental effects, costs and social aspects. In most cases, the antagonistic targets of cost minimisation, reduction of environmental effects and high convenience for the user (mainly of the waste collection scheme) cannot be fulfilled by one scenario. More likely is a constellation in which high costs are linked with high environmental standards and high convenience, whereas low-cost scenarios turn out to be less environmental favourable or less convenient.

A Strategic Environmental Assessment (SEA) is an approach to incorporating environmental considerations in the development of plans and programs. It can also be regarded as a decision support process, especially when applied to the development of a plan. Details on SEA are defined under the Directive 2001/42/EC, which must be implemented by the Member States by July 2004. Participation is an essential element of the SEA process. Besides that, participatory processes have been established with the directive 2003/35/EC, the public access information directive 2003/4/EG and with the methodology of Sustainability Impact Assessment [Kirkpatrick and Lee, 1999].

SEA for waste management up until now has been conducted on a voluntary basis in only a few cases [http://europa.eu.int/comm/environment/eia/sea-studies-and-reports/sea-case-studies.htm]. This case study provides an overview of the SEA for waste management in the province of Salzburg, Austria. The process took place in 2003.

2 SEA CASE STUDY

Initial position

In the province of Salzburg, a mainly rural region in Austria with approximately 500,000 inhabitants, a new plan for municipal waste management needed to be developed. This plan had to determine the goals and implementation of waste management for the future. To enable a broad technical and public discussion, a SEA was started in January 2003. The possible consequences for the environment, society and economy needed to be taken into account during this process. An SEA can be applied to already formulated plans and programs as well as to plans and program in the preparatory phase. In the case of Salzburg, the SEA process was applied while the waste management plan was under development. Thus the environmental report from the SEA will serve as a basis to formulate the waste management plan.

2.1 Description of SEA procedure

With the use of a participatory process it should be ensured that different interests are used to build up synergies as well as partnerships and hence find sustainable solutions as a conjoint decision [Büchl-Krammerstätter, 2003]. Participation expands the programme information and it should help to clarify and stabilise communication and power relationship between stakeholders. As the decision, which stakeholders are included in the process as well as the kind of participation are critical points [Kapoor, 2000] all relevant stakeholders in all decision-making phases were included with equal say.

All the affected parties (Table 1) and two expert teams were involved throughout the process. One of the expert teams developed the waste management scenarios and conducted the LCA modelling, while the other expert team concentrated on the assessment. After defining the work program of the process, all major decisions (e.g. framework, selection of criteria and scenarios) had to be taken by a majority of the project team. Main tasks of the core group were to prepare the process, to prepare and guide the workshops and to prepare the environmental report. Tasks of the project group were to discuss and agree on the rules for the workshops, to discuss and agree to the chosen methodology and assessment criteria, to discuss preliminary results and to discuss and agree on the final results, laid down in the environmental report.

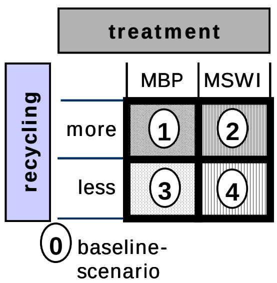
The aim of the process was to pinpoint the pros and cons of the different scenarios rather than identify the "best solution." As described by Finnveden et al., [2003], an SEA can have the function of supporting a choice between two or among several alternatives or of identifying the critical aspects of studied alternatives. Here the latter was the case and the final decision about the future strategy will be taken by the government of the province of Salzburg.

The SEA was divided into five main steps, which were each addressed in a workshop. In the first step the scope of the SEA was defined and general rules as well as a code of behaviour for the following process were outlined. Definitions for the time and the spatial horizon were established. The waste streams to be considered for or excluded from the analysis were listed.

organisation	no of delegates	core group	project team
Provincial gouvernement	1	x	x
Provincial administration	7	x	x
Austria federal ministry of agriculture and forrestry, environment and water management	2	x	x
Expert teams (researchers)	8	x	x
Federation of Austrian Industry	1		x
Economic chamber of Salzburg	1		x
Chamber of labour	1		x
Environmental ombudsman	1		x
Municipalities	7		x
Moderator	1		x

Table 1: participants and their representation

In a second step methodologies for the definition and selection of scenarios and for the assessment criteria were defined. For each waste stream (such as waste paper, biogenic waste, residual waste etc.), the different options were considered for collection, recycling and treatment. Figure 1 illustrates an aray showing four basic scenarios (1-4), supplemented by the baseline scenario 0, built by combining more or less recycling of recyclables (like waste paper or glass) with treatment options (MBP, MSWI) for waste types like residual waste. After listing up the options by the expert teams for each of the waste streams, in close co-operation with the project team, final options were selected and assigned to the scenarios as a common decision of all parties. Thus in this context, a scenario means a combination of options for the single waste streams, including the effects on other waste streams. Here, discussions in small groups pointed out to be a very useful method. Eight scenarios were ultimately developed (step 3).



MBP: mechanical biological pre-treatment

MSWI: municipal solid waste incinerator

Figure 1. General framework for scenarios

Additionally, a baseline scenario was defined to represent the status quo but included a prognosis for the waste quantities in 2012.

In adjustment with the environmental authorities and based on a decision of the project team in a workshop different subjects of protection were determined. The criteria selection (step 4) and the choice of judgment criteria is variable. Individuals may differ greatly, and they may also (unknowingly) be redundant. The choice of judgement criteria and their relative weight in assessing alternatives can be delicate. Feedback in the criteria selection process is both balancing and reinforcing [Nandalal and Simonovic, 2002]. As the choice of criteria should be acceptable to all participants [Balcomb and Curtner, 2000] the whole project team was involved in this process. To identify the criteria and indicators relevant to the assessment, a as complete as possible list of potential impacts was worked out by the expert group, which could be activated by a waste management measure and which have a potential impact to the defined subjects of protection. This list was presented to the project team and a choice was taken. The subjects of protection as well as the possible environmental, social and economic impacts of different waste management procedures were afterwards merged into a so-called "relevance matrix". For each array in this matrix, the relevance of the environmental impact to the affected area of protection was screened. The arrays in which a real impact was detected on the subject of protection provided the building blocks for the rating matrix shown in Figure 2. If available, quantitative indicators (like GWP, AP, etc.) were used. In some parts it was necessary to resort to qualitative indicators. Some of the indicators were used in more than one category. For example, the amount of residues from waste treatment processes has influence both in the category “environmental effects” as well as in “social effects”.

Table 2 shows the selected criteria. For the environmental effects, impact categories were mainly used from the LCA modelling (like HTP, EP, AP, GWP etc.). Quantity and quality of residues from waste treatment processes were used as an indicator for environmental effects of final storage. Additional indicators such as traffic flow (from waste transport), hazardous incidents and land use were used on a qualitative basis. Indicators like traffic flow could also have been used on a quantitative basis. But our previous research [Salhofer et al., 2003], [Wassermann 2003] showed, that it needs a somewhat detailed approach to model the traffic flow from waste transport. Regarding the available time and resources, traffic flow was reduced to a qualitative (non quantified) indicator, whereby more recycling is related to more traffic and vice versa.

For the economic effects, we calculated the cost effects for the waste producers. Regional added value and synergy effects were included only qualitatively. Again, this was due to the time and financial restraints in the process.

For the social effects, almost no additional quantitative indicators could be used. Typical criteria -- odour, noise or user convenience -- were described and assessed on a qualitative basis. Here residues were used as indicator for landfill volume (influence on landscape) and autarky (are adequate disposal facilities available in the region?).

Table 2: Selected criteria by category (simplified)

Category	Quantitative criteria	Qualitative criteria
environmental effects	HTP, TETP, AETP, AP, EP, POCP, GWP, residues	traffic flow, hazardous incidents, land use
economic effects	cost effects for waste producers	regional added value, synergy effects (treatment sites)
social effects	residues, cost for waste producers	appearance, traffic flow, regional jobs provided, odour, noise, convenience, autarky

HTP … human toxicity potential, TETP … terrestrial ecotoxicity potential, AETP … aquatic ecotoxicity potential, AP … acidification potential, EP … eutrophication potential, POCP … photochemical ozone creation potentials, GWP … greenhouse warming potential

After identifying the criteria, the expert teams conducted the assessment of the eight scenarios (step 5). For each array, a comparison was made with the baseline scenario. For the quantitative criteria, a range of ±10% was considered neutral, while a larger difference was assessed as positive (+) or negative (-). Single criteria, which were considered as very sensible, lower threshold values were used. These cases were discussed and decided by the project team. For the qualitative criteria, a comparison was made with the baseline scenario based on arguments documented in the environmental report [Koblmüller et al., 2004].

The assets and drawbacks of each scenario should be shown in a traceable and understandable way. In most multi-criteria methods, the relative importance of criteria is made more precise by some numerical weighting. Although there are a lot of more or less objective possibilities for multi-criteria weighting within the single categories as well as between the categories like normalisation, ABC-Method, AHP-Method, CBA and others [Saaty, 1980, Al-Kloub, 1996, Roy, 1990], in this case it seems rather difficult to evaluate among various criteria for the following reasons.

For weighting it is necessary to translate different quantitative as well as qualitative criteria in uniform quantitative terms. The chosen methodology and the often subjective determined importance of each weight will show the priority set into environmental, social and economic categories as well into the indicators itself by the responsible persons. This subjective preferences mostly depend on the decisions makers. However, each of the objectives is not necessarily equally important to different decision makers [Chambal et al., 2003]. The specific situation in Salzburg, where the SEA took place, of upcoming elections where a political change of the persons in charge was expected, led to the decision to abandon any weighting.

Therefore it was decided to create a rating matrix for each scenario as shown in Figure 2. This allows the user to recognise the advantages and disadvantages of each scenario without a preference for one scenario as the best or worst solution. A description of the process, the assumptions made, the scenarios and the assessment results were documented in the draft environmental report. To conclude the environmental report, comments from the participants in the process are being worked out now. After this the final environmental report will be published. After that the governmental authorities will utilise the environmental report as a basis to formulate the final waste management plan.

impact of waste management measures / Factors/Objectives			air pollution	liquid pollutant emission	noise	residues	traffic	utilisation of resources	sensitivity of WMS	costs
environment	human beings	human health, well-being	+	-	+/-		+		+/-	
	flora, fauna	habitats, biodiversity	+/-	+/-			+		+/-	
	environment	soil	+/-			+/-				
		water	+/-	+/-						
		air	-				+	+/-		
		climate	-				+	+/-		
	resources	raw materials						+/-		
		surface area				+/-				
ec	economy	waste producer								+
		national economy	+/-			+/-			-	
society	society	utilisation interests	+/-		+/-	+/-	+		+/-	
		landscape and cultural heritage				+/-				
		disposal autarky				+/-			+/-	
		job provision								
		convenience	+/-		+/-				+/-	+
		local / regional practicability	+/-		+/-	+/-	+			+

WMS: waste management system

Figure 2. Example of the rating matrix for one specific scenario (simplified).

3 CRITICAL POINTS

During the process several critical points occurred, which will be discussed in the following.

3.1 Selection of criteria

The chosen set of criteria can be classified by their degree of quantification. In detail they are:

- quantitative criteria such as HTP, GWP etc., as generally used in LCA modelling
- qualitative, but figure-based criteria, such as waste transport-related traffic, noise or job effects and
- qualitative, non figure-based criteria, such as synergy effects, autarky and availability of facilities.

Additionally, the criteria vary in their site specifics and spatial dimension:

- local effects are covered by criteria such as land use, noise, residues, regional added value, regional job provision etc. while
- global effects are covered by GWP, HTP, AP etc.

As this list shows, qualitative criteria are used mainly for local effects, while for global effects quantitative indicators are used. The main reason for that is the choice of an LCA approach for the quantitative indicators, which commonly addresses more the global effects. For the qualitative indicators, the local conditions were respected more, but at the same time the analysis was much less detailed. For example, odour and noise were regarded only at the basis of a simple classification (more composting triggers more odour, more recycling causes more noise etc.) In summary, the selection of the indicators reflects the conflicting approaches of the involved parties in the process. While the members of the expert teams were more interested in scientifically acknowledged indicators, the other participants (industry, municipalities,…) placed more emphasis on criteria that reflected current discussions taking place in society (eg. traffic problems) and were more relevant to regional politics.

3.2 Method of assessment

After defining the criteria for assessment, the next important step is to decide whether to weight the criteria.

For the LCA-based environmental criteria, proper assessment methods are available. Two different methods can be distinguished: problem-oriented methods [Guinée et al., 2001] and damage-oriented methods [Goedkoop et al., 2000]. Damage-oriented methods allow for the aggregation between impact categories in terms of common factors. Problem-oriented methods such as CML aim to simplify the complexity of hundreds of mass flows into a few environmental areas of interest. Here the aggregation of results is possible through normalisation or other approaches like the Swiss Eco-factors 1997 [BUWAL, 1998] where the aggregation to one single value is possible.

For the economic effects, a macroeconomic approach is possible but includes practical obstacles, e.g. the regional breakdown of data, the estimation of the regional effects of investments etc.

For social criteria, the aggregation of criteria is possible based on methods such as a value benefit analysis or multi-criteria assessment approaches [Noble, 2004].

Environmental and economic effects can be integrated with a cost benefit analysis. Limitations include the difficulty of estimating the monetary value of the environmental impacts [cf. Fatta and Moll, 2003] which can lead to an overestimation of economic effects and an underestimation of environmental effects.

A weighting for all the criteria in the selected categories of environment, economy and society is also possible with the use of multi-criteria analysis.

In this study as indicated before no weighting took place. As a consequence, an assessment result was obtained for each criterion and field. This led to a large number of single results, for which it was difficult to make an overview and summary. In particular, there were problems with communicating such a large number of results to all the participants in the meeting.

3.3 Result representation and interpretation

After assessing the scenarios, it is necessary to interpret and represent the results in a non-technical self-explanatory form. Especially in our case, where a mixture of quantitative and qualitative assessment methods was chosen, problems occurred during this step. It was decided to provide the results of the analysis in the qualitative way shown in Figure 2. For the qualitative indicators this was quite easy. However, for the quantitative indicators the classification of the results into "positive", "neutral" and "negative" compared with the baseline scenario became a critical point. This part of the evaluation involved forming a judgement on whether or not a predicted effect will be environmentally significant.

Although different methods for solving this problem are known (like normalisation), it was decided to use threshold values. The chosen method meant that the relative differences between scenarios could be large, while the absolute effects is on a very low level. This led to discussions about the height of the threshold values. The relative importance and magnitude of the results for each system in comparison to a reference system could provide more robust results but in the end it is not as self-explanatory and requires additional time and effort.

4 CONCLUSIONS

In our case study, the gap between the requirements of the participatory process and a more scientific approach turned out to be the major stumbling block. Participants from local authorities, industry etc. take a more practical approach (which is helpful in analysing the local situation and circumstances in detail) but have little interest in methodology. In this case agreements on the selected methods were made at an early stage without long discussions. Participation intensified once the initial results were visible. In that stage, some of the participants tried to influence the selection of criteria and modify the assessment method, thus influencing the result.

Summarising, the process of strategic environmental assessment turned out to be helpful, to identify pros and cons of the scenarios analysed, although not all questions could be addressed in a scientific sound way.

5. ACKNOWLEDGEMENTS

This study underlying this paper was funded by the governmental authorities of the province of Salzburg. The authors want to thank Mr. Wilfried Mayr and Dr. Markus Graggaber for the good co-operation.

6. REFERENCES

Al-Kloub, B., T. Al-Shemmeri, and A. Pearman, The Role of Weights in Multi-Criteria Decision Aid, and the Ranking of Water Projects in Jordan, *European Journal of Operational Research, 99*(2),278-288, 1996

Balcomb, J.D., and A. Curtner, Multi-Criteria Decision-Making, Process for Buildings, U.S. Department of Energy, 2000

Björklund, A., Environmental systems analysis of waste management – Experiences from applications of the ORWARE model, Ph.D. thesis, Royal Institute of Technology, Stockholm, 2000.

Büchl-Krannerstätter, K., Bürgerbeteiligung und Umweltschutz, Netzwerkbau 02/03, 2003

BUWAL, Bewertung in Ökobilanzen mit der Methode der ökologischen Knappheit – Ökofaktoren 1997. Schriftenreihe Umwelt Nr. 297, 1998.

Chambal, S., Shoviak, M. and A.E. Thal, Decision analysis methodology to evaluate integrated solid waste management alternatives, *Environmental Modeling and Assessment 8*(1), 25–34, 2003.

Fatta, D. and S. Moll, Assessment of information related to waste and material flows. European Environmental Agency, Technical Report 96, Copenhagen 2003.

Finnveden, G., M. Nilson, J. Johansson, et al., Strategic environmental assessment methodologies – application within the energy sector. *Environmental Impact Assessment Review 23,* 91 – 123, 2003.

Goedkoop, M., S. Effting and M. Collignon, The Eco-indicator 99 – A damage oriented method for Life Cycle Impact Assessment. PRé product ecology consultants, second edition, 2000.

Guineé J., M. Gorrèe, R. Heijungs, et al., Life cycle assessment – An operational guide to the ISO standards. Ministry of Housing, Spatial Planning and the Environment and Centre of Environmental Science - Leiden University, 2001.

Kapoor, I., Towards Participatory Environmental Management, *Journal of Environmental Management 63*, 269 – 279, 2001

Kirkpatrick, C., and N. Lee, Special issue: Integrated appraisal and decision making, *Environmental Impact Assessment Review, 19(3)*, 227 – 232, 1999.

Koblmüller, M., W. Konrad, C. Pladerer, P. Karner, M. Loidl, S. Salhofer, G. Wassermann, E. Binner, J. Stubenvoll, Strategische Umweltprüfung Abfallwirt-schaftsplan Salzburg, vorläufiger Umwelt-bericht, Salzburger Landesregierung, Abt. 16, Salzburg 2004.

Nandalal, K.D.W., and P. S. Simonovic, State-of-the-Art Report on Systems Analysis Methods for Resolution of Conflicts in Water Resources Management, Report Prepared for Division of Water Sciences, UNESCO, 2002

Noble, B., Strategic environmental assessment quality assurance: evaluation and improving the consistency of judgements in assessment panels. *Environmental Impact Assessment Review 24*, 3 – 25, 2004.

Roy, B. The outranking approach and the foundations of ELECTRE methods. *Theory and Decision 31*, 49-73 1991.

Saaty, T.L., The Analytic Hierarchy Process. McGraw-Hill, New York, 1980

Salhofer, S., F. Schneider and G. Wassermann, The Environmental Impact of Transports in Waste Disposal Systems – a Recycling of Refrigerators in an Austrian Case Study. Proceedings of the Ninth International Waste Management and Landfill Symposium; Cagliari 2003

Wassermann, G., The Use of GIS for Optimising Waste Collection. In [eds]: Dhir, R.K., M.D. Newlands and J.E. Halliday: Sustainable Waste Management. Thomas Telford Publishing, London 2003

Land Use change for Salinity Management: A Participatory Model

Nigel Hall[a], Mark Oliver[b], Tony Jakeman[a], Allan Nicholson[c] and Bill Watson[d]

[a]*Integrated Catchment Assessment and Management (iCAM), The Australian National University, Canberra ACT 0200, Australia, nhall@effect.net.au*
[b]*Rural Economic Services, ACT Australia*
[c]*Department of Infrastructure, Planning and Natural Resources NSW, Orange, Australia*
[d]*WAREC, ACT Australia*

Abstract: This paper describes the use of spreadsheet models to help farmers and their advisors to make decisions on land and water use to manage dryland salinity. Salinity management requires an understanding of catchment data and processes. Modelling forms part of the process of understanding the catchment and the turning of data into useable information for salt management. The TARGET project is a NSW government pilot program to support integrated catchment management in selected catchments in New South Wales. A major feature of the program is the simultaneous progress of research and implementation of salinity management measures in a context of adaptive learning. The research described here took place at the same time that extension staff and cooperating farmers were planning and implementing salt management procedures. The dual focus of the project meant that communication was of central importance and this affected the type of modelling carried out. Because of time and data constraints the development of an integrated model of the biophysical and economic system, including spatial and temporal feedbacks, would have had limited value. Instead a partial model was developed that reflected the financial consequences of land use changes and was which was transparent to farmers. The biophysical feedback mechanisms and the external costs and benefits that they imply were external to the model and based on subjective analysis by experts in the field. The paper presents selected analytical results and shows that modelling that is accessible to farmers can best assist salinity management in a context where farmers, advisors, scientists and economists are working together.

Keywords: Salinity management; land use; farm management, spreadsheet model.

1. INTRODUCTION

One of the biggest challenges in model building is working with policy makers, extension staff and farmers in ways that build confidence and contribute to change in policies and farm practices.

The aim of the modelling in this case was to assist policy makers and farmers in making on-farm investment decisions to manage salinity in the catchments of the Lachlan and Macquarie Rivers in New South Wales (NSW) in the Southeastern part of Australia.

This paper describes the process of building a model to analyse the financial consequences of current farming practices and proposed management actions and presents some results from the analysis.

The project involved simultaneous research and implementation over a period of several years. The more usual approach is to do research first then implement the findings but financial and political considerations prevented this. In the event, the approach worked well, because the practical experience of the extension team in the first year agreed with the findings of the modelling team. This led to improved project implementation in the second and subsequent years.

1.1 Context of the modelling

An assessment of salt trends in the Murray-Darling Basin by Williamson et al. [1997] highlighted the severity of salinity problems confronting the Central West Region of New South Wales (catchments of the Macquarie, Lachlan and Castlereagh Rivers). For example, it was predicted that the Macquarie River at Narromine would be unfit for human consumption 30 percent of the time by 2020, and 55 percent of the time by 2050.

Significant efforts will be required to halt or reverse salinity and water quality problems. In many cases, a change at an individual farm level is unlikely to result in much change to what is usually a regional scale problem. According to Hajkowicz et al. [2001] effective solutions may require changes to land use practices and production activities over whole catchments or drainage basins.

In some cases the benefits of management actions accrue to the broader community (eg. biodiversity benefits). This externality in benefits may lead to market failure unless the community is willing to compensate farmers for the costs incurred in protecting, for example, biodiversity.

The *Tools to Achieve Landscape Redesign Giving Environmental Economic Targets Project* (TARGET) is a cornerstone project of the NSW Salinity Management Strategy. A major objective of the TARGET project is to facilitate large-scale land use change in catchment areas that have been identified as being major contributors to Murray-Darling Basin salinity by providing community funding and support.

The TARGET project was funded as part of the National Heritage Trust Murray-Darling 2001 program with joint funding from the Commonwealth of Australia and the State Salinity Strategy in New South Wales.

Management of the TARGET project is the responsibility the Department of Infrastructure, Planning and Natural Resources (DIPNR) in NSW. DIPNR proposed a number of on-farm management actions to target natural resource and environmental hazards, primarily salinity, in each of four selected sub-catchments. Advisory and financial support was provided to farmers in the selected catchments to implement the suggested management strategies.

The selected strategies were largely no-regrets actions which have low cost, if implemented, and include increased use of native, perennial and saline pastures, establishment of farm forestry and saline forestry plantations, increased use of conservation farming practices, intercropping and, increased fencing off of waterways and of remnant vegetation.

2. MODELLING ISSUES

2.1 Farm level models

Modelling of systems can be carried out at a variety of levels or scales. In the farm and natural resource modelling context, the key levels are field or paddock, farm and catchment or region.

The farm is the common management unit for agricultural land use. It is mostly at this level that economic, social and management variables are included in models.

Salinity modelling at the catchment level is particularly important for assessing the implications of hydrological processes beyond the individual farm.

The modelling needs of the project are at farm and catchment level; a representative farm model was adopted to meet these two needs. That is, the model uses a single reference farm to represent the farms in a catchment.

In any group of actual farms, each is likely to have a different resource limitation. For example, one farm may be short of land, another short of capital and a third short of labour. These limitations help to determine the cropping and management decisions for each individual farm. In the case of a representative farm however, these differences are evened out, which may bias the representation of decision-making. This aggregation error is a technical problem with using an average farm that needs to be acknowledged but is unavoidable except under highly restrictive conditions described by Buckwell and Hazel [1972].

2.2 Time structure

Some models represent a single period while others simulate large numbers of periods. Multi-period farm management models typically are run over 20 to 30 years. This is important for long-term investments such as tree crops and to take account of gradual changes often associated with natural resource processes, for example, changes to watertables and salinity levels.

Integrated economic and hydrological models may define different periods for different systems; for example, hydrogeological data may be daily and income data annual. This can be complex to model and have heavy data needs if, for example, daily rainfall over a period of years is to be included.

2.3 Modelling approach

Oliver et al. [2002] reviewed the existing salinity management models in Australia. These included mathematical programming models, simulation models and spreadsheet models.

An example of a mathematical programming approach is the MIDAS family of models, which are whole-farm, profit maximising linear programming models developed in Western Australia. MIDAS models use detailed biological and economic relationships to analyse interactions between enterprises on farms. Pannell [1996] has used MIDAS in salinity research in Western Australia. This type of model was considered to be too complex and too location specific to be ideal for the TARGET project.

Simulation models attempt to reproduce the structure of decisions and feedback in farming and natural systems and are usually based on a specific simulation language. Researchers with a natural science background often prefer simulation models because they allow relative freedom to represent environmental processes in reasonable detail. In general, simulation models take better account of biological and hydrological feedbacks than either mathematical programming or spreadsheet models.

The Australian Bureau of Agricultural and Resource Economics (ABARE), in cooperation with the MDBC and CSIRO, developed a simulation modelling framework that incorporates the relationships between land use, vegetation cover, surface and ground water hydrology and agricultural returns Bell and Heaney [2000]. This model is however based on a large catchment overview that was unsuitable for the small specific sub-catchments used for the TARGET study.

The development of spreadsheet systems such as Excel has made it relatively easy to develop models capable of calculating or solving a variety of financial and statistical functions. Although spreadsheet models are relatively easy to build they often have a relatively short operational life. An example is the FARMULA model, developed in Western Australia and used in a number of salinity analyses by Morrissey et al. [1996]. This model has not been redeveloped since 1996 and is no longer operational.

The workings of spreadsheets are easy to understand and allow relatively quick construction of models. These attributes can be helpful for extension work or where the model is to be used for practical catchment planning. A simple spreadsheet model can facilitate communication and open up discussion that might be inhibited by more technically sophisticated modelling approaches.

Because of these advantages, and the fact that no other models could be used without significant adaptation, the project team concluded that a spreadsheet model should be built from scratch for this project.

2.4 Representative farm

The farm model uses a constructed representative farm to compare costs and returns from each selected management action. This representative farm has been constructed to be broadly representative of properties with salinity management problems in the catchment without breaching the confidentiality of any individual farmer.

The physical characteristics of the representative farm (such as farm size, crop/pasture areas and enterprise types) are based on the results of the producer profiles studies and discussions within the whole project team. The main enterprises used by a majority of respondents were incorporated into the representative farm. Livestock and crop enterprises used by a minority of producers were not included.

Some of the financial characteristics (such as debt level and capital expenditure on plant and improvements) were derived from the producer surveys. Enterprise specific information, particularly variable costs, was based on published gross margin data such as that in NSW-Agriculture [2002]. Forestry information gathered by Hall [2002] was also drawn on.

3. THE TARGET MODEL

3.1 The team approach

A model is a representation of ideas and hypotheses about how a system works. In an integrated modelling system the modeller attempts to incorporate all the information known to the researchers in the model.

Integrated models offer the promise of solving the salinity management problems of a catchment in a single operation. However, they are expensive in both time and data requirements and may become a 'black box' system that is not well understood or trusted by farmers or catchment managers.

Our approach was to use the shared understandings of the multidisciplinary research team as the background to development of one or more simpler numerical models. In TARGET the modelling is an integral part of a process of research and application to control salinity in specified catchments.

This approach allowed use of a straightforward farm management model without attempting to model the biophysical interactions endogenously. The multidisciplinary team as a whole took responsibility for the integration of the modelling.

The modelling team used the judgement of other team members with relevant expertise to take account of the biophysical and social aspects of salinity management in the catchments. There was very limited biophysical data, at farm level, in most of the catchments studied. The farm sustainability survey also found a number of non-financial impediments to salinity management including strong preferences for particular farming systems and family situations. These issues were taken account of in the specification of actions to manage salinity that were restricted to those that were acceptable to farmers and were expected to have the desired biophysical effects.

The project structure relevant to the modelling is shown in figure 1. The Department of Infrastructure, Planning and Natural Resources (DIPNR) have a continuing relationship with both farmers and catchment managers and a strategic role in the research. DIPNR appointed a Project Board that included departmental officers and farmers to supervise the research and ensure that it was integrated with departmental and community aims.

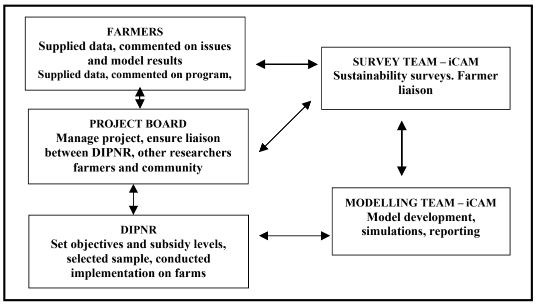


Figure 1. Integrated Team Modelling approach

The Integrated Catchment Assessment and Management group at Australian National University contracted the modelling and survey teams. The survey team included a farm management economist and a hydrogeologist. It was found that involving both disciplines in the farm interviews led to valuable mutual understanding that benefited the modelling and was appreciated by the farmers.

The team visited farmers selected for the survey by DIPNR to investigate their sustainability as well as collect data needed for the model. Sustainability was based on an assessment of the stocks and flows of key sub-systems identified by Watson et al. [2003]. This survey provided farm data for the modelling that was based on the actual catchments studied and could be related back to actual farms. This relating had to be done through the survey team because of farmers' sensitivity to their private data being known to others, including DIPNR. This sensitivity limited their access to the model.

The modelling team, who were also responsible for the survey processing, were briefed on their task by the survey team and DIPNR officers. There was a continuing interaction between the teams, the Board and DIPNR as the model was developed and validated.

The analysis was determined by DIPNR, the survey team and the modellers, in consultation with the Board, with the aim of analysing strategies which DIPNR was encouraging as part of the implementation phase. Farmer feedback on the analysis during the surveys and in Board discussions also influenced the analysis and so the modelling process.

For example, the areas of tree planting for each catchment and strategy were determined by the whole team. This team included farmers, through the Board and surveys, and DIPNA officers, implementing the salinity management programs. In this way the analyses were expected to be realistic and relevant to actual management.

3.2 Model structure

The TARGET model was developed as a multi-enterprise, multi-period, whole-farm analysis tool with an emphasis on 'what if' types of analysis. Most financial inputs (eg. prices and costs) and production inputs (eg. yields, lambing rates) can be readily varied on a yearly basis.

The spreadsheet model consists of seven main worksheets that accommodate a broad range of farm enterprises including a cattle enterprise, two sheep enterprises, up to six broadacre winter crops, fodder crops and fodder production, up to four pasture types and two forestry enterprises. There are also two ancillary worksheets comprising sheep and cattle stocking rate assumptions.

The cattle worksheet, for example, calculates opening and closing numbers by stock category (eg. steer) as well as by age group. Sales, purchases, joinings, births and deaths can be adjusted on a yearly basis if required. The worksheet also calculates total stock sales revenue as well as sales revenue by age group and category. In addition the worksheet calculates up to nine categories of variable costs, total variable costs and costs by age group and category. The other enterprise worksheets are similar in their coverage.

The physical summary worksheet is linked to the cattle, sheep, crop/pasture and forestry worksheets and summarises totals for sheep and cattle numbers, DSEs, crop, pasture and tree areas on a yearly basis over 40 years as well as providing an internal consistency check to ensure maximum areas and stock numbers set by the user are not exceeded.

The financial results worksheet is linked to the cattle, sheep, crop/pasture, forestry and overhead/capital worksheets and summarises sales revenue and total variable costs for each livestock, crop/pasture and forestry enterprise. It also provides a yearly cash flow budget over 40 years. This whole farm cash flow budget shows income from each enterprise as well as other sources, variable costs for each enterprise, overhead and capital costs and calculates NPV ands yearly cumulative debt level.

Most of the enterprise production coefficients, input costs and prices can be varied on a yearly basis to allow the researcher to take account of feedback from environmental degradation over time as well as test the sensitivity of the model to key variables.

The model's analysis period extends to 40 years in order to account for long-term enterprises such as farm forestry. The analysis viewpoint is effectively that of a property manager looking forward into the future. That future will include uncertainty with respect to prices, weather and government policies; therefore, the actual outcomes will not necessarily correspond to the expectations now held. Uncertainty is not internal to the model, like the hydrological issues it is discussed as part of the team approach.

4. RESULTS AND DISCUSSION

The economic analysis used to evaluate the profitability of each management action is the Net Present Value of cash flow on the farm over the analysis period. Six salinity management actions were selected for the analysis:

- Increase perennial pastures to reduce accessions to groundwater by replacing crops or annual pasture,
- Plant saline pastures to use saline areas,
- Fencing of remnant vegetation for conservation,
- Fencing of waterways for conservation,
- Establishing farm forestry to reduce accessions to groundwater and so reduce the spread of salinity,
- Establishing saline agroforestry to use saline land and draw down groundwater in discharge areas,

Selected results are shown in Table 1 for one of the four catchments studied. The others showed similar results. Most of the management actions considered would reduce the Net Present Value of cash flow compared to a continuation of current land use into the future. However, with the exception of farm forestry and fencing waterways, all actions produced farm incomes within one percent of the base scenario. These results include the effect of assistance to farmers under the TARGET program. Planting more perennial pasture was the only activity predicted to increase farm incomes.

Table 1 Modelling results for Warrangong catchment

Salinity mitigation measure	% change in Net Present Value
Increase perennial pasture area	0.2
Increase saline pasture area	-0.1
Fence-off remnant vegetation	-0.5
Fence-off waterways	-1.1
Establish farm forestry	-7.7
Utilise saline agroforestry	-0.1

The results presented in table 1 take no account of the environmental benefits that may flow from implementation of any of the management options. The low cost of most of the proposed actions means that the benefits of most of the management actions would not need to be large to make them worth adopting.

The results of the study were presented to the Board, to DIPNA, to the farmers in each catchment and to a two-day workshop open to the public. The surveys and modelling were accepted as providing valuable information about the costs of salt management procedures. Individual farmers did not however interact directly with the model at any of the meetings.

The income losses predicted for most of the proposed actions to manage salt suggest that they would be unattractive to farmers. This was consistent with the experience of DIPNR officers implementing the measures in the field, who found a low level of interest in many of the proposed actions.

Although the representative farm analysis showed that most of the activities were marginal or unprofitable, individual farmers were prepared to carry out particular practices. This reflected both personal preferences for risk and enterprises and the different financial structures of individual farms that were not reflected in the representative farms used for modelling because of aggregation error.

After the first year of the project, DIPNR decided to alter their approach away from a general offer of financial assistance, towards a tender system. Farmers are asked to tender for carrying out specific management activities within a particular sub-catchment. This change in policy was only partly the result of the modelling but the modelling results contributed to the outcome.

5. REFERENCES

Bell, R and A Heaney, *A Basin Scale Model for Assessing Salinity Management Options: Model Documentation*, ABARE Technical Working Paper 2000.1, 2000.

Buckwell, A and P Hazel, Implications of aggregation bias for the construction of static and dynamic linear programming supply models, *Journal of Agricultural Economics* 23(2) 119-24, 1972.

Hajkowicz, S, T Hatton, W Meyer and M Young, *Concepts of Landscape Redesign: A Discussion Paper*, CSIRO Land and Water, Glen Osmond, South Australia, 2001.

Hall, N, *Forestry Economics for the Lachlan and Macquarie Catchments*, Integrated Catchment Assessment and Management (iCAM) Centre, Canberra, 2002.

Morrissey, H, R Van Hilst, B Watson and M Oliver, *Jojoba: Farm Level Analysis and Market Assessment*, ABARE report prepared for the Murray-Darling Basin Commission, Canberra, 1996.

NSW-Agriculture, *Farm Enterprise Budget, 1st X Lambs*, Sydney, 2002.

Oliver, M, N Hall and W Watson, *Farm Economic Analysis Models: Scoping Study*, iCAM client report 2003 TARGET 10, Integrated Catchment Assessment and Management (iCAM) Centre, ANU, Canberra, 2002.

Pannell, D, Lessons from a decade of whole-farm modelling in Western Australia, *Review of Agricultural Economics* 18 373-83, 1996.

Watson, W, R Evans, J Powell and M Oliver, *Macquarie and Castlereagh Catchment Report*, Report no 2003 TARGET 14, Integrated Catchment Assessment and Management (iCAM) Centre, Australian National University, Canberra, Canberra, 2003.

Williamson, D, G Gates, G Robinson, G Linke, M Seker and W Evans, *Salt Trends. Historic trend in salt concentration and saltload of stream flow in the Murray-Darling drainage division*, *Dryland Technical Report*, 1, *Murray-Darling Basin Commission*, Canberra, 1997.

Participatory Multi-agent Systems Modeling for Collective Watershed Management: The Use of Role Playing Game.

P. Promburom [a]

[a] *Multiple Cropping Center, Chiang Mai University, Chiang Mai 50200, Thailand.*

Abstract: Scarce farm land and water resources in the highland watersheds of northern Thailand coupled with multiple users have led to conflicts among stakeholders who play important roles in the system dynamics. Integrating companion modeling and multi-agent systems (MAS) can facilitate adaptive learning processes to result in a decentralized collective management strategy that meets the balanced needs of all parties. However, this requires innovative methods and tools, and coordination from all stakeholders involved in the process. This paper presents the results of a preliminary study on conducting role-playing games (RPG) in order to verify the researcher's perceptions of an interested highland watershed, where a human-/agroecosystem is located within the multi-layered politics of resource management. Two RPG's were conducted with interested stakeholders using simplified rules and environment. Performing the role allowed players to improve knowledge and understanding of both space-and-time-dynamic processes of the whole system. Information obtained from the games supplemented with interviews mutually improved earlier knowledge of researcher and resulted in the "post-perception" which will be used in further participatory MAS modeling processes.

Keywords: multi-agent systems modeling; watershed; collective decision-making; companion modeling; role-playing game.

1. INTRODUCTION

The human-/agroecosystem of upper northern Thailand is characterized by mountainous tropical forest ecosystem, where various ethnic groups are practicing agriculture for staple food and cash crops. Since the 1950s, drastic changes have occurred in land use patterns, resulting from political and marketing factors, coupled with an increase in population density. This has had a substantial effect on natural resource viability and the integrity of watershed systems.

This compelled the Thai government to impose land use constraint laws and policies to preserve forest area in the highlands. Thus, it has produced conflict among multiple stakeholders who differ in goals and strategies, and play important roles in the use and management of land and water resources in the watershed area.

Number of integrated natural resource management projects was implemented in watershed area of northern Thailand using dynamic and multi-agent system (MAS) model. However, most of the model conceptualization, design, development, and validation phases were implemented by the researchers [Letcher et al. 2002], and roles of local and government institutions were merely included in the model [Becu et al., 2003b].

This study aims at coupling role-playing games (RPG) with computer MAS models to tackle natural resource management problems in a watershed area. It involves multiple political layers and stakeholders e.g., forester and forest policy, land developer and soil conservation policy, and local forest resource management organization.

The paper describes the use of RPG and field interviews to verify the researcher's pre-conceptualization, and enhance co-learning processes among stakeholders of a highland watershed system in northern Thailand, where complex resource managements issue are settled.

2. COMPANION MODELING FOR NATURAL RESOURCE MANAGEMENT

The multi-agent systems (MAS) approach and computational modeling techniques have been progressively developed to explore and understand individual behavior and interaction among agents and the environment that represent the complexity of the whole system [Gilbert and Troitzsch, 1999].

They have been increasingly used to deal with ecological and socioeconomic issues arising from the management of scarce resources by multiple users. Integrating MAS with other biophysical or economic models and spatial database tools can enhance the adaptive learning capability of all stakeholders regarding their roles and effects on ecological system dynamics. This has tremendous potential for assisting decision-makers in understanding and managing landscapes [Gimblett, 2002].

In the field of complex common-pool resource management, many studies have focused on strengthening the adaptive capacity of involved stakeholders. Some of the key issues that contribute to the failure and success of sustainable resource management are dialogue among multi stakeholders, multi-layered institutions, tools and methods that facilitate experiment, knowledge improving, and co-management process [Dietz et al., 2003; Borrini-Feyerabend et al., 2000].

Recent integration of companion modeling with participatory approaches aims at empowering interested stakeholders through the acquisition of a clear understanding and a long-term vision of their system dynamics. Thus, this allows them to cooperate and manage their natural resources collectively [Barreteau, 2003a]. Coupling RPG with MAS modeling has been applied to improve understanding of complex phenomena and to develop, modify, and validate MAS models. This can facilitate negotiation and collective decision-making among stakeholders [Barreteau, 2003b; D'Aquino et al., 2002].

3. PRE-PERCEPTION OF THE SYSTEM

3.1 Overview of Maehae Watershed

The Maehae watershed comprises two sub-watershed areas in northern Thailand. It is located 80 km southwest of Chiang Mai, one of the major forest-covered areas in Thailand. This highland slope complex area is about 3,288 ha with 70% of pine mixed with evergreen and dry-dipterocarp forests. There are 14 villages and 550 households, scatter over three districts. The two major ethnic groups, the Karen and Hmong, are practicing agricultural activities in both traditional and high-value cash crops and fruit orchards, which have been actively introduced and supported by the Royal Project Foundation (RPF) development center.

The highland watershed areas in the north have been generally perceived as a fragile, vulnerable, susceptible national asset and subject to protection and management by government. Highland dwellers and agricultural activities in this area have contributed to highland land and water resources degradation. Meanwhile, the new Thai constitution in 1997 provided a range of new policies to empower stakeholders and local institutions to participate in managing their own local resources in a sustainable way. The Maehae watershed also falls into this category where common resources are located within the multiple political layers of resource management.

3.2 Pre-conceptualization

In mid-2003, data were collected using secondary information from previous studies done by local research institutes. Semi-structured interviews with various local key informants and government agencies were also conducted to complement conceptualization of the Maehae system. Pre-system analysis resulted in a list of key stakeholders and their important roles in using and managing land and water resources in this watershed area.

Based on pre-analysis steps, key stakeholders and theirs roles were identified. The farmers are likely facing insecure ownership of their lands. Because most of the cultivated lands are under the national forest reserved boundary. Hence, they are claimed as legally protected areas. The RPF, Land Development Department officer (LDD), and Royal Forestry Department officer (RFD) are key government agencies working in the area. RPF development center is actively introducing and supporting cash crops and fruit cultivation to increase farmers' income. LDD and RFD are responsible for natural resource conservation. The LDD promotes soil conservation practices to reduce soil erosion. The RFD promotes forest resource rehabilitation through the collaboration of local people. Occasionally, the conflicts over resources uses have occurred. For examples, encroaching the restricted forest area, disagreement on water sharing.

This pre-perception on environmental components, stakeholders, their actions and associations that influence the Maehae system dynamics was transformed and developed into a prototype MAS model using Unified Modeling Language (UML) static class and simple sequence diagrams. The preliminary design of the "world" representing the Maehae watershed system consists of three major components, corresponding to the stakeholders, their ecological environment, and the local institutions. Stakeholders share and intervene in common resources with different objectives and perceptions. Local institutions are formal and informal groups or organizations representing stakeholders who share similar interests [Promburom et al., 2004]. Figure 1 illustrates a simplified conceptualization of the Maehae watershed system. The solid arrow line represents either one- or two-way association between stakeholders, while dash line and its gradient shows the perception and understanding level toward an interested context.

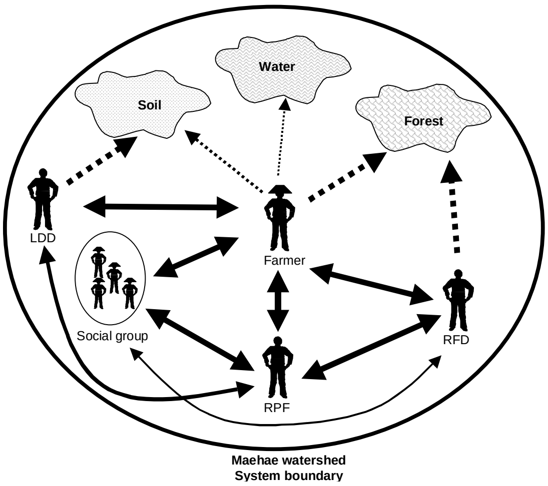


Figure 1. Pre-perception of the Maehae watershed system.

4. RPG AND COLLECTIVE LEARNING

Before developing the model, the researcher's pre-perception was tested in the field using the simplified role-play games. Two main objectives of conducting the games are: a.) To verify and improve the researcher's knowledge, b.) To initiate collective learning of stakeholders on system components and dynamic processes.

Two games were designed and played with local farmers. In the first game, two participants were assigned to perform as government agencies and the rest acted as local farmers. In the second game, a real local RFD officer was invited to play according to his real task.

In the evening and the day after the game sessions, the research team interviewed players individually at home. The interview issues covered comparison of the player's real life with the game, reasons for the role that the player performed during the game, perception and experience of other key stakeholders' roles, and general context of the Maehae watershed.

4.1 The First Role-Playing Game

In late 2003, the first game was designed as a simplified version of a complex previously conceptualized model. Simplifications were made regarding the heterogeneity of the landscape and stakeholders. Some common rules in access and management of the land were flexibly defined but most social rules were left to the player themselves due to the different level of household resource availability and farming strategy.

The first game was conducted with 12 participants, eight of them represent three different types of farmer, rich (type A), middle income (type B), and poor farmers (type C). There were 3 type A, 5 type B and 2 type C respectively. The other two participants were assigned to perform the roles of RFD and LDD.

The 3-D block model was used to represent a simplified typical highland watershed with various slope classes. The model was painted to represent three categories of landscapes corresponding to foothill, mid hill and top hill areas.

At the beginning of the game, each farmer received a different amount of cultivated plots allocated on varying slopes and initial cash to invest in cultivation. Each yearly time step, the individual farmer can freely allocate different crops to the given lands. Each farmer was allowed to open new plots according to the respective strategies. The RFD player was assigned a task to maintain forest area above threshold level of 40%. Thus, RFD has the right to withdraw any new opened plot. Likewise, the LDD player should try to promote soil conservation practices to reduce soil erosion. The reason of swapping the RFD and LDD roles is to make the farmers feel free to perform the given roles.

The aim of this is to see coordination and negotiation that may occur during the game among individual farmers or farmer group and RFD players.

At the end of each time step, the random climate condition was announced. This will affect production and soil erosion of crop plots. Then, the crop allocations on the 3-D block model were collected and used for calculating farmers' household income balance. Lastly, the moderator aggregated and announced the amount of erosion and remaining forest area to all players. This aimed at encouraging RFD and LDD players to actively play their roles for the next gaming session. During the game process, facilitators and the moderator observed some interesting actions and interactions among players.

4.2 Lessons from the first RPG

It can be observed that two poor farmers decided to open new plots at the time step 1 and 2. This significantly decreased forest area, thus encouraging RFD to play the forest protection role actively. In the next step, the RFD took out the new opened plots of the two poor farmers. This process made communication and negotiation between the RFD and poor farmers emerged. The result was that the RFD took a new plot from each poor farmer and allowed the rest to remain until the end of the game.

The LDD player tried to convince farmers to adopt soil conservation practices after two time steps, as he was concerned by the increased amount of soil erosion announced on the public board. He either went out to the 3D block model and communicated with farmers, or assimilates information within farmers' group with same ethnicity. This was clarified during follow-up interviews conducted after finishing the game that they rarely communicated and negotiated across communities and even less between two ethnic groups.

A collective manner of trying to compromise with RFD and LDD was shown. Forest area and soil erosion increased during the beginning steps and then declined to a steady stage toward the end of the game [Figure 2]. This contradicted the pre-perception, in which it was expected that the one who plays the role of the poor farmer will encroach on forest area to claim more land to increase production that fulfills household needs. The interviews confirm that 15 villages have been coordinating the forest conservation network for more than 10 years to manage and protect forest areas. Rules and regulations on forest resource accesses were set up and agreed upon for all members. This is to lower the degree of forest law enforcement, since most of the agricultural area fall into forest reserved area. This is the co-initiative networking among communities with closely support from the local RFD officer. Thus, it made the players reflected upon the collaborative action in the game.

Furthermore, most of the players did not directly know the role of the LDD but they experienced some of the soil conservation practices implemented through RPF. However, collective decision-making on suppressing soil erosion has emerged during the game. During the discussion right after the game, some of players indicated that the increase in soil erosion urged them to cooperate with LDD. Both the LDD and the farmer players expressed the new knowledge gained about the soil conservation roles of the LDD.

There is no strong evidence to support the real change in this behaviour. However the field observation and the interview confirm that farmers are concerning about soil fertility by preparing the cultivated-bed-plot against slope to prevent "good soil lost".

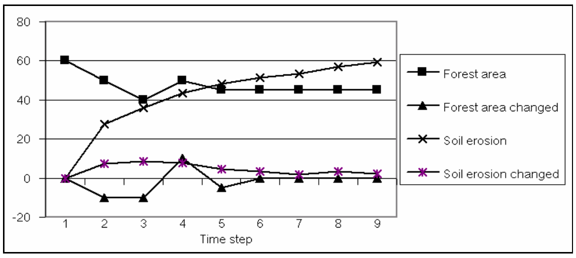


Figure 2. Forest area and soil erosion changes during the first RPG.

4.3 The Second Role-playing Game

The second game was conducted one month after the first game. This aimed at clarifying the understanding on how farmers adapt when faced with limited land resources and forest protection policy. Moreover, this tried to reproduce the history of changes in agricultural pattern. There were eight farmer players, four of whom had participated in the first game, and the rest came from different villages.

At this time, the local forest officer was invited to perform this role corresponding to his own duty. One player was assigned to perform the LDD role because the real LDD agent has rarely contacted or communicated directly with farmers. Some rules were changed according to the stated objectives and the comments from players in the first game to make the game closer to common phenomena. These are; a.) Chance of climate condition for good:normal:drought is 1:1:3; b.) There are no high value cash crops and fruit orchards during time step 1 and 2; c.) Product price ranked by good, medium, and low, will be randomly chosen. This will affect the household's account balance calculation.

4.4 Lessons from the second RPG

During the game, poor farmer players tried to get more land for cultivation in time step 1, 2, and 3. When forest area declined to 40%, which was the alarm level for RFD (this was the given task for RFD described to all players before starting the game). This revealed the information flow within the group and instantly made collective self-management emerge without any forced action from RFD player. This revealed the players' point of view toward the forest resource situation and management. The regulation is so embedded in the minds of the players that the regulator does not need to force them to take action. The performance toward soil erosion showed similar coordination, which was closely consistent with the first game [Figure 3].

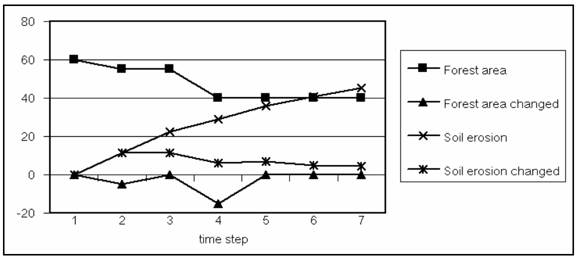


Figure 3. Forest area and soil erosion changes during the second role-play game.

During the interview, most of players expressed that the first two time steps were similar to the situation in the past. Before RFD was established in 1978, agricultural productivity was low. Thus, people needed more land than nowadays to produce crops and generate income. The study of Ekasingh et al. [Ekasingh et al., 2001] confirms this land use change. The discussion after the game supported this historical scene. Furthermore, younger, more educated generations had more employment opportunity. The dependency of household income on agriculture has been gradually decreased.

5 Post-perception after Role-play Game

The rules, flow, and atmosphere of the game can provoke players to react to situations individually and collectively. This allows them to extend their vision and understanding beyond their existing scopes. The game makes them perceive that there are multiple stakeholders taking action in the same system context with differing objectives. Furthermore, this also provides views on interaction between system components and consequences of inter-scale linkage between farm and watershed levels.

The second game imitated the historical scenes of the Maehae watershed and then continued with present situations. This replayed agrarian transformation processes, involved stakeholders, influence factors, and causes and consequences to the players. It can be seen that these two RPG facilitated collective learning processes of players and provided the understanding on complex space-and-time dynamic processes through a simple exercise.

On the researcher's side, RPG can help verifying previous perceptions by allowing players to react toward given rules and environments. Individual decision-making in the game was clarified during the interview, thus added to the researcher's knowledge. One of the important outcomes from RPG was the emergence of a collective manner which stemmed from individual decision-making to tackle common problems; for instance, players tried to suppress soil erosion and maintain a given forest area threshold.

Information and lessons learned from RPG and the follow-up interviews were analyzed altogether with additional key informant interviews, then compared with the pre-perception. The post-perception diagram in Figure 4 illustrates the new outlook toward the Maehae watershed system. Major changes are perceptions of stakeholders toward resources, associations and flow of information among stakeholders, another additional stakeholder, and external factors that may influence system changes in the future. All perceptions and degree of association which varied from pre-perception are represented using gray lines.

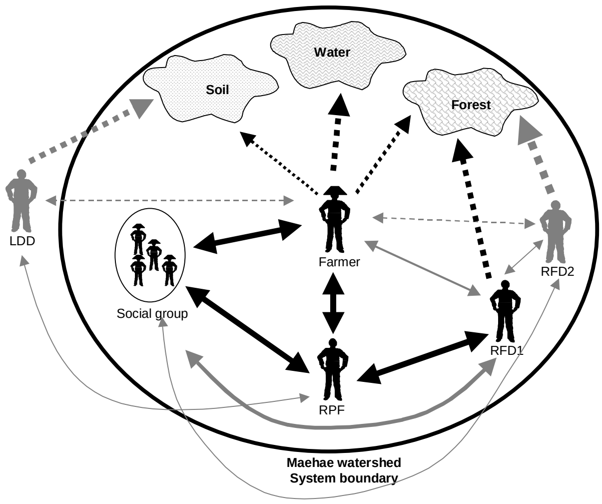


Figure 4. Post-perception diagram of Maehae watershed system.

Most players were directly familiar with RPF and local RFD officers, so-called RFD1 in Figure 4. The Forest network is a social group that strongly influences local forest management among communities in Maehae (as emerged during RPG). Therefore, from the players' point of view, forest degradation is not a problem for Maehae community.

The LDD officer became a stakeholder outside the system boundary. In fact, the regional LDD will propose a plan and budget to restrain soil erosion in highland area. Then, this will be implemented and promoted through collaboration with the RPF staffs.

The RFD2 is a new stakeholder representing forest officers from the forest protection division. He takes charge in protecting and arresting the one who illegally acts against national forest reserve law, which is stricter than the RFD1. The RFD2 communicates indirectly with farmers but through the social group. The forest protection division is now proposing the national park expansion plan to cover Maehae watershed area. This would lead to more forest law enforcement and restrictions.

6. CONCLUSION AND PERSPECTIVES

This preliminary work corresponds to the first iterative steps using a companion modeling approach [Barreteau, 2003a] to support and encourage participatory and collective management of natural resources at the watershed level in northern Thailand. The RPG can bring a better understanding on how individuals behave and interact with the environment and how this may affect the dynamics of the systems. This provides room for putting together the missing parts and the dynamics of the Maehae system that improve the knowledge of both the researcher and other interested parties. Moreover, shared representation, which cannot obtain from individual interview, can emerge through RPG.

The further research steps are developing MAS model combining biophysical and social dynamic components. This model will be run and tested with stakeholders for validation and verification purposes. Furthermore, participatory scenario elicitation will be conducted. It would be interesting to use another RPG to test these possible scenarios suggested by stakeholders.

Although the RPG can enhance collective learning process among researcher and stakeholders, but it is limited by cost, time, and players arrangements. Therefore, this RPG will be further coupled with the computer MAS model to be used as a shared representation among stakeholders to iteratively simulate land use and resources dynamics under alternatives desirable scenarios of resource management.

7. REFERENCES

Barreteau O., Our Companion Modeling Approach. *Journal of Artificial Societies and Social Simulation,* 6(2). Electronic document: http://jasss.soc.surrey.uk/6/2/1.html, 2003a.

Barreteau O., The Joint Use of Role-Playing Games and Models Regarding Negotiation Processes: Characterization of Associations. *Journal of Artificial Societies and Social Simulation,* 6(2). Electronic document: http://jasss.soc.surrey.uk/6/2/3.html, 2003b.

Becu N, P. Perez, B. Walker, O. Barreteau and C. Le Page, Agent-based simulation of a small catchment water management in northern Thailand: description of the CatchScape model. *Ecol. Modelling* 15(2-3), 319-331, 2003.

Borrini-Feyerabend, G., M.T. Farvar, J.C. Nguinguiri, and V.A. Ndangang, *Co-management of Natural Resources: Organising, Negotiating, and Learning-by-Doing.* GTZ and IUCN, 95pp., Kasparek Verlag, Heidelberg (Germany), 2000.

D'Aquino P., C. Le Page, and F. Bousquet, The SelfCormas Experiment: Aiding Policy and Land-Use Management by Linking Role-Playing Games, GIS, and ABM in the Senegal River Valley. In: D.C. Parker, T. Berger, and S.M. Manson, editors. Agent-Based Models of Land-Use and Land-Cover Change: Report and Review of an International Workshop. Irvine, California, USA. October 4-7, 2001. 70-72, 2002.

Dietz T., E. Ostrom, and P.C. Stern., The struggle to govern the commons. *Science,* 302, 1907-1912, 2003.

Ekasingh M, C. Samranpong, W. Weerajit. Land use and cover changes cetection in Mae Hae and Nong Hoi. Annual Academic Conference 2001, Chiang Mai, Thailand, Royal Project Foundation. 286-307, 2001.

Gilbert N., K.G. Troitzsch. *Simulation for the Social Scientist*. Open University Press, 288 pp., Buckingham, Philadelphia, 1999.

Gimblett H.R.. *Integrating Geographic Information Systems and Agent-based Modeling Techniques for Simulating Social and Ecological Processes*. Oxford University Press, 344 pp., 2002.

Letcher R.A. W.S. Merritt, B.F.W. Croke, A.J. Jakeman, C. Buller. Integrated Water Resources Assessment and Management (IWRAM) Project: integrated toolbox. iCAM Working Paper, 28 pp., 2002.

Promburom, P., M. Ekasingh, B. Ekasingh and C. Saengchyoswat, Multi-agent systems for collective management of a northern Thailand watershed: model abstraction and design. In "Companion Modeling and Multi-Agent Systems for Integrated Natural Resource Management in Asia. IRRI, Philippines, 2004. (in press)

Participatory Spatial Modeling and the Septic Dilemma

Alexey Voinov, Erica J. Gaddis, and Helena Vladich

Gund Institute for Ecological Economics, University of Vermont, USA

Abstract: Whereas point sources of nutrients are quite well known and controlled, there is growing concern about non-point sources, especially those that are related to individual homeowners and citizens' practices. They seem to be the hardest to manage and reduce. On-site sewage disposal systems (OSDS) are among the major contributors to the nutrient pollution of surface and ground waters. In Calvert County, Maryland, up to 25% of the non point source nitrogen pollution originates from septic systems. A participatory landscape modeling approach has been used to analyze and visualize the impact of septic systems on the water quality entering the estuary. The landscape model tracks the fate of nutrients released from septic tanks and other non-point sources. A series of stakeholder workshops have been arranged to demonstrate, using the model, how septic discharges contribute to the water pollution in the estuary. Our results suggest that septic tanks are a less significant contributor to surface water nitrogen pollution in the short-term than was previously assumed whereas fertilizer runoff may be more important than previously thought. We are exploring how this participatory process can be used to influence decision-making and management policies in the County to reduce all sources of nitrogen to local waters.

Keywords: Spatial modeling; Sewage disposal systems; Stakeholders; Landscape; Decentralized wastewater

1. INTRODUCTION

A new chapter in environmental management is upon us. Whereas previously the major dichotomy was between point and non-point sources, now the perceived source of environmental degradation shifts from 'large companies' and entities, which can be regulated to individuals (us) who make independent choices. The largest source of water pollution in many parts of the United States comes from individual homeowners, small farmers, and small businesses. These non-point sources are more dispersed, and difficult to quantify, than pollution coming from large agricultural tracts. This makes it nearly impossible to regulate. As a result there is a growing need and interest in solutions that engage citizens. Education and participatory environmental management requires tools that can be used for visualization of options and evaluation of complex systems. We believe that modeling tools will be instrumental in achieving good decisions using the participatory management framework.

Excessive nutrient loads to the Chesapeake Bay from surrounding cities and rural counties has led to eutrophication especially in small harbors and inlets [EPA 2002]. The Maryland Tributary Strategies, Chesapeake Bay 2000 Agreement and Calvert County Comprehensive Plan [MDDNR 2000], calls for reductions in nutrients entering the Bay to reduce impacts on aquatic natural resources. Though the goal set for phosphorous appears achievable, reductions in nitrogen lag well behind the goal. Most sewage in rural residential areas in Maryland is treated by on-site disposal systems, or septic tanks. For Calvert County, the Maryland Department of Planning has estimated that 25% of the non-point source nitrogen pollution to local waters originates from septic systems. Therefore it appears that if Calvert County is to meet nutrient reduction goals, then nitrogen from septic systems must be addressed.

The Gund Institute for Ecological Economics has developed a spatially explicit Landscape Modeling Framework (LMF) that can be used to estimate the relative impact of different point and non-point sources of nutrients on waters throughout a watershed [Costanza 2003]. The goal of the current project was to apply this framework to the Solomon's Harbor watershed, the most densely populated watershed in rural Calvert County. The specific goals of this project include:

1. Understand whether upgrading septic tanks can make a difference in nitrogen pollution of the Harbor in either the short or long-term.
2. Determine how housing density and distribution affects nitrogen loading.
3. Most importantly, engage stakeholders in meaningful dialogue about behavior patterns and local citizen-initiated decisions.

By applying modeling tools we provide visualizations and data in a compelling and clear way. As a result, we hope to support decisions that will help to achieve nutrient reduction goals. These results can then be applied to other watersheds in Calvert County.

2. LANDSCAPE MODELING FRAMEWORK

2.1 Spatial Modeling Environment

The LMF couples the dynamic nature of ecological and hydrologic process models with GIS technology. The modeled landscape is partitioned into a spatial grid of square unit cells. Unit models, composed of existing modules, are run in each cell. Modules are archived and described in a Library of Hydro-Ecological Modules [LHEM 2004]. The hydrology module simulates water flow vertically within the cell. Phosphorus and nitrogen are cycled through plant growth and organic matter decomposition modules (Figure 1).

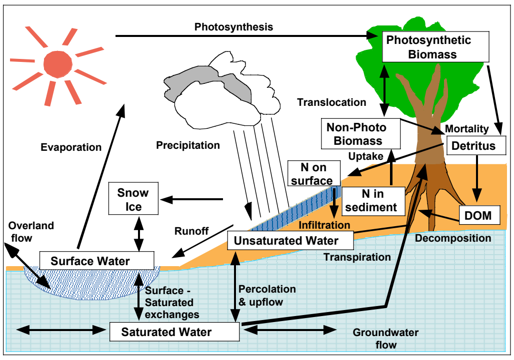


Figure 1. Main variables and flows considered in the unit model.

While the unit model simulates ecological processes within a unit cell, horizontal fluxes are within the domain of the broader spatial implementation of the unit model that forms the landscape model. Such fluxes are driven by cell-cell head differences of surface water and of ground water in saturated storage [Voinov et al. 1999]. Nutrients and other compounds are carried by water transport across the landscape. This spatial implementation is achieved within the framework of the Spatial Modeling Environment (SME) [Maxwell and Costanza 1995, Maxwell and Costanza 1997a, Maxwell and Costanza 1997b, Maxwell and Costanza 1994, SME3 2003]. SME links local unit models with GIS spatial data and algorithms of horizontal transport. Feedbacks among the biological, chemical and physical model components are important structural attributes of this framework [Maxwell 1999, Maxwell and Costanza 1995, Voinov et al. 2004]. Thus, when run within the LMF, the landscape evolves to reflect changing hydrology, water quality, and material flows between adjacent cells. A database of parameters serves as input to assembled models, which represent different habitat types within a landscape, including those dominated by human activity [Voinov et al. 2004]. Further modules are currently being developed to allow dynamic simulations in terms of economic variables such as social, natural and built capitals.

2.2 Study Area and Case Studies

We have previously applied the LMF to several watersheds in Maryland [Costanza et al. 2002, Voinov et al. 1999a, Voinov et al. 1999b, Voinov et al. 1999], including the Hunting Creek watershed also in Calvert County. In this project we focus on the most densely populated area in Calvert County that drains into Solomon's Harbor (Fig.2). Only a small portion of the watershed is serviced by a sewer system, whereas the rest of the area is entirely on septics.

The existing design of septic tanks provides for practically no removal of nitrogen, and all the discharge is leached into the groundwater. Alternative septic designs are expensive, especially as retrofits. Our challenge was to provide the county and citizens with the information required to make good decisions that would be both affordable and effective at reducing nitrogen loads to the harbor. We used the modeling approach to help compare various scenarios and understand what priorities should be set and how septic tanks, atmospheric deposition, and fertilizer contribute to the nutrient loading of the harbor.

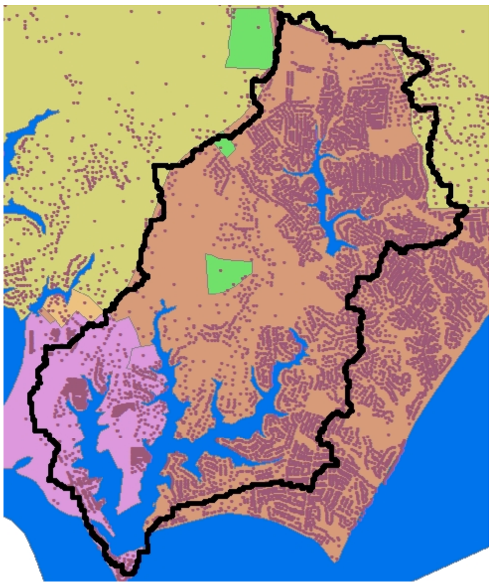


Figure 2. Study area. Dots are individual residences. The lower magenta area is serviced by a central sewer; remaining area is on septic tanks.

2.3 Model calibration and validation

Water quality and flow data were not available to calibrate the model for Solomon's Harbor. Instead, the model was calibrated for the nearby Hunting Creek watershed [Seppelt and Voinov 2002, Voinov et al. 1999a, Voinov et al. 1999b] using flow and nitrogen data collected by the USGS from 1990 – 1995 [USGS 2000]. However, we felt that applying that model to a different watershed should not be used to predict actual nitrogen concentrations, but was appropriate for use in comparing scenarios and relative nitrogen runoff.

3. STAKEHOLDERS AND DECISION-MAKING

From the start of this project we have focused on the application and the use of modeling tools to support decision-making and community education. We have designed a web page and organized a series of community stakeholder meetings to engage residents in the process [OSDS 2004] .

3.1 Stakeholder meetings

Four (of five) stakeholder meetings have been held so far with members of the Solomon's Harbor community. The first meeting attracted almost one hundred people, representing diverse interests of concerned citizens, real estate agents, developers, state environmental regulators, county planners, septic tank companies, non-governmental organizations, and representatives from the team. The meeting focused on discussion of relative sources of nitrogen to Solomon's Harbor and community concerns with respect to water quality. There were several prepared [OSDS 2004] about septic tank processes and nitrogen transport in watersheds, but the group was most eager to engage in active discussion. This discussion was effectively facilitated by the Green Mountain Institute for Environmental Democracy. It is generally agreed in the community that septic tanks are a significant contributor to the nitrogen load in nearby Solomon's Harbor. One of the goals of our modeling research was to test this assumption and determine relative contributions of various nitrogen sources to the harbor.

The second and third meetings were gatherings of a smaller task team (~12 people) that had volunteered to collaborate closely with the modelers during the entire term of the project. This group included several citizens, a real-estate agent, the county planner, and a representative from a septic company. We have focused on the tradeoffs of decentralized wastewater alternatives. One of the main goals of the second meeting was to demonstrate which technologies complete the nitrogen reduction process (nitrification and denitrification). Many septic technologies make claims of nitrogen reduction when they only facilitate the process of nitrification, which does nothing to actually reduce total nitrogen loading (in nitrate form) to groundwater. A simple spreadsheet was developed for selecting alternatives based on the tradeoffs between cost and nitrogen reduction. A survey was distributed to assess citizens' concerns and interest in changing the practice of septic treatment.

3.2 Scenario development

During the third meeting, considerable effort was spent deriving scenarios for septic tank reduction based on survey results, open discussion, and input from all interest groups. An interesting question emerged from this discussion: Given limited resources, is it better to focus on scenarios which we suspect will have the greatest impact on water quality or those scenarios which are most easily implemented politically? Scenarios are very different for each perspective. For example, scenarios which are likely to have the greatest impact on water quality are:

1. Upgrade or remove all septic tanks (central sewer).
2. Upgrade all septic tanks for nitrogen removal within a specified distance (60 m, 150 m, and 300 m) from surface waters.

Alternatively, the stakeholders felt that the following scenarios would be more easily implemented and thus should be focused on:

1. Upgrade all septic tanks in houses newer than 1993. (All such houses have 2 chamber tanks, which permit a less expensive upgrade).
2. Upgrade all septic tanks at the time a home is sold (11% per year).

A consensus was reached to test all of the scenarios using the landscape modeling framework.

4. MODELING RESULTS

4.1 Relative nitrogen loads on watershed from anthropogenic sources

Relative loads of nitrogen to the entire watershed were calculated over five years (1990 – 1995) using time-specific data. We have a good estimate of atmospheric deposition [NCDC 2000], septic loading of nitrogen [USEPA 2000], and fertilizer usage by farmers. It is considerably more difficult, however, to estimate fertilizer use by residents in suburban neighborhoods. An original estimate had been 5 kg/ha, which would correspond to 22% of the total load of nitrogen to the watershed. This is a relatively low estimate, and to test this assumption, we examined the recommendations listed by

Scott's fertilizer [Scotts 2004] for Kentucky bluegrass in Maryland. In order to be relatively conservative, we assumed that only 1/4 of the residents in the county followed these recommendations and that 1/5 of the residential area was covered with lawn. Based on these assumptions residential fertilizer usage could be as high as 50 kg/ha, thus accounting for 63% of the total nitrogen load to the watershed (Fig. 3). Estimates of fertilizer usage determined by the LTER study in Baltimore, Maryland are 15 – 25 kg/ha [Band 2004]. Thus, a medium level (15 kg/ha) of fertilizer usage was assumed for the purposes of running the scenarios. This would account for 38% of the total nitrogen load to the watershed (Fig. 3). All of the relative comparisons of scenarios described in the sections to follow are connected to this assumption.

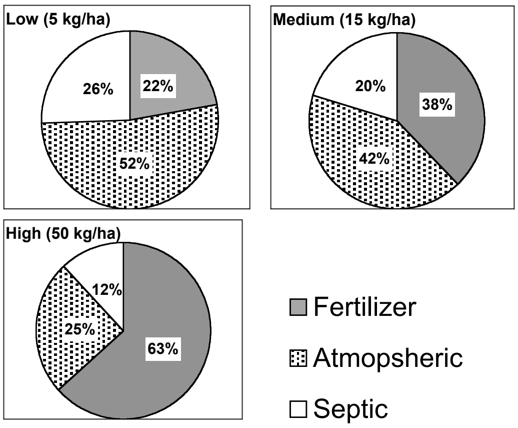


Figure 3. Relative sources of nitrogen to Solomon's Harbor watershed based on fertilizer use assumptions.

4.2 Effects of each nitrogen source on total loading to surface waters

The proportional contribution of nitrogen from anthropogenic sources to the entire watershed differs from the proportional contribution of each source of nitrogen that migrates *to the harbor*. Nitrogen from atmospheric deposition, for example, is deposited on the surface of the landscape, most often during rain events. Fertilizer on the other hand, is added (theoretically) periodically to the landscape in quantities that provide for plant uptake. As a result, a higher percentage of the nitrogen, which comes from atmospheric deposition, is mobile and likely to runoff into nearby surface waters compared to nitrogen from fertilizer.

Nitrogen deposited from septic tanks is discharged relatively deep in the soil and migrates to shallow aquifers. Migration of nitrogen via this pathway is dependent on the movement of groundwater, which can be quite slow especially in the relatively flat

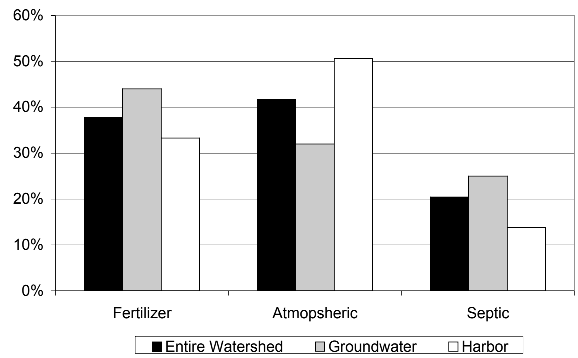


Figure 4. Proportion of nitrogen from 3 sources to Solomon's Harbor Watershed, Groundwater, and Habor over 5 years.

and homogenous area of Calvert County. As a result, nitrogen from septic tanks can be expected to have a smaller total contribution to surface waters in the short-term, including the harbor, than fertilizer or atmospheric deposition. The resulting expectation is that nitrates will accumulate in groundwater, which has been noted in Calvert County. This explains why removal of atmospheric deposition and fertilizer have a greater impact on water quality in Solomon's Harbor over 5 years than does removal of septic nitrogen (Fig. 4).

4.3 Results of Scenario modeling

Results of the scenario runs recommended by the stakeholder group are presented in Figure 5. We should anticipate a very small reduction after one year, and an equally small one in year two, unless we upgrade all septics. By year 3 we see a small increase in the reductions, however the maximum reduction in load is less than 13%, when all the septic tanks are upgraded. This delayed response is related to the slow movement of groundwater in this relatively flat watershed.

Upgrading all septics is a very costly plan and it is unlikely that it would be accepted by the public. All alternative scenarios give even smaller effects. We expected distance to closest stream to be an important factor in septic nitrogen loading, as it is for fertilizer loading. However, this appears not to be the case. The reduction in nitrogen as a result of extended buffers from surface waters appears to be the result of the total number of houses taken off of septics rather than their relative distances from surface waters. This is in part because, unlike surface water runoff, groundwater accumulates nitrogen but has no mechanism for reduction.

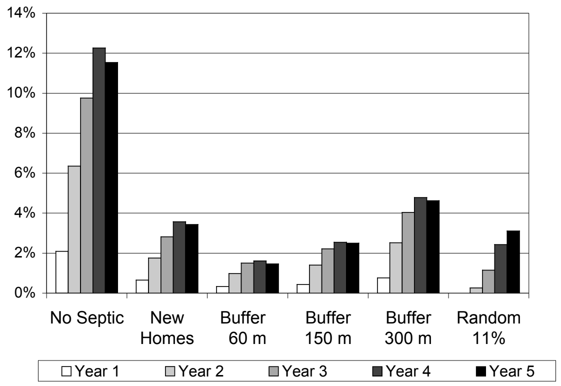


Figure 5. Projected reduction of nitrogen flow to Solomon's Harbor for different scenarios

4.4 Effects of modeling results on stakeholder involvement and policy decisions

An interesting issue with significant impacts on the participatory modeling approach has emerged during this project. As it became clearer that fertilizer and atmospheric deposition have a significantly larger effect (more than the community thought) on nitrogen loads in Solomon's Harbor, we began to wonder whether the expense of any of the proposed septic management scenarios would have a real effect on the trophic status of the harbor. Clearly, cleaning up septic tanks is a good environmental decision regardless since it would improve groundwater quality and to some degree surface water quality.

Atmospheric deposition cannot be directly influenced by local citizens. Fertilizer can be influenced but through educational initiatives rather than policy changes. Ultimately, this requires involvement of other governmental and citizen groups beyond the Department of Planning and Zoning which is currently leading the initiative to reduce nitrogen in the harbor. We are thus faced with the dilemma of presenting our results such that residents are not made to feel helpless toward the situation but also such that we do not give false hopes of improved water quality due to the upgrade of select septic tanks in the watershed.

A similar situation arose in the small town of St. Albans, Vermont, which has been dealing with the problem of phosphorus runoff to nearby St. Albans bay. In the 1980s, the community spent large amounts of money to install a wastewater treatment plant to remove phosphorus, and loads have been significantly reduced. However, the continued non-point source loads as well as the internal loading from historic sediments in the bay as meant that water quality has not improved and is now not expected to improve for at least 20 more years. As a result, many residents are quite frustrated.

All this has been discussed during the fourth meeting with the small group, in which results from the modeling exercises were presented. The model itself is too large to run scenarios during the course of a meeting (each scenario takes 2 -5 hrs to run). The general outcome was some sense of confusion since the results presented were clearly not quite expected. The results were in some contradiction with the previous rough estimates performed by W. Boynton [OSDS 2004]. The group was therefore quite eager to help with gathering more information on fertilizer applications and was willing to work on developing best strategies of communicating the results with the larger pool of stakeholders. It is clearly an exciting educational opportunity for stakeholders to be involved in the actual process of fact-finding and decision-making.

5. CONCLUSIONS

From the start of this project we have focused primarily on the applications of modeling tools rather than on their development and refinement. We had a fairly well tested modeling framework, which is flexible enough to apply in different situations. The scope of the project did not offer us much time and resources to fine-tune our tools for the scale and watershed of interest here. It was a challenge to generate some meaningful results in a short time that could be used to make certain changes in management practices, and policies. As a result we are restricted to making comparative conclusions rather than absolute predictions.

Application of modeling results is always a challenge. A model is a simplification of the real world, which always excludes a multitude of factors that may become important for policy development. For example, based on our models we may conclude that the role of septic discharge is actually pretty low. The overall input from septic systems is the lowest among all the anthropogenic nitrogen sources. In addition, the discharge is leached into groundwater, which then becomes contaminated and eventually affects the quality of the surface water. There is a huge buffering capacity of groundwater, which means that it takes a long time for the effects to surface and it also takes an equally long (or even longer) time to see the effect of management policies that decrease the amount of nitrogen in the septic discharge.

Therefore, if we do insist on policies that will mandate improvements in septic tank performance we are likely to end up with a lack of observable improvement, in the short-term. On the other hand, managing septic loads is most feasible to implement at the local level. Atmospheric pollution is clearly the factor that contributes the most nitrogen load to the watershed and harbor, but it is unlikely that local governments can do much to control it. Most of it comes from transboundary long-distance transfer. Fertilizers are clearly the second most important factor (and depending on assumed fertilizer usage could surpass atmospheric), but in a free market democratic society it is practically impossible to mandate fertilizer usage by individual homeowners. We could easily imagine tax incentives or tradable rationing to limit the overall fertilizer application, but this is unlikely to be implemented by the County. The most effective approach is an educational campaign used to convince residents to reduce the amount of fertilizer used on their lawns.

We have found dynamic spatial modeling tools to be an effective tool in stakeholder discussions of complex non-point source pollution issues. The tools allow stakeholders to visualize and assess the tradeoffs between short-term and long-term pollution issues and their relative costs and difficulties. The interaction with the stakeholder community was an exciting experience that led us to several insights. Based on the community support of this project, we hope that the educational and policy changes derived from the results of the modeling experiments will result in real change in nitrogen loads to Solomon's Harbor.

6. ACKNOWLEDGEMENTS

Our thanks are due to Dave Brownlee, the Calvert County Principal Environmental Planner for initiating the project and providing essential data and support. We thank Ken Jones for helpful comments and mediation in the stakeholder meetings. Most of our excitement about the stakeholder participation is due to John Gray, a citizen whose help went far beyond the volunteer efforts we anticipated. Funding was provided by a grant from the Maryland Department of Natural Resources with cost sharing from the University of Vermont. The contents of this paper do not in any way reflect the views, opinions, or policies of the Department of Natural Resources.

7. REFERENCES

Band, L. Personal communication, 2004.

Costanza, R., A. Voinov, R. Boumans, T. Maxwell, F. Villa, H. Voinov and L. Wainger, Integrated ecological economic modeling of the Patuxent River watershed, Maryland, *Ecological Monographs*, *72*(2), 203-231, 2002.

Costanza, R., and A. Voinov, *Landscape simulation modeling: A spatially explicit, dynamic approach,* Springer-Verlag, 330 pp., New York, Springer-Verlag, 2003.

EPA, *The state of Chesapeake Bay: A report to the citizens of the Bay region*, Chesapeake Bay Program, EPA 903-R-02-002, 2002.

LHEM, *Library of Hydro Ecological Modules*, http://www.uvm.edu/giee/LHEM, 2004.

Maxwell, T. and R. Costanza, *Spatial ecosystem modeling in a distributed computational environment*, In: Bergh, J. v. d. and J. v. d. Straaten, Toward sustainable development: Concepts, methods, and policy, 111-138, Island Press, 1994.

Maxwell, T. and R. Costanza, Distributed modular spatial ecosystem modeling, *International Journal of Computer Simulation: Special Issue on Advanced Simulation Methodologies*, *5*(3), 247-262, 1995.

Maxwell, T. and R. Costanza, A language for modular spatio-temporal simulation, *Ecological Modelling*, *103*(2, 3), 105-113, 1997a.

Maxwell, T. and R. Costanza, An open geographic modelling environment, *Simulation Journal*, *68*(3), 175-185, 1997b.

Maxwell, T., A Parsi-model approach to modular simulation., *Environmental Modeling and Software*, *14*, 511 - 517, 1999.

MDDNR, *Maryland Tributary Strategies, Chesapeake Bay 2000 Agreement and Calvert County Comprehensive Plan*, http://www.dnr.state.md.us/bay/tribstrat/patuxent/patuxent.html, 2000.

NCDC, *Daily climate data*, http://www.ncdc.noaa.gov/oa/ncdc.html, 2000.

OSDS, *Onsite sewage disposal systems (OSDS) in Calvert County*, http://giee.uvm.edu/AV/OSDS, 2004.

Scotts, *Annual lawn care program*, http://www.scotts.com/, 2004.

Seppelt, R. and A. Voinov, Optimization methodology for land use patterns using spatially explicit landscape models, *Ecological Modelling*, *151*(2-3), 125-145, 2002.

SME3, *Spatial Modelling Environment*, http://giee.uvm.edu/SME3, 2003.

USEPA, *Onsite Wastewater treatment system manual*, US Environmental Protection Agency. National Risk Management Research Laboratory., EPA/625/R-00/008, 2000.

USGS, *Surface-Water data for the nation*, http://waterdata.usgs.gov/nwis/sw, 2000.

Voinov, A., H. Voinov and R. Costanza., Landscape modeling of surface water flow: 2nd Patuxent watershed case study., *Ecological Modelling*, *119*, 211-230, 1999.

Voinov, A., R. Costanza, R. Boumans, T. Maxwell and H. Voinov, *Patuxent Landscape Model*, http://www.uvm.edu/giee/PLM, 1999a.

Voinov, A., R. Costanza, L. Wainger, R. Boumans, F. Villa, T. Maxwell and H. Voinov, Patuxent landscape model: Integrated ecological economic modeling of a watershed, *Environmental Modelling and Software*, *14*, 473-491, 1999b.

Voinov, A., C. Fitz, R. Boumans and R. Costanza, Modular ecosystem modeling, *Environmental Modelling and Software*, *19*, 285-304, 2004.

How can Integrative Modelling Contribute to Sustainable Regional Development in Practice? The Potential of Group Model Building –Results of a Case Study

Dipl.-Ing. Ruth Förster and **Prof. Dr. rer. oec. Susanne Kytzia**

IRL-Institute for Spatial and Landscape Planning, Chair of Regional Resources Management, Swiss Federal Institute of Technology (ETH), ETH-Hönggerberg, Postbox 171, CH-8093 Zürich, ruth.foerster@irl.baug.ethz.ch

Abstract: Planning and related decision making in the context of sustainable regional development can be understood as social learning process, involving different actors in order to cope with the complexity and uncertainty of the situation. Formal computer-models are aiming to support scientific-sound planning and decision making. But they are substantially unused in practice. In order to improve their use in practice, literature revealed that the contact to the specific application context and especially to their potential users has to be ensured. Based on a literature review it can be concluded, that (1) on the one hand the model(ling) takes over different functions (e.g. as mediators, experts etc.) in the social interaction processes (e.g. knowledge exchange, negotiations) of heterogeneous actors. (2) On the other hand different social interactions influence the way the model is built and later used. An ongoing research project aims to contribute to a better understanding of these interwoven influences in the context of planning/decision making for sustainable regional development. The project is focusing on the social practices of the different actors involved in the modelling/planning process. First findings of an ongoing explorative, accompanying case study are presented. Concerning (1) it can be stated, that the model is still a "phantom" and its applications are vague. The "substitutes" used at this stage, "scenarios" and "Cause-Effect-Interactions Models", help to initiate a knowledge exchange and to build a common ground for discussions between the Project Team and the stakeholders but also in between the Project Team. Concerning (2) it has been observed that the process is characterised by different and changing roles of the stakeholders from practice involved. By further investigations it has to be clarified: How do these points evolve in the course of the process and why? Which effects do this observed points have for the further process, the Model and its use? How do they relate to the design of the process and the model-qualities?

Keywords: Group model building; integrative modelling; functions of models; sustainable regional development; social practice, explorative qualitative social research

1. INTRODUCTION

Planning and related decision making in the context of sustainable regional [1] development can be understood as social learning process, involving different actors (e.g. scientists, planners, politicians, citizens) in order to cope with the complexity and uncertainty of the situation. Heterogeneous knowledge, values or interests are included in the course of the process, helping to foster the legitimisation and acceptance of plans and enhancing their chances for implementation into practice e.g. Selle [1996], Haeley [1997]. There is a demand for scientific-sound planning and decision support in this context e.g. Batty [1994], Heeb & Roux [2002].

[1] Scale, defined by functionalities in a system under consideration rather than by political boundaries, taking into account the interrelation of land-use changes and other resources (financial, natural, human...)

Especially, formal[2] computer-models are thought to contribute to a scientific-foundation by supporting a better understanding of complex system behavior (descriptive use) and its development options e.g. via scenarios (predictive use) (e.g. Batty [1994], Wegener [2003]).

In the last decade many different types of computer-models have been produced in order to support decisions or planning in a variety of contexts. Overviews see e.g. Batty [1994], Wegener [2003] for integrated land use and transport models (ILTM), Geertman & Stillwell [2003] for Decision or Planning Support Systems (DSS, PSS).

Nevertheless, these models - as far as discussed in literature - seem to be substantially unused or not to be used in practice like intended see e.g. Geertmans & Stillwell [2003], Uran & Janssen [2003],. Timmermans [2003] and Wegener [2003].

A main reason for this is that contact to the application context respective the potential model users is often missing or lost during the modelling process, see e.g. Geertmans & Stillwell [2003], Timmermans [2003], Wegener [2003], Rouwette et al. [2002]. This can be related to (a) model qualities, e.g. model type, model structure, user interface and (b) the process design (see e.g. Geertmans & Stillwell [2003], Timmermans [2003], Wegener [2003]). To overcome the mismatch between offer of formal computer-models and their application in practice the model has to be "contextualised" within its specific application context.

Different authors stress, that the involvement of potential model-users at an early stage of the modelling process is important to adopt the model to the application context and to foster its use in the "real world", e.g. Uran and Janssen [2003] for DSS or Vennix [1996], Rouwette [2002] for system dynamic modelling (SDM).

Combinations of different formal models with actor involvement have been especially designed in order to support social interactions (learning processes, group negotiations etc.). Studies show that they help to make implicit perceptions, theories explicit and thus communicateable (e.g. Rouwette [2002], Barreteau et al. [2001]). The modelling is used in a procedural way, putting the learning process on the fore (see Barreteau et al. [2003], Rouwette [2002]). The resulting model has an explorative character in contrast to descriptive or predictive models with the resulting product on the fore (see Förster et al [2003]).

[2] The relations between the different system-elements are formalised by mathematical functions delivering quantified results

Other studies investigating the use of computer-models on the interface between theory and practice, discuss that the models may take over different functions in the social interaction processes e.g. as trading zones, mediators or autonomous agent (Morgan & Morrison [1999], Sismondo [1999]). Based on this we conclude that the use of formal computer-models as scientific-sound decision support tools is mainly influenced by (1) the social interactions of the actors involved and (2) the functions taken over (mostly unintendly or unknowingly) in the course of the social interactions related to the modelling and the planning/decision making processes.

Therefore it is of utmost importance to understand in general these interwoven influences on a "micro-level", investigating the social practice of the actors in the course of the processes. So far empirical studies on this topic are lacking.

An ongoing research project wants to deliver hypotheses to the following questions:

(1) Which functions takes the modelling resp. model over in the course of the social interaction processes in which phases during planning/decision making?

(2) Which functions do different social interactions between different actors take over in different phases of the modelling process resp. for the model?

(3) How do the functions relate to (a) model qualities (e.g. structure of the model, modelling techniques used) called "model-technical" aspects and to (b) the process how the model is built?

The results presented in this paper are first answers to question (1) and (2) are based on an ongoing, <u>explorative</u> case study which shall deliver hypotheses for the further planned two cases within the research project.

2. DESCRIPTION OF THE CASE STUDY

As first case the involvement of stakeholders for building an integrative, formal computer-model ("the Model") within an ongoing research project is chosen. The project started in 2002 and is planned to end in 2005. The Project Team (core) consists of eight scientists of different disciplinary backgrounds. The Model is supposed to help to identify driving forces for sustainable development in a specific Alpine Region. It shall become a "prototype of an integral interactive decision making tool" (according to the project plan). The Model will integrate different aspects of resource use (financial, ecological) and land use changes. It shall deliver different future regional development scenarios during a time period of fifty years. Therefore it

combines three different modelling moduls: two spatially implicit modelling techniques, i.e. material and energy flow analysis, Input-Output analysis with a modul combining spatially explicit modelling techniques (e.g. using celluar automata).
At an early stage of the project different representatives of stakeholder groups from the Alpine Region have been involved in an especially designed process (here referred to as "group model building").
Two types of stakeholder involvement have been designed: (1) An "Accompanying Group" ("Begleitgruppe") consisting of nine representatives of the following six stakeholder groups: tourism & gastronomy, building & construction, trade, farming, youth & schools, citizens. (2) "Satellite Groups" ("Satellitengruppe") consisting of between six to eight representatives of each stakeholder group mentioned above.
Each representative of (1) has helped to find participants for (2) and took also part in (2).
For each group one moderated workshop has been performed. The themes and goals of the workshops of (1) and (2) were different (s. Chapter 4.).
The moderated workshops took two to three hours and were designed and organised by people of the Project Team. In each workshop members of the Project Team were present.
The Satellite Group - workshops (six in total) were especially designed to identify important elements and cause-effect-interactions of the "Regional Alpine Systems" qualitatively.
The method used is adapted from Frederic Vester's so called "Cause-Effect-Interaction Models" ("Wirkungsmodelle") (Vester & Hessler [1988], Vester [1998], Gomez & Probst [1999]). The method aims to help to make different perceptions, theories on the system under consideration explicit.
The results of the different workshops are going to be synthesized by the Project Team and shall be fed back into the scenario development and model building process. A final workshop of the Accompanying Group to discuss the results of the synthesis is planned for autumn 2004. After this workshop, it will be decided how to go on with the modelling and stakeholder involvement.
The events are summarised in Table 1.

3. METHODS AND PROCEDURE

The first case study of the overall research project is explorative and shall deliver hypotheses for the further research. It is focusing on the communicative social interactions during the model building process. Furthermore, the selection criteria for the further planned two case studies will be made more precise.
According to our two research questions (see paragraph 1), we are interested in the social practice of modelling/planning (i.e. how it is done), which is displayed by the discourse of the different actors involved. Therefore selected group discussions (workshops, meetings) of the Project Team, the Accompanying Group and the Satellite Groups have been accompanied by participatory observation between May 2003 and February 2004. Of each of the following GMB-phases at least one event is selected: a) Preparation of stakeholders' involvement, b) Performance of workshops, c) Discussion of results & synthesis d) Discussion of Results with Stakeholders by Project Team e) Decision how to proceed and finalising the Model, f) Use of the Model.

Phase) Who	**Theme**
a) GD: Project Team & Moderator	Preparation "Kick off" 1. Workshop Accompanying Group and Satellite Groups
b) GD: 1. Workshop Accompanying Group	Kick - Off: Stakeholder involvement Feedback scenarios, and planned process
a) GD: Project Team & Moderator	Test Workshop: Satellite Group: "Wirkungsmodell"
b) GD: 6 workshops, each Satellite Group (1 selected: Tourism & Gastronomy)	Develop and discuss stakeholder specific "Wirkungsmodell"
Interview: Representative Project Team	What learnt so far from the process? What used so far, how?
c) GD: Project Team & Moderator	Synthesis of workshop results of Satellite Groups I
c) GD: Project Team & Moderator	Synthesis of workshop results of Satellite Groups II
c) GD: Project Team & Moderator	Synthesis of workshop results of Satellite Groups III
d) GD: 2. Workshop Accompanying Group	Presentation and Discussion of Synthesis
e) Project Team	Integration of Results into scenarios and modelling, decision how to proceed with the stakeholder involvement
f) Project Team/Stakeholders (potential users)	Use of Resulting Model

Table 1: Overview on events in chronological order
GD: Group-Discussion:
□: investigated in this paper
Grey laid underneath: events planned

The participatory observation will be complemented by guideline-interviews and document analysis. An overview on the events observed, investigated in this paper and planned to be observed is shown in Table 1. The meetings have been recorded on minidisc, transliterated and analysed according to explorative qualitative empirical social research methods (see e.g. Lamnek [1995a/b], Flick et al. [2002]).
For the generation of hypotheses we follow an explorative approach based on grounded theory by Glaser and Strauss (see e.g. Flick et al. [2002]).
Therefore the transliterations are worked through 2-3 times, with "equally distributed attention" in order to identify interesting points, reappearing patterns concerning the two research questions - generally asking: Who/By whom? How? What? Why? In which situation?
Since the openness towards the transliterated material is "triggered" by the researcher's perceptions of the problem under consideration, also the questions related to this are declared. The questions are adapted iteratively in the course of the research process. Due to literature review (see paragraph 1) and own considerations these are in a very compressed form the following, started with:
(1) How is the communication in the heterogeneous group going on? Examples:

- How are relations between different persons set? (science - practice)...

(2) Which functions are dedicated to the end product, the Model, concerning its prospective use?
(3) Which functions are taken by or dedicated to the Model within the ongoing social communicative interaction processes?
Examples:

- prescriptive, predictive, explorative use
- integrating different types of knowledge, values or interests or not
- supporting/hindering communication processes (e.g. knowledge exchange, negotiations, group decisions)...

(4) Which (different) functions are taken over by or dedicated to the persons involved for the model, the modelling process?
(5) How do the observed points change throughout the process?...
Because of our explorative (iterative) research concept intermediate results are to be discussed in regard to the further research process. They will also be contrasted with theoretical concepts.

4. RESULTS

4.1 Communication Process - Overall

The communication process is challenging for the stakeholders and scientists. The scientific, technical information fed in by the Project Team (e.g. on model structure, scenarios etc.) is not a priori known to "lay-people". Nevertheless, most of the stakeholders anticipate the technical terms used by the scientists (e.g. scenarios etc.). On the other hand the Project Team partly tries to popularise their research work e.g. by using headlines from newspapers on the problems investigated. Most stakeholders participate actively within the discussion and show a strong interest in the Project. They see their involvement as positive signal for the applicability of the research results in practice. This is mostly done in comparison to a research project, judged as negative example, which took place a few years ago. All stakeholders see an individual or role-specific "use" of a resulting Model. Thus the discussion climate is mostly benevolent towards the project and the Project Team, right from the beginning of the stakeholder involvement.

4.2 Functions of Model(ling)

What "the resulting Model" will be, its application(s) and potential users stay vague during this stage. The functions dedicated to the model are partly conflicting.

Different functions are dedicated to the resulting "Model" by the Project Team and by the stakeholders. In the project plan "a prototype of an interactive integral decision support tool for planers" as application oriented result of the research project is planned. Throughout the participation process the Project Team started to discuss and distinguish two model types: a so called "decision support tool", which is meant to deliver prognostic results ("What <u>will</u> happen if something changes") on one hand and a so called "learning tool" ("Lerntool"), delivering results for scenarios ("What <u>may</u> happen if something changes") on the other hand. At this stage of the project (emerging from phase a), Table 1) they decided to promote a learning tool for "the public" of the specific Alpine Region. The Project Team communicates this towards the stakeholders during the workshops and stresses the potentials and limitations of the resulting Model. Further they declare, that they have to serve two masters, practice and science.
Possible applications of the model are discussed between stakeholders and the Project Team explicitly. Within the "Accompanying Group"

individual stakeholders demand for decision support on a strategic level, delivering predictions.
The expression “model” is used in a vague way. It stands for example for “Cause-Effect-Interactions Model“ (“Wirkungsmodell”) a structural, mental model or the “physical” computer-model. Other expressions used for the Model are, e.g.: scenario, system.
Scenarios and “Cause-Effect-Interactions Models” (“Wirkungsmodelle”) are used to “represent” the model and train the thinking in scenarios (how the model works): “What may happen if?”
The “physical” computer-model does not exist yet as whole and no computer-representation is shown. Instead, as representations for the Model and its results scenarios and “Cause-Effect-Interactions Models” are introduced in the workshops. The scenarios are explained as: “What may happen if”. Discussion on these helps to overcome the yet somehow vague idea of the Model and its application. This happens in the context of the Accompanying and Satellite Group workshops but also in the discussions of the Project Team.
Furthermore scenarios are seen as “trigger” to build an integrative model, for knowledge integration over different disciplines but also between scientific and local/specific knowledge of the stakeholders.

4.3 Functions of Social Interactions

Especially, the roles of the different people involved are interesting.

The functions respective roles of the people involved are different and change.

Roughly five types of people (groups) involved can be distinguished so far:

- Project Team: scientists, different disciplines
- Moderator: moderator, scientist, consultant for Project Team, “translator” between scientists and stakeholders
- Accompanying Group: stakeholders, expert lay-people, representatives
- Satellite Group: Stakeholders
- The public: potential users of the “learning tool”.

The roles taken over by each type are varying and partly not explicitly declared within the group discussions. Within each type, the functions taken over by individuals can also be differentiated. For example the *Roles of stakeholders involved*
The roles taken over by the members of the “Accompanying Group” change, they are partly speaking as stakeholders as dedicated by the Project Team, partly as private person. The change of the role to the private person is only partly made explicit during the discussion, the Moderator does not comment or react on the role change either.
Furthermore, the Accompanying Group distinguishes between them and the public - the potential model users, taking the role of an “expert-lay” group, which knows more about the model(ling) as the public.
The roles dedicated to the stakeholders involved by the Project Team are different and manifold, for example. The members of the “Accompanying Group” are meant:

- to give input, feeback for building up two possible development scenarios of the model from an overall stakeholder perspective (a kind of validation, legitimisation demanded).
- to answer the question from an overall stakeholder perspective: How does this specific “Alpine Region” function? (knowledge deliverer)
- to declare interests, demands and possible application contexts towards the resulting model as stakeholders (political function)
- to be gatekeepers and multiplicators for the Satellite Groups and to support the organisation of the Satellite Groups
- to be multiplicators for “the public”, i.e. potential users of the model within their stakeholder community
- to be important representatives of stakeholders, which may influence the “Regional Alpine System” especially in regard to resources or land use.

In general, the stakeholder involvement shall help to produce application oriented results and help to foster acceptance of the results by potential model users.

5. CONCLUSIONS AND OUTLOOK

It can be stated, that the process is characterised by different and changing roles for the stakeholders involved which are partly not explicitly declared. Within further investigations it has to be clarified:

- Do the roles of the people taken over change during the whole modelling up to the implementation process? How and why?
- Which effects have the changing and partly not explicit declared roles?
- Especially, do the roles become more precise?
- Do role conflicts appear?

Especially the role of the “Accompanying Group” and its individuals is manifold and seems at the moment to be important for the communication and acceptance of the model by “the public”. The Accompanying Group seems to become a lay expert-group, anticipating terms used by the scientist, being trained in scenario-thinking in contrast to the public, i.e. the potential user.

These roles have to be observed further within the proceeding of the modelling process.

The Model is still a “phantom” not physically available and its application, addressees are vague. The “substitutes” used at this stage are scenarios and “Cause-Effect-Interactions models”. They help to initiate a discussion between the Project Team and the stakeholders but also in between the Project Team. They can be seen as “trigger” for knowledge exchange (in sense of a more process-oriented use of the model)

Within further investigations it has to be clarified:

- Which effects does the use of the scenarios, “Cause-Effect-Interactions Models” have for the further process and especially for the acceptance and applicability of the resulting model?
- How is the information, knowledge gathered from the stakeholders further processed and the process made visible for the stakeholders?
- And which effect has this?
- What of the Model has to be presented explained to the stakeholders, in order that they can/will use it?

Overall: the strong tension and challenge to deliver results for the scientific community and an applicable learning tool for practice can be seen. Strategies to handle this tension followed by the Project Team in this early stage of the modelling process can overall be characterised by oscillating between giving limitations and potentials of the Model, working qualitatively with the stakeholders and staying somehow vague and flexible.

6. ACKNOWLEDGEMENTS

We want to thank the Project Team for the very supporting cooperation, Prof. Sabine Massen for the methodological input for the explorative empirical social research and the reviewers of the paper for their substantial feedback.

7. REFERENCES

Batty, Michael, A chronicle of scientific planning: The Anglo-American modeling experience, *Journal of the American Planning Association* 60/1, 7–17, 1994.

Barreteau, Olivier, François Bousquet and Jean-Marie Attonaty, Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems, *Journal of Artificial Societies and Social Simulation*, *4*(2), 2001 (http://jasss.soc.sureey.ac.uk/435.html).

Barreteau, Olivier et al. Our Companion Modelling Approach, *Journal of Artificial Societies and Social Simulation*, *6*(1), 2003 (http://jasss.soc.sureey.ac.uk/6/2/1.html).

Flick, Uwe, Ernst von Kardoff and Ines Steinke (Eds.) Qualitative Forschung - ein Handbuch. Reinbek: Rowohlt, 2000.

Förster, Ruth, Markus Maibach, Christian Pohl und Susanne Kytzia, Was könn(t)en integrative Computer-Modelle für eine nachhaltige Siedlungsentwicklung leisten? *GAIA 12*(4), 321-324, 2003.

Geertman, Stan and John Stillwell, 1 Planning Support Systems – An Introduction, in S. Geertman, J. Stillwell (Ed.): Planning Support Systems in Practice, Springer, Berlin p. 1–22, 2003.

Gomez, P. and G. Probst, Die Praxis des ganzheitlichen Problemlösens: vernetzt denken, unternehmerisch handeln, persönlich überzeugen. 3. Auflage. Bern, 1999.

Healey, Patsy, Collaborative planning: shaping places in fragmented societies. Macmillan. London, 1997.

Heeb, Johannes and Michel Roux, Gemeinsam Landschaften gestalten. Werkzeuge für gesellschaftliches Lernen. LBL Verlag, Lindau, 2002.

Lamnek, Siegfried, Qualitative Sozialforschung. Band 1: Methodologie. 3. korrigierte Auflage. München - Weinheim: Psychologie Verlags Union, 1995a.

Lamnek, Siegfried, Qualitative Sozialforschung. Band 2: Methoden und Techniken. 3. korrigierte Auflage. Weinheim: Psychologie Verlags Union, 1995b.

Morgan, Mary S. and Margaret Morrison (Eds.), MODELS AS MEDIATORS. Perspectives on Natural and Social Science. Cambridge University Press. Cambridge UK, 1999.

Timmermans, Harry, The Saga of Integrated Land Use-Transport Modeling – How Many More Dreams Before We Wake Up?, C*onference Keynote Paper, 10th International Conference on Travel Behavior Research,* Lucerne, August 10–15, 2003.

Rouwette, Etiënne A.J.A et al., Group model building: a review of assessment studies. In: *System Dynamics Review 18*(1), 5-45, 2002.

Selle, Klaus, Planung und Kommunikation. In: Selle, Klaus (Ed.) Planung und Kommunikation. Wiesbaden and Berlin: Bauverlag Ed. 11-22, 1996.

Sismondo, Sergio, Models, Simulations, and Their Objects. In: Sismondo, Sergio and Gissis, Snait (Eds.): Modelling and Simulation. Cambridge, New York: Cambridge University Press, 247-260, 1999.

Uran, O. and R. Janssen, Why are spatial decision support systems not used? Some experiences from the Netherlands", *Computers, Environment and Urban Systems (27)*, 511–526, 2003.

Vennix, J.A.M., Group Model Building. Facilitating Team Learning Using System Dynamics. Wiley, Chichester, 1996.

Vester, F., Ballungsgebiete in der Krise. Vom Verstehen und Planen menschlicher Lebensräume. aktualisierte Neuauflage. München, 1998.

Vester, F. and A. von Hessler, Sensitivitätsmodell. Forschungsbericht 80-10104034 im Auftrag des Umweltbundesamt August 1980. 2. unveränderte Auflage, 1988.

Wegener, Michael, OVERVIEW OF LAND-USE TRANSPORT MODELS. Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management. 'sendai mediatheque', Sendai, Japan. May 27-29, 21 pages (CD-Rom), 2003.

Obstacles in Launching a Participatory Group Discussion and Modelling Process

S. Panebianco and C. Pahl-Wostl
Institute of Environmental Systems Research, University of Osnabrueck, Germany
(silke.panebianco@usf.uni-osnabrueck.de)

Abstract: The combination of model-building and group discussions seems to be a promising approach to support learning processes among stakeholders involved in management problems since it allows combining factual analysis and qualitative and quantitative model simulations with an analysis of subjective perceptions and mental models. However, the step of launching such a participatory process might fail due a lack of willingness of the stakeholders to participate in the process. This paper reports about an attempt to involve a specific group of stakeholders and the problems encountered during this process. Although the stakeholders mentioned different reasons, the most important but primarily hidden reason appeared to be the political sensitivity of the issue. In this situation, a sensible alternative for the group model-building process seems to be the construction of the model by the researcher on the basis of single interviews with the stakeholders. Subsequently, the completed model can be offered to the stakeholders as a simulation tool for testing development scenarios and management measures which is assumed to support a learning process and raise the awareness of stakeholders regarding mutual dependencies and the possible need for collective action. Finally, the paper discusses recommendations drawn form these experiences.

Keywords: Group Model-building; Stakeholder participation; Wastewater Disposal

1 INTRODUCTION

The sewage disposal in former East Germany faces a great challenge: the realization of an appropriate public or private disposal system. At the Institute of Environmental Systems Research in Osnabrueck a project is being conducted to investigate this problem within a regional case-study by using agent-based simulation. Initially, an additional aim was defined which proved to be unfeasible during the first organizational steps: The intention was to organize an intensive participatory discussion and group modelling process with the relevant stakeholders to stimulate a learning process among them and to investigate the requirements for collective choice processes.

The paper discusses the reasons why launching this group process failed. We present the alternative approach chosen for building and using the model. Finally, we discuss from these experiences how to deal with this kind of problems and what should be taken into account for planning participative group discussion and modelling projects.

2 TWOFOLD PROJECT GOAL: MODEL-BUILDING AND GROUP DISCUSSION

After the German reunification the waste water management in the region to be investigated in former East Germany faced a great challenge. Sewage treatment plants were limited so far to industrialized cities, and their purification capabilities were mostly insufficient. Where no public infrastructure existed private collecting pits were used, which often now turn out to be leaky. Therefore, in the last decade, the municipalities spent a huge amount of money to construct an appropriate public infrastructure with modern waste water plants and canalization. But today new problems appear: A legal sewage treatment is not yet fully realized, but further investments would exceed the financial resources of the municipalities, subsidies are reduced and the rising fees cause public opposition. Due to the region's negative demographic development and reduced water consumption some sewage plants are not working to full capacity. Future developments could further weaken the system. Now the questions arise, whether it is reasonable to continue connecting settlements to existing sewage plants respectively to build new plants and, where no public infrastructure is planned, how people can be forced to invest into private sewage plants.

To investigate the dynamics and effects of socio-technological developments in infrastructure systems and to test the impacts of different management strategies, a research project is conducted within the framework of a PhD-thesis at the Institute of Environmental Systems Research at the University of Osnabrueck, Germany. The system dynamics in the development of the sewage disposal are based on the actions of the different stakeholders involved, their mutual influences and the implications on other, non-personal system variables. This information is to be elicited directly from the regional stakeholders in a participative manner in addition to secondary literature and theory. To formalize and investigate the system we decided to apply an agent-based modelling approach, since the actors (e.g. ministry, sewage associations or citizens) and their behaviour can be represented as autonomous, interacting agents. The agent-based modelling approach gives the opportunity

- to improve the understanding of the system dynamics, i.e. the behaviour and interactions of the stakeholders,
- to investigate the impact of different assumptions about the actors' behaviour,
- to assess the consequences of changing conditions on the further development of the sewage disposal system and
- to test the effects of possible management strategies and measures.

The goal of investigating the issue using agent-based modelling is not only to derive an increased understanding of the dynamics of the system. The project aims in particular at using the insights gained from modelling to stimulate a learning process among the regional stakeholders. Thus, the above mentioned insights from modelling should become an integral part of a critical discussion in the stakeholder group. On the one hand, this can be done by using the completed model as an experimental tool, for instance within a workshop or presentation at the end of the project [see e.g. Gilbert and Troitzsch, 1999 and Herz and Blätte, 2000]. However, apart from learning by <u>using the results</u> from a model, it is emphasized in different publications that the potential for learning is particularly high for the persons involved in <u>building</u> the model [Vennix, 1996; DeGeus, 1990; Pahl-Wostl, 2002b]. The model building process becomes part of a process of social learning. Social learning is to be fostered by integrating model building and analysing into a broader group discussion process. Mental models and subjective perceptions of stakeholders are elicited and subjected to a discussion in the group. Stakeholders should realize their differences and learn to deal with them constructively. They should become aware of their interdependence and develop plans for collective action. During this process, the stakeholders build an ownership of the models which should result in an increase of the likelihood for the results from model simulation to feed into management decisions. A number of projects have recently been conducted which report on positive experiences in participative model-building with the stakeholders relevant for the problems at hand leading at best to social learning processes concerning potential management strategies for problems [see as project examples Barreteau et al., 2001; D'Aquino et al., 2003; Hare et al., 2003; Pahl-Wostl, 2002b]. Following these experiences we intended to organize a participative group discussion process integrating model-building and application with the regional stakeholders of sewage disposal. The intention was to bring together the stakeholders - or some representatives - relevant in the development of sewage disposal. Within several group sessions, they should have had the opportunity to define the actual problems from their point of view, to reflect about the factors and relations of the system and to discuss about possible scenarios and suitable measures. The agent-based model built on the basis of the information provided in these discussions and particularly in a special group model building session was planned to be used within the stakeholder group as a tool to support the discussion. Altogether, this approach would have implied an intensive process of stakeholder participation.

However, such a participatory process requires a considerable degree of trust both within the stakeholder group and between the stakeholders and the research team as pointed out by Pahl-Wostl and Hare [2004]. The effort to build this level of trust should not be underestimated. A lack of trust proved to be a key problem in setting up a participatory process in the project we analyse in this paper.

3 DIFFICULTIES IN LAUNCHING THE INTENDED PARTICPATIVE PROCESS

After a first approach to the issue of the project by a rough problem- and stakeholder-analysis encompassing the analysis of documents (publications, internet, media) and first consultations of external experts we started to contact the stakeholders. As we envisaged the Ministry as a superordinated key actor to become a potential promoter we contacted it first and asked for its support making clear that we would be flexible concerning the concrete project design. In general, in these first contacts we introduced the issue we wanted to examine and asked people about their interest in providing some information and discussing the subject with us or other actors. During the contacts we were rather cautious introducing the modelling aim since it appeared to be difficult to explain the meaning of modelling on the phone. But this proved to be not the main problem: It became apparent already from the very first interviews that an intensive participation process in the intended manner would not be possible since some actors and particularly the Environmental Ministry showed some reservations concerning their involvement in the issue in general or their participation in the discussions in particular. Several reasons were mentioned:

- Awareness of influences: Particularly the Environmental Ministry did not consider it appropriate to participate as a central actor in a discursive process about the region's disposal system, since the municipalities or the sewage administration unions respectively are formally responsible for the sewage disposal. Influences by determining the political and legal conditions and by providing subsidies for special technologies were not estimated to be very important.
- Problem and option awareness: Some stakeholders did not see any need for debate. From the Ministry's point of view, existing ecological and managerial investigations provided a sufficient scientific foundation for deciding about further investments. A few sewage administration unions argued that due to the restrictions posed by their previous investments or by the legal conditions their actions are rather determined, they stated to have no options to decide about.
- Time constraints: Other important arguments were the time constraints and the limited personal capacities which prevent the participation in a time consuming discussion process.

Most of the stakeholders contacted right at the beginning of the project mentioned one or several of these aspects in order to explain why they were not willing to participate in the process. We did not manage to convince them of the value of a discussion process. However, contacts with additional stakeholders and experts confirmed our impression that the political sensitivity of the issue was the actual, hidden reason of some stakeholders' caution and rejection: In that region, the issue of sewage disposal is a very sensitive subject. The undesirable developments led to a quite aggressive and polemical public discussion in the media. Citizens' groups were set up which organized actions up to a hunger strike. The public accusations are directed to the political representatives and to the administration of the state, the sewage administration unions and the municipalities. These actors are accused of corruption, lobbyism and a lack of democracy. On the other hand, the citizens' groups, the media and the companies offering alternative technologies are reproached with arguing factually incorrect and ignoring the legal and economic circumstances of the past and the present. We assume that primarily due to these negative experiences some actors now reject to meet some others in a discussion. Those who are supposed to be responsible for the problems by the public seem to avoid the issue and wait until the public calms down.

However, regarding the political sensitivity of the subject, a discussion process among the actors would run the risk of becoming uncontrollable and less expedient and would probably overtax the personal skills of the researchers involved in this project.

The problem of political sensitivity of this project is obviously not an exception as it is addressed in several publications reporting on participative processes. Glicken [2000], for instance, points out that public participation processes "often focus on issues with a high emotional content". This increases the significance of value-based knowledge, i.e. moral and normative valuations within the debate, especially if issues are complex and stakes are high. Since decisions concerning environmental problems are usually based on cognitive or scientific knowledge, participation might involve people with a very different kind of knowledge and argumentation which might cause communication problems. Vennix [1999] as another example focus on messy problems in group model-building. Messy problems encompass "situations in which there are large differences of opinion on the problem or even on the question of whether there is a problem." Such problems are characterized by the typical behaviour of defensiveness due to the fear of losing face and by the mutual expectations of not being understood by other participants. Although both authors refer here to problems occurring during group discussions, the anticipation of these problems might explain the unwillingness of people to attend a group discussion process.

4 SOLUTION: ADAPTATION OF PROJECT DESIGN

Due to the sensitivity of the subject and the difficulties in bringing the stakeholders together, the intended project design was challenged. The additional goal to initiate a common group discussion and learning process as integral part of the model-building process was given up. However, the aim of investigating socio-technological developments in infrastructure systems by means of agent-based modelling could be maintained anyway. But we realized that the project required a shift in emphasis and approach towards analyzing causes for conflict and building up a certain degree of trust und mutual understanding in a stepwise process. Hence, the project design was changed in the following manner. Instead of a common discussion and group model-building processes encompassing all important stakeholders, in a first phase the model is being built by the researcher on the basis of individual interviews. Besides, regional and national statistics, literature about the waste water disposal as well as behavioural and diffusion theories provide decisive information for the construction of the model. This approach is inspired by a number of other projects in which models are based as well on information from interviews, observations, data, literature or theory without an explicit group model-building discussion within the stakeholder group [see for example Gimblett et al., 2002 or Berger, 2001]. In our project nearly all stakeholders agreed to give such interviews. People seem to be more familiar with individual interviews and assess them as being more controllable than broader discussions with more stakeholders. Since the investigation was not initiated by one particular stakeholder group but by a "neutral" third party, the interviewees were quite open and argued in a way they probably would not have done in the presence of other stakeholders. They tried to convince the interviewer from their point of view. Only the representative of the Environmental Ministry was not willing to meet until first results of the other interviews could be presented. Presumably this was due to the initial goals of the project presented during the first contact. Since the Ministry did not agree to the initially intended approach, it reacted cautiously to the new approach as well.

The interviews were conducted as in-deep semi-structured interviews with a few representatives of the most important stakeholders chosen from the stakeholder-analysis. Although the small number of interviewees raises a lack of representativeness, they provide a rather deep insight and enables the emergence of new aspects. The execution and analysis of the interviews have to meet the special requirements of agent-based modelling. For instance, information is needed about the aspects to be represented in the model. Hence, one should know which problems are most important from the stakeholders' point of view as emphasised by Hermans [2001], and how they evaluate the quality of waste water disposal. The interviews confirmed that the stakeholders have different concepts about reasons and goals. The interrelations and influences between the actors, their behaviour and the arguments influencing their behaviour have to be derived as precisely as possible from the interviews since they are the base for the model-rules. However, the interviews included the following items:

- Development of waste water disposal
- Definition of a "good wastewater disposal"
- Problems and solutions
- Influences on the development
- Role of the interviewee (tasks, actions, reasons)
- Further development of the general conditions and the sewage disposal itself

The model is currently in the phase of implementation. It will represent a virtual world similar to the structure of the region under investigation regarding the geographical distribution of population and settlements. On the base of an initial setting the model is expected to retrace – with a certain level of abstraction – the previous development of the sewage disposal system including the construction of public wastewater plants and canalisation or the dissemination of different types of private technologies. Additionally, it will give the opportunity to change parameters, behavioural assumptions, general conditions and policy measures in order to assess the influence of these aspects on variables like water pollution or costs.

In a second phase the model built by the researcher is to be presented to the stakeholders. As this phase has not started yet, we are unfortunately not yet able to report on the success of this second part of the approach. But according to some requests during the interviews concerning the results, the interviewees are rather interested in the presentation, since the model is the result of their efforts. It might be possible to organize this feedback not only in meetings with single stakeholders, but also in little, homogeneous groups, e.g. with sewage administration unions and water agencies of the municipalities in order to stimulate a common discussion at least among these actors. In the feedback step the stakeholders shall be asked to assess the correctness of the model and to test some scenarios and management strategies. Since the model unites the view and the information of the different stakeholders in one encompassing system and will show possible developments and the implications of the implemented management strategies, we expect that it stimulates a learning process at least for some stakeholders. One can anticipate that the combination of the different current strategies of extending the public sewage disposal system with an ongoing negative population development will offer interesting conclusions and provide evidence where cooperative action is required. In the interviews, the negative demographic development was frequently mentioned as a central problem, but at the same time sensible strategies to react on this challenge seem to be unclear and contradictory. Additionally, it would be very valuable if the model could illustrate the importance of cooperation between different stakeholders and at best stimulate it. For example, the model will show the dependencies between the reconstruction or improvement of private waste water plants offering a great potential for water protection, the support of the realization by the private owners through measures of different actors like subsidies from the Ministry, administrative orders and the diffusion of information by the citizens' groups and the media.

5 CONCLUSION: WHAT TO LEARN FROM THESE EXPERIENCES

Based on our experiences with launching a participative modelling and discussion process the following aspects seem to be particularly important.

The problem that arose in this project can finally be summarized as a discrepancy between the initially intended methodological project approach and the problem situation under consideration. The project was initially method driven since we started with the vision of realizing a participative group modelling-process. But the project approach has to be closely adapted to the nature of the problem situation [Glicken, 2000]. There does not exist any general pattern of participative modelling to apply regardless of the special problem situation and stakeholder-constellation. The methodology should not determine how the problem is approached, but it should be the other way around. This applies particularly to situations with great social dynamics, where some methods may be inappropriate or even unfeasible.

We decided about a particular concept of project design and participative approach only after a rather brief analysis of the problem and the local situation. At this early stage the actual nature of the problem and the stakeholder constellation was obviously misinterpreted. As emphasised by Varvasovszky and Brugha [2000] or Glicken [2000] a prerequisite for choosing a suitable approach is that the stakeholder constellation and the problem situation are sufficiently known. Therefore, a detailed stakeholder and problem analysis should be conducted It would be quite useful to conduct such an analysis before the start of a project. The current practice of funding agencies (e.g. European projects) to strongly recommend the involvement of stakeholders as project partners already during the proposal preparation stage has to be judged as very positive in the light of the experiences reported from this project and probably made also during other projects. However, it would also be quite useful to provide financial resources for initial scoping studies.

Additionally, it is important to maintain methodological flexibility, especially if there are some uncertainties concerning the best approach to be applied. Even during the project some adaptations of the approach or the introduction of new methods might become necessary or sensible, either because unexpected problems and constellations arise or because people ask for them. Certainly, this requires knowledge about a broad range of different methods as for example Glicken [2000] and Vennix [1999] point out. Especially in the field of stakeholder participation mistakes are committed easily and can have negative impacts on the further course of the process.

For an external researcher it might be difficult to motivate the stakeholders to actively participate in a project. Personal contacts, particularly a promoter from among the actors often prove to be very helpful. However, the promoter has to be accepted by the stakeholders, otherwise the project goal might be expected to be biased towards his interests and people will not participate. In our case it turned out to be particularly difficult to find an appropriate promoter. The Environmental Ministry was envisaged as the promoter but it was neither willing nor suitable because of its partiality. Fortunately another administrative body partly took over this role as it suggested contacting some particular stakeholders. Beyond this we did not find any appropriate promoter, since the field is sharply divided into advocates of either public or private sewage disposal.

The aim to bring together the actors as soon as possible was among others caused by the limited time available within the duration of the project. Planning and conducting all the steps of such an intensive participative group discussion and modelling process including stakeholder and problem analysis, organizing the participative process as well as model-building is rather laborious. Thus, such projects need to be provided with sufficient personnel and time. Furthermore, as pointed out in several publications [e.g. Vennix, 1999; Glicken, 2000] appropriate expertise and skills are a very important prerequisite for conducting these different steps successfully. The expenditure to acquire these skills is great and frequently underestimated. Given the typical duration of projects of about three years and the limited financial scope the ambition of a project has to be matched with the resources available.

The applicability of a participative method depends among others on the stage of the problem and the problem awareness of the actors [Pahl-Wostl, 2002a]. In our case, the problem is already in an advanced stage indicated by high public problem awareness, visibility of negative consequences and open conflict between stakeholders. Hence, a mediation process seems to be most appropriate. In other projects being in another problem stage (e.g. first problem definition, strategy development or implementation of already confirmed measures) combined with a lower level of conflict, the initialization of participative processes is assumed to be less critical. However, if a participatory process of group model-building is intended but proves to be unfeasible due to reservations of the stakeholders, the approach applied in this project might be a sensible alternative. Instead of a group model-building process, the model can be constructed by the researcher on the basis of individual interviews with the stakeholders. Afterwards, the completed model can be handed to the stakeholders as a simulation tool for testing development scenarios. This can support a learning process as originally intended in the group model-building phase and build trust in the method, the role of the scientist as facilitator and analyst but as well in the need to engage in a participatory process with other stakeholders.

Finally, we recommend that each participatory model building project should be accompanied by a careful documentation of the stakeholder process and a critical evaluation of the methods employed. In particular one should more often report problems and failures to advance the state of the art and to improve the project design.

6 REFERENCES

Barreteau, O. et al., Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems, *Journal of Artificial Societies and Social Simulation, 4(2),* 2001.

Berger, T., Agent-based spatial models applied to agriculture: A simulation tool for technology diffusion, resource use changes and policy analysis, *Agricultural Economics, 25*, 245-260, 2001.

De Geus, A., Modeling to predict or to learn? *European Journal for Operations Research, 59*, 1-5, 1992.

D'Aquino, P. et al., Using self-designed role-playing games and a multi-agent system to empower a local decision-making process for land use management: The SelfCormas experiment in Senegal, *Journal of Artificial Societies and Social Simulation, 6(3),* 2003.

Gilbert, N. and K.G. Troitzsch, Simulation for the social scientist, Buckingham, Open University Press, 1999.

Gimblett, H.R. et al., An intelligent agent-based model for simulating and evaluating river trip scenarios along the Colorado River in Grand Canyon National Park, Gimblett, H.R. (Ed.), Integrating geographic information systems and agent-based modeling techniques for simulating social and ecological processes, Oxford, 245-275, 2002.

Glicken, J., Getting stakeholder participation 'right': a discussion of participatory processes and possible pitfalls, *Environmental Science & Policy, 3*, 305-310, 2000.

Hare, M.P., R.A. Letcher and A.J. Jakeman, Participatory modelling in natural resource management: a comparison of four case studies. *Integrated Assessment, 4(2),* 62-72, 2003.

Hermans, L.M., Using stakeholder analysis to increase the effectiveness and relevance of water resources system modelling, Schumann, A.H. et al. (Ed.), Regional Management of

Water Resources. Wallingford, 183-189, 2001.

Herz, D. and A. Blätte, Simulation und Planspiel in der Sozialwissenschaften. Eine Bestandsaufnahme der internationalen Diskussion, Münster, LIT Verlag, 2000.

Pahl-Wostl, C., Participative and stakeholder-based policy design, evaluation and modeling processes, *Integrated Assessment, 3(1)*, 3-14, 2002a.

Pahl-Wostl, C., Towards sustainability in the water sector: The importance of human actors and processes of social learning. *Aquatic Sciences, 64*, 394-411, 2002b.

Pahl-Wostl, C. and M. Hare, Processes of social learning in integrated resources management. *Journal of Community and Applied Social Psychology, 14*, 1-14, 2004.

Varvasovszky, Z. and R. Brugha, How to do (or not to do)… a stakeholder analysis, *Health Policy and Planning, 15(3)*, 338-345, 2000.

Vennix, J.A.M., Group model-building: tackling messy problems, *System Dynamics Review, 15(4)*, 379-401, 1999.

Vennix, J.A.M., Group model building: facilitating team learning using system dynamics, Chichester, John Wiley & Sons, 1996.

A COMMUNITY DECISION SUPPORT SYSTEM TO ENHANCE STAKEHOLDERS' PARTICIPATION IN WATER RESOURCES MANAGEMENT

R. Giordano [*], G. Passarella, V.F. Uricchio, M. Vurro
Water Research Institute – I.R.S.A., Italian National Research Council – C.N.R., Bari, Italy
[*] e-mail: rafgiord@yahoo.it

Abstract: The stakeholders' involvement in any decision making process is a key point in the Integrated Water Management (IWM). A successful watershed management process has to be participatory, allowing the stakeholders working together to set criteria for sustainable management, to identify priorities and constraints, to evaluate possible solutions, to recommend technologies and policies, and, finally, to monitor and evaluate any possible impact. For these reasons, any kind of support for handling a fair, rational and efficient debate and for achieving agreements and compromises is strongly desirable. In this contribution, a Community Decision Support System, capable to assist individuals and groups in representing and communicating their own perspectives, is proposed. Furthermore, the system can identify conflicts among stakeholders assuming a *multi-level* perspective. In this research work, the definition of "fuzzy semantic distance" between the judgments expressed by each stakeholder is used as a clustering method . The resulting clusters are, then, used for a cooperative solution of the problem.

Keywords: Community Decision-making; Negotiation Support System; Group Cognitive Mapping; Fuzzy Clustering.

1. INTRODUCTION

In water resources management domain, increasing interest is posed to the stakeholders' participation. In this perspective, mutual learning, conflict management, and iterative and adaptive decision-making process can play an important role as means to address complexity [Hjorsto, 2004]. To enhance public participation in water management it's fundamental to allow all possible stakeholders, both individuals and organizations, to participate in the decision process. Thus, conflicts analysis and resolution have to be carried out adopting a *multi-level* approach, firstly involving individuals. In our contribution a Community Decision Support System is proposed. Such a system is able to support discussion and collaboration, it helps participants to structure their problem, to learn about possible alternatives, their constraints and implications and supports them in the specification of their own preferences.

Thus, the participatory process has to embrace the problem structuring phase. Many efforts have been made to support problem structuring in complex situation. Among this approaches, the Soft OR [Hjorsto, 2004] seems particularly interesting to enhance public participation. In the public participation context, the Strategic Options Development and Analysis (SODA) methodology can aid to structure multiple conflicting aspects and set individual's views into context. The cognitive mapping is at the core of the method. A Cognitive Map can be defined as a map made up of concepts linked to form chains of action-oriented argumentation [Eden and Ackermann, 2004]. In our research work, Cognitive Maps are firstly used to capture parts of individual stakeholder's point of view.

To identify conflicts in a multi-level perspective and facilitate the negotiation and the definition of the community's perspective of the problem and preferences, the system can support in creating the, so called, "communities of interests", which gather all the stakeholders having similar needs. Thus, a clustering procedure able to create clusters among the stakeholders' interests is proposed. Such a methodology is based on the definition of a fuzzy semantic similarity measure that has to be applied to the individuals' cognitive maps considering the opinions expressed by each stakeholder on the critical aspects of the problem.

This contribution is organized as follows: in the second section some aspects concerning the participation and the conflicts arising in water management are described; the third section is devoted to the description of the system's performances and architecture; in the fourth section a case study is presented.

2. ENVIRONMENTAL RESOURCES MANAGEMENT AND NEGOTIATION PROCESS

The awareness of the importance of shared decision process in complex domains, like water management, derives from the importance of stakeholders' role in such processes: if they are not involved at all in any alternative constructions and evaluations, then the decision process outcomes could be controversial and the proposed solutions could generate strong opposition, making those solutions unfeasible. Moreover, stakeholders' influence in the decision process is not only determined by the single stakeholder's attributes but also by the way in which different stakeholders' groups interact forming interaction networks [Hare and Pahl-Wostl, 2002]. The role of the participatory process in water management is also established by the European Community Water Framework, which strongly encourages the active involvement of all the affected parties in the resource management [Pahl-Wostl, 2001].
Conflicts of interest over water resources can be greatly due to the variety in quality demands and the number of stakeholders, which are affected, in different ways, by decisions concerning the use of the resources. Thus, water management should involve processes in which stakeholders jointly *negotiate* how they will manage environmental resources [Johnson et al., 2001]. Support for handling a fair, rational and efficient debate and for achieving agreements and compromises is required.
The literature about the negotiation support in natural resources management seems waivering between two positions: on one hand many approaches propose a negotiation support system based on stakeholders' modelling techniques and agent based simulations. The model is, then, used by the agencies to structure the negotiation in a manner that is likely to facilitate an agreement. On the other hand, it focuses on the communication among stakeholders as a basis for consensual outcomes [Becu et al., 2003]. In such a case, the models are helpful in negotiation because they provide stakeholders with potential consequences of various choices involved [Barreteau et al., 2003].
Among the approaches aiming to simulate negotiation, the agent-based modelling seems really interesting. In fact, it permits the coupling of environmental and social systems, allowing to model disaggregated human decision making in environmental management [Hare and Deadman, 2004]. An agent is characterized by a set of rules that govern both the individual behaviour and the interactions with the other agents. To define these rules some approaches start from observation of human societies and try to extract regularities among behaviour [Pahl-Wostl and Ebenhoh, 2004; Pahl-Wostl, 2002].
In our work we move from the concept that the negotiation is a process of social interaction and communication. In this perspective, conflict identification plays an important role, providing a means of understanding stakeholders' interests. In our works a methodology for conflict identification based on a fuzzy similarity measure is proposed.

3. FUZZY COMMUNITY DECISION SUPPORT SYSTEM

If negotiation is mainly a communicative action, Water Community Decision Support System (WCDSS) has to facilitate the exchange of information concerning a particular problem in water management among different community members. Hence, a water community panel is provided. Such a panel allows the members to subject to other community members a particular water management problem. Thus, a community member, individual or organization, can define a problem that could be considered relevant for water community. To support this phase of community decision process, a "problem structuring support module" have been included in system architecture. In our research work, Cognitive Maps are used to capture parts of stakeholders' point of view. To help user in defining his/her own Cognitive Map, an user friendly interface has been designed. Such interface drives the user step by step during the map creation. The first phase is the "concept identification", that is, after giving a short definition of the problem, the system asks to the user to define the important concepts for that problem. The interface provides information on what "concepts" mean, how to define them, etc. At the end of this phase, the system shows to the user all the concepts in a graphical way and asks to him/her to identify possible links between the different concepts simply drawing an arrow. The user can define the link's strength choosing among three terms (weak, strong and very strong). After that, the cognitive map is shown to the user, which can change both concepts and links until he/she feels that the map actually represents the problem.
Thus, after the first step an individual's cognitive map is defined. The map is stored in the community panel and an user's problem description becomes available for other members of the community.
In this work, the cognitive maps analysis has been made using Decision Explorer (DE)

(www.banxia.com), a software package developed by the University of Strathclyde and largely adopted for map design and analysis. DE allows us to compute the *domain* and *centrality* of a concept, which provide information about its importance. More in detail, the domain measures the importance of the concepts by assessing their potency, i.e. the number of direct links (both as input and output). The centrality measures, instead, the importance by considering both direct and indirect links [Albino et al., 2002]. Thus, *key concepts* of user's map can be defined by using concepts with high degree of domain and centrality. To increase the user's confidence in system results, the key concepts are shown to the user that can suggest some changes. Moreover, the system asks to the user to group key concepts to create sets, that is, groups of concepts that deal with a specific issue or topic. The relevance of each set (i.e. the number of concepts per set and the importance of contained concepts) is a further measure of the importance that different issues have for different individuals. The user assigns a name or label to any different sets.

When other community members log on to the system, a community panel module provide them information about the problems already "annotated" on water community panel. If they are interested to these problems, the system supports them in constructing *their own* problem definitions. They can also modify the already annotated cognitive maps , adding or deleting concepts, or changing the links. At the end of this stage, different problem definitions are stored in the system and all the information about the stakeholders' interests are known.

As stated before, conflicts in environmental resources management can emerge at different level. In this work a first phase of conflict identification and resolution is performed using individuals' cognitive maps, but the concept of "community of interests" has been also considered. These communities could be defined as groups of people that share similar interests. To create these communities, the proposed system uses the sets of key concepts contained in all individual's maps. To define the communities, the following formula has been adopted:

$$S(x,y) = \frac{\sum_i w_x(i)\cdot C_x(i)\cdot C_y(i) + \sum_i w_y(i)\cdot C_x(i)\cdot C_y(i)}{\sum_i w_x(i) + \sum_i w_y(i)} \qquad (1)$$

where: $w_x(i)$ is the relevance of i-th sets f concepts according to the opinion of stakeholder x; $w_y(i)$ is the relevance of i-th sets f concepts according to the opinion of stakeholder y; $C_x(i)$ is equal to 1 if the stakeholder x considers sets i or it is equal to 0 if not; $C_y(i)$ is equal to 1 if the stakeholder y considers sets i, or it is equal to 0 if not. The value of $S(x,y)$ is in the range [0,1].

Therefore, the interests of the stakeholders x and y are similar if $S(x,y)$ assumes a high value, that is, both if cognitive maps have many common sets and relevance of common sets is high. In the following figure, the membership function to the set "Similar" is shown.

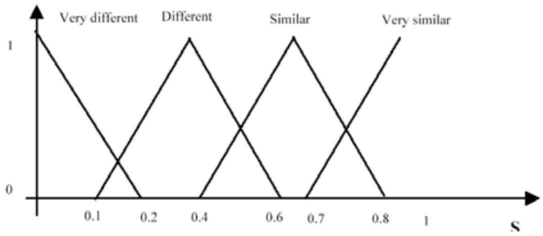


Figure 1. Fuzzy Semantic Similarity Measure

The negotiation within each community allows us to define the "aggregated" cognitive maps (e.g. "environmentalist cognitive map"). Referring to the agent-based modeling of negotiation process (see section 2), these aggregated maps could be compared to the "average" behavior of the "typical" agent. In our approach, the average behavior is defined by a negotiation process among individual stakeholders.

In a community decision process, the alternatives are not defined *a priori*, rather they emerge during the process because of the interaction among participants. Thus, after the communities of interests have been defined, the stakeholders can negotiate to define alternatives with the members of the same community.

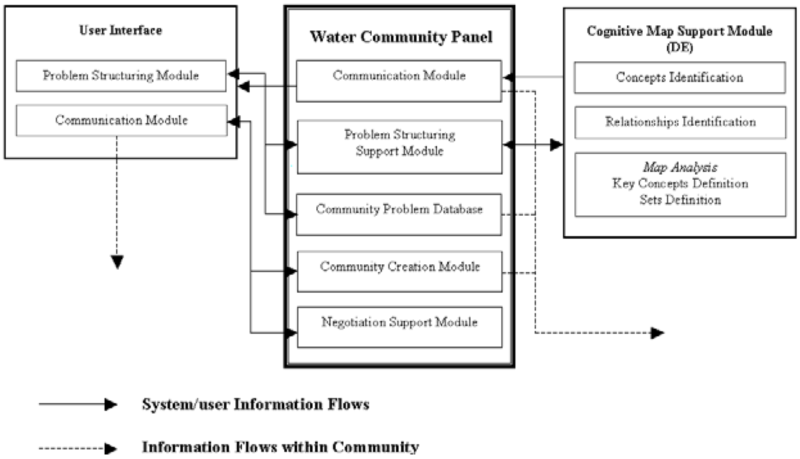


Figure 2. System Architecture

At the end of the first phase of negotiation, each community has its own proposed alternatives. At this point a second level of conflict has to be identified. During this phase new groups can be created considering the agreement among the communities. These groups can be called "coalitions". To create these coalitions, the communities' opinions about alternative have been used. In this phase, the fuzzy set theory has been adopted since stakeholders' opinions are expressed in linguistic terms (e.g., very good, good, moderate, etc). To define the possibility of creating coalitions among the different communities of interests, the following fuzzy semantic distance has been used [Munda, 1995]:

$$S_d (A, B) = \Sigma \mid \mu_A(x_i) - \mu_B(x_i) \mid$$

where, $S_d(A, B)$ defines the similarity degree between two fuzzy sets *A* and *B*. In our work, S_d represents the similarity between the opinions expressed by two different communities; $\mu_A(x_i)$ is the membership degree of *i*-th alternative to the fuzzy set "Good alternative" based on judgment expressed by community *A* and $\mu_B(x_i)$ concerns the judgment expressed by community *B*.

Considering the opinions of all communities of interests, the system creates an agreement matrix, highlighting the communities that can create a coalition, that is, communities with a high degree of similarity. A new fuzzy set called "Similar" have been created and the similarity between two communities can assume four different linguistic values (i.e.: very similar, similar, different and very different) according to the value of S_d.

After this stage, it becomes fundamental to start the negotiation process among coalitions, whose results are an improvement of the agreement on any management action.

The architecture of the proposed system is shown in Figure 2.

4. CONFLICTS NEGOTIATION IN WATER RESOURCES MANAGEMENT

Our research work deals with conflicts identification and resolution in water resources management in the Candelaro River basin, located in the north of the Apulia Region. The aim is to create a Negotiation Support System to be included as a module in a DSS architecture able to facilitate the integrated water resources management in this basin. In this perspective, we test our work applying the methodology for conflicts identification in a case study concerning the water management in scarcity condition. In this phase of the work, the user interface has not been yet developed. Thus, we built individuals' cognitive maps by interviewing different possible stakeholders.

More in detail, we interviewed the chief of the Local Water Management Agency and the users of the irrigation network (farmers) to define their cognitive maps. As described in the previous section, the degree of similarity among their interests has been identified. Following the proposed methodology, the first phase of problem structuring concerned the concepts identification. Thus, we supported the interviewees to identify concepts by explaining them what concepts mean in our methodology and providing them with some example. At the end of this phase, the interviewee was asked to define the links between the concepts and to define the strength of these links. In the cognitive map, we used different graphical representation for the links according to their strength. The cognitive map of the chief of the Local Water Management Agency is shown in figure 3.

important). Between parenthesis the relevance of each sets is reported. The relevance is defined

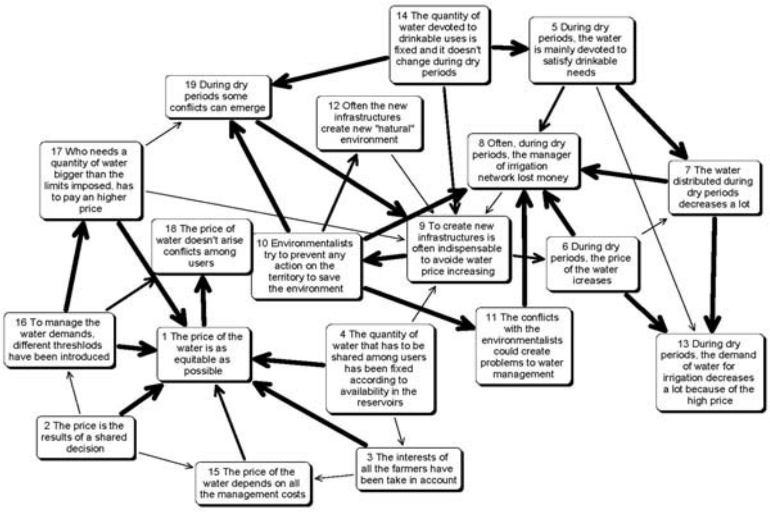


Figure 3. Local Water Manager's cognitive map.

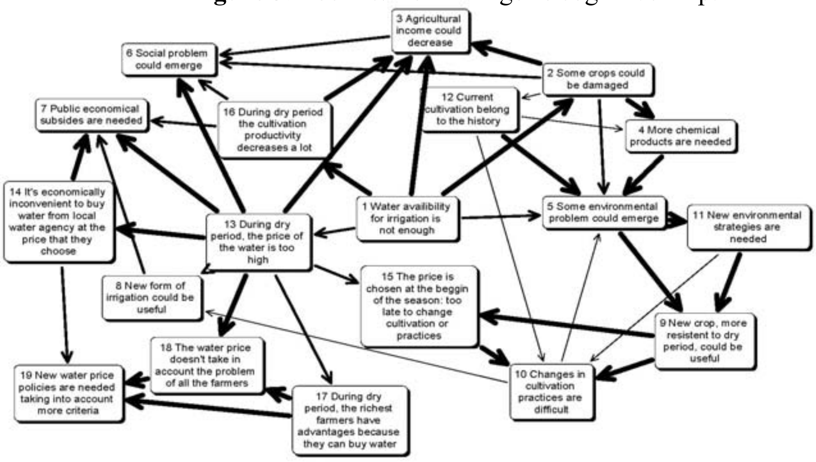


Figure 4. Irrigation network user's cognitive map

To define the *key concepts* of this map, we considere the concepts with a high number of links characterized by a high degree of strength. In the following, the key concepts of previous map are listed in a descending order of importance: 1) To create new infrastructures is often indispensable to avoid water price increasing; 2) The price of the water is as equitable as possible; 3) Often, during dry periods, the manager of irrigation network lost money; 4) Environmentalists try to prevent any action on the territory to save the environment; 5) During dry periods, the water is mainly devoted to satisfy drinkable needs; 6) During dry periods, the price of the water increases.

After this step, the interviewee was asked to group the concepts and to assign to each set a label. The sets identified by the chief of water agency are: 1) water price (very important); 2) infrastructure developing (important); 3) economic problem in scarcity condition (not

considering the number and the importance of concepts included in each set.

We interviewed also an user of the irrigation network and, following the same methodology, we built his cognitive map (Figure 4). Grouping the concepts, three sets have been defined by the user: 1) water price (very important); 2) damages to the cultivation (very important); 3) strategies to safe the cultivation (important).

The degree of similarity between the interests expressed by the chief of local water agency and the irrigation network user can be calculated using the formula reported in section 3. According to this formula, the degree of similarity is: $S(x,y) = 0.43$. Therefore, using the membership function proposed in previous section, the interests expressed by the two stakeholders can be considered *similar*. In fact, both of them consider the "price of the water" as a very relevant issue during the dry period. Of course, they consider this issue from different point of view, but it could be considered as a point to start the negotiation process.

6. CONCLUSIONS AND FUTURE DEVELOPMENT

In this contribution a Community Decision Support System able to enhance the stakeholders participation in water resources management has been proposed. In the definition of such a system, we move from the idea that the community decision-making process is not only a "voting" process, in which the community members can only judge the different alternatives already defined by a central authority. On the contrary, from a community decision-making perspective, each member can highlight a problem relevant for the community and the alternatives have to be created in a collaborative environment. Therefore, the proposed system supports individuals to structure their problem perspective and to proposed it to the other community members. Since now, only the module for conflicts identification has been developed. To support negotiation among stakeholders, many other modules need to be defined and it is going to be done in order to complete the system architecture.

Moreover, the future developments of our research have to deal with some disadvantages emerged in this first phase of the experience. Mainly the drawbacks are related to the human language ambiguity. In fact, as stated in the previous sections of this work, the definition of similarity measure for conflicts identification is based on the comparison among the labels assigned by each stakeholder to the set of concepts. Unfortunately, different users can assign different labels to similar sets. Therefore, important information could be lost and the results of conflicts identification phase could be wrong, misleading the negotiation process. During *vis-à-vis* interviews we overcame this drawback leading the problem structuring process. That is, when the two interviewees assigned the labels to their sets, we suggested some small changes to the labels if the sets were similar according to the contained concepts. Such an operation was easy in our experiment since the interviewees were only two. Thus, it was not difficult to analyse the sets and to suggest changes. On the other hand, the proposed system has to facilitate the negotiation within a whole community that could mean, perhaps, hundreds of cognitive maps. Hence, many other studies aiming to overcome this drawback using argument analysis, fuzzy set theory, Artificial Intelligence, etc., are needed.

Furthermore, the role of Internet as a tool for the democratisation of the decision-making process has to be investigated.

7. REFERENCES

Barreteau, O., Garin, P., Dumontier, A., Abrami, G., Agent-Based Facilitation of Water Allocation : Case Study in the Drome River Valley. *Group Decision and Negotiation,* 12, 441-461, 2003.

Becu, N., Perez, P., Walker, A., Barreteau, O., Le Page, C., Agent based simulation of a small catchment water in Northern Thailand: Description of the CATCHSCAPE model. *Ecological Modelling*, 170, 319-331, 2003.

Eden, C., Ackermann, F., Cognitive mapping expert views for policy analysis in the public sector, *European Journal of Operational Research*, 152, 615-630, 2004.

Hare, M., Deadman, P., Further towards a taxonomy of agent-based simulation models in environmental management. *Mathematics and Computers in Simulation*, 25-40, 2004.

Hare, M., Pahl-Wostl, C., Stakeholder Categorization in Participatory Integrated Assessment Processes. *Integrated Assessment*, 3, 50-62, 2002.

Hiorsto, C.N., Enhancing public participation in natural resource management using Soft OR—an application of strategic option development and analysis in tactical forest planning, *European Journal of Operational Research*, 152, 667-683, 2004.

Johnson, N., Ravnborg, H.M., Werstermann, O., Probst, K., User participation in watershed management and research, *Water Policy*, 3, 507-520, 2001.

Mingers, J., Rosenhead, J., Problem structuring methods in action, *European Journal of Operational Research*, 152, 530-544, 2004.

Munda, G., *Multicriteria evaluations in a fuzzy environment*, Phisica, Verlag, The Hague, 1995.

Pahl-Wostl, C., Participative and Stakeholder-Based Policy Design, Evaluation and Modeling Processes. *Integrated Assessment*, 3, 3-14, 2002.

Pahl-Wostl, C., Ebenhöh, E., An Adaptive Toolbox Model: a pluralistic modelling approach for human behaviour based on observation. *Journal of Artificial Societies and Social Simulation*, 7 (1).

Collaborative Decision-making Processes for Maintaining Biodiversity: Two Australian Case Studies

C. Pettit[a], **D. Pullar**[b]

[a] *School of Mathematical and Geospatial Science, RMIT University, Melbourne, Australia (Chris.pettit@rmit.edu.au)*

[b] *School of Geography, Planning and Architecture, University of Queensland, Australia*

Abstract: There have been many models developed by scientists to assist decision-makers in making socio-economic and environmental decisions. It is now recognised that there is a shift in the dominant paradigm to making decisions *with* stakeholders, rather than making decisions *for* stakeholders. Our paper investigates two case studies where group model building has been undertaken for maintaining biodiversity in Australia. The first case study focuses on preservation and management of green spaces and biodiversity in metropolitan Melbourne under the umbrella of the Melbourne 2030 planning strategy. A geographical information system is used to collate a number of spatial datasets encompassing a range of cultural and natural assets data layers including: existing open spaces, waterways, threatened fauna and flora, ecological vegetation covers, registered cultural heritage sites, and existing land parcel zoning. Group model building is incorporated into the study through eliciting weightings and ratings of importance for each datasets from urban planners to formulate different urban green system scenarios. The second case study focuses on modelling ecoregions from spatial datasets for the state of Queensland. The modelling combines collaborative expert knowledge and a vast amount of environmental data to build biogeographical classifications of regions. An information elicitation process is used to capture expert knowledge of ecoregions as geographical descriptions, and to transform this into prior probability distributions that characterise regions in terms of environmental variables. This prior information is combined with measured data on the environmental variables within a Bayesian modelling technique to produce the final classified regions. We describe how linked views between descriptive information, mapping and statistical plots are used to decide upon representative regions that satisfy a number of criteria for biodiversity and conservation. This paper discusses the advantages and problems encountered when undertaking group model building. Future research will extend the group model building approach to include interested individuals and community groups.

Keywords: GIS; environmental planning; group model building.

1. INTRODUCTION

Traditionally, physical planners and environmental modellers in isolation have undertaken land use planning. However, this resulted in outcomes not acceptable to stakeholders and the wider community. Today the dominant paradigm has moved towards planning *with* the community rather than *for* the community, as advocated by Forester [1999]. This paper examines two Australian case studies, which focus on land use planning and biodiversity issues: i) in urban areas, and ii) in a regional context.

The first case study analysis deploys both geographical information system (GIS) and planning support system (PSS) technologies in developing green system scenarios. The *What if?* PSS tool has been used to elicit decision factors and their weightings of importance from environmental planners within the City of Darebin in order to run and re-run urban green space scenario simulations.

The second case study involved running workshops with a scientific panel to develop an ecoregional classification for the state of Queensland. Visual spatial exploration tools were developed to enable data communication and model exploration. The study aimed to elicit map-based information from experts (mainly ecologists and biologists) on prior probabilities for the distribution of environmental variables that characterized bioregions within a Bayesian modelling framework [Pullar et al., 2004].

This paper discusses the methodology used to formulate the different scaled scenarios, examines the results of urban biodiversity scenarios (case study I) and ecoregional classifications (case study II), and offers directions of how approaches will be refined through the incorporation of further feedback from planners, experts, and the ultimately the community.

2. METHODOLOGY

Miller et al. [1998] have developed a collaborative approach for analysing green systems in the urban context, which integrates suitability analysis with GIS technology for identifying suitable sites for greenway development in the town of Prescott Valley, Arizona, USA. However, there is a paucity of research examining urban green spaces systems in Australia. Our paper incorporates a case study of work in progress, which focuses on the development of a collaborative land suitability approach for developing green systems scenarios for the City of Darebin, Melbourne.

2.1 Scenario Modelling

Spatial modelling is based upon the formulation and evaluation of future scenarios. In the context of decision support systems a scenario can be defined as:

"*A description of the current situation, of a possible or desirable future state as well as of the series of events that lead from the current state of affairs to this future state.*" [Veenklaas and van den Berg, 1994, p99].

PSS are a specialised form of spatial decision support systems (SDSS) focusing on planning (predominantly land use) related decisions. PSS have been described as geo-information tools used to assist in public and private planning processes (or parts thereof) across a range of spatial scales and within a specific planning context [Stillwell, 2003]. In December 2003, the 4th Place Matters - workshop was held in San Francisco, USA. (www.placematters.us). This workshop focused principally on the application of four collaborative PSS tools including:

1. What if?
2. CommunityViz
3. Index
4. Place3S

The workshop enabled participants to explore the application of these PSSs in a charette environment, (a charette is defined as a collaborative hands-on learning experience), to see how planners and community facilitators to formulate different collaborative visions could use them. Direct involvement by the community in the early stages of plan formulation through the interaction of advanced spatial planning tools is an approach which is now being widely applied in urban planning [Synder, 2001].

In formulating the green system scenarios for the City of Darebin case study the *What if?* PSS developed by Klosterman [1999] was used. *What if?* is a collaborative GIS based PSS used to derive land suitability maps, calculate future land use demand, and formulate possible scenarios. In this research only the suitability module of *What if?* was used to formulate urban green system scenarios.

2.2 Approaches to Group Decision Modelling

There are many techniques for group decision modelling, which integrate multiple criteria decision-making and GIS based techniques [Malczewski 1999; Jankowski and Nyerges 2001]. Some of these include the *passive* use of technology where PSS operators observe and model the results of scenarios created by participants using non-technical media such as paper maps and land use stickers. The *passive* scenario modelling approach was demonstrated at the PlaceMatters workshop 2003, where Place3S PSS software, was used in a charette to formulate and evaluate potential development scenarios for a fictitious study area – Edge City, as illustrated in Figure 1.

Figure 1. Place3S passive group decision modelling

Other approaches to group decision modelling may be undertaken using more *active* techniques, which allow the participants to directly interact with PSS software. Such can be the case with the *What if?* PSS, where participants enter their weightings of importance directly into the software and visualise the scenario results immediately after the geocomputational processing is completed. The later, *active* technology approach is utilised in both the case studies presented in this paper.

Generally, the factor weighting and ratings employed in active group decision models need to be determined by informed decision-makers and stakeholders. These individuals and groups may include environment planners and managers from local and state governments, community groups, and the public using formal approaches such as interview, public consultation, and discussions. For example, in Pettit's [2003] land use planning scenarios for Hervey Bay, the local council planners were consulted to determine the appropriate weighting and ratings. This approach was applied in developing preliminary urban green systems scenarios for the City of Darebin. Similarly, when dealing with scientific based decisions there is a need for the input of knowledge by experts in influencing the outcome of a classification. In the ecoregion case study expert knowledge from a panel is used to select model parameters, adjust parameter distributions, and assist in separating structural characteristics of observed data from unstructured random effects. This is explained in detail for the ecoregional analysis in the next section.

3. A CASE STUDY APPROACH

3.1 Urban Biodiversity in the City of Darebin

Green systems such as parks and waterways are an important component of the urban environment. Preserving and enhancing green systems is critical not only for sustaining and improving the quality of life of urban residents but also for conserving biodiversity. In the Metropolitan v 2030 Strategy, the importance of sustainable green systems is expressed in the vision for the city in 2030 in the form of green wedges [DOI, 2002].

The city of Darebin is one of 31 Council comprising metropolitan Melbourne Geographically. Darebin is located between Melbourne’s central business district (CBD) and the City of Whittlesea growth area. The City has diverse open space networks that are facing pressure from urban development. Certain areas with environmental and cultural significance in the City of Darebin need to be protected including: native grasslands, waterways (such as Merri and Darebin Creeks), and significant remnant vegetation (such as river red gums at Mount Cooper). Darebin has nearly 0.6 square metres of open space per person, a level comparable to many other inner Melbourne municipalities.

Before the urban green system scenarios were formulated for the City of Darebin a geodatabase was created. Datasets comprising this database includes:

1. Green systems data – open spaces, waterways and water areas;
2. Biodiversity data – flora, ecological vegetation classes, threatened fauna;
3. Cultural heritage data – registered heritage sites; and
4. Planning scheme and land parcel (cadastral) data.

A number of buffering and union GIS operations were performed on these data layers. For example, to accommodate the analysis of suitability of land use by protecting the natural and cultural heritage, green systems, biodiversity, and cultural heritage areas were buffered into 5 zones (0-25, 25-50, 50-75, 75-100, >100 metres) to classify the conservation importance of these areas in sustainable land use. The resultant datasets were used to construct the geodatabase for importing into *What if?*.

A number of urban green systems scenarios were developed to consider the principal of sustainable built environment, and to analyse and determine the supply of land (both quantity and location) suitable for protecting and enhancing the existing green system and biodiversity. Land suitability analysis was undertaken using a weighted linear combination (WLC), multiple criteria analysis (MCA) model, based on the sieve mapping overlay technique [McHarg, 1969]. The MCA model is a simple mathematical procedure which multiplies each decision factor’s rating by the overall weighting of importance assigned by the user and then performs an additive operation, combining all suitability factors, to derive a final potential cost surface, also known as a suitability map. By using this simple MCA technique the underlying mathematical model is easily understood by most planners, decision-makers and ultimately the community.

The decision factors have been formulated from a number of spatial datasets listed above using standardised factor values related to buffer distances and weightings determined by priorities in accordance to the existing Darebin Council planning policies. In summary, the land suitability analysis module in *What if?* enabled multiple decision factors to be synthesised in order to determine the relative suitability of different locations for a particular land use (‘open space’ in this study). Table 1 contains the weightings and rating assigned by the City of Darebin environmental planners used to formulate one of the sustainable urban green system scenarios.

Category		Considered suitability factors and ratings							
Factors		Fauna	Flora	Open space	Water features	Heritage	EVC	Land Use	
Factor weight		3	3	3	3	-	-	2	
Buffers (m)	0-25	5	5	3	5	-	-	Park	5
	25-50	5	4	3	4	-	-	bus	2
	50-75	5	3	2	3	-	-	ind	2
	75-100	5	2	2	2	-	-	infra	3
	Outside	1	1	1	1	-	-	res	2

Table 1. Urban Green System Scenario weightings and ratings.

These weightings and ratings, combined with permissible land use conversion values, were used to formulate the urban green space scenario illustrated in Figure 2. This particular scenario is based on a policy of preserving open space, primarily for biodiversity purposes, rather than cultural amenity.

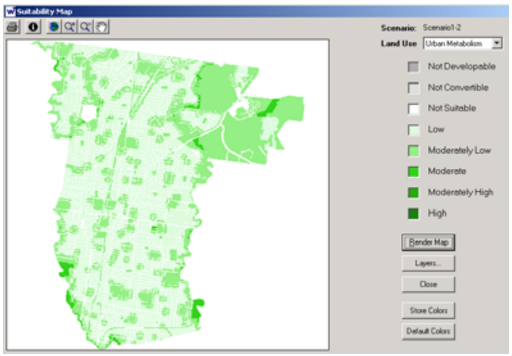


Figure 2. Biodiversity Green System Scenario for the City of Darebin.

3.2 Modelling Ecoregions in Queensland

Ecoregions define recognisable areas that embody broad environmental and landscape structures. Queensland, like most states in Australia, has developed a hierarchical classification of ecoregions for conservation planning, natural resource management and funding allocation. Because of their significance in decision-making and legislation it is essential that ecoregions be defined in an objective and scientifically defensible manner. This lead to a project between the state Environmental Protection Agency and the University of Queensland to develop a ecosystem classification that combined expert (qualitative) knowledge from ecologists and (quantitative) data analysis of environmental data. The resulting classification is described in Pullar et al. [2004]. The main feature of this classification is that qualitative forms of knowledge are interpreted as quantitative data in a rigorous statistical model. Forms of knowledge include rules of thumb, expert advice, scientific publications, or reports. The Bayesian modelling approach supports the inclusion of this knowledge as informative priors, which is balanced with hard data in the model result. The procedure of transforming knowledge into probability statements for informative priors in modelling is called information elicitation. In this case study, information elicitation is carried out through a focused workshop that brings together scientists to define ecoregions. The experts often express their knowledge geographically by describing regions that, in their opinion, characterise a unique pattern of biotic and landscape qualities. For instance, coastal lowland regions with Melaleuca open forest. Through information elicitation we obtain geographical descriptions for ecoregions, and analyse these locations with measurable environmental data, such as climatic, terrain, and soil data. The experts can interactively explore and adjust environmental data distributions and their confidence in these to conform to their opinions of ecoregions. This is transformed into probability distributions and an overall weight is given for this prior information. The Bayesian modelling computes definitions for ecoregions producing outputs as classified maps, uncertainty maps and charts showing the posterior probability distributions.

A GIS application has been written for elicitation of ecoregion classifications. The three main components of the system are: i) an exploratory interface for users to select areas that define ecoregions by querying locations and attributes (biotic and landscape categorical variables), ii) data visualisation tool that allows users to adjust the abiotic variables used to characterise these ecoregions, and iii) a Markov Chain Monte Carlo (MCMC) simulation algorithms that uses a Gibbs sampling technique [Gilks et al., 1996] to classify ecoregions as density mixture distributions. Figure 3 shows an example of an interface for the data visualisation tool. Experts can adjust class breaks of data distribution for several variables to define a model-based cluster corresponding to an ecoregion.

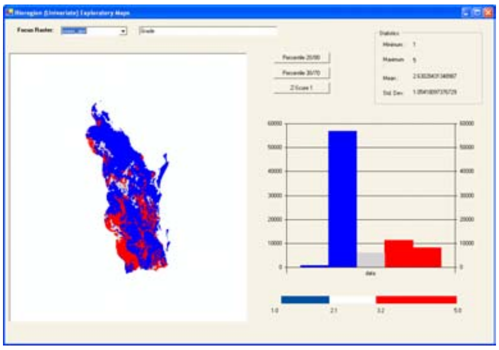


Figure 3. Data visualisation of environmental (abiotic) variables from geographical selection.

4. ADVANTAGES & WEAKNESSES

The land suitability analysis approach to urban green systems scenario modelling has many advantages. It is objective and effective; decision factors can be clearly defined and adjusted according to the objectives. It is transparent as it enables planners, decision-makers and ultimately community members to formulate scenarios by allowing participants to select effective decision factors, assign various weightings of importance to different decision factors, and examine the results by spatial demonstration (suitability maps and reports). Weightings and ratings can easily be adjusted to re-run scenarios based on different policy emphasise.

This type of modelling approach can help to optimise the allocation of additional land to enhance existing green systems in built-up areas such as in the City of Darebin. However, one of the weaknesses of the model is the constraint imposed by jurisdictional boundaries. This means that the formulation of an integrated urban green system is restricted, which negatively impacts on the effectiveness of biodiversity corridors, a critical issue in urban environmental management. For example, Darebin and Merri Creeks flow through three councils – Darebin, Moreland and Banyule. For biodiversity protection of such riparian habitats cross-jurisdictional land use management practices need to be put in place.

Another weakness in the urban green systems scenario modelling approach is it considers the future land use suitability only from the perspective of biodiversity conservation and green systems enhancement. However, in reality many other complex issues need to be considered including social and economic dimensions pertaining to surrounding land uses and values.

The advantages and weaknesses in the ecoregional classification approach are discussed in Pullar et al. [2004]. In this case the output of expert collaboration is to produce a regionalisation that is used as a framework for environmental decision-making. Differences of opinion can arise because certain types of species are seen as endemic to an area, or experts believe an ecological community needs to be identified within a region to remain intact. If there are conflicts in opinion these can be broken down by geographical descriptions to pinpoint where the differences lie. It is generally agreed that this does provide a better way for informing the decision-making process, and the resulting classifications are considered more defensible from an ecological and legal standpoint.

5. FUTURE WORK

In this study, the factor weightings and ratings are only considered from environmental planners from the City of Darebin. However, there is a need to involve stakeholders, and community groups in the selection of model inputs, factor weightings and ratings. This will be explored in future studies.

Further work is required in improving the usability of collaborative PSS systems such as *What if?*. For example work is currently being undertaken to improve the weighting and rating capability of the suitability module through the incorporation of slider bars to enable users to set decision factor trade-offs, as illustrated in Figure 4.

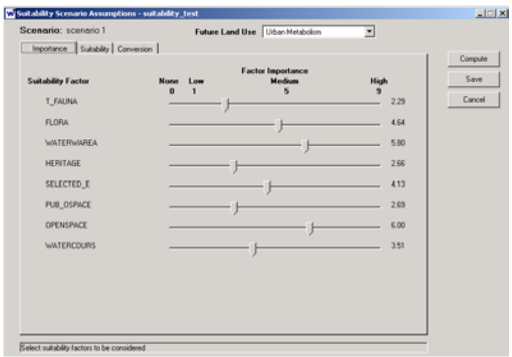


Figure 4. Future enhancements to What if? - slider bar decision factor settings.

6. CONCLUSIONS

One of the current research foci of scenario modelling has been the integration of maps with multiple criteria decision models. Progress in this area has been slow due to a limited role played by maps as decision support tools [Jankowski et al., 2001]. This research examines the application of two modelling approaches, which endeavour to integrate mapping with multiple criteria decision-making techniques.

The urban green systems scenario modelling for the City of Darebin is an example of how GIS and PSS technology can be used to collaboratively preserve and enhance biodiversity in urban areas. Using a collaborative scenario modelling approach, decisions on sustainable land use can be made by potentially incorporating the opinions of different experts, decision-makers, and ultimately the wider community.

It is anticipated that a refined scenario modelling approach will eventually be widely used by planners, decision-makers and the community to assist in achieving the Melbourne 2030 vision. Such a modelling approach could help local councils comprising metropolitan Melbourne to implement an integrated green systems strategy for preserving and enhancing urban biodiversity and conservation.

The ecoregion classification demonstrates a different approach to modelling. The aim is to produce an ecoregion classification based upon environmental data. However, much of the knowledge about ecoregions cannot be readily expressed quantitatively. The case study demonstrates that expert knowledge can be expressed as geographical descriptions, which are transformed into probability measures used in a rigorous statistical model. The use of a Bayesian modelling approach highlights the possibility of combining qualitative and quantitative information to produce model results that include uncertainty values.

7. REFERENCES

DOI., Melbourne 2030: Planning for Sustainable Growth. Melbourne, Victorian Department of Infrastructure, 202pp.,Melbourne, 2002.

Forester, J., *The deliberative practitioner: encouraging participatory planning processes*, MIT Press, 321pp., Cambridge, Mass, 1999.

Gilks, W. R., S., D.J. Richardson and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 512pp., London, 1996.

Jankowski, P., N. Andrienko and G. Andrienko, Map-centred exploratory approach to multiple criteria spatial decision-making, *International Journal of Geographical Information Science* 15(2), 101-127, 2001.

Jankowski, P. and T. Nyerges, Geographical Information Systems for Group Decision Making, Taylor & Francis, 240pp., London, 2001.

Klosterman, R. E., The What if? Collaborative planning support system, *Environment and Planning B: Planning and Design*, *26*, 393-408, 1999.

Malczewski, J., *GIS and Multicriteria Decision Analysis*, John Wiley & Sons, 392pp., Brisbane, 1999.

McHarg, I. L., *Design with Nature*, Doubleday Natural History Press, 208pp., New York, 1969.

Miller, W., W. M. G. Collins, F. R. Steiner and E. Cook, An approach for greenway suitability analysis, *Landscape and Urban Planning, 42*, 91-105, 1998.

Pettit, C., Formulating a Sustainable Development Land Use Strategy for Hervey Bay 2021, paper presented at the 8th International Conference on Computers in Urban Planning and Urban Management, CUPUM, Sendai, Japan, May 27-29, 2003.

Pullar, D., S. Low Choy and W. Rochester, Ecoregion Classification using a Bayesian Approach and Centre-focused Clusters, paper presented at the 2nd International Environmental Modelling and Software Society Conference, IEMSS University of Osnabruck, Germany, June14-17, 2004.

Stillwell, J., Planning Support Systems in Practice, paper presented at the 8th Conference on Computers in Urban Planning and Urban Management, CUPUM, Sendai, Japan, May 27-29, 2003.

Synder, K., Tools for Community Design and Decision-making, in *Planning Support Systems in Practice,* S. Geertman and J. Stillwell (eds.), Springer, 99-120, Berlin, 2001.

Veenklaas, F. R. and L. M. van den Berg, Scenario building: art, craft or just a fashionable whim?, paper presented at the Symposium Scenario Studies for the Rural Environment, Kluwer Academic Publishing, Wageningen, Netherlands, September 12-15, 1994.

Role of Politics and Economy in the Worsening of Water Related Problems in Kerala, India

K. Shadananan Nair

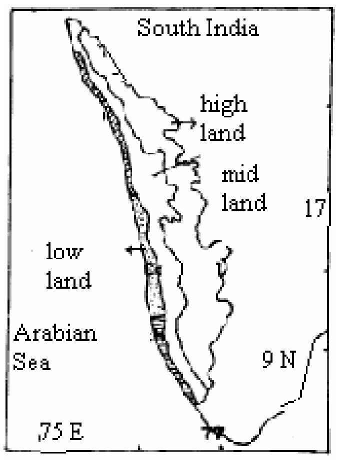
Cochin University of Science & Technology, Vallayil House, North Gate, Vaikom - 686 141
Kottayam Dt., Kerala, India,
e-mail: nair59@yahoo.com

Abstract: Globally, water resources are being fast depleted or deteriorated, while the demands are increasing manifold. Water resources management is becoming a greater challenge in the developing countries with changing environmental, political, social and economic conditions. Politics and shadow economy are among major hurdles in resource management, making implementation of projects and policy guidelines often a failure. The State of Kerala in India is rich in water resources, but it experiences serious seasonal water shortages because of inefficient water conservation and management practices, an after effect of the slow government machinery and corruption. Delays in project completion lead to substantial increase in costs and many projects are left half the way, lacking finance. Disputes over water sharing among different users and upstream-downstream users continue without amicable settlements, just because of political reasons. Human impact on the resources is tremendous in the State. There are rules and regulations to avoid degradation of resources from waste and chemical input, sand quarrying in riverbeds and catchments, deforestation in watersheds and overdraft of groundwater. These rules and judicial orders become farce to those with strong political and financial influence. An appropriate and frequently updated water policy, a strong political will to implement policy guidelines and suggestions, and efficient measures to prevent misappropriation of money allotted to water related projects can help overcome water crisis in Kerala.

***Keywords*:** Water Scarcity; Kerala; Management; Politics; Shadow economy

1. INTRODUCTION

The State of Kerala is a narrow strip of land with a length of 590Km and average width of 65Km that lies in the southwest coast of India, parallel to the bordering Arabian Sea. Though its area is only 38,863Km^2, which is about 1.2 percent of the total surface area of India, it supports a population of more than 32 million, which is 3.1 percent of the total Indian population [Census of India, 2001]. The disproportion between its area and population is reflected in the density of settlement. The present population density is 820 persons/ Km^2. This is a challenging factor in providing adequate water supplies.



Kerala consists of three distinct geographic divisions (Fig. 1): 1. The high lands, which slope

Fig.1 Kerala - Physiography

down from the Western Ghats Mountain with an average height of 900 metres and is a plantation zone. 2.The midlands lying between the Western Ghats and the western lowlands, made up of undulating hills and valleys and is a rich cultivation area. 3. The low lands or the coastal area made up of river deltas and backwater. Orography of the Western Ghats that fully border the eastern side of Kerala produces heavy rainfall of more than 300cm in the State (Fig. 2)**,** 270cm of which falls during

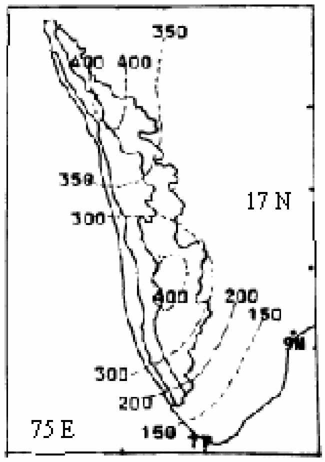


Figure. 2 Kerala – Rainfall pattern

June - December. Northern parts of the State receive most of the rainfall from the southwest monsoon whereas the northeast monsoon is also pronounced in the southernmost parts. Intensity of rainfall ranges from around 2cms/rainy day in the southern districts to around 3cms/rainy day in the north [Nair, 1987]. The State is rich in water resources, with 44 rivers and their tributaries and other freshwater bodies**.** These rivers have an average length of 80Km and a watershed of 700 Km^2. In spite of this wide network of water bodies, Kerala faces serious water shortages on certain occasions as a result of irregularities in temporal and spatial distribution of rainfall, slope of the terrain and improper measures of conservation and management of water resources. Water related problems in Kerala are multi-faceted and it ranges from its peculiar geography to management and politics, which are briefly mentioned below.

1.1 Physiography and Climate

Being a humid tropical region, rainfall over Kerala is of high intensity and highly seasonal in nature. Most of the rainfall is during the monsoons. Winter monsoon is the end of active rainy season, which is characterised by thunderclouds that release all the rain in few minutes, giving less time for infiltration. The pre-monsoon clouds also are of similar nature. All the rivers originate in the Western Ghats Mountain region. Steep slope of the terrain permits the water to flow fast to the Arabian Sea before it could be stored or effectively harnessed. Large storage systems in their watersheds are not safe and advisable in the steep slopes, as small earthquakes have been reported recently.

2. THREATS TO WATER RESOURCES

2.1 Deforestation

Large-scale deforestation in the Western Ghats around the watersheds and introduction of plantation crops in highlands replacing the natural vegetation reduced the storage capacity of soil and resulted in surface soil erosion in watersheds and sedimentation in rivers. This affected summer flow in rivers and some of the once perennial rivers and rivulets became seasonal in the last few decades.

2.2 Sand Quarrying and River Bank Agriculture

Sand quarrying in rivers and watersheds are killing the rivers. Such activities lead to bank erosion, lowering of water table and create several environmental problems. Ground water level in some of the watersheds has gone down by nearly one metre in the last two decades. Agricultural practices in the riverbanks (and also inside the dry riverbeds) during non-rainy months also add to bank erosion and sedimentation in rivers.

2.3 Degradation of Water Resources

All the 44 rivers in Kerala are highly polluted by untreated domestic and industrial wastes and pesticides and fertilizers from the agriculture. Most of the industries are near the thickly populated riversides, often near cities and towns. There is no efficient water treatment system in industries and cities. Pollution level in some of the sites is far above permissible limits. Water quality in River Chitrapuzha that surrounds Kochi, the largest city in Kerala is among the poorest in India.

2.4 Land Reclamation and Construction

Sand filling of ponds, farmlands, wetlands and other water bodies affects natural water flow and groundwater recharge. Construction of new roads and buildings has blocked many canals, which were important for navigation and freshwater. Vast areas of wetlands and paddy fields have been converted into settlement and industrial areas in the post-independent era.

2.5 Increasing Urbanization

Shrinking areas of paddy fields, higher wages, and new system of education that doesn't give practical applications of what one learn reduced rural unemployment and promoted more urbanization. Most of the people diverted from traditional jobs, which were mainly caste-based. Cities in Kerala are not well planned to accommodate large population and therefore water supply becomes inadequate and often interrupted. Providing water in sufficient quantity and in time is increasingly becoming a problem. Competition in suburban areas where there are more settlements of the poor leads to conflicts. Dimensions of drainages are insufficient. Domestic wastes often block drainage channels and waterways. Water logging not only affect availability of safe water, but also creates serious health issues.

2.6 Overdraft and Misuse

Overdraft, careless use and improper maintenance of delivery system contribute much to the water scarcity. Water used for gardening and even washing vehicles in big households use several times the water needed for sustaining the life of a poor family. Number of deep bore wells is terribly increasing. Overdraft of groundwater has invited salinity intrusion far inland in certain locations in coastal areas. There is no way of accounting the theft and illegal use of water.

2.7 Inefficiency in Planning and Management

Government machinery is very slow and there is a lack of cooperation among the different departments involved in project implementation. This causes unnecessary delays in project completion. One major irrigation project in which misappropriation of millions of money was suspected was recently given up half the way after two decades of arguments and legal fight among political parties. One major project to protect the wetlands agriculture in central Kerala is now creating adverse effect. There was much deviation from the original project proposed by the Netherlands Agency. This not only affected targeted rice production, but also deteriorated the surface and groundwater in the whole wetland area, as natural flushing process was obstructed by the artificial barrage.

3. IMPACT OF SHADOW ECONOMY AND POLITICS IN WATER RESOURCES MANAGEMENT

The large-scale inequality in finance and the widening gap between the rich and poor leads to social issues, which is reflected in water resources also. One fifth of the population of Kerala is working outside the State and a good number of them are outside India, especially in Arabian Gulf countries. Conversion rate of the currencies make some families very rich. There is also a class including politicians and government officials who collects money through illegal means. In addition, liberalization policy opened doors to more private industries and salaries in some of them are as competent as that of an outside agency. Their changing life style of the elite class needs supply of water many times that required by an ordinary family. About 40% of the water is consumed by a class of 10 % of the population and for the millions of poor, safe water and its timely availability is far away from reality [Malayala Manorama, 2004]. Careless use and misuse of water are often part of modern and lavish life.

Bribery and corruption have become almost part of the society in daily life. Some officials in the Government sector are very corrupt. For the timely action on domestic water supply and maintenance, the government staffs ask special money. In all the water related projects, misappropriation has become very common and several judicial enquiries and legal procedures are going on. Such corruption makes delays in projects and creates necessity of more investment than estimated. Some of the projects were abandoned half the way, after spending millions. A number of major and medium irrigation projects taken up by the State

Government as far back as forty-five years still remain incomplete, leading to high cost overruns and forcing the Government to abandon many of them. There are 16 projects such as the Kallada (started in 1961), Pazhassi (started in 1964) and Kanhirapuzha (started in 1961) that remain incomplete mid-way. The delays have escalated the costs of many of them has multiplied 50 times, without commensurate increase in irrigation potential [Nair, 2002]. Though the reports on officials being caught red-handed while accepting bribes, the social evil still continue uncontrolled. The vigilance cases registered in water related projects include all classes of officials, including ministers. Fifty-three vigilance cases have been registered about irregularities in the Kallada irrigation project alone. In one case the accused have been punished. Enquiry in 17 cases has been completed. In 11 cases, charge sheets have been submitted to the courts [People's Democracy, 2001]. In January 2002, the Vigilance Special Court sentenced four former engineers and a private contractor to four years' rigorous imprisonment and a fine for causing a total loss of about Indian Rupees 2.2 Million to the exchequer in the execution of works relating to the Project. The prosecution case against the engineers was that they had hatched a conspiracy to make exorbitant payments to the contractor for the blasting and removal of laterite stone after drawing up a supplementary agreement though the contractor was bound to undertake the same work for a much lower rate as per the original agreement [The Hindu, 2002].

Interests in water resources projects are sometimes more money motivated than development. Seeking project assistance from external agencies is considered as a new development strategy and all local bodies are eager to attract big externally aided projects in the name of social security [Ramadas, 2004]**.** Through agencies for contracts for the construction and maintenance and commission from the purchase of raw material or equipment officials and politicians gain a lot. The competition for acquiring the reclaimed land, in real, relative's or proxy names is really astounding.

Political interference in all sectors makes implementation of rules and regulations difficult. There is ban on sand filling of paddy fields and wetlands and control on sand mining and use of deep bore wells. But, in all the rules, there are loopholes for the rich and politically influential people to escape. As a result, area of paddy fields has been largely reduced, discouraging farming. The State now depends on neighbouring sates for food and vegetables. Shrinking paddy fields and disappearing public ponds and wells affected a large population that depended on local water bodies to sustain their life. Higher wages discouraged marginal landowner's from cleaning the ponds or wells in their premises. Some political parties persuaded the workers to demand unjustifiably higher amounts. This, in turn gradually created rural unemployment. Government recently introduced strong restrictions in sand quarrying and land reclamation. Sand quarrying needs a licence that is to be issued for a specific location for a specific dimension that doesn't affect the environment much. In reality, no regulations are practiced when scientists, local administration and government officials become more and more corrupt. Levelling of wetlands and paddy fields also require government sanction. Sanction is given to fill few square metres of land that is certified not suitable for agriculture and for those who has no other piece of land available for house construction. Many people first get the sanction with the help of fake documents and money and fill extensive area, just with a business mind. As a result, water is becoming a scarce resource even in rural areas. Similar forces play in the large-scale deforestation in watersheds.

Industries do not always obey rules and regulations of the government or the pollution control board that affect their production. They are clever to secure safety award from certain agencies to escape through the gaps in law and also to make officials and political leaders silent with gifts or money. Local people's fight with a giant multinational soft drink company against overdraft of groundwater is still continuing in a rural area 'Plachimada'. Water level in the whole village is going down fast and there is a government ban on the company to use water above a limit. But, still the company has not responded to the official orders. Big Companies such as The Hindustan Paper Corporation, Mc Dowell, and a number of small and large factories around the metropolitan city of Kochi release tremendous amount of poisonous chemicals into the rivers Muvattupuzha and Periyar [Nair, 2004].

Inefficient management of water is the most important hazard in water related development activities. Appointment of staff or members is based on political pull, money and caste. As a result, there is the lack of really skilled administrators or technicians. Many of them do not have the ability to assess the situations or to foresee few years ahead.

There exist disputes over water sharing among different regions within the state and also with neighbouring states. Politicians and representatives of the ruling parties, especially of the decisive ones in the allies arrange schemes to provide water to their regions of interest, diverting water from really needing surrounding areas. This occasionally results in violent protests.

Water dispute between Kerala and the east lying State Tamil Nadu is worsening. Tamil Nadu lies in the rainshadow region of the Western Ghats mountain and rainfall there is less than one third that of Kerala. Kerala is not willing to sign new agreements for sharing water or to renew the old agreements. Though the state receives three times global average rainfall and majority of the water is wastefully flowing into the sea, it claims that there is no surplus water available for sharing. It is quite unfortunate that the national water policy of India doesn’t give any guideline for amicable settlement of water disputes. According to the constitution of India water is a state matter and Central administration has limitations in interfering. Any amendment is not easily possible, as the Government is supported by political parties of regional interest. Representatives from water rich regions always oppose any suggestions not of their regional interest.

4. CONCLUSIONS AND SUGGESTIONS

Population of Kerala and associated needs in water are fast increasing, whereas the resources are fast depleting. Management and conservation of water resources become more and more difficult and complicated.

What Kerala needs is an appropriate water policy at the state level and in agreement with the national water policy of India. Water development projects are to be completed time bound. There should be an effective mechanism to monitor this, with transparency in all dealings. Proposals for new projects should be considered only after completing the ongoing major ones. State has to concentrate on minor and lift irrigation projects which will have low gestation period as well as cost. Active public participation in development committees can help avoid misuse of money and corruption to a certain extent. For this, facilities and financial freedom may be provided to the committees under the panchayats/municipalities and their wards, the lowermost level of administration where peoples’ cooperation is already available as ‘neighbours gathering’. They can be trained to spread awareness in local level water conservation such as rainwater harvesting and enhancement of groundwater recharge at houses, cleaning of public wells and ponds, afforestation and so on. To avoid the overuse and misuse, water may be priced slab-wise and water must be provided free to the extreme poor. Traditional methods of water conservation and agriculture are to be revived and initiated from the domestic level. Such methods are environment friendly and involve nominal expenditure and capable of solving problems associated with regional water scarcities. An example is the making of heaps of sand to trap water at the ending stage of rainy season to enhance infiltration, percolation and groundwater recharge. These methods became uncommon mainly because of hiking labour costs and change of particular classes of people from traditional jobs. Government should promote use of small machinery by providing some subsidies on purchase. Another fascinating and significant ancient tradition of conserving land and water is protecting patches of forests and water bodies in the name of local deities. The ‘sarpa kavu’ (Snake forests or sacred groves), the miniature forest to worship holy snakes (and certain other deities) was once an integral part of agricultural plots and many households in Kerala and they still exist in isolation. This ecosystem consists of many species of trees, shrubs and rare herbs of high medicinal value. A well-protected pond, which helps a lot in recharging and preserving water, is an essential part of this forest. Every year, there use to be ‘puja’ (offerings) to the snakes and deities and before offering ‘puja’ the ponds are cleaned. Quality and quantity of water in nearby wells are largely influenced by this ecosystem. Felling of the forests and filling of ponds have resulted in falling water table and added to water shortage in non-rainy months [Nair, 2003]. The Sate should take necessary measures to protect this rich ecosystem by providing incentives to the people who preserve it.

The State needs a leadership with strong and impartial political will to implement rules and regulations. In the present scenario, development requires private participation and for that the political leadership and trade union leaders should compel the militants workers opposing privatisation to accept the ground realities The taxpayer should be more vigilant and reactive to social evils like corruption and learn the meaning of real democracy where power is basically vested on the voters.

REFERENCES

Government of India, Census of India 2001, *Report published on website*, 2001

Malayala Manorama, February 29 edition, *5*, 2004

Nair, G.K., 2002, Kerala: Focus on lift irrigation, big projects out, *Business Line*, 21/3/2002

Nair, Shadananan, K. Hydrometeorological studies of Kerala State in relation to the Western Ghats region, *Unpublished PhD thesis submitted to the Cochin University of Science & Technology*, 180pp, 1987.

Nair, Shadananan, K. 2003, Role of water in the development of civilization in India – A review of ancient literature, traditional practices and beliefs, Proc. International Symp. 'The Basis Of Civilisation—Water Science?' Rome, Italy, 3-6 December, 2003.

Nair, Shadananan, K., 2004, Impact of Natural and Human Induced Environmental Changes on the Water Availability in Kerala, *Unpublished Project Report*.

People's Democracy, Vol. XXV, No. 16, April 22, 2001

Ramadas, T.N., 2004, Relevance of traditional technologies for local area rural development – an alternate path of Kerala's development (A watershed approach), Proc. *16th Kerala Science Congress*, 267-270, 2004.

The Hindu, KIP case: HC adjourns plea, January 2, 2002

The Adaptive Integrated Data Information System (AIDIS) for a Holistic River Basin Assessment in the Challenge Program "Water and Food (CPWF)"

Wolfgang-Albert Flügel

Department *of Geoinformatics, Hydrology and Modeling (DGHM), Friedrich-Schiller University (FSU-Jena), Jena, Germany*

Abstract: Integrated Water Resources Management (IWRM) requires a sophisticated and scale bridging data assessment and management comprising geo-spatial data components, measured or simulated time series and socio-economic information. Such efficient data management is also of paramount importance for decision makers in semi-arid regions where water is scarce, and crop production in irrigation agriculture is consuming a major portion of this limited resource. To assess innovative methods of Integrated Water Resources Management (IWRM) and water productivity in irrigation agriculture but preserving environmental sustainability the Challenge Program “Water and Food” has been launched by the CGIAR at the Baseline Conference in Nairobi, Kenia in November 2003. Applying ‘frontier science’ to support decision making for increased food production without consuming more water, the program needs to integrate multidisciplinary information and research knowledge related to IWRM comprising remote sensing, GIS layers and socio-economic information. For the data organization and management of the CP the Adaptive Integrated Data Information System (AIDIS) has been developed by the DGHM which is presented in this paper. It applied a object-oriented data model, accounts for defined development objectives and is based on open source software (OSS).

Keywords: *Integrated water resources management (IWRM); Object-relational data model, Geo-spatial data mangement; Distributed client/server network; River Basin Information System; Hydrological Response Units; Modeling*

1. INTRODUCTION

Integrated Water Resources Management (IWRM) nowadays is a decision making process which must be based on sufficient and distributed information [STAUDENRAUSCH & FLÜGEL, 2001]. The latter comprises geo-spatial data components, results of measured and simulated dynamics both from the natural environment and the human development of a river basin (Fig. 1).

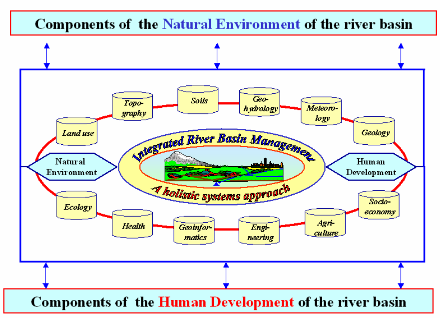


Fig. 1: Multidisciplinary data suppliers for IWRM

Population growth, rise of incomes, industrialization and urban development are priority criteria for IWRM decision making. In the past years increasing attention has been given to the water demand claimed for the protection of environmental systems and the services they provide for the people directly or indirectly depending on their existence. The term ‘sustainable management’ has been introduced to describe the attempt to balance the use of the river basin resources in favor of human development without compromising and destroying the environment.

There is agreement that this attempt has not always been implemented successfully, and the competition for water frequently becomes emotionally loaded in irrigation agriculture where water is used for crop production and as a priority resource to sustain human livelihoods.

Water resources subject for IWRM within a river basin are provided by rainfall, glaciers, snow covered mountains, lakes, rivers, wetlands, groundwater aquifers and marine waters [COSTANZA et al. 1997]. In their regional setup and interaction they comprise the natural river basin water potential, which in turn is the base IWRM and economic development relies on.

2. SYSTEM STATUS INFORMATION

A prerequisite for IWRM in the sense of an sustainable development is sufficient information about the distributed status of the river basin subsystems. It is required for a holistic system analysis [FLÜGEL, 2000] and related data and information must be differentiated according to space and time, 'hard' and 'soft' data or knowledge:

(i) Space related information is specifying the spatial heterogeneity within the river basin. It is used for distributed modeling and the analysis of transport systems such as the interaction between upstream and downstream basin regions, i.e. in sediment transport and deposition.

(ii) Time related information can either be connected with a point of observation as it is the case for a time series, or can describe the change of a landscape such as land use obtained from remote sensing. Examples in relation to IWRM are hydro-meteorological measurements or records of irrigation water consumption.

(iii) Beside these 'hard' data used to quantify the status of a system, 'soft' information are frequently obtained by socio-economic questionnaires. They are less quantitative but describe the system status, aiding decision making from the human perspective. Peoples judgement about the status and quality of water supply are an example for such information.

(iv) 'Knowledge' is another important part of information that assess and specifies a river basin. The interpretation of this term 'knowledge' might often be different, but it is understood that such information should be available for decision makers to improve the system status. Classifying the erosion potential [FLÜGEL & MÄRKER, 2003] based on measurable landscape criteria is an example for such a 'knowledge' information.

In summary system analysis is based on different data and information which describes and quantify the spatially distributed system status. The latter, however, is the integration of the situation within the respective subsystems of the river basin. The higher the basin's heterogeneity and socio-economic development, the higher are the needs to organize the information in such a way that they can be made available for decision makers.

3. THE CGIAR CHALLENGE PROGRAM (CP) "WATER AND FOOD"

The Challenge Program (CP) „Water and Food (CPWF)", which has been launched by the Consultative Group of International Agricultural Research (CGIAR) in November 2003 requires such holistic system analysis and respective data and information strategies. The program is addressing deficits in respect to IWRM and crop production [http://www.waterforfood.org/].

In the context of a semi-arid river basin with predominantly irrigation agriculture the problem can be highlighted as follows: **(1)** A growing population has to be supplied with sufficient food and water as a basic requirement to alleviate poverty and sustain livelihoods. **(2)** Water diversion to irrigation agriculture will be under increasing stress and requires sophisticated management and decision making. **(3)** The worldwide recognized necessity to reserve water for the environment is becoming a further priority factor for basin water management. The challenge is to divert sufficient water for growing crops and simultaneously to compromise the demands of other important water users.

3.1 Objective and Research Scales

The general objective of the CPWF therefore is to support an increase crop production for a growing population but to maintain the level of present use of water for agriculture and thereby stimulate socio-economic development and to preserve the environment.

This objective is dealt with by the CPWF in five research themes active in a set of macro-scale benchmark river basins (BB), which provide data, information, and methodical knowledge about generic and local basin components and their transport process dynamics.

The five themes are organized within three hierarchical scale levels:

- The micro-scale level comprises the process systems describing the *plant-field-farm-system* level.
- The meso-scale level is represented by the *river basin*. Special focus is given on **(i)** the availability and quality of water from different sources such as surface water, groundwater and precipitation, and **(ii)** the interactions and trade-offs among and across water consumers.
- The macro-scale is the *external environment* in which the river basin is situated, and comprises the national and global environment. Special attention is given **(i)** on the water sector in

relation to global circulation dynamics and anticipated global climate change, and **(ii)** on macro-economic policy that impact the water sector such as trade in food and fiber, or energy policies.

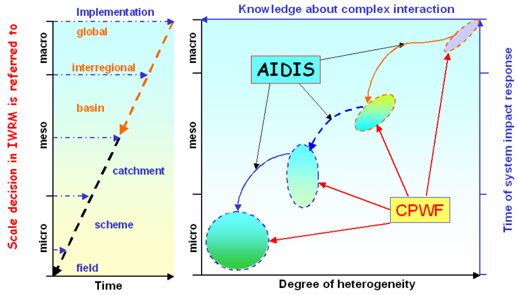


Fig. 2: Scale bridging data integration of AIDIS in the CPWF

3.2 Benchmark Basins

The geographic focus of the CPWF is provided by the nine Benchmark Basins (BB), which act as 'real world' research laboratories and server nodes for the AIDIS network implementation. As shown in Fig. 3 they have a global dimension in their geographic distribution, and provide a representative assembly of the "real world" water allocation problems the Challenge Program is focussing on.

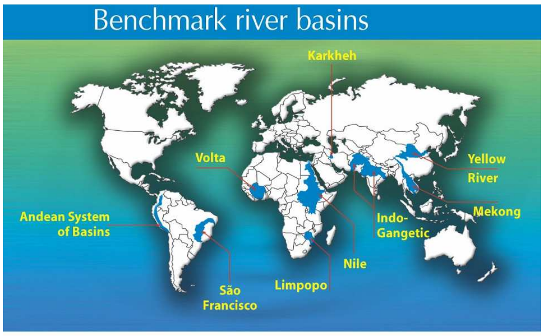


Fig. 3: Benchmark River Basins of the CPWF

The BBs are: **(1)** Yellow River (E-Asia), **(2)** Mekong River (SE-Asia), **(3)** Indus-Gangetic System (Asia), **(4)** Karkeh River (W-Asia), **(5)** Limpopo River (S-Africa), **(6)** Nile River (E-N-Africa), **(7)** Sao Francisco (S-America **(8)** the Volta River (W-Africa) and **(9)** several Andean Rivers (S-America and M-America).

River basins stretching over different climatic zones such as the Nile River have been chosen by purpose to study and simulate process dynamics interactive on different scales and various levels across geographical regions. It is foreseen that research projects will establish a number of pilot study areas on a micro- and meso-scale level representing the natural and socio-economic heterogeneity of the respective BB.

Research activities in each BB of the CPWF will be guided and advised by Benchmark Basin Coordinators (BBC) stationed in a National Agricultural Research Service (NARES). Thse institutions will also host the data base server and manage the system within a distributed server network.

4. OBJECTIVES OF AIDIS

The CPWF has adopted the principle of integrated systems analysis and is carrying out research of interactive process dynamics across scales. It consequently will be collecting and managing a diverse assembly of data and information comprising beside others **(i)** time series related to water quantity and quality, **(ii)** information about climate change induced by increasing concentrations of greenhouse gases, **(iii)** socio-economic census data, and **(iv)** information about environmental systems. From this diversity the following **conceptual objectives** of an Adaptive Data Information System (AIDIS) have been identified:

(i) The system must have a geo-spatial extension as 'hard' and 'soft' data have a geographic reference that can be specified by a Cartesian coordinate system.

(ii) The structure of the data and information administration must reflect the differentiation of the river basin's 'real world' subsystems. The latter in turn comprise **(a)** generic components based on the natural environment within the global setup and **(b)** basin related components which reflect the regional basin setup.

(iii) The system has to reflect the hierarchical structure of the transport systems within the natural environment and human development (NEHD) in respect to energy, water, solute and sediment transport or economic productivities.

(iv) It must have the ability to integrate the different kinds of data and information in such a way that they are available for the decision making process and for the design of "what-if-scenarios" for dynamic system modeling.

(v) Information and knowledge from past research should be organized in a River Basin Information System (RBIS) accounting for the conceptual.

(vi) The system's structure must be flexible enough so that extensions can be made depending on the progress of the system's understanding within the researchers and decision makers community.

(vii) To ensure compatibility between distributed AIDIS installations the metadata associated with the respective geo-spatial datasets must be encoded according to the ISO19115 standard (ISO, 2003).

In addition to these conceptual objectives some important **technical objectives** must be met if AIDIS should be successfully applied in present research programs:

(viii) Arc-View shape files and other map formats should be supported in terms of import, export and visualization

(ix) Distributed installations must be realized in a Web-based 'thin client-server' design.

(x) Replication strategies between the distributed server installations must account for specific needs in respect to data security and distribution policies.

(xi) Software for the data base management system (DBMS) and its components should be available on demand.

A literature survey [ANZLIC, 2000; FGDC, 1997; UNRCC-AP, 1997; UNRCC-Americas, 2001; PCGIAP, 1998] revealed that there are numerous attempts to organize and coordinate national, regional and global geo-spatial data. However, integrating such diverse data and information as required by the CPWF, and realizing the conceptual and technical objectives specified above has not been reported so far in the context of such a program.

It was therefore decided by the CPWF and the International Water Management Institute (IWMI) to initiate such a development [FLÜGEL & RIJSBERMAN, 2003] in cooperation with the DGHM at the Friedrich-Schiller-University (FSU-Jena) in Germany [http://www.geogr.uni-jena.de]. A prototype has been presented at the CPWF-Baseline Conference in Nairobi, Kenia in November 2003, which is undergoing continuous updates and extensions since then.

5. AIDIS SOFTWARE STRUCTURE

Improved strategies of IWRM – one of the CPWF key objectives – requires decision making based on reliable data, system information, and model simulation [DAVID ET AL. 1997]. They have been organized and are easily accessible in the Adaptive Integrated Data Information System (AIDIS), that accounts for the conceptual and technical objectives in the following way:

- AIDIS is based on open source software components which are complemented by the developments done by the DGHM from the FSU-Jena, Germany [http://www.geogr.uni-jena.de].
- The object-relational model PostgreSQL [WORSLEY & DRAKE, 2002] has been selected as a relational data base management system (RDBMS). It is highly extensible, and permits the use of the procedural Standard Query Language (SQL).
- PostGIS [http://postgis.refractions.net/] is used as the geo-spatial extension. It provides import and export of Arc-View© shape files and OpenGIS "Simple Feature Specification for SQL" [http://www.opengis.org/specs/].
- The APACHE [http://httpd.apache.org/] Web-Server is used for the client/server communication. The visualization of maps is done by means of the Minnesota Map Server (MMS) [http://mapserver.gis.umn.edu/].

Because of its different software components users might experience problems when installing and operating the system. Furthermore the use of the MMS requires some training and the design of separate map templates to control the visualization and data queries. And, last but not least, the data, information and metadata stored in AIDIS must be linked with the attribute data imported with the various GIS maps.

For these purposes the DGHM developed the following complement AIDIS components:

- A Data Base Networking System (DBNS) automatically installs all required software components including the Linux operating system on the server. Detailed log files are written during the installation for error checking. Furthermore the DBNS automatically handles the replication between the distributed AIDIS nodes in the network and the AIDIS common to the CPWF (Fig. 4).
- The River Basin Information System (RBIS) is based on the MMS and makes use of the OpenGIS functionalities provided by PostGIS. RBIS is also handling the map related data import and export via the Web and thereby allows to generate new GIS maps for dissemination i.e. within a Clearinghouse Network (CN). At present more GIS functionality is developed for RBIS by means of the OpenGIS "Simple

Feature Specification for SQL" [http://www.opengis.org/specs/].

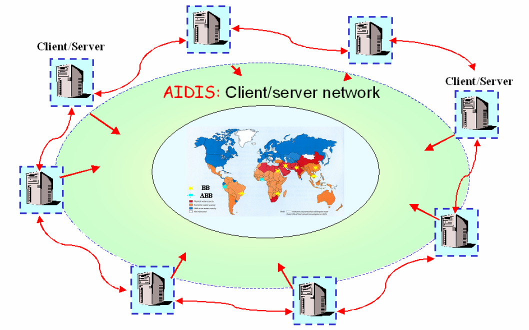


Fig. 4: Distributed Client/Server Network Nodes in the BB

The modular component structure of AIDIS provides sufficient flexibility for distributed node installations as required by the CPWF. It also accounts for the different security requirements in the different BB where data providers have their own dissemination policies which can be implemented by the AIDIS administrator in each individual BB node. The replication strategy implemented by the DBNS is only accessing such data and information from the various BB nodes that have been assigned for global distribution in the AIDS common to the CPWF.

5.1 AIDIS data model components

The data model of AIDIS accounts for the remaining conceptual objectives. It was designed as a object-relational model, and has been implemented by means of the Unified Modelling Language (UML). This model provides the required flexibility to adapt AIDIS to different research programs and benchmark river basins as regional node installations. It furthermore permits the integration and parameterization of dynamic simulation models as well as the coding of knowledge and expertise obtained from previous research projects. The latter, however, require a structural classification to be assessed via a relational table structure.

The knowledge component reveals the research related concept of AIDIS, which is not only a sophisticated RDBMS but also provides an development interface for integrating frontier research. In this regard AIDIS has potential to be extended towards a knowledge based Decision Support System (DSS) for IWRM and river basin management.

The following data model objects have been implemented as AIDIS packages so far:

(i) Common to Data Base (CDB) comprises all tables related to metadata. The metadata philosophy of AIDIS is based on the fact, that all data and information obtained from the river basin have a geographical reference. The metadata model therefore is based on the implementation of the ISO19115 standard [ISO, 2003], and respective code tables used to specify features, parameters or units.

(ii) Spatial Data (SD) are comprising GIS vector and grid maps as well as satellite images. This package is linked with the tables generated from the AIDIS PostGIS installation and the imported shape files geometries and attributes.

(iii) Thematic Data (TD) provides access to 'hard' data measured, simulated or observed in the field or the lab, and also includes census information.

(iv) River Basin Information System (RBIS) comprises information about the natural environment and the human development (NEHD) of the river basin such as water distribution and management in irrigation agriculture.

(v) Model Information System (MIS) provides metadata about models that have been used in the research program or are integrated into AIDIS for parameterization, such as is done at present for the J2000 rainfall-runoff model [KRAUSE, 2001].

(vi) Data Derived Evaluation (DDE) containes scientific knowledge derived from published data evaluation together with indicators classifying the spatial heterogeneity within the basin.

5.2 Parameterization and Regionalization

Models such as PRMS/MMS [LEAVESLEY ET AL., 1983] and J2000 [KRAUSE, 2001] are distributed and object oriented. They integrate well researched process algorithms which have been developed for the micro-scale, meso-scale or macro-scale, and can directly be parameterized from AIDIS. The challenge, however, is to link hydrological and climate model tools in such a way that their simulation output is complement model input of the next higher or lower model scale hierarchy.

The development of sound regionalization concepts proceeded from lumped (basin is one unit) to fully distributed (basin consists of several connected units), and evolved hand in hand with the development of hydrological water balance models. Distributed models apply spatial process entities scattered over the basin either as geometric grid cells or as Hydrological Response Units (HRU) i.e. defined by FLÜGEL (1996). HRUs are identified according to process based criteria accounting for the components of the natural and human environment. They are delineated by means of sophisticated GIS analysis [FLÜGEL, 1997], geo-referenced via their coordinates, and networked via their topographic topology [STAUDENRAUSCH, 2001].

The HRU approach has been applied successfully in various climatic and topographic regions in Europe, the US and Africa [FLÜGEL, 1996; KRAUSE, 2001; STAUDENRAUSCH and FLÜGEL, 2001] and the attributes provided with the GIS data layers can directly be used in AIDIS to quantify their respective system status [FLÜGEL, 2000].

6. CONCLUSION

Integrated Water Resources Management (IWRM) requires decision making processes based on best knowledge of the river basin. This information comprises geo-spatial data components, resulting from measured and simulated systems dynamics of the natural environment (NE) and human development (ND).

As this demand could not be satisfied from existing products the Adaptive Integrated Data Information System (AIDIS) has been developed and applied for the Challenge Program "Water and Food" (CPWF). AIDIS is based on open source software and complement developments provided by the DGHM of the FSU-Jena, Germany. A object-relational data model is used for AIDIS which permits a modular component structure required for the regionalization of the distributed system dynamics.

AIDIS is under continuous development by the DGHM which also provides a comprehensive training program. As the data model and the software components are adaptive to various river basin setups AIDIS will also been implemented in EU-projects and regional research programs.

7. REFERENCES

ANZLIC Discussion and background papers. – In: Proceedings of the ANZLIC Clearinghouse Workshop, 3-4 May, Adelaide, Australia, 2000

CONSTANZA ET AL. The value of the world's ecosystem services and natural capital. *Nature* 387 15 May 1997

DAVID, O., BUSCH, C, KÜSPERT, K AND FLÜGEL, W.-A. Objektorientierte Datenbanken in der Hydrologie: Anwendungsbeispiel OMS. - In: GEIGER, W., JAESCHKE, A., RENTZ, O., SIMON, E., SPENGLER, TH., ZILLOX, L. & ZUNDEL, T. (Eds.): Umweltinformatik'97, 11. Intern. Sympos. D. Ges. F. Informatik, GI, Bd. 1, Vol. 1, 225-234, 1997

FGDC Framework, Introduction and Guide. Book of Federal Geographic Data Committee, Washington, 1997

FLÜGEL, W.-A. Hydrological Response Units (HRU's) as modelling entities for hydrological river basin simulation and their methodological potential for modelling complex environmental process systems - DIE ERDE, 127, 42-62, 1996

FLÜGEL, W.-A. Combining GIS with regional hydrological modelling using Hydrological Response Units (HRUs) - An application from Germany. - Mathematics and Computers in Simulation, 43, 305-312, 1997

FLÜGEL, W.-A. Systembezogene Entwicklung regionaler hydrologischer Modellsysteme. - Wasser & Boden, 52. Jg., H. 3, 14-17, 2000

FLÜGEL, W.-A. and MÄRKER, M. The Response Units Concept and its Application for the Assessment of Hydrologically Related Erosion Processes in Semiarid Catchments of Southern Africa. – ASTM-STP1420, 163-177, 2003

FLÜGEL, W.-A. & RIJSBERMAN, F. The Challenge Program "Water and Food" for river basin scale water resources assessment. – Proc. MODSIM'03, Vol. 1, 434-439, 2003

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ISO 19115:2003(E): Geographic Information – Metadata. – 1st edition, Geneva, Switzerland, 2003

KRAUSE, P. Das hydrologische Modellsystem J2000, Schriften FZ-Jülich, Bd. 29, 2001

LEAVESLEY, G..H., LICHTY, R.W., TROUTMAN, B.M. AND SAINDON, L.G., Precipitation Runoff Modelling System: User's manual, Water Resources Investigations 83-4238, USGS, Denver, Colorado, 1983

WORSLEY, J. C. & DRAKE, D. Practical PostgreSQL. – O'Reilly & Associates, 2002

PCGIAP a spatial data infrastructure for the Asian and Pacific region. PCGIAP Publication No. 1, Canberra, 1998

STAUDENRAUSCH, H. Untersuchungen zur hydrologischen Topologie von Landschaftsobjekten für die distributive Flussgebietsmodellierung. PhD FSU Jena, http://www.db-thueringen.de, 2001.

STAUDENRAUSCH, H. and FLÜGEL, W.-A. Development of an Integrated Water Resources Management System in Southern African Catchments. – Phys. Chem. Earth (B), Vol. 26 , No. 7-8, 561-564, 2001

UNRCC-AP Resolutions of the 14th United Nations Regional Cartographic Conference for Asia and the Pacific, 3-4 February, Bangkok, 1997

UNRCC-Americas Proceedings of the United Nations Regional Cartographic Conference for Americas, January, USA, 2001

A Conceptual Model about the Application of Adaptive Management for Sustainable Development

M. Knoflacher and U. Gigler

ARC systems research, Seibersdorf, Austria

Abstract: This conceptual model is based on theoretical implications and analyses of empirical examples, and focuses on how functional processes in the environmental system affect decisions taken in the social system. Complex functional relationships at several temporal and spatial scales such as climate conditions or pollutant emissions characterise functional processes in the environmental system. Actions of humans in the environmental system are crucially constrained by limited information about system conditions, and limited predictability of system development. Both constraints are not equally valid for all functional processes, and all interactions between humans and the environmental system, because diverse human activities and interests lead to an array of interpretations regarding the environmental system. An active comparative exchange of partial views about the environmental system can enlarge social knowledge. These requirements cannot be satisfied solely by scientific investigations; observations and experiences of non-scientific human actors also have to be considered. Integrating the different types of knowledge by applying systems analysis methods delivers the basis for developing general management plans e.g. in urban planning, although uncertainty in decision-making cannot be eliminated fully through integration alone. It is nevertheless an essential precondition for applying the adaptive management instrument because it delivers, in addition to an improved understanding among stakeholders, an overview of functions such as impact-effect relationships and scales relevant for human actors in the environmental system. The adaptive management process is necessary in order to establish a functioning feedback system, to determine the type of organisational framework needed, and to ascertain which actor groups to involve in decisions and assessment procedures within the framework of sustainable development.

Keywords: conceptual model, systems theory, interaction loop, adaptive management, sustainability

1. INTRODUCTION

1.1 General introduction

Highly complex situations on the interface between human and environmental systems often present major challenges to those attempting to manage such systems. Examples include managers of natural resources who are charged with having to harmonise often conflicting objectives such as preserving a particular ecosystem and managing it sustainably while simultaneously extracting resources (e.g. forest). Urban planners, developers, housing specialists, social workers or others involved in e.g. regenerating derelict urban areas also face multiple challenges that can include reducing crime rates and unemployment by providing new jobs, engaging in environmental clean-up while attempting to attract industry and commerce, and physically improving a particular part of town thereby also improving a city's image. These and many other situations exhibit similar characteristics. All situations affect numerous stakeholders and at the same time are influenced by a multitude of interested parties [Ander et al., 2001; Tomerius, 2000]. Environmental and human systems are inherently complex and attempts to understand them often fail, because of too many unknowns. Nevertheless, in order to be able to manage natural or urban areas more successfully, improved ways of dealing with information and information flows need to be found in settings that are characterised by complexity, uncertainty, and unexpected situations [Holling, 1978]. A conceptual model showing interactions between human and natural systems illustrates the conditions necessary for adaptive and flexible management approaches. Only then will it be possible to make informed and sensible decisions that allow those responsible to tackle the multiple issues at hand and achieve more sustainable outcomes.

1.2 Adaptive management

Adaptive management is an integrated, multi-disciplinary approach for managing natural resources such as wetlands or forests and can also be applied in other fields such as urban planning or environmental management. The instrument was conceived to develop more effective and more resilient policies [Holling 1978] acknowledging that natural systems always change as a result of human intervention and therefore require an adaptive approach that is capable of responding to such changes [Gunderson, 1999]. The approach attempts to find viable solutions in situations where many stakeholders with differing objectives facing limited information must make decisions.

Key elements of the instrument include the use of experiments which allows managers and stakeholders to learn from those experiments and assess successful or failing approaches [Walters and Holling, 1990]. Comprehensive monitoring throughout the management process is another crucial element that informs managers about the area under observation [Grumbine, 1996; Lessard, 1998]. In this particular approach, it is essential to employ system-relevant indicators ideally before, throughout, and for a period following the management process. Another very central feature of adaptive management is that all relevant stakeholders should be involved in the process, remain informed, provide input and take part in decision-making [McLain and Lee, 1996]. Adaptive management thus attempts to make learning through feedback more efficient and part of the management process, promotes flexibility, and involves all relevant stakeholders.

2. EMPIRICAL INVESTIGATIONS

Empirical evidence from urban regeneration processes in different European cities shows that certain key aspects substantially contribute to successful revitalisation examples. Those include cooperation among stakeholders, a clear vision, provisions to remain flexible and adaptive throughout the process, and a regional and market-oriented approach [Ander et al., 2001; Seewer and Menzi, 1999]. Regenerating large, old, industrial inner city sites at a minimum requires large sums of money, a clear commitment by the owner(s) that the site should be redeveloped, a clear management strategy, and a time-frame. Additionally, those engaging in revitalisation need to strike a balance between investing in environmental clean-up, attracting industry and commercial enterprises to ensure employment opportunities and fulfilling short and long range goals of the city. The following examples illustrate the need to apply features of adaptive management in complex regeneration examples.

Planners and developers in Gothenburg, Sweden charged with redeveloping a former industrial site initially set out to attract strictly industrial users. In the span of the following decade, the first strategy being unsuccessful, they developed a vision and adapted their strategy and plan to one that was more responsive to the needs of Gothenburg and the region surrounding it. Instead of focusing on industry alone thereby homogenising the site, they promoted a more diverse and mixed use approach [Ander et al., 2001]. Project promoters also understood the need to observe market developments and adapt to potential changes over the long-term, in this case several decades, in contrast to traditional linear planning approaches. Thus, the stakeholders' initial experiment failed and they incorporated what they learned into the new vision and plans, simultaneously observed market conditions, dealt with the inherent uncertainty and acted according to the new requirements.

The same case also demonstrates the importance of involving all relevant stakeholders in the management process from the very beginning and fostering public-private partnerships (Ekman, pers. comm., 2003). Representatives from public and private organisations communicated needs, changes, and constraints throughout the process and cooperated closely on revitalising the site. Because the stakeholders had known one another for years working on common objectives, a certain level of trust had been built that was instrumental in successfully revitalising the site. Each stakeholder was informed about the planning process or took part in it, and most importantly, they each participated in decision-making.

A water resource management case study in the United States illustrates that lack of cooperation and trust can result in a disfunctional programme [Gigler, 1998]. Disagreements over monitoring plans and responsibilities, frequent institutional reorganisations with an unclear assignment of tasks, and personal disagreements among staff all led to deepening distrust among those responsible for the monitoring programme and resulted in poor execution of the prescribed monitoring and data analysis. Due to conflicts between staff, preserving the headwaters of two streams and their associated wetlands in an acceptable state and at a crucial time in spite of development in the vicinity was severely hampered.

3. THEORETICAL BACKGROUND

3.1 General relationships

Sustainable development depends on a long term balance between the ecological and the human social system. The realisation of this anthropocentric concept crucially depends on understanding the characteristics of particular systems and their interdependencies. Environmental systems are driven by the dynamic counteracting hierarchies of free energy and self organisation [Knoflacher et al., 2003], (Figure 1). Different balance levels of the counteracting processes can be classified as partial systems of the environmental system [Knoflacher, 2002].

During evolution, balance was achieved through losses and emergence of species and their ability to adapt to changing conditions continuously [Cockburn, 1995)]. The challenge of the sustainable development concept in this context is that it does not accept any losses of large portions of a population because of e.g. harmful environmental conditions [WCED, 1987].

The technological potential of the human population represents a benefit as well as a risk for sustainable development (Figure 1). In some cases, it enables human liberation from the energetic hierarchy. This enables human welfare and the development of particular characteristics of human societies. However, it also increases the risk of adverse effects because of critical changes of environmental conditions or because of self destruction during conflicts in the human population.

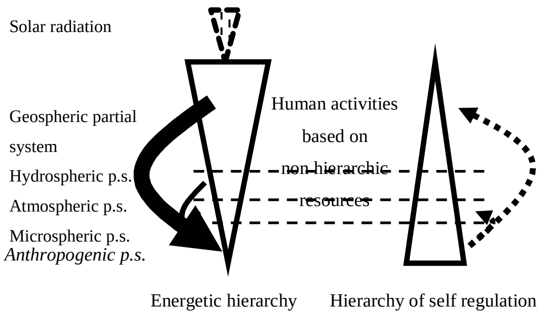


Figure 1. Systems framework conditions for the sustainable development concept.

These systems framework conditions illustrate that there is a strong need for a better understanding of the affected partial systems. The entire human population also needs to make use of its high potential to adapt to dynamic system constraints within a certain region.

Interactions between the human social system and the environmental system are asymmetric. Basics for human life are the exchange of energy and chemical compounds between the human social system and the environmental system. These exchanges are increasing entropy in the environmental system, which can only be compensated by self organising processes in ecosystems. Of particular interest for sustainable development is how humans perceive information coming from the environmental systems and their impacts on environmental systems beyond basic interactions.

3.2 System characteristics

In spite of voluminous literature about characteristics of environmental systems [Odum, 1983; Mason and Moore, 1985; White et al., 1992; Joergensen, 1992] and social systems [Giddens, 1993; Luhmann, 1994; Habermas, 1995; Bourdieu, 1997] relatively few publications deal with interactions between information flows and physical impacts. This is particularly valid for considering conditions of complexity in this context.

Environmental systems are at least complex in their structure and functions. Structural complexity can be found in the composition of system elements at a huge bandwidth of spatial and temporal scales, also including relationships among different elements. A typical example is the composition of a natural forest ecosystem with biotic and abiotic components and their relationships. Functional complexity is characterised by the different qualities entailed in processes within environmental systems. It is expressed in the example mentioned e.g. in energetic webs combined with material flows.

Human social systems are at least complex in their functions. Recent human societies are characterised by interactions of human actors with different responsibilities and competences [Luhmann, 1994]. In addition, human actors can change their functional membership to distinct functional groups. A typical example for that is an individual person switching between the professional and the private role.

Complex interactions also result from systems complexity. This will be familiar to anybody who has attempted to bring environmental effects into a consistent order. But it becomes also apparent through numerous and parallel running individual interactions e.g. within a large region. Several thousands of people can work at the same time in their garden, are cutting wood in the forest, or drive cars. Each interaction is based on an individual decision and will be related to individual targets. Such individual interactions can be influenced by general laws only to a limited extent for the following reasons. 1. Due to complexity of structural conditions; each individual is living under different framework conditions, in part because of social interactions. 2. Due to variability of individual targets and values; e.g. different values a forest has for forest owners versus tourists. 3. Because of probability to get punished if the general rule is neglected. This probability will be strongly reduced with an increasing number of adverse activities taking place and with a decreasing difference in individually caused effects.

4. THE CONCEPTUAL MODEL

4.1 The basic interaction loop

Individual human interactions with the environmental system can be interpreted as a regulation loop (Figure 2).

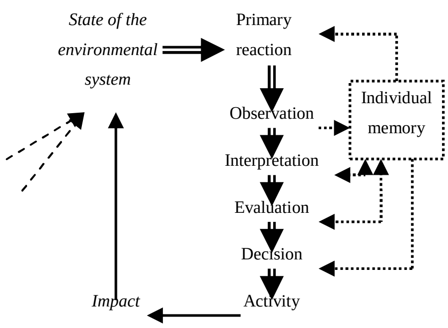


Figure 2. Basic feedback loop of human interactions with the environmental system.

Essential preconditions for a first reaction to environmental state conditions are a person's physiological and psychological perceptibility. Physiological perceptibility depends on specific characteristics of the human sensory system [Schmidt and Thews, 1980]. Psychological perceptibility depends on the specific awareness of an individual, influenced by former experiences, individual objectives, and information from the social system [Hoffmann, 1976; Popper, 1995]. It is presented in Figure 2 as individual memory.

A more detailed observation of the actual environmental state depends on information from the first reaction and on the general interest of the individual. It delivers comprehensive information about the environmental system in relation to an individual's expectations. This information is used for individual interpretation of actual conditions of the environmental system, and potential reasons for it. Interactions with other persons can modify the individual interpretation.

An individual evaluation differs from interpretation. In this step results of the interpretation are compared with individual objectives that take into account additional information about social framework conditions. An individual decision depends on outcomes of the evaluation. Basically, a decision has to be made about whether to act. Additional decisions about specific activities would follow if the decision was to move forward.

The subsequent activity is not directly related to the individual. At a minimum, the individual only has to give instructions to other persons for implementing an action. At a maximum, the individual has to act personally.

The reaction of the environmental system to specific impacts depends on the qualitative and quantitative characteristics of the impact and the specific characteristics of the affected environmental system, and the occurrence of additional impacts caused by other persons.

Non-linear relationships can be expected between the individual impact and effects in the ecosystem. Time lags and delocalisation effects can camouflage substantially the real effects in the environmental system. Counter- intuitive reactions of the environmental system can cause misinterpretations. Hence, the basic interaction loop can only support a first guess about the actual state of the environmental system and the effects of a certain impact. Through repeated interactions under different circumstances, it is possible that the accuracy of the first guess can be improved gradually, but not to complete certainty. Uncertainty will increase again, if attempts are made to predict the long term effects of an impact [Knoflacher, 2002].

4.2 Adaptive connections of the basic loop

Different experiences and interests exist within a larger population, because of different individual tasks and objectives. As a result, one can expect variability in individual memory and first reactions. However, the variability is limited because of common biological and cultural constraints [Berger and Luckmann, 1991]. By applying scientific methods, a slight extension of the breadth of individual memory and primary reactions can be achieved [Speck, 1980].

Consequences of this variability are different perceptions and interpretations of the environmental system. The a priori exchange of information among different actor groups is very limited because of social barriers [Luhmann, 1994].

Traditional technocratic approaches are overcoming this problem by a hierarchical interpretation of the environmental system, where scientific interpretation ranks at the top. Functionally, that reduces the potential variability of interaction loops to one dominant loop with severe consequences. Considering the multiple interactions that are possible between the environmental and the social system, this approach neglects structural complexity. Most of the affected actor groups will not benefit from scientific findings, because of different acting scales and different contextual conditions. Well known social reactions are that actor groups negatively react to decisions and doubt expert opinion.

The challenge in attempting to overcome these problems is the difference of spatial and temporal scales of different individual basic loops. Hence, harmonisation is only possible at certain points in basic loops. Different actors within a region can only react adaptively, if a common understanding of relevant environmental system properties is achieved. This in turn needs a transformation of individual interpretations (I_i) to a common interpretation (I_C) of the system state.

$$I_i \rightarrow I_C \quad (1)$$

Several methods can be applied for identifying a common interpretation among key persons of different actor groups [Kruse et al., 1996; Seifert, 1999; Geißler and Rückert, 2000]. Results of these processes have to include quantitative and qualitative properties as well as relevant indicators for different actors of the environmental system that is being considered.

Systems analyses of these results have to be carried out in order to identify potential risks for further development, and optimised solutions. This task should also be carried out by involving representatives of the affected actor groups to avoid misinterpretations, and in particular to develop a common understanding about constraints and risks. Formally individual memory (IM_i) is extended to more integrated individual memory $(IM_i^{'})$ in this process.

$$IM_i \rightarrow IM_i^{'} \quad (2)$$

This process can result in a structural adaptation of all relevant interactions between environmental and social systems. A formal criterion for fulfilling this objective is to integrate all relevant actor groups in the whole process.

Clarifying functional roles of different actors provides the framework for defining targets and indicators for future activities. Crucial for acceptance of individual targets is the agreement on a common target for future development of the region. The common target should be observable and easy to understand for all relevant actors.

Subsequent definitions of interaction rules among involved actors are necessary to implement outcomes. Interaction rules can be defined as agreements or contracts that depend on actor requirements. In this context, it has to be clear, that accuracy in forecasting future developments is limited because of basic characteristics of ecological and social systems [Green et al., 2003].

All agreements about interactions and targets are therefore only first guesses in relation to any future development. Therefore it is a big challenge to agree on dates and reasons for common assessment procedures in the future. Such assessment procedures should support adaptability to changes in framework conditions by considering the general strategy, and methodically they should be based on principles of adaptive connections in basic loops.

4. CONCLUSIONS

Adaptive management offers a very promising approach for managing human and natural systems as demonstrated in the case studies described. The interaction loop provides the theoretical basis for understanding typical interactions between humans and the environment. Only by introducing the adaptive element, however, does it become possible to accurately portray and understand typical interactions between humans and the environmental system. Acting adaptively is therefore imperative when dealing with complex, highly uncertain situations with many unknown variables.

Remaining flexible throughout the development process, experimenting, involving stakeholders and adapting to the needs dictated by partners, the market or the regulatory framework was an explicit objective of at least three of the four case studies. The studies demonstrate that managing regeneration adaptively enables stakeholders to deal with inherent uncertainty more successfully and can contribute to sustainable outcomes even in a setting where sustainability is not an explicit goal.

5. ACKNOWLEDGMENTS

The empirical work for this paper was funded through the MASURIN project (contract no. EVK4-2001-00054) by the European Union.

REFERENCES

Ander, et al., Norra Älvstranden: The Process, City of Gothenburg, 46 pp., Sweden, 2001.

Berger, P. and T. Luckmann, *The Social Construction of Reality*, Penguin Books, London, 1991.

Bourdieu, P., *Sozialer Sinn*, Suhrkamp, Frankfurt, 1997.

Ebeling, W., J. Freund, and F. Schweitzer, *Komplexe Strukturen: Entropie und Information*, B. C. Teubner, Stuttgart, 1998.

Ekman, J., , pers. communication, project leader, Norra Älvstranden Utveckling AB, 2003.

Geißler, P. and K. Rückert (Hrsg.), *Mediation – die neue Streitkultur*, Psychosozial, Gießen, 2000.

Giddens, A., *Sociology*, Polity Press, Cambridge, 1993.

Green, P.J.,. Hjort N.L., and S. Richardson (eds.), *Highly Structured Systems*, Oxford University Press, Oxford, 2003.

Grumbine, R.E., Reflections on "What is Ecosystem Management", *Conservation Biology*, 11(1), 41-47.University Press, Oxford, 1996.

Gunderson, L., Resilience, Flexibility and Adaptive Management – Antidotes for Spurious Certitude?, *Conservation Ecology*, 3(1), art. 7, 1999.

Habermas, J., *Theorie des kommunikativen Handelns*, Suhrkamp, Frankfurt, 1995.

Hoffmann, H.-J., *Psychologie und Massenkommunikation*, de Gruyter, 1976.

Holling C.S., editor. 1978, *Adaptive Environmental Assessment and Management*, John Wiley and Sons, , 377 pp., London.

Joergensen, S. E., *Integration of Ecosystem Theories: A Pattern*, Kluwer, Dordrecht, 1992.

Knoflacher, M. , A Conceptual Model for Integrated Assessment, in: Integrated Assessment and Decision Support Proceedings of the 1st Biennial Meeting of the International Environmental Modelling and Software Society; IEMSS Congress, 3, 121-126, Lugano, 2002.

Knoflacher, H.M., U. Gigler., T. Tötzer, and B. Naefe, Assessment of sustainability – can it be standardised? Proceedings EasyEco, www.sustainability.at/easy, 2003.

Kruse, L., C.-F., Graumann and E.-D. Lantermann, *Ökologische Psychologie*, Beltz, München, 1996.

Lessard G., 1998, An adaptive approach to planning and decision-making, *Landscape and Urban Planning*, 44, 81-87

Luhmann, N., *Soziale Systeme*, Suhrkamp, Frankfurt am Main, 1994.

Mason, B., and C.B. Moore, *Grundzüge der Geochemie*, Enke, Stuttgart, 1985.

McLain, R.J. and R.G. Lee, Adaptive Management, Promises and Pitfalls, *Environmental Management*, 20(4), 437-448, 1996.

Odum, E.P., *Grundlagen der Ökologie*, Thieme, Stuttgart, 1983.

Popper, K. R., *Objektive Erkenntnis*, Campe, Hamburg, 1993.

Schmid, R.F. and G. Thews, *Physiologie des Menschen*, Springer, Berlin, 1980.

Seewer, U. and R. Menzi, Raumplanung und Industriebrachen, in: Dossier: Raumbeobachtung, Bundesamt für Raumplanung, Bern, Schweiz, 1999.

Seifert, J.W., *Moderation & Kommunikation*, Gabal, Offenbach, 1999.

Speck, J. (Hrsg.), *Handbuch wissenschaftstheoretischer Begriffe*, Vandenhoeck, Göttingen, 1980.

Tomerius, S., Recycling Derelict Land in U.S. and German Cities – Transatlantic Sharing of Approaches, Strategies and Visions., 2000, http://www.difu.de/english/occasional/recycling.shtml

Walters, C.J. and C.S., Holling, Large-scale Management Experiments and Learning by Doing, *Ecology*, 71(6), 2060-2068, 1990.

WCED, *Our common future*, Oxford University Press, Oxford, 1987.

White, I.D., D. Mottershead, and S.J. Harrison, *Environmental Systems*, Chapman & Hall, London, 1992.

A Model Based on Cellular Automata for the Simulation of the Dynamics of Plant Populations

Stefania Bandini [a], Giulio Pavesi[b]

[a]Dept. of Computer Science, Systems and Communication
University of Milano–Bicocca
Milan, Italy
bandini@disco.unimib.it

[b]D.I.Co.
University of Milan
Milan, Italy
pavesi@dico.unimi.it

Abstract: We present a Cellular Automata based model for the simulation of the dynamics of plant populations. The evolution of a plant population in a given area mainly depends on the resources available on the territory (in turn influenced by other factors like sunlight, rain, temperature, and so on) and how different individuals (plants) compete for them. Traditional methods used in this field are continuous and based on differential equations that model the global evolution of the system: unfortunately, most of the data needed to provide reliable parameters for these models are usually scarce and difficult to obtain. The model we present is instead thought in a bottom–up fashion, and is based on a two-dimensional Cellular Automaton, whose cells, arranged on a square grid, represent portions of a given territory. Some resources are present on the area, divided among the cells. A tree is represented in the model by a set of parameters, defining its species, its size (that is, the size of its parts such as limbs, trunk, and roots), the amount of resources it needs to survive, to grow, and/or reproduce itself (produce fruits). The model has been applied to the simulation of populations consisting of robiniae (black locust), oak, and pine trees, on the foothills of the italian alps, with encouraging results reproducing real experimentally observed population trends.

Keywords: Cellular Automata; discrete models; plant population dynamics.

1 INTRODUCTION

Modeling the dynamics of plant populations living in a given area is a widely studied and extremely challenging problem, as described for example by M.G.Barbour et al. [1998] and J.Silverton and D.Charlesworth [2001]. The main difficulty lies in the acquisition of data for the definition of the parameters of the models, that must cover very long time periods, especially in the case of perennial plants. Such data must include the resources available on the territory, and those needed by plants to sprout, survive, grow, and reproduce themselves. In fact, the evolution of a plant population is mainly influenced by the resources available (i.e. sunlight, water, substances present in the soil), and how the different individuals compete for them. Traditional models are continuous and based on differential equations like those introduced and employed (among many others) by J.L.Uso-Domenech et al. [1995]; Q.Zeng and X.Zeng [1996a, b]; C.Damgaard et al. [2002], and usually model the evolution of the system with global parameters such as the total number of trees and their overall biomass. More recently, Cellular Automata have been introduced to study this problem as for example by R.L.Colasanti and J.P.Grime [1993]; H.Baltzer et al. [1998], but usually their application was limited to the evolution of single infesting species.

In this paper, we present a discrete model based on two–dimensional Cellular Automata, that allows the simulation of the evolution of heterogeneous plant populations composed by different perennial species as in real woods and forests. The evolution of the system is thus modeled in a bottom–up fashion, that is, is the result of the interactions among single individuals and their competition for the resources available on the territory, as for example discussed by D.Tilman [1994] and J.Ehrlen [2000]. In this paper, we show how the model has been applied to the simulation of populations consisting of robiniae (black locust), oak, and pine trees on the foothills of the italian alps, with encouraging results reproducing real conditions. However, we believe that its generality and flexibility make it suitable for the simulation of different case studies and conditions, with no changes in its basic structure.

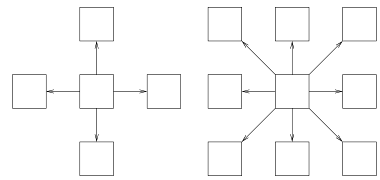


Figure 1: Von Neumann (left) and Moore (right) neighborhoods. In CA, the update of the state of a cell (the central one in the figure) depends on the state of the cell itself and on the state of cells defined as its neighbors.

2 CELLULAR AUTOMATA

A *Cellular Automaton* (CA) consists of a regular discrete lattice of *cells*, where each cell is characterized by a state belonging to a finite set of states. Each cell evolves (changes its state) according to a given *update rule*, which depends only on the state of the cell and a finite number of neighboring cells. All the cells of the automaton follow the same update rule. The automaton evolves through a sequence of discrete time steps, where all cells update their state simultaneously. For a more general introduction on CA we refer the reader to, for example, the book by E.Goles and S.Martinez [1990].

The idea of our model is that the cells of the CA represent portions of a given area. Each cell contains some resources, and if conditions are favorable, can host a tree. A tree is represented in the model by a set of parameters, defining its species, its size (that is, the size of its parts such as limbs, trunk, and roots), the amount of each resource it needs to survive, to grow, and reproduce itself (that is, produce fruits). A single tree has been "decomposed" in different parts in order to reproduce the effect of environmental influences. In fact, the environment and the resources available determine how the overall biomass of the tree is divided among the different parts composing it, as discussed by D.Tilman [1988].

When a tree produces fruits, some seeds are scattered in the neighboring cells. A seedling can sprout in a cell when the latter contains a seed, no other tree, and a sufficient amount of each resource. In this case, a tree is born, and the state of the cell now comprises also all the parameters defining the tree present in it (otherwise set to zero, if no tree is present). Then, the cell has also to contain enough resources to sustain the growth of the plant. The quantity of resources needed varies according to the species of the tree and its size. When a tree starts growing, its increasing mass begins to need a larger amount of resources, that can also be taken from the neighbors of the cell where it is located. Thus, the sprouting or the growth of other trees in its proximity is negatively influenced, that is, the tree starts competing for resources with the others. Whenever a tree cannot find enough resources to survive, it dies.

3 THE MODEL

We now give a more formal description of the model. The cells of the automaton are arranged on a two–dimensional square grid. The state of each cell is defined by a flag denoting whether or not it contains a tree, and a set of variables denoting the amount of each resource present in the cell, as well as the features of the tree (possibly) growing in it. The update rule of the automaton mainly depends on the presence of a tree in a cell. In case a tree is present, part of the resources of a cell (and in the neighboring ones, if the tree is large enough) are absorbed by the tree. Every cell also produces at each update step a given amount of each resource (that in any case cannot exceed a maximum threshold value). The production of resources in the cells is determined by a set of global parameters that reproduce environmental factors such as rain, presence of animals in the area, and so on. The effect of the presence of a tree in a cell on the neighboring ones has been modeled by making resources flow from richer cells to poorer ones (that usually contain less resources since a part of them is consumed by the tree). The resources we explicitly included in the model are water, light, nitrogen, and potassium. Both von Neumann and Moore neighborhoods (shown in Fig. 1) have been considered in the simulations. The CA can be thus defined as:

$$\mathbf{CA} = \langle R, N, Q, f, I \rangle$$

where:

- $R = \{(i,j) | 1 \leq i \leq N, 1 \leq j \leq M\}$ is a two–dimensional $N \times M$ lattice;
- H is the neighborhood, that can be either the von Neumann or Moore neighborhood;
- Q is the finite set of cell state values;
- $f : Q \times Q^{|H|} \rightarrow Q$ is the state transition function;
- $I : R \rightarrow Q$ is the initialization function.

3.1 The Cells

Each cell of the automaton reproduces a square portion of terrain with a side ranging from three to five meters. As mentioned before, each cell contains some resources, and can host a tree. Thus, the possible states of a cell must define:

- The type of terrain the cell reproduces;
- The resources present in the cell;
- The amount of resources the cell produces at each update step, and the maximum amount of resources it can contain, according to its type;
- Whether a tree is present in the cell, or not;
- If a tree is present:
  - the size of the tree;
  - the amount of each resource it needs at each update step to survive and grow;
  - the amount of each resource stored by the tree at previous update steps;
- Seeds scattered by trees living in the area.

If we assume that k types of resource and l different tree species are present in the area, the finite set of states Q can be defined as follows:

$$Q = \{\mathbf{R}, \mathbf{M}, \mathbf{P}, T, \mathbf{Z}_T, \mathbf{N}_T, \mathbf{U}_T^G, \mathbf{U}_T^S, \mathbf{R}_T, \mathbf{M}_T, \mathbf{G}_T, \mathbf{S}\}$$

where:

- $\mathbf{R} = \{r_1, \ldots, r_k\}$ is a vector defining the amount of each resource present in the cell;
- $\mathbf{M} = \{m_1, \ldots, m_k\}$ is the maximum amount of each resource that can be contained by the cell;
- $\mathbf{P} = \{p_1, \ldots, p_k\}$ is the amount of each resource produced by the cell at each update step;
- T is a flag indicating whether a tree is present in the cell or not;
- $\mathbf{Z}_T = \{z_T^r, z_T^t, z_T^l, z_T^f\}$ is a vector defining the size of the different parts of the tree (in our model, roots, trunk, leaves, and fruits);
- $\mathbf{N}_T = \{n_T[1], \ldots, n_T[k]\}$ are the amounts of each resource the tree takes from the cell at each update step (in turn depending on its size $\mathbf{Z}_T$);
- $\mathbf{U}_T^G = \{u_T^G[1], \ldots, u_T^G[k]\}$ is the vector defining the amount of each resource needed at each update step by the tree to *grow*;
- $\mathbf{U}_T^S = \{u_T^S[1], \ldots, u_T^S[k]\}$ is a vector defining the minimum amount of each resource the tree needs at each update step to *survive*; for each i, $1 \leq i \leq k$, we have $u_T^S[i] < u_T^G[i] < n_T[i]$;
- $\mathbf{R}_T = \{r_T[1], \ldots, r_T[k]\}$ is the amount of each resource stored by the tree at previous update steps;
- $\mathbf{M}_T$ is a vector of threshold values for different parameters defining the tree, such as maximum size, maximum age, minimum age for reproduction, maximum number of seeds produced for each mass unity of fruits, and so on. These threshold values can be fixed or picked at random in a given range when a new tree is created;
- $\mathbf{G}_T = \{g_T^r, g_T^t, g_T^l, g_T^f\}$ is a vector defining the *growth rate* of each of the parts of the tree, that is, how much each part of the tree grows when enough resources are available;
- $\mathbf{S} = \{s_1, \ldots, s_l\}$ is a vector defining the number of seeds present in the cell for each of the l species growing in the territory.

3.2 The Update Rule

At each update step of the automaton, the tree present in each cell (if any) takes the resources it needs from the cell itself and uses them to survive, grow (if enough resources are available), and produce seeds. If the resources available in the cell exceed its needs, the tree stores some resources. Conversely, if the resources available in the cell are not sufficient, the tree uses resources stored at previous update steps. If also the resources stored are not sufficient for the tree to survive, the tree dies. A newborn plant can sprout in a vacant cell, if the latter contains a seed of its species, and again enough resources.

Moreover, we defined the update rule in order to reproduce the increasing influence that a growing tree can have on neighboring cells. For example, its roots can extend beyond the limits of the cell hosting it. Or, when it gets taller, it shades an increasingly wider area around itself, thus having a negative influence on the growth of other trees in its neighborhood. We modeled the impact of a tree in a given position on its neighborhood by making resources flow from richer cells to poorer ones. In other words, a cell hosting a large tree is poor on resources, since the tree at each update step takes most (or all) of them. If the neighboring cells are vacant, their resources remain unused, and thus are richer than the one hosting the tree. Therefore, if we let resources flow from richer cells to poorer neighbors, the effect is that in practice a large tree starts to collect resources also from neighboring cells. Notice that if we include sunlight among the resources contained by a cell, we can model in this way also the "shade" effect. Also, in this way it is possible to render more fertile areas located in the proximity of rivers or lakes, since water contained into them spills in neighboring terrain cells. Seeds are also introduced in the model as a resource that moves from cell to cell. Thus, a trees can scatter their seeds in the surrounding area.

Each update step of the automaton covers a given period of time. As shown here, the rules are suitable for one year updates (that is, each cell update reproduces the evolution of the population in one year), but they can nevertheless be further fine–grained in order to model for example single seasons, by changing the amount of resources produced by cells at each step (see the Resource Production rule below).

Now, let $C(i,j)$ be the cell located at position (i,j) in the lattice. With $\mathbf{R}(i,j)$ we will denote the resource vector of cell $C(i,j)$, with $\mathbf{M}(i,j)$ the maximum resource values, and so on. The transition function can be divided in four sub–steps, defined as follows.

Tree sustenance. If a tree is present in cell $C(i,j)$, it takes from it a given quantity (defined by $\mathbf{N}_T(i,j)$) of each available resource $\mathbf{R}(i,j)$. If, for some resource i, the amount available $r_i(i,j)$ is lower than the corresponding value in $\mathbf{N}_T(i,j)$, then the tree takes the whole quantity $r_i(i,j)$. The amount of resources taken depends on the size of the tree $\mathbf{Z}_T(i,j)$. Then, if enough resources (those taken at this step, plus the resources stored at previous steps), are available, as defined by vector $\mathbf{U}_T^G(i,j)$, the tree grows, that is, each part grows according to the growth rate vector $\mathbf{G}_T(i,j)$ associated with the tree. Else, the resources might be just sufficient for the tree to survive (vector $\mathbf{U}_T^S(i,j)$). In this case, the tree parameters are left unchanged. In both cases, the tree "burns" an amount of each resource, as defined by vector $\mathbf{U}_T^G(i,j)$ or $\mathbf{U}_T^S(i,j)$. All the unused resources collected at this step are stored and added to vector $\mathbf{R}_T(i,j)$. Otherwise, if the overall amount (stored plus collected) of at least one resource is under the "survival threshold" of the tree, the latter dies. The tree also dies when it reaches its maximum age defined in vector $\mathbf{M}_T(i,j)$. All the resources that are not absorbed by the tree can remain in a cell, or disappear.

Tree reproduction. We have two cases to consider: a tree is present in the cell, or the cell is vacant. In the former case, the tree may produce some seeds (if it is old enough, and according to the size of its fruits $z_T^f(i,j)$), that are used to update the corresponding variable in the seed vector $\mathbf{S}(i,j)$. Also, new trees cannot sprout from seeds contained in a cell if a tree is already present. Instead, a cell can be vacant and contain some seeds. If the resources present in the cell are sufficient (quantities defined as global parameters for each tree species) a new tree is born. If seeds from different species are present in the cell, the winning species is chosen at random, with probability proportional to the number of its seeds.

Resource production. In the third sub–step, each cell produces a given amount of resources, according to its production vector $\mathbf{P}(i,j)$. In any case, the amount of each resource contained in the cell cannot exceed the corresponding maximum value defined by vector $\mathbf{M}(i,j)$.

Resource flow. In this step, resources are balanced among neighboring cells, in order to let resources flow from richer to poorer cells. Let $r_h(i,j)$ be the amount of resource h contained by cell $C(i,j)$, and assume that we are using the von Neumann neighborhood. $r'_h(i,j)$, the amount of resource i after this update sub–step, is defined as:

$$r'_h(i,j) = \frac{r_h(i,j)}{2} + \frac{r_h(i+1,j) + r_h(i-1,j)}{8} + \frac{r_h(i,j+1) + r_h(i,j-1)}{8}$$

In other words, we can see each cell as divided in four parts, each one containing the amount $r_h(i,j)/4$ of resource h, and corresponding to one of the neighbors. The amount of resource h contained in each part is balanced with the corresponding part of the neighbors. In case we adopt the Moore neighborhood, we can imagine the cells as split into eight portions. The effect is that, if cell $C(i,j)$ is richer on resource h than its neighbors, part of its content will spill into them. As mentioned before, $r'_h(i,j)$ cannot exceed the corresponding maximum value defined for the cell ($m_h(i,j)$). In this case, we set $r'_h(i,j) = m_h(i,j)$. The same rule is applied to each of the components of the seeds vector $\mathbf{S}(i,j)$.

3.3 The Initial Configuration

The initial configuration of the CA can be defined by the user, by setting appropriate resource parameters for each cell. Also, some trees might be already present on the territory, with all the variables defining them set. Or, the territory might be empty, with some seeds scattered here and there (clearly, if no tree and no seeds are present, nothing happens when the automaton is started).

4 THE USER INTERFACE

The model has been implemented in C++ under Windows NT. The user interface permits to define explicitly:

- Different types of cell, according to the maximum amount of resources the cell can contain and the amount of resources it produces, in order to resemble the features of different types of terrain. Moreover, it also possible to reproduce rivers (by setting high values for water content and production, and zero maximum content values for other resources), rocky terrain (with very low values for all the resources), roads (zero values for all the resources), rivers and lakes (containing only water) and so on;

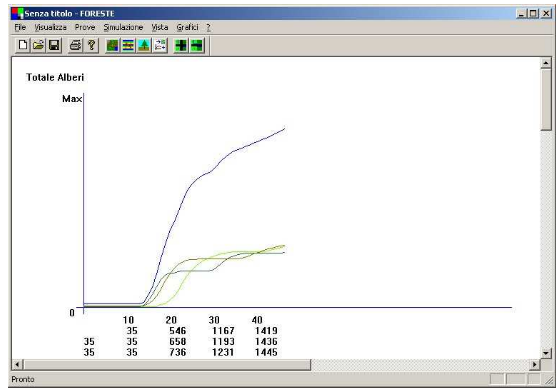


Figure 2: The user interface showing the total number of trees, and the number of trees of each species present in the area for the example shown in Fig. 4.

- Different tree species according to the amount of resources needed at each update step, to the growth rate of the different parts, that is, how resources are distributed among the different parts, the quantity of seeds produced. There is no upper bound on the number of species that can be defined;
- Additional resources, other than those shown in this article;
- The initial configuration of the automaton.

The interface shows step–by–step the evolution of the system, giving a straightforward image of the growth of the trees. Moreover, it is possible to show the distribution of the resources on the territory at each step, and the overall results of the simulation (total number of trees, trees for each species, total biomass, biomass of each single species and single tree, and so on), as shown in Fig. 2,3, and 4. In our experiments we could easily implement on standard desktop PCs CA consisting of thousands of cells, corresponding to several hectares of land, and a single update of the automaton (including the graphic layout) took a few seconds. Thus, the model seems to be suitable for the simulation of case studies of feasible size.

5 CONCLUSIONS

In this paper we presented a model based on CA for the simulation of the dynamics of plant populations. Our simulations, reproducing populations of black locusts, oaks and pine trees living on the foothills of the italian alps have shown results qualitatively similar to real case studies.

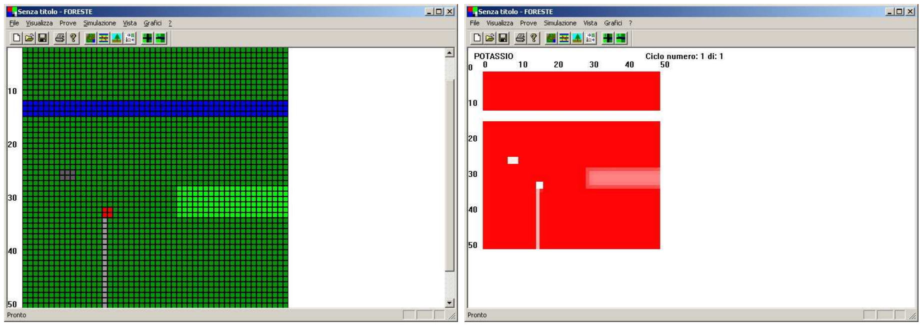


Figure 3: An initial configuration of the automaton (left). The dark strip represents a river. The image to the right shows the initial distribution of potassium in the cells. Darker areas are richer on potassium.

Figure 4: Example of the user interface, showing three different stages of the evolution of a plant population composed by black locusts, oaks, and pine trees, starting from the initial configuration of Fig 3.

We believe that the flexibility of the model, that allows the user to define explicitly different types of terrain and tree species, can provide an useful tool for the simulation of real case studies and a better understanding of the main factors influencing the dynamics of plant populations.

REFERENCES

C.Damgaard, J.Weiner, and H.Nagashima. Modelling individual growth and competition in plant populations. *Journal of Ecology*, 90:666–671, 2002.

D.Tilman. *Dynamics and Structures of Plant Communities*. Princeton University Press, 1988.

D.Tilman. Competition and biodiversity in spatially structured habitats. *Ecology*, 75:2–16, 1994.

E.Goles and S.Martinez. *Neural and Automata Networks*. Kluiwer Academic Publishers, 1990.

H.Baltzer, W.P.Braun, and W.Kohler. Cellular automata models for vegetation dynamics. *Ecological Modelling*, 107:113–125, 1998.

J.Ehrlen. The dynamics of plant populations: does the history of individuals matter? *Ecology*, 81: 1675–1684, 2000.

J.L.Uso-Domenech, Y.Villacampa-Esteve, G.Stubing-Martinez, T.Karjalainen, and M.P.Ramo. Mariola: A model for calculating the response of mediterranean bush ecosystem to climatic variations. *Ecological Modelling*, 80: 113–129, 1995.

J.Silverton and D.Charlesworth. *Introduction to Plant Population Biology*. Blackwell, 2001.

M.G.Barbour, J.H.Burk, W.D.Pitts, F.S.Gilliam, and M.W.Schwartz. *Terrestrial Plant Ecology*. Benjamin Cummings, 1998.

Q.Zeng and X.Zeng. An analytical dynamic model of grass field ecosystem with two variables. *Ecological Modelling*, 85:187–196, 1996a.

Q.Zeng and X.Zeng. Two variables dynamic model of grass field ecosystem with seasonal variation. *Ecological Modelling*, 85:197–202, 1996b.

R.L.Colasanti and J.P.Grime. Resource dynamics and vegetation processes: a deterministic model using two–dimensional cellular automata. *Functional Ecology*, 7:169–176, 1993.

Modelling Self-Organization Processes in Socio-Economic and Ecological Systems for Supporting the Adaptive Management of Forests

Alexander Kaufmann and Ernst Gebetsroither

ARC systems research GmbH, A-2444 Seibersdorf, Austria
alexander.kaufmann@arcs.ac.at, ernst.gebetsroither@arcs.ac.at

Abstract: Managing the numerous and interrelated processes between man and nature in order to use renewable resources in a sustainable way is confronted with conflicting objectives, external effects, complex interdependencies, uncertainty and other features that make it nearly impossible to come to unambiguous optimal decisions. Self-organization in socio-economic and ecological systems - the process of structuring a system by the elements of the system themselves without hierarchical or external control - is often the reason for ambiguity and uncertainty. Adaptive management is an approach to deal with these challenges. This natural resource management method is permanently monitoring both socio-economic and ecological systems in order to be able to react rapidly on any development pushing the systems into an undesired direction. Understanding and simulating the underlying self-organization processes helps to make the adaptive management of renewable resources both more effective and more efficient. In this paper we present a simple conceptual model of self-organization in socio-economic and ecological systems to understand better their effects on the interrelations between them, applied to the special case of forestry.

Keywords: self-organization; agent-based modelling; forest succession model

1. INTRODUCTION

The paper is related to the on-going discussion on sustainable use of renewable resources. This is a complex problem that encompasses socio-economic as well as ecosystem processes. It is characterized by complexity, a high degree of uncertainty, and information deficits and asymmetries [Holling, 1978]. Self-organizing processes in the biosphere and anthroposphere restrict, often even eliminate the possibility for prediction and control and are a major reason for the difficulties that arise in managing natural resources. For the effective management of renewable resources it is necessary to design a more flexible, feedback-driven, and continually adaptive management process [Grumbine, 1994; Holling, 1978] and to broaden the scope by including socio-economic processes in addition to those in ecosystems [McLain and Lee, 1996; Holling, 1978; Peterson et al., 1997; Haney and Power, 1996]. We want to contribute to this so-called 'adaptive management' model by gaining an in-depth understanding of self-organization processes in socio-economic and ecological systems. Self-organization processes extracted from agent-based models provide additional information to field surveys and expert knowledge to identify the key parameters and processes underlying the dynamic behaviour of the system. To find this key parameters and processes is very important, because explicitly monitoring them can help to recognize ongoing changes earlier and adapt the management in time accordingly.

We present a preliminary, still very simple version of a self-organization model concerning the use of forest resources. The total model consists of two separate models, one for socio-economic, the other for ecological self-organization, which are interlinked. The paper presents a work in progress, therefore the final chapter will give an outlook of how the model could be further developed.

2. SOCIO-ECONOMIC SUBSYSTEM

In this chapter we describe the agents included in the socio-economic model and give an overview of the interrelations between these agents.

2.1 Agents: forestry and timber industries

The agents in the socio-economic model belong to three major categories: forestry, communities and timber demanding industries. Other potentially important agents like tourism or hunting are not included in the basic model, while agents like policy, demand for timber products and competing sources of timber supply are considered as exogenous forces.

In general, any socio-economic agent, be it an individual person or an organization, is defined by a strategy (S) and a set of production factors (F). In our model there are only agents with a single strategy. A strategy consists of an objective (Z) and a set of routines (R), comprising:

- Expectation models (E) concerning the effects of actions on and the behaviour of the agent's environment with specific timeframes.
- Action rules (A) employing available production factors.
- Evaluation standards (V) comparing the achieved results with the stated targets.
- Update rules (U) regarding when and which objectives have to be adjusted and new routines have to be applied in case of unsatisfactory results of V.

Agents evolve by changing their strategies. In our model an agent can adjust his objective (e.g. by lowering the targeted profit margin) but not his basic rationality (e.g. firms always aim at profits). Furthermore, an agent can adjust his set of routines, replacing one or several of them. Routines may be changed either due to the insufficient performance (the objective is not achieved) or in the course of forming an new organization, a case that is not considered in our present model.

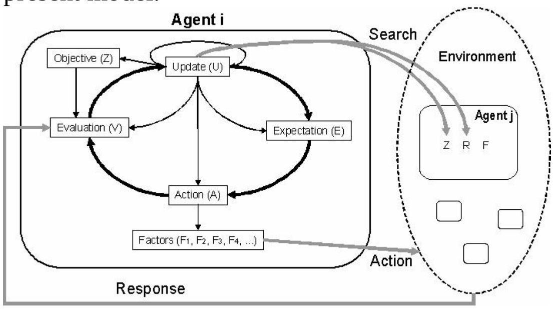


Figure 1. Evolution of a socio-economic agent

Evolution is based on an expectation - action - evaluation - update cycle which goes on continuously in any agent. If the evaluation of the results of past actions is unsatisfactory in relation to the stated targets, then the update rules are invoked which can change all of the routines (also the update routine itself) as well as the objective. Action rules are invoked by certain expectations and are executed by employing disposable production factors. In the case of a negative evaluation outcome, the agent looks for new routines. The search of new routines can either be global or restricted. We apply both types of search. Global search means that an agent screens all routines currently in use in his environment. In the case of restricted search the agent deliberately screens or can only access a selection of routines.

The agents in our model may differ with regard to the objectives as well the routines applied. It is impossible to describe the whole pool of objectives and routines in detail in this paper. The following chapters give only a short overview of the main characteristics of objectives and routines that are relevant for the agents.

2.1.1 Forestry

This is a heterogeneous category of agents, comprising small forest owners with little commercial interest as well as big forestry companies, run by a professional management.

The objective of any forestry agent is to make a profit from selling timber (reduced by a certain amount for own use, e.g. as fuelwood). The subsistence goal may be significant in the case of small owners while it is usually negligible in the case of forestry companies. Another source of heterogeneity are the actual profit target values. Some agents want to increase profits while others are satisfied keeping their profits constant.

Any forestry agent has at least three models forming expectations. One refers to the demand per quality of timber, another to the price per quality, and a third concerns the annual yield per quality (i.e. a combined stock-growth model of the forest in yield). Quality is (simplified) defined by species and shape of a tree.

In our present model we do not consider investment, i.e. enlarging the own forest area. This would be linked with organizational change which we have ruled out generally in the first model. So agents produce timber on a given area by applying different forest management methods. In the first version of the model we consider only the following methods: plantation, clear cutting (with reforestation or natural rejuvenation) and selective harvesting (with reforestation only if a relatively large number of trees are removed). A further simplification is that each method is only characterized by a specific combination of harvesting/removal and reforestation (and the related costs), neglecting maintenance efforts and technical issues of removal (e.g. machines used).

Evaluation is the comparison of realized and targeted profits. Update rules are invoked if the achievement of goals is not satisfactory. The update rules specify when objectives or routines have to be changed (thresholds), which are going to be changed and how they will be adjusted (e.g. by certain search rules).

2.1.2 Timber demanding industries

We consider the three most important timber demanding industries which are furniture manufacturing, construction and paper production.

All agents state certain profit targets which differ regarding margin and development (similar to the forestry agents). Each firm is active in one market only - either furniture, construction or paper. There are no diversified firms.

The agents use models predicting demand for wood products of different quality, prices of the respective goods and prices of the required timber. These models are based on some kind of past experience (e.g. average demand and price over a certain period), more elaborate forecasting methods are not included in the present model. Demand for wood products is exogenously given (e.g. constant, increasing market demand regimes).

Each firm has a plant-specific production function with a certain maximum capacity. Due to the same reasons as in the case of forestry we do not yet consider the possibility of investing in new capacity. Two industries have segmented markets - furniture and construction. We assume that both can produce high- and low-quality goods. In the case of furniture, for example, high quality products are made of beech or oak, lower quality uses pine, spruce or birch. For both, however, straight and long timber is required. In construction high quality (e.g. straight and long beech, oak, larch, fir) is required for buildings, floors, and similar products whereas low quality (branches, residue) is sufficient for boardings and particle boards. Paper production needs only cheap and fast growing timber like poplar, birch or spruce. In the present model we do not consider technological change. There is neither an increase in productivity nor the possibility to substitute lower for higher quality timber. Because we do not consider any differences regarding technology between the firms, costs of production other than the cost of buying timber can be neglected. What matters, on the contrary, are costs of transportation between the seller of the timber and the own manufacturing plant.

For evaluation and update rules the same as in the case of forestry applies in the general sense.

2.1.3 Communities

This class of agents comprises all administrative units at the local level. We do not consider the political process behind the behaviour of mayors and councils but regard the agent 'community' as if representing the collective interests of its population.

In our model the communities are only interested in the protection function of forests, i.e. preventing erosion or reducing the risk of avalanches in alpine regions. Other functions of forests like recreation and hunting are not considered in the first version of our model. Furthermore, we assume that no community practices commercial forestry. In order to fulfil the protection function, a forest has to be kept in a certain state (e.g. density and coverage, age and species structure). Communities expect that a certain state of a forest reduces the risk of natural disasters (of different frequency and magnitude) like landslides or avalanches.
A community can execute two actions: It can prescribe and enforce detailed management standards (e.g. ban of clear cutting), and it can directly influence the structure of those forests that are in communal property by respective maintenance and reforestation. Most forests in our model are privately owned, however, so communities have to rely primarily on standards as means of influence. If frequency and magnitude of natural disasters exceed the accepted limits, the respective community adjusts standards or maintenance and reforestation practices.

2.2 Interrelations between agents

The central interrelations between socio-economic agents in our first model are between timber suppliers and timber consumers, forming a timber market. In addition, communities influence forestry by setting standards according to the need for protection (see Figure 2).

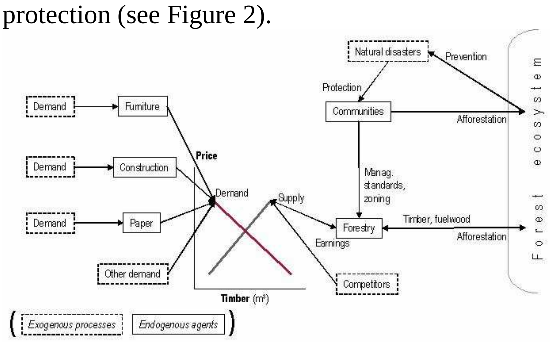


Figure 2. The socio-economic model determining the use of forest resources

2.2.1 Timber market

In our model the market is characterized by imperfect information and strategic behaviour. Prices are influenced by all the agents supplying as well as demanding timber. Each agent's behaviour depends on and affects all the other agents. Such market conditions can lead to a wide range of patterns, equilibrium is a possible but not a necessary outcome. Price and supply of timber are determined in a self-organized way:

Each forestry agent expects price (p) and demand (d) for each timber quality (q) of his available stock (v) of trees. The costs of extraction (c_x) depend on the management method. He ranks the quality classes according to the respective profits (p_q v_q - c_x v_q). Based on this ranking he removes trees from his stock, beginning with the most profitable class, up to expected demand or available stock (whatever is the actual limit). Summing up all these offers from the agents as well as from exogenous competitors leads to the total supply of timber differentiated by quality. Supply is then matched with demand (see next paragraph) and contracts are concluded when the offered price equals the accepted price for timber of a certain quality. Offers which cannot find any firm willing to buy at the proposed price are stored (reducing the volume to be removed in the next round). Based on the results, agents adjust their routines and the next cycle begins.

Each firm of the timber demanding industries expects demand (d) for and price (p) of its products, the price of timber of the required quality which, because it is the only input, equals the cost of production (c_q). Furniture manufacturers and construction firms can have two products, low- and high quality. They produce the more profitable good up to expected demand or maximum production capacity (whatever is the actual limit). If capacity exceeds demand for the more profitable product then the remaining capacity is used for the second good. Sales minus costs gives the profit. Costs comprise only timber and its transportation to the plant (p_q d_q - d_q (c_q + c_t)). Summing up all the timber demand from the agents as well as from exogenous demand leads to the total demand for timber differentiated by quality. Demand is then matched with supply (see previous paragraph) and contracts are concluded when the accepted price equals the price offered by the forestry agents. After having completed all the contracts, the firms produce the goods and try to sell them. This second-tier demand is exogenously given and determines the agents' success which enters the decision process in the next round. Firms can reduce their costs either by asserting lower prices of timber or by choosing more closely located forestry agents, reducing transportation cost c_t.

2.2.2 Protection function

The protection function is affected by a simple cycle: Extraction, reforestation (by forestry as well as communities) and natural growth determine the structure of the forest. This, in turn, determines its capacity to prevent natural disasters. This feeds back into the decision of communities whether to change standards and reforestation or not.

3. FOREST SUBSYSTEM

Natural ecosystems are generally undergoing changes and responding to changes, they are open, in flux and are affected by a series of often stochastic factors. For a sustainable management process scientific analysis and monitoring are therefore necessary [Schaffer, 1997]. An agent-based model can help to analyse the self-organization process during the forest ecosystem succession. Together with the monitoring of the key parameters within the forest ecosystem an adapted management can be established.

3.1 Agents in the forest model

Different trees with different main characteristics are the agents of the forest model. These agents are spatially fixed which means that they are not able to move during development, but their influence due to seed dispersal and competition with other trees goes beyond their local position. Within a forest succession dynamic there are differing types of species with differing characteristics. Pioneer species which have advantages at the beginning of settlement or resettlement of an area (e.g. a gap within a forest ecosystem after a windthrow) cannot compete with species later on in the succession dynamic. Existing old forest stands therefore often do not represent the biodiversity explored during the dynamic. For example, if we look at Austria's forest inventory, many species seem not necessary to be modelled because of their marginal occurrences. However, if we want to simulate the self-organization processes, they are of course, important too. At this stage of the model we include about 10 different tree species, but the future might show that we have to include more.

3.2 Self-organization in a forest ecosystem

A definition for self-organization in biological systems is given in Camazine, Deneubourg et al. [2001]: "Self-organization is a process in which pattern at the global level of a system emerges solely from numerous interactions among the lower-level components of the system." Within a forest ecosystem trees can be seen as lower-level components which act mainly local while their interaction shapes global patterns. One main problem of analysing changes in forest ecosystems is that visible changes of patterns often take a very long time. A computer-based simulation model therefore can provide a tool to observe these trends earlier and help to find more sustainable management decisions. One major goal is hence to capture the main events and mechanisms that determine the temporal and spatial dynamics within a forest succession under given environmental circumstances.

We use a simple but, as several studies have shown [Shugart and West, 1981; Green, 1989], efficient approach which is based on the forest model JABOWA III from Botkin et al. [Botkin, 1993]. In the model the trees are characterized by diameter (breast height diameter, BHD), height and species. Every species is defined by a set of parameters and they compete against the other to get more light, water and nitrogen. For every agent, every tree, the general growth equation is calculated (Equation 1).

$$\Delta GGF = f(D_{max}, H_{max}, b_2, b_3, D, G) * f(\text{environment}) \qquad (1)$$

This calculates the maximum growth potential (maximum changes) for each tree species, as a function of few parameters like the max. diameter or max. height this tree species can reach, reduced by local environmental responses, f(environment), which is a factor between 0 and 1. The environmental responses are correlated with the available light and the site conditions (Equation 2)

$$f(\text{environment}) = f(\text{light}) * Qi \qquad (2)$$

The site conditions include the general temperature response function TFi, the "wilt" factor WiFi, an index of the drought conditions a tree can withstand, the "soil wetness" factor WeFi, an index of the amount of water saturation of the soil a tree can withstand and the index of tree response to nitrogen content of the soil NFi (Equation 3).

$$Qi = TFi * WiFi * WeFi * NFi \qquad (3)$$

In a natural forest ecosystem dead trees belong to the common view and are a main entity, whereas they are often missing in a managed forest. This is important because some seeds chiefly germinate on old dead trees. Within the model mortality will be simulated accordingly to JABOWA III in two different ways.
First, there is an inherent risk of death for any tree, independent of the competition with other trees, e.g. the death due to a windthrow, or, also very important as interrelation to the socio-economic subsystem, via harvesting, which will be discussed later on.
Second, there is competition-induced death. Trees that grow badly over a certain period of time (e.g. ten years) have a higher probability to die than well growing trees.

These are the responses of an existing tree to the environment, but the natural reproduction of trees, i.e. rejuvenation, is also very important. A tree has to reach a kind-specific age for seed production and produces characteristic seeds. To simplify the model, we distinguish only between shade-tolerant, intermediate and shade-intolerant seeds. Different to the forest model developed by Botkin et al. [1993] (JABOWA III), we don't assume that there are always enough trees to produce seeds independent of the management practice and how many old trees are in the area able to produce seeds. We think that this fits more to the reality and as other investigations have shown, special events as increased seed production in the same time with increased open areas might influence strongly the dynamic [Wiegand, Milton and Wissel 1995]. In the first approach we assume concentric dispersal of the seeds from the producing tree, but the model could be improved by correlating the seed dispersion with the main wind direction.

Self-organization within the forest succession model emerges due to indirect local competition of the trees for light (f(environment)) and space (seed dispersal). An indirect competition for nutrients and water is not implemented in the present version as one tree does not influences the general availability of nitrogen and water of a neighbouring tree.

3.3 Parameterization and validation

Parameterization of models is one of the crucial steps in the development of a model. One reason to use a relatively simple model approach is to find valuable data for parameterization. Main inputs for the forest self-organization model are the tree characteristics, the site conditions and information about stochastic events like windthrow, seed dispersal or tree mortality. At the present stage of the model there are no feedbacks from forest to the occurrences of calamities like from bark-beetle. The validation of the dynamic behaviour of the forest subsystem has to be done intensively to prove the reliability of the simple model approach.

The validation process itself can be described in a three step procedure:
First, the potential growth curve of each tree species has to be simulated independently from competition with other trees.
Second, the different environmental response function of each tree species has to be checked independently.
Third, the competition between different tree species has to be analysed.

The results of these three steps have to be discussed with experts and compared with the development of existing natural forest ecosystems in Austria before the connection with the socio-economic model.

The environmental influences described above can be summarized as *external natural influences,* which are often stochastic events, difficult or even impossible to predict. They can occur in unmanaged natural as well as in managed forests. The interrelation to the socio-economic subsystem is simulated due to the so called *external human influences*. The positive feedback between *Ability to compete with other trees*, *Resources* and the *General growth function* characterizes the self-organization process (Figure 3).

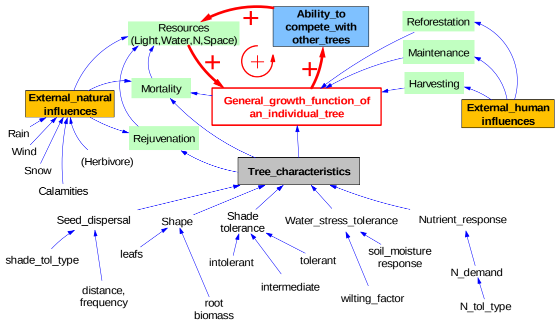


Figure 3. General forest model scheme

4. THE TOTAL MODEL

The total model helps to identify the key self-organization processes in the socio-economic and the ecological systems and the feedback relations between them. This knowledge enables adaptive management to better target its activities on the most crucial processes and at the right time. Three types of interaction link both models − removal (harvesting), reforestation and maintenance. In the present version of the model only the first two are considered (except for a rudimentary role regarding the protection function for communities, see chapter 2.1.3).

Removal, harvesting: Timber for different purposes like furniture, construction and paper production require different tree species with different qualities like age, shape or BHD. Demand, determined in the socio-economic subsystem, provides the information which trees at what location and by which harvesting method will be removed. In the present model there are only very simplified versions of management methods (plantation, selective harvesting and clear cutting, see chapter 2.1.1). Further impacts of harvesting (e.g. on the soil) are not yet considered.

Reforestation: From the socio-economic subsystem information is provided which tree species when and where should be planted. Reforestation has no direct influence on the general growth function (ΔGGF) of existing trees, but affects them indirectly, because competition for resources by newly planted trees also affect existing trees.

5. OUTLOOK

The model presented in this paper is simplified and we don't know yet whether our self-organization model will really enable us to analyse the key processes for the adaptive management of renewable resources. Further necessary improvements could be:

- Adding new types of agents (e.g. hunting, tourism, deer, bark-beetle, wild boar).
- Including other forest functions like recreation.
- Incorporating the evolution of agents (e.g. setting up new organizations) in the model.
- Making the agents' strategies more realistic.
- Considering the role of technology and technological change.
- Integrating GIS data (e.g. topography, elevation) to specify the environmental response functions (e.g. temperature, light, soil moisture).
- Choice of an appropriate modelling software e.g. NetLogo, RePast, Swarm.

6. REFERENCES

Botkin, D. B., Forest Dynamics, An Ecological Model, Oxford University Press Inc, ISBN 0-19-506555-7, pp. 309, New York, 1993.

Camazine, S., J-L. Deneubourg, N.R. Franks, J. Sneyd, G. Theraulaz, E. Bonabeau, Self-Organization in Biological Systems, Princeton Studies in Complexity 2001, Princeton University Press, ISBN 0-691-01211-3, pp .538, Princeton.

Green, D.G., 1989, Simulated effects of fire, dispersal and spatial pattern on competition within forest mosaics. Vegetatio 82(2): 139-153.

Grumbine, R.E., What is Ecosystem Management?, *Conservation Biology*, *8*(1), 27-38, 1994.

Haney, A., and R.L. Power, R.L., Adaptive Management for Sound Ecosystem Management, *Environmental Management*, *20*(6), 879-886, 1996.

Holling, C.S. (ed.), *Adaptive Environmental Assessment and Management*, John Wiley and Sons, London, 1978.

McLain, R.J., and R.G. Lee, Adaptive Management, Promises and Pitfalls, *Environmental Management*, *20*(4), 437-448, 1996.

Peterson G. et al., Uncertainty, Climate Change, and Adaptive Management, *Conservation Ecology*,*1*(2), 1997.

Shugart, H.H, and D.C. West, Long-term dynamics of forest ecosystems. American Scientist 69(6):647-652, 1981.

Wiegand, T., Milton, S.J. and C. Wissel, A Simulation Model for Shrub Ecosystem in Semiarid Karoo, South Africa, Ecology, Volume 76, Issue 7: 2205-2221, 1995.

Situated Cellular Agents for Crowd Simulation and Visualization

S. Bandini, S. Manzoni, G. Vizzari [a]

[a]Dipartimento di Informatica, Sistemistica e Comunicazione, Università di Milano–Bicocca,
Via Bicocca degli Arcimboldi 8, 20126, Milano, Italy
giuseppe.vizzari@disco.unimib.it

Abstract: The paper presents a Multi Agent Systems (MAS) approach to crowd modelling, based on the Situated Cellular Agents (SCA) model. This is a special class of Multilayered Multi Agent Situated System (MMASS), rooted on Cellular Automata, providing an explicit spatial representation and the definition of adjacency geometries. The model also defines a concept of autonomous agent, provided with an internal architecture and individual state and behaviour, capable of different means of space–mediated interaction (synchronous, between adjacent agents, and asynchronous among distant entities). Heterogenous entities may be modelled through the specification of different agent types, defining different behaviours and perceptive capabilities. After a brief description of the model, its application to simple crowd behaviours will be given (e.g. lane and group formation), and an application providing the integration of a bidimensional simulator based on this model and a 3D modelling application (3D Studio) will also be described. The adoption of this kind of system allows to specify, simulate and evaluate a design solution, but also to easily produce a realistic visualization of the simulation, in order to facilitate the communication with involved actors. In fact, while expert decision–makers often require only abstract and analytical results deriving from the simulation, other people involved in the decision–making process related to the design may be helped by other forms of graphical representation.

Keywords: multi–agent systems, simulation, crowd behaviour.

1 INTRODUCTION

The design of different kinds of environmental structures, at different detail levels, from the corridors or emergency exits of a building to the road system on urban or regional scale, requires some kind of simulation system, in order to evaluate strategies and designs before they are actually implemented. There are different approaches to simulation, based on different theoretical models, ranging from analytical ones [Helbing [1991]] to those based on Cellular Automata [Wolfram [1986]]. Rather than tackling simulation problems in a global manner, defining centralized solution methods that manage various aspects of the modelled system, they can be suitably reformulated in terms of local interacting agents, that try to achieve their own goals, by means of coordination or competition schemes. The solution, global system behaviour, is obtained as an emergent effect of agents' individual local behavior [Ferber and Drogoul [1992]]. In this framework, approaches based on Multi Agent System (MAS) principles [Ferber [1991]] that propose to focus on interaction aspects of agent groups and crowds have been defined. These works have shown that intelligent group behavior and solution to complex problems can be obtained as the result of interactions between agents characterized by a simple internal model [Drogoul [1995]]. A MAS could thus represent a mean of modelling those self–organizing systems that carry out the planning operation to obtain the solution of the design problem. MASs could even be useful in another kind of iterative design process, centered on simulation [Caneparo and Robiglio [2003]]. The latter could provide a first phase in which the designer makes some preliminary choices about the environment, the entities that inhabit it and their behaviour, then a cycle of simulations is performed and the design can thus be evaluated. If necessary it can be suitably modified, manually by the designer or in a semi–automatic way. Then a simulation, evaluation and adaptation cycle could be iterated until the design produces results that are considered acceptable.
The acceptance of the design is generally based on some kind of quantitative analysis of the simulation results performed by experts, but especially in

this area there are often stakeholders and people involved in the decision process that are not able to understand this kind of information. Where graphs and tables may not be effective, a realistic visualization of simulation dynamics can integrate them and allow non–experts to better understand the effects of the design on system dynamics.
The aim of this paper is to describe the Situated Cellular Automata (SCA) model, a particular class of Multilayered Multi Agent Situated Systems (MMASS)[Bandini et al. [2002a]], that has been designed for applications requiring an explicit representation of the spatial structure of the environment. This structure can be regular or irregular and agents' behaviour is strongly influenced by their position, as it is determined as a consequence of synchronous interaction with other adjacent entities (i.e. *reaction*) or according to the perception of signals asynchronously emitted by at–a–distance agents (i.e. *field diffusion*). Such a remote interaction represents a mean of modelling the concept of locality, while reaction can represent a direct cooperation between neighbours. The following Section briefly describes the SCA model, focusing on agent interaction, while Section 3 describes how to exploit it in order to represent crowd behaviours. Later two sample applications will be shown, respectively in indoor and outdoor situations. Conclusions and future developments will end the paper.

2 SCA MODEL

The SCA model defines MAS whose entities are situated in an environment whose structure (i.e. *space*) is defined as an undirected graph of sites. This Section is not meant to represent a thorough and formal description of the SCA model, but it will be focused on interaction mechanisms that it defines. In fact it will be shown that it is possible to focus on interactions and specify very simple reactive agents, obtaining realistic behaviours.
Agents may interact in two different ways: if they are adjacent in the environment structure they can perform an agreement process in order to synchronously change their state with a *reaction* operation; instead field emission–diffusion–perception mechanism allows at–a–distance, space–mediated, asynchronous communication mean. According to its own state and behaviour specification an agent may emit a field, defining its parameters (type and intensity value). Field values propagate throughout the space according to the diffusion function, specified by the field type. The $Perception_\tau$ function, characterizing each agent type τ, defines the perceptive capabilities of an agent. It defines whether

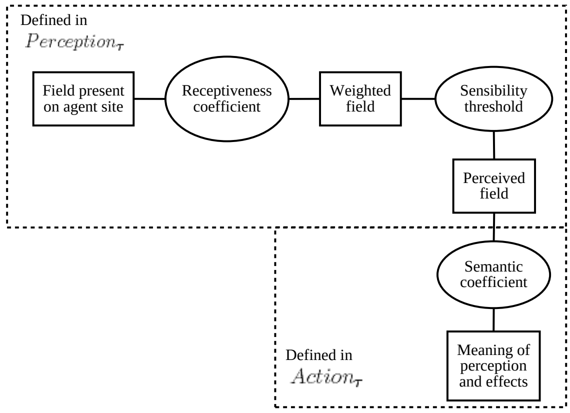


Figure 1: A schematic description of the perception mechanism.

or not a certain field is perceived by an agent or is neglected, as its value at the site is such that it is considered too faint. Field perception constitutes a fundamental aspect of the perception–deliberation–action mechanism that specifies SCA *agent behavior*. This mechanism describes agents as characterized by a set of possible actions, and a mechanism for the selection of the action to be undertaken based on the internal state and the position of the agents themselves. The set of possible actions (i.e. $Actions_\tau$) specifies whether and how agents of type τ change their state and/or position, how they interact with other agents, and how neighboring agents can influence them. In general, actions have two possible purposes: they can be undertaken by an agent in order to modify its state or position (*intra–agent actions*), or to interact with other agents in both synchronous or asynchronous ways (*inter–agent actions*). In the following more details of agents interaction and behaviour will be given.

2.1 Modelling signals through fields

Each agent is provided with a set of sensors that allow its interaction with the environment and other agents. At the same time, agents can constitute the source of given fields acting within the space (e.g. noise emitted by a talking agent). Each field is characterized by the set of values that the field can assume during its propagation throughout the space, a diffusion function, and field composition and field comparison functions that define field manipulation. The diffusion function of a certain field type computes the value of a field on a given space site taking into account in which site and with which value it has been generated (since the spatial structure is generally not regular and paths of different length can connect each pair of sites, it may return a number of values depending on the number of paths connecting the source site with each other site). The composition function expresses how fields of the same type have to be combined (for instance, in order to obtain the unique value of that field type at a site), and the comparison function that matches field values. For instance, in order to verify whether an agent can perceive a field type, its value at a site, modulated according to a receptiveness coefficient, and agent sensitivity threshold are compared by this function. A general agent interaction mechanism based on the field emission–diffusion–perception mechanism can be defined through the specification, for each interaction, of:

• a *field source* that can correspond to an agent (that may have a very limited behavior, modelling thus an object). For instance, fields can be emitted by agents to indicate their availability to fulfil given tasks (e.g. a guide agent in a museum, a policeman in a city, a public phone);

• a *function* to define field diffusion throughout the environment structure and that specifies how field values have to be modulated (e.g. when an agent moves far from a group of agents its view must be reduced and it is no more visible when he goes out from a room);

• a *field sensor* and *perception function* associated to each agent that allows the representation of different and dynamic agent perceptive abilities. Agent perception can be dependent on agent state, goals, and context (e.g. agent sensitivity to the presence of a fire exit must be higher in an emergency situation). Fields themselves are neutral even if they can have related information in addition to their intensity; they are only signals, with an indication on how they diffuse in the environment, how they can be compared and composed. Different agent types may be able to perceive them or not and, in the first case, they may have completely different reaction. Therefore, the effect of attracting or repelling objects or agents is obtained with the suitable definition of field types and behaviours related to certain agent types.

2.2 Perception and action

Agents are entities that are situated in an environment and that, according to an internal representation of their state, perceive their local environment (*perception*) and select within a set of possible actions (*deliberation*) in order to move, modify their state, and interact with other agents (*action*). Perception, deliberation and action characterize SCAs that can differ according to their type. Field perception for each agent and field type can be informally defined through the specification of the set of *fields* that agents are able to perceive (i.e. maximal agent capabilities, e.g. deaf agents can not perceive sound fields); a *sensitivity threshold* that indicates, according to agent state, the minimum field value that the agent is able to perceive (e.g. when they are involved in a conversation, agents are less sensitive to surrounding feeble noises); a *sensitivity coefficient* that modulates field values according to agent state (e.g. when an agent is in a hurry, it is more sensitive to exit and elevator signs). The perception mechanism is summarized in Figure 1: its first part is related to the physical possibility of the agent to perceive a signal in a certain situation, while the second one refers to the semantic value that the agent gives to the perception in the current circumstances. Agent behaviour can be specified using the L*MASS [Bandini et al. [2002b]] language that defines the following primitives:

• $reaction(s, a_{p_1}, a_{p_2}, \ldots, a_{p_n}, s')$: this primitive defined for agent a situated in the site p allows it to synchronously interact with agents $a_{p_1}, a_{p_2}, \ldots, a_{p_n}$, situated in $p_1, p_2, \ldots, p_n$ adjacent to p, that have agreed to take part in the interaction; the effect of this interaction is the change of its state from s to s';

• $emit(s, f, p)$: the $emit$ primitive allows an agent to start the diffusion of field f on p, that is the site it is placed on;

• $trigger(s, f_i, s')$: this primitive specifies that an agent must change its state from s to s' when it perceives a field f_i;

• $transport(p, f_i, q)$: the $transport$ primitive allows to define agent movement from site p to site q (that must be adjacent) upon reception of field f_i.

For all these primitive, some conditions (i.e. sort of guards on the execution of the related operation) on agent state and perceived fields can also be specified. In some cases these two parameters can be insufficient, as they are just related to a single site, therefore these conditions can include the intensity of fields present in adjacent sites. For instance, in order to specify a $transport$ operation, this is necessary to model the behaviour of an agent wishing to move to the adjacent site with the highest intensity of a certain field. In order to specify that a certain agent of type τ_a can attract agents of type τ_b one must respectively define a field type $F_{a \leftarrow b}$, specifying required parameters, insert a specific $emit$ action in $Action_{\tau_a}$ and a $transport$ operation in $Action_{\tau_b}$ indicating that the related agent of type τ_b should move towards the adjacent site with the highest value for field type $F_{a \leftarrow b}$. More precisely the $trasport$ action would be specified as follows:

$$
\begin{aligned}
&action : transport(p, f_{a \leftarrow b}, q)\\
&condit : position(p), empty(q), near(p, q),\\
&\qquad\quad perceive(f_{a \leftarrow b}), best(q)\\
&effect : position(q), empty(p)
\end{aligned}
$$

where p and q are sites, $position(p)$ specifies that the related agents is placed in p, $empty(q)$ indicates that site q is not occupied by other agents, $near(p, q)$ specifies that the arguments are adjacent sites (i.e. connected by an edge in the spatial structure) and $perceive(f_{a \leftarrow b})$ indicates that the agent is able to perceive a field $f_{a \leftarrow b} \in F_{a \leftarrow b}$. The additional condition $best(q)$ is verified when for all sites r adjacent to p and currently empty, the intensity of field type $F_{a \leftarrow b}$ is lower or equal than is site q. Repulsion requires the same operations, with a difference in the $transport$ action, whose destination is the site with the lowest value for the repelling field type. More complex conditions for the transport operation can cause interesting effects, such as an agent that keeps at a certain distance from the source of a specific field type, following thus its movement but avoiding contact. The definition of different field sources and types (or, equivalently, the inclusion of an indication of the related source in the information related to fields) allows to define different way–points, intermediate steps in a movement script. An agent may be perceptive to the first field type and move towards its source. When the perceived field intensity reaches a certain level, in other words when the distance has reduce under a certain degree, the agent may change its goal, becoming perceptive to the field emitted by the next way–point. In order to obtain more complex behaviours, related for instance to agents interests, goals, and more autonomous behaviours requires a more composite field definition, that should encapsulate more information than their simple intensity, and different agent actions. A formal description of the introduced modelling elements can be found in [Bandini et al. [2004 – to appeara]].

3 MODELLING WITH SCA

As previously specified in Section 2, the SCA model provides a discrete representation of the environment. The latter provides an abstraction of the actual spatial structure in which the simulation takes place. In order to exploit the SCA model, the first step is thus to describe the simulation scenario in terms of a discrete and possibly irregular network of nodes. Figure 2 shows the 3D representation of

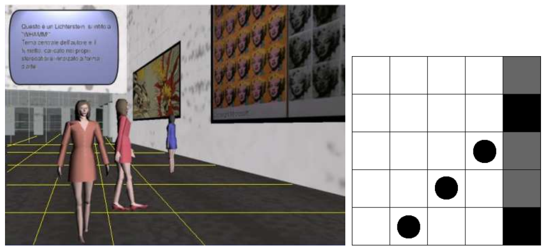


Figure 2: 3D representation of museum room and the related 2D representation with a grid structure.

a museum room and the related abstraction with a grid structure. Black squares are occupied by walls, grey ones represents artworks, while agents are represented with black circles. The decision on the granularity of the tessellation depends on the goal of the simulation and the features of the scenario: for instance, if the simulation goal is to evaluate the design of a corridor in an evacuation situation, the tessellation should reflect the actual dimensions of a human body and its space occupancy.

The second step is to describe the behaviour of various entities placed in the environment, in order to obtain a realistic system dynamic. With reference to the same figure, we could model artworks and doors as sources of fields that can have an attractive effect over agents roaming in the environment, according to their own internal state and goals. Agents may be perceptive to fields emitted by artworks placed at a distance lower than a specified threshold, and be attracted for a specific time interval, but only if they still have not already observed it. They may also be perceptive to doors and passageways, with a different priority: for instance they may decide to move from one room to another, following the field emitted by the door, after they have observed every artwork present in the room. The placement of field sources must be decided taking into account the diffusion function related to field types: in order to obtain a realistic behaviour of agents they should not be able to perceive the presence of an artwork if they do not see it. This consideration indicates that the diffusion of some signals can exploit the actual 3D model of the environment for the diffusion of specific field types, while for other ones (e.g. audible signals) the bidimensional abstraction can be enough to obtain believable system behaviour. The definition of fields and diffusion functions, with reference to the representation of the environment, is just one side of behaviour modelling. The other one is related to the specification of the actions undertaken by agents, for instance, as a reaction to the perception of a certain signal. Different entities may

Figure 3: A screenshot of a 3D application showing a procession formation.

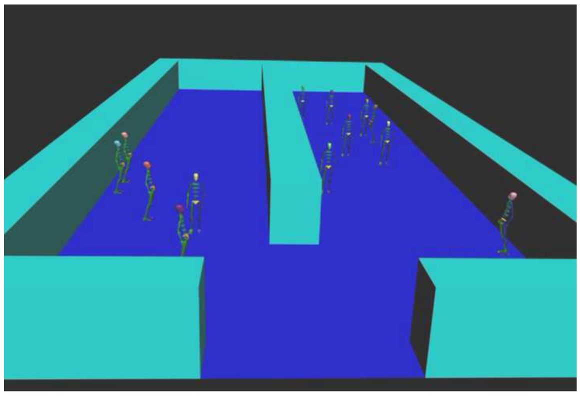


Figure 4: A screenshot of the animation produced by the 3D modelling tool in an indoor scenario.

react in a completely different way to the perception of the same field, and even the same agent can select different actions according to its own state. For instance, in order to specify that an agent should follow a specific path, the related way–points could be associated to field sources. The agent could be sensitive to the field emitted by the closest way–point and moving towards it, becoming sensitive to the next one once its distance becomes lower than a certain threshold. Moreover the way–points could be exploited by different agent types related to different paths. In other words the order of the points to visit could be defined in the behaviour specification related to a specific agent type, and sources could be just relevant points indicating their presence through the emission of a presence field. In the same way effects of attraction and repulsion can be defined in order to fine tune the behaviour of various entities roaming in the environment according to its infrastructure. Agents movement anyway can also be based on the behaviour of other active entities. In other words agents may be at the same time affected by fields but also sources of signals affecting other entities. For instance, a specific agent may be the source of a presence field that is considered attractive by a specific agent type. In this way crowds can concentrate around a leader and follow him/her in a procession (see Figure 3). In a similar way lanes and queues can be obtained specifying that every agent is only sensitive to the signals emitted by the preceding one, and having leaders that guide the crowd, following specified paths.

4 Sample experiments

One of the applications developed to implement SCA based simulations exploits a simulator based on a bidimensional spatial structure representation and an existing commercial 3D modelling instrument (3D Studio MAX). The simulator has been developed as an experimentation and exploitation of a long term project for a platform for MMASS based simulations [Bandini et al. [2004 – to appearb]]. This software is based on the Java platform and its goal is to implement basic elements and mechanisms of the MMASS model in order to allow a user to rapidly use these components to build a simulation.

This simulator produces results that can undergo a quantitative analysis whose results can be easily understood by experts of the area of application. In different situations it can be useful, for sake of communication with non–experts, to obtain a more effective visualization of simulation dynamics. To do so, the bidimensional simulator produces a log–file provided with a fixed–record structure, in which every record is related to a node of the spatial structure or the position of an agent with reference to this structure. Initially, the simulator prints the structure of the environment, then the starting position of each agent. For every iteration the new position of every agent is also printed. This file is parsed by a 3D Studio Max script which generates a plane and walls related to the spatial structure, nodes related to sites, and bipeds related to agents. Splines are then generated starting from the discrete positions assumed by various agents, and represent bipeds' movement. This process introduces modifications to trajectories defined by the bidimensional simulator whose sense is to give a more realistic movement to agents' avatars. This application was tested in an abstract indoor situation, related to an evacuation scenario (see Figure 4).

The interaction mechanism between the simulator and the 3D modelling tool is currently being mod-

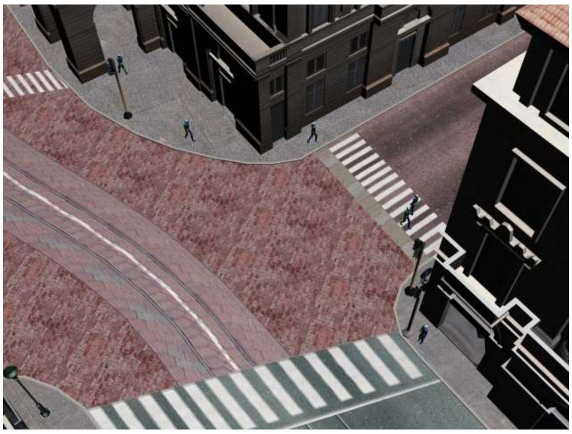


Figure 5: A screenshot of the animation produced by the 3D modelling tool in an outdoor scenario.

ified in order to allow an easy integration with existing models of the environment. To do so it will be possible to draw the bidimensional abstraction of the environment directly on images obtained by the model. The development of this tool for the specification of the environment is currently on–going and will allow an easy adoption even in a realistic outdoor scenario. A screenshot of an animation produced in this kind of scenario is shown in Figure 5[1].

5 CONCLUSIONS

This paper has presented the application of the SCA model to crowd modelling, simulation and visualization. The model provides the possibility to explicitly define a spatial structure of the environment in which the simulation takes place. Relevant objects are modelled as sources of fields, signals that diffuse in the environment and can be perceived by agents. The reaction to the perception of these fields is defined by agent type, which also specifies perceptive capabilities with reference to their state.
Currently the SCA model is being applied to simulation supporting localization, design of environments, but it is also being considered as an instrument for urban and environmental planning.
The applications that were described in this paper provide the integration of a simulator and a 3D modelling tool, but the design of an integrated simulator and 3D engine for the development of real–time applications exploiting elements of the model is currently under–way.

[1] The 3D model of Scala Square appears courtesy of Geosim systems.

REFERENCES

Bandini, S., S. Manzoni, and C. Simone. Dealing with space in multi–agent system: a model for situated mas. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2002)*, pages 1183–1190. ACM Press, 2002a.

Bandini, S., S. Manzoni, and C. Simone. Heterogeneous agents situated in heterogeneous spaces. *Applied Artificial Intelligence*, 16(9-10):831–852, 2002b.

Bandini, S., S. Manzoni, and G. Vizzari. Situated cellular agents: a model to simulate crowding dynamics. *IEICE Transactions on Information and Systems: Special Issues on Cellular Automata*, 2004 – to appeara.

Bandini, S., S. Manzoni, and G. Vizzari. Towards a specification and execution environment for simulations based on mmass: Managing at–a–distance interaction. In *Fourth International Symposium From Agent Theory to Agent Implementation*, 2004 – to appearb.

Caneparo, L. and M. Robiglio. Urbanlab: Agent–based simulation of urban and regional dynamics. In *Digital Design: Research and Practice: Proceedings of the 10th International Conference on Computer Aided Architectural Design Futures*. Kluwer Academic Publisher, 2003.

Drogoul, A. When ants play chess (or can strategies emerge from tactical behaviors?). In *Proceedings of 5th European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW93)*, volume 957 of *Lecture Notes in Computer Science*, pages 13–27. Springer–Verlag, 1995.

Ferber, J. *Multi–Agent Systems*. Addison–Wesley, 1991.

Ferber, J. and A. Drogoul. *Using Reactive Multi-Agent Systems in Simulation and Problem Solving*, chapter Distributed Artificial Intelligence: Theory and Praxis, pages 53–80. Kluwer Academic Publishers, 1992.

Helbing, D. A mathematical model for the behavior of pedestrians. *Behavioral Science*, (36):298–310, 1991.

Wolfram, S. *Theory and Applications of Cellular Automata*. World Press, 1986.

Uncertainties in LCA of Plant-Growth Regulators and Implications on Decision-Making

Georg Geisler, Stefanie Hellweg, and Konrad Hungerbühler,

Swiss Federal Institute of Technology, HCI, 8093 Zurich, Switzerland, stefanie.hellweg@chem.ethz.ch

Abstract: Uncertainty assessment in LCA enables the evaluation of the significance of results, which is important for providing sound decision-support. In this work, an LCA was performed on two plant-growth regulators considering various sources of uncertainty: In the LCI, uncertainties of imprecise measurements of elementary flows, temporal and spatial variation, and different production processes were assessed. In the characterisation phase, the uncertainties of substance properties and the composition of sum-parameters were considered. These uncertainties were expressed as probability distributions and assessed via stochastic modelling (Monte-Carlo Simulation). For most LCI- and LCIA-data, generic uncertainty ranges were used. Uncertainties due to assumptions on the production efficiency were reflected by a best-case and a worst-case scenario. Contributions to variance of all uncertain input parameters were calculated. One plant-growth regulator was defined as significantly better than the other, if the impact score was lower in 90% of the simulations. The results showed that differences in median impact scores of a factor of 1.6 were sufficient in the impact categories global warming, acidification, and eutrophication for a significant distinction of the products. The applied doses and the elementary flows of basic-chemical production and energy supply had the highest contribution to variance in these impact categories. By contrast, dispersions are large concerning the toxicity impact categories and the photooxidant creation potential. This can be mainly attributed to the high contribution to variance of sum-parameters and characterisation factors. The implications of these uncertainties on the decision-making process are discussed. Moreover, tentative rules of thumb for estimating the significance of results are put forward. Finally, a format is proposed how complex results of uncertainty assessments may be presented for decision-support.

***Keywords**:* decision making; Life Cycle Assessment (LCA); pesticides; significance; uncertainty

1 INTRODUCTION

Quantifying uncertainty in LCA is an important step towards reliable and transparent decision support. The theoretical foundation and tools for uncertainty analysis in LCA are published [Huijbregts, 1998a, 1998b, Huijbregts, 2001, Weidema, 1996]. However, there is still a lack of case studies analysing uncertainty in LCA results and implications for decision support.

A prerequisite for an uncertainty analysis is the availability of information quantifying the uncertainty. Unfortunately, specific factors for uncertainty in individual LCI or LCIA parameters are only rarely available today. To handle this lack of specific uncertainty data, the use of generic uncertainty factors has been proposed for groups of parameters (e.g. air emissions, characterisation factors). Concerning the LCI, such generic uncertainty factors were derived by Finnveden and Lindfors [1998] in a comparison of LCI datasets on PVC production from different sources. For characterisation factors of LCIA methods, generic uncertainty factors have been published by method developers (e.g. Huijbregts [2003] concerning the CML-baseline method [Guinée, 2001]).

In this work, we assess the uncertainty of an LCA comparing two plant-protection products using generic uncertainty factors. A simple format for the presentation of uncertain LCA results is proposed. It is discussed to what extent a full uncertainty analysis is necessary to obtain reliable results in routine application of LCA, taking into account implications of uncertainty for decision making. Rules of thumb for simplified significance criteria are suggested.

2 CASE STUDY

A case study on two plant-protection products is used to illustrate consequences of uncertainty in LCA for decision making of pesticide producers. Both products are assessed for their use as plant-growth regulators in winter wheat. The product Moddus contains trinexapac-ethyl as active substance and is relatively new on the market (since 1990). The product Stuntan is established since 1960. Stuntan is a fictive name representing a range of similar products from different suppliers that contain chlorocholine chloride as active substance. The functional unit is the dose applied to 1 ha of crop, as recommended by pesticide registration authorities [BfL, 2002]. Such a product comparison is of interest e.g. for pesticide producers to benchmark new products against established ones.

Published LCIs were used regarding the supply of basic chemicals and energy as well as transport, distribution and tractor operation [Geisler, 2003a]. A specific estimation procedure [Geisler, 2003b] was applied to inventory LCIs of fine chemical production, namely the supply of active substances, formulation ingredients, and their precursors. Uncertainty in this estimation procedure concerning the efficiency of chemical production is depicted by a best and a worst-case scenario (see below). The full LCIs concerning the production of the active substances are published in Geisler [2003a], LCIs for formulation ingredients are documented in Geisler et al. [2003b]. The LCIA was carried out using relevant impact categories of the CML-baseline method [Guinée, 2001].

3 METHODS

The LCA was calculated in Excel using a matrix-inversion algorithm proposed by Cano-Ruiz [2000]. Parameter uncertainty was propagated through this algorithm into impact scores using Monte-Carlo simulation (@Risk [2001], Latin Hypercube sampling, 30'000 iterations). We used correlated sampling for parameters that appear in the life cycles of both plant-protection products. Scenario uncertainty was depicted by calculating one Monte-Carlo simulation for each scenario. The influence of individual parameters on the uncertainty of the impact scores was quantified with the contribution to variance [Fenner, 2001] of each parameter. The contribution to variance measures the influence of a parameter on the results distribution in terms of dispersion and absolute magnitude.

To evaluate the product comparison, we calculated the quotient of impact scores of the two alternatives:

$$Q = I_{Moddus} / I_{Stuntan} \quad (1)$$

where Q is the quotient of impact scores (dimensionless) and I is an impact score (unit of the impact category). In calculating such a quotient, uncertainty applying to both alternatives cancels out to an extent. Percentile distributions of Q were obtained as output of the Monte-Carlo simulations. Significant differences between the two alternatives were assumed, if 90 % of the values of Q were above or below unity.

We assumed a lognormal distribution for most parameters because it yields only positive values and because its long tail in high values is deemed appropriate for LCA parameters [Huijbregts, 2003]. Lognormal distributions were parameterised using dispersion factors [Huijbregts, 2003, Slob, 1994]:

$$k = X_i(0.975) / median_i \quad (2)$$

where k is the dispersion factor, i is the uncertain parameter, and X is the 97.5^{th} percentile of i. The range of uncertainty (uncertainty range, UR, dimensionless) in quotients of impact scores (Equation 1) is expressed as 90 % confidence interval:

$$UR = X_i(0.95)/X_i(0.05) \quad (3)$$

Sources of uncertainty included in the assessment are shown in Table 1. Concerning LCI flows, uncertainty and different sources of variability are depicted as one single generic dispersion factor per group of flows.

It was of interest here to derive such factors for processes of basic chemical production, because they exhibit a major contribution to the LCI of the production of active substances in the case study [Geisler, 2003b]. Therefore, we derived generic dispersion factors from the differences between elementary flows for the production of benzene and sodium hydroxide. Six and nine LCIs were compared for the production of benzene and sodium hydroxide, respectively [Geisler, 2004]. Calculating dispersion factors for comparable elementary flows in these different LCIs yields information on all sources of uncertainty in the LCI assessed here (Table 1).

A specific model uncertainty in the LCI stems from the use of the estimation procedure for LCIs of chemical production-processes in the supply of the active substances and formulation ingredients [Geisler, 2003b]: Knowledge on the efficiency of production processes is uncertain. Since pesticide producers have an influence on the production efficiency of precursors, we specifically wanted to illustrate the consequences of neglecting environmental objectives in supply chain management. Therefore, production efficiency in the chemical industry was assessed in a best and a worst-case scenario (Table 1) [Geisler, 2003b]. Finally, for those LCI data acquired specifically for this study (e.g., applied doses of the products) probability distributions were fitted.

Table 1: Sources of uncertainty [Huijbregts, 1998] covered in this work.

Source\phase	LCI	LCIA
Parameter uncertainty	Imprecise calculation of flows; Unknown composition of sum parameters	Imprecise knowledge on substance/ environmental properties
Model uncertainty	Assumptions on production efficiency	N/a [a]
Uncertainty due to choices	Different allocation methods, system boundaries, etc.	N/a [a]
Temporal variability	Variation of parameter values between years	N/a [a]
Spatial variability	Variation of parameter values between production sites	N/a [a]
Variability between objects/sources	Different production processes for the same product	Variability in exposure parameters

[a] N/a – not assessed.

To depict uncertainty in characterisation factors of the CML-baseline method [Guinée, 2001], generic dispersion factors were used as published by Huijbregts [2003]. Sources of uncertainty comprised in these factors are parameter uncertainty and variability in exposure assessment parameters (e.g. human characteristics, Table 1). Sum parameters carry a specific uncertainty, because their composition is not known quantitatively. Therefore, uniform distributions were defined for the characterisation factors of sum parameters. The minimum and maximum value of each uniform distribution was defined by the minimum and maximum characterisation factor, respectively, of the range of substances a sum parameter comprises. Additionally to this uncertainty in the composition of sum parameters, the uncertainty of characterisation factors themselves was modelled for sum parameters as for any other characterisation factor.

4 RESULTS

4.1 Case-Study Results

In Table 2, the probability of the quotient of impact scores (Equation 1) to be larger than one and the uncertainty ranges are shown. The spreads in the distributions are caused by uncertainty in LCI flows and in characterisation factors. Uncertainty ranges are considerably higher regarding the toxicity impact-categories than for the other midpoints. Significant differences between the two products occur only in the worst-case scenario, with regard to acidification, photooxidant creation and human toxicity impacts: Moddus shows significantly higher impact scores than Stuntan according to the significance criterion chosen (see Methods).

The applied doses of the two plant-protection products have high contributions to variance in all impact categories (Table 2), because the applied dose is the reference flow of the functional unit. Therefore, uncertainty in this parameter has an effect on all other parameters in the life cycles compared. The doses are uncertain, because in pesticide registration, a dose range is set permitting some flexibility to the farmer. The utilisation of this dose range by farmers is influenced by various factors, e.g. differences in prices between products. Uncertainty in the LCIs of basic chemical and energy supply, expressed as dispersion factors, also contributes considerably to variance. Concerning the toxicity impact categories, the uncertain composition of sum parameters plays a major role for uncertainty. With regard to single substances, the characterisation factor for emissions of chlorocholine chloride to air and water has high contributions to variance in freshwater ecotoxicity impact-scores. This contribution to variance of impacts of chlorocholine chloride explains the large uncertainty range in freshwater ecotoxicity in Table 2, because the generic uncertainty factors for the characterisation factors of chlorocholine chloride are as high as 50 (emission to air) and 100 (emission to water) [Geisler 2003a]. Additionally, air emissions of substrates in chemical production exhibit a considerable contribution to variance in the worst-case scenario, where the emission factor for such substances is relatively high [Geisler, 2003b]. Due to the unavailability of mammalian no-effect data for these substrates, we applied a worst-case no-effect value [Geisler, 2003a] to calculate characterisation factors for the human toxicity potential in USES-LCA [Huijbregts, 1999]. It is common practice in chemical industry in Western Europe to combust off-gases containing such highly toxic substances [Geisler, 2003b]. The high contribution to variance exhibited by these substrate emissions therefore gives a conceptual idea of the consequences of such emissions. Uncertainty in tractor operations largely cancels out in the product comparison due to correlated sampling. Remaining sources of uncertainty have small contributions to variance below 6 %.

4.2 Using Uncertain Results in Decision Making

Considering uncertainty in decision making is important, but may substantially increase the complexity of results. The presentation of uncertain results to decision makers may however be facilitated by a simplified representation, such as presented in Table 3.

The use of highly uncertain LCA results for decision support may not be advisable. For instance, it would not be desirable suppressing the development of the new product Moddus on the grounds that the LCA shows no progress compared to the established product Stuntan concerning freshwater ecotoxicity, as long as method uncertainty is a major cause for this insignificance. Therefore, impact-category results carrying extremely high uncertainty should be marked as such (Table 3).

Table 2: Probability (%) of the quotient of impact scores to be larger than one, with asterisks designating significant differences between the products and contribution to variance (CTV) of groups of parameters.

Impact category	Global warming	Acidification	Eutrophication	Photoox. creation	Human toxicity	Freshwater ecotoxicity	Terrestrial ecotoxicity
P(Q>1), best case	86	53	33	51	64	29	71
P(Q>1), worst case	88	99*	75	90*	93*	61	84
UR, best case	2.2	2.3	2.5	3.8	5.6	15	40
UR, worst case	2.7	2.8	3.5	3.3	120	45	40
Highest CTV	Applied doses	Applied doses	Basic chemical and energy supply	Applied doses	Sum parameters	Sum parameters	Applied doses
2^{nd} highest CTV	Basic chemical and energy supply	Basic chemical and energy supply	Applied doses	Sum parameters	Applied doses	Chlorocholine chloride to water	Sum parameters, substrates to air

Table 3: Simplified representation of the results of the product comparison under uncertainty (👎 means that the impact score of Moddus is significantly higher than that of Stuntan, -- means that the results are insignificant, and ~ denotes high method uncertainty).

Production efficiency scenario	Implication for supply chain	Likeliness	Global warming	Acidification	Eutrophication	Photo-oxidant creation	Human toxicity	Freshwater ecotoxicity	Terrestrial ecotoxicity
Best case	High environmental standards	High	--	--	--	--	--	~	--
Worst case	Low environmental standards	Low	--	👎	--	👎	👎	--	~

4.3 Significance: Rules of Thumb

It is impossible to fully predict uncertainty in LCA results without conducting a quantitative uncertainty analysis. Since full uncertainty analyses are time-consuming, rules of thumb concerning the significance of LCA results would be helpful in LCA practice.

In deterministic case studies, only expert judgement is available to set significance criteria for results. We found estimated significance criteria in published case studies ranging between 1.1 and 2 (e.g. [Frischknecht, 1996, Ross, 2003]) expressed as quotients of impact scores and only once as high as 10 [Finnveden, 1998], expressed as quotients of elementary flows. In our case study, median quotients (Equation 1) are assumed to approximate deterministic results. Significant differences demanded median quotients larger than 3 concerning toxicity impact-categories, and around 1.6 concerning other impact categories. Compared to our findings, expert judgements common in literature overestimated the significance of LCA results. Our results suggest that a median quotient of impact scores larger than two may be considered on the safe side of being significant, concerning the impact categories global warming, acidification, eutrophication and photooxidant creation. This rule of thumb is supported by inherent characteristics of these impact categories regarding uncertainty, namely few impact pathways and a small number of elementary flows contributing to these categories [Geisler, 2004]. Case-study results exhibiting smaller differences should be evaluated for significance with a full uncertainty analysis. Regarding toxicity impact-categories, no rule of thumb is proposed, because large dispersion factors of individual parameters cause highly varying uncertainty in individual toxicity impact-scores. A detailed uncertainty analysis seems indispensable for reliable decision support concerning toxicity impacts assessed by the CML-baseline method.

The case study used here to establish the rule of thumb compared products with relatively similar life cycles. This was also the case in the work of Huijbregts [2003]. Larger differences between life cycles, e.g. mechanical compared to chemical weed control, often lead to larger differences in impact scores of alternatives. However, strongly differing life cycles also imply weaker correlations among input distributions in these life cycles, leading to larger uncertainty ranges (Equation 3). These two trends counteract each other in their effect on the significance of differences between impact scores of alternatives.

We conclude that in the absence of better data, the rule of thumb may be used in LCA if a full uncertainty analysis is outside the scope of the study. It should however be born in mind, that the rule of thumb proposed here has not been verified yet for products with very different life-cycles.

5 DISCUSSION AND CONCLUSION

The comparison of the plant-growth regulators showed no significant differences in most impact categories. With regard to freshwater ecotoxicity impacts, larger differences between Moddus and Stuntan are superimposed by exceptionally large uncertainty (two orders of magnitude between 5th and 95th percentile of the quotient). High uncertainty also superimposes relatively large differences between impact scores regarding terrestrial ecotoxicity impacts in the worst-case scenario. Measures to reduce uncertainty should be taken before these toxicity impact-scores are used for decision support.

In spite of large uncertainty in some impact categories, the case-study results give the important information that Moddus is not significantly environmentally preferable to Stuntan, regarding the more likely best-case scenario (Table 3). Another useful recommendation from the case study may be the inclusion of environmental objectives in the supply-chain management of pesticide producers, to avoid the worst-case scenario (Table 3).

The ranges of uncertainty of the case-study results can be compared to those published by Huijbregts [2003] for a case-study comparing housing insulation options. Uncertainty ranges in Huijbregts [2003] are considerably smaller than in this work, especially with regard to toxicity impact-scores. One reason for this is that sum parameters play a crucial role for the high uncertainty range (Equation 3) in toxicity impact-scores in our work, while they were of minor importance in the case study of Huijbregts [2003]. Second, higher dispersion factors were assumed in this work compared to Huijbregts [2003] for LCI parameters and characterisation factors. These differences in parameter uncertainty are mainly due to substantial efforts of Huijbregts [Huijbregts, 2003] to acquire specific dispersion factors for parameters in the LCA with high contribution to variance. Such an iterative approach is however not generally practicable in LCA, because it is very labour intensive and necessitates access to substance data and models used in the calculation of characterisation factors.

The rule of thumb for the significance of LCA results (Section 4.3) is useful when a quantitative uncertainty analysis is not feasible (e.g., due to time restrictions). However, routine uncertainty analyses should be aimed at in future LCA practice. To this end, data and guidelines on the definition of uncertainty in LCI and LCIA are needed. To further facilitate an uncertainty assessment, LCI and LCIA data providers should supply quantitative uncertainty information including correlation estimates for individual parameters. Ecoinvent [2003] has already made a first step toward this direction by estimating uncertainty ranges for LCI parameters. In contrast to parameter uncertainty, a large variety of choices and sources of model uncertainty are less accessible to quantitative analysis. It is suggested that choice and model uncertainty of specific interest for goals and scopes of case studies be modelled quantitatively. Choices and model uncertainty generally applying to LCA should be made transparent to decision makers. This enables decision makers to explicitly accept choices and models employed as being an adequate basis for decision support.

References

@Risk 4.5 - Risk Analysis Add-in for Microsoft Excel, Newfield, NY, USA: Palisade Corporation, 2001.

BfL, Pflanzenschutzmittelverzeichnis (*Directory of Plant Protection Products*), Bern: Bundesamt für Landwirtschaft, 2002.

Cano Ruiz JA, Decision Support Tools for Environmentally Conscious Chemical Process Design, Dissertation, Massachusetts Institute of Technology, 2000.

ecoinvent Centre, ecoinvent data v1.01, Final reports ecoinvent 2000 No. 1-15., Dübendorf: Swiss Centre for Life Cycle Inventories, 2003.

Fenner, K, Transformation Products in Environmental Risk Assessment: Joint and Secondary Persistence as New Indicators for the Overall Hazard of Chemical Pollutants, Dissertation, Swiss Federal Institute of Technology No. 14400, Zurich, 2001.

Finnveden G and L-G Lindfors, Data quality of Life Cycle Inventory data - rules of thumb, *International Journal of Life Cycle Assessment*, 3(2), 65-66, 1998.

Frischknecht RE, U Bollens, S Bosshart, M Ciot, L Ciseri, G Doka, R Hischier, A Martin, R Dones, and U Ganter, Ökoinventare von Energiesystemen *(LCIs of Energy Systems)*, 3rd ed, Bern: Swiss Federal Office of Energy, 1996.

Geisler G, Life Cycle Assessment in the Development of Plant Protection Products: Methodological Improvements and Case Study, Swiss Federal Institute of Technology (ETH), Zurich, 2003a.

Geisler G, T Hofstetter, and K Hungerbühler, Production of Fine and Speciality Chemicals: Procedure for the Estimation of LCIs, *International Journal of LCA*, OnlineFirst http://dx.doi.org/10.1065/lca2003.10.139, 2003b.

Geisler G, S Hellweg, and K Hungerbühler, Uncertainty Analysis in Life Cycle Assessment (LCA): Case Study on Plant-Protection Products and Implications for Decision Making, *International Journal of LCA*, submitted, 2004.

Guinée JB, M Gorrée, R Heijungs, G Huppes, R Kleijn, A de Koning, L van Oers, A Wegener Sleeswijk, S Suh, and HAU de Haes, CML-Guide to Life Cycle Assessment, Leiden: Centre of Environmental Studies, Leiden University (CML), 2001.

Huijbregts MAJ, Application of Uncertainty and Variability in LCA, Part I: General Framework, *International Journal of Life Cycle Assessment*, 3(5), 275-280, 1998a.

Huijbregts MAJ, Application of Uncertainty and Variability in LCA, Part II: Dealing with parameter uncertainty and uncertainty due to choices in Life Cycle Assessment, *International Journal of Life Cycle Assessment*, 3(6), 343-351, 1998b.

Huijbregts MAJ, Priority assessment of toxic substances in the frame of LCA - The multi-media fate, exposure and effect model USES-LCA, Amsterdam, NL: University of Amsterdam, 1999.

Huijbregts MAJ, G Norris, R Bretz, A Ciroth, B Maurice, B von Bahr, B Weidema, and ASH de Beaufort, Framework for modelling data uncertainty in life cycle inventories, *International Journal of Life Cycle Assessment*, 6(3), 127-132, 2001.

Huijbregts MAJ, W Gilijamse, AMJ Ragas, and L Reijnders, Evaluating uncertainty in environmental Life-Cycle Assessment, *Environmental Science and Technology*, 37, 2600-2608, 2003.

Ross S and D Evans, The environmental effect of reusing and recycling a plastic-based packaging system, *Journal of Cleaner Production*, 11(561-571, 2003.

Slob W, Uncertainty analysis in multiplicative models, *Risk Analysis*, 14(4), 571 - 576, 1994.

Weidema BP and MS Wesnaes, Data quality management for life cycle inventories - an example of using data quality indicators, *Journal of Cleaner Production*, 4(3-4), 167 - 174, 1996.

The Influence of Agricultural Data Uncertainty in the Life Cycle Assessment of Biodegradable Hydraulic Lubricants

R. Ferret, G. Mendoza and M. Castilla

Foundation Tekniker, Research and Development Center, 20600 Eibar, SPAIN.

Abstract: From a Life Cycle Assessment study of hydraulic lubricants from vegetable and mineral oil, it has been found that the agricultural step in the biodegradable hydraulic lubricant production has the main environmental impact. The aim of this study is to develop an uncertainty analysis of agricultural data used in the biodegradable hydraulic lubricant inventory from this LCA to determine the influence of data uncertainty in the environmental LCA result. Two parameters have been selected for the uncertainty analysis: fertilization practices and machinery operations. A variation parameter analysis has been carried out obtaining that assumptions and simplifications made (pre-treatment soil operations and fertilization rates), which clearly influence on the different impact categories studied. A Monte Carlo analysis has been performed showing Eutrification Potential category is the most affecting by input data uncertainty propagation, which means the fuel consumption and fertilization rates input data need to be selected from a more accuracy source.

Keywords: Uncertainty analysis, Life Cycle Inventory (LCI), data quality, Monte Carlo simulation, agriculture, parameter variation.

1. INTRODUCTION

Thousands of tons of oil are discharged into nature causing important environmental problems and leading to the need of alternative lubricant products with minimal environmental impact. To support the key environmental decision in replacing conventional mineral lubricants with one based on vegetable oil, a cradle-to-grave Life Cycle Assessment (LCA) must be performed following the stages of the both processes, which are shown in the Figure 1.

The process for obtaining mineral lubricants is already clearly defined and well-known. On the other hand, the manufacturing process of vegetable oil based lubricants is not so defined, as the agricultural step has not been so much established.

Thus, the major environmental impact contribution in the biodegradable lubricant process comes from the agricultural step. However, data quality for the agricultural production is one of the main sources of uncertainty, specially regarding the establishment of valid data sets describing crop production in a consistent way, which includes the choice of representative regions as well as adequate scenarios.

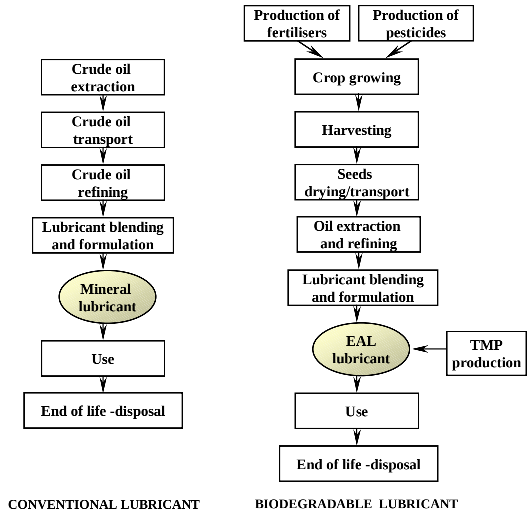


Figure 1. Process stages of two systems compared.

Nowadays, it is well recognized that different sources of uncertainty are present in life cycle inventories and impact assessments, but modelling data uncertainty is not a common practice in life cycle inventories. Validity of results obtained in a Life Cycle Assessment gains by a data quality evaluation.

A LCA study has been carried out on hydraulic lubricants from mineral and vegetable oil in which the functional unit has been defined as 20.000 working hours, taking into account the life expectancy of a lubricant in a typical hydraulic system. From this LCA has been proven that crop production represents the most environmentally unfriendly step in the biodegradable lubricant manufacturing.

The goal of this study, therefore, is to develop an uncertainty analysis of agricultural data used in the biodegradable hydraulic lubricant inventory from this comparative LCA to determine the influence of data uncertainty in the environmental comparative LCA result.

2. MODELLING THE AGRICULTURAL SYSTEM

Concentrating on the crop production step, it is important to consider the influence of agricultural field activities as well as fertilizers and pesticides production or transportation stages [Brentrup, 2000]. Therefore, parameters like yield, fertilization rates, nitrate leaching, ammonia volatilisation or energy consumption do not contribute in the same way in the final environmental impact.

The scheme of the involved steps in a crop production are shown in the Figure 2.

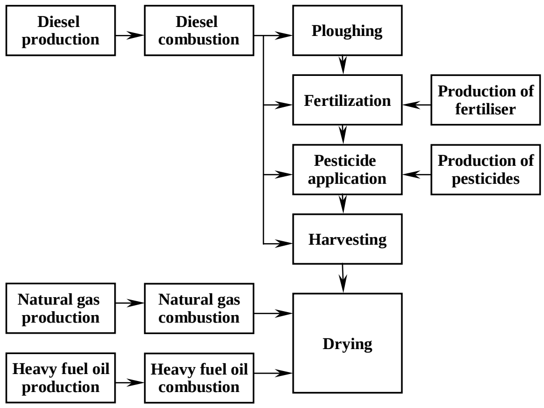


Figure 2. Stages considered in the agriculture step of the study.

Taking into account the preliminary results obtained from the LCA study and in order to carry out the uncertainty analysis, two main parameters has been selected from agricultural step: fertilization rates and energy consumption, which produce the highest environmental impact. In case of energy consumption, the input parameter has been expressed as diesel consumption in the machinery operations.

2.1. Energy consumptions

The machinery operations in the vegetable base hydraulic lubricant production have shown a significant impact on the overall results of LCA. These operations include: pre-treatment, ploughing, harrowing, harvesting, etc. This machinery use is referred to the fuel consumption; which means the fuel combustion is an input variable in the analysis.

The production of machines as such and the effect of use on the life of the machine have been excluded in the study.

The calculation of the use of fuel has been estimated once the farming operations have been established for the crop production which is used in the vegetable oil obtention. This estimation has been based on the average power needed to carry out each operation and on the fuel used to supply this power [Lee, 2000.] Finally, data for emissions associated to fuel combustion in machining operations have been taken from Dalgaard [2003].

In reference to energy consumptions, the energy used to dry grains before storage has to be consider. Seeds are dried by been heated with a natural gas-fired burner which usually operates at moderate temperature to prevent cracking. Therefore, the amount of water to be removed from the crop has been considered as the factor determining fuel consumption [Weidema, 1999].

2.2. Fertilization considerations

LCI average rates estimated from European sources [FAOstat] have been taken due to the data heterogeneity in crops fertilization.

As it has been said before, agriculture is by far the main source of emissions, nearly 90% of the global emissions identified in the comparative LCA analysis [Audsley, 1997] and this impact is directly associated to total NH_3 and NO_x emissions as a consequence of crops fertilization.

N-related emissions have been shown a high influence on many complex interactions between soil type and climatic conditions and, on the other hand, on parameters determined by means the agricultural management practices. That is the reason why it is often difficult to derive exact rates for N release to air and water.

Because of lack of agricultural data to quantify the loss of N fertilizer from ammonia volatilisation and emission of nitric oxide (NO_x) through nitrification after fertilizer is applied to fields, emission prediction models have been considered from Bouwman [1996]. This author proposed an emissions factor for N_2O from mineral and organic fertilizers, simplifying the complex dependencies of the fertilization parameters.

This study has not been taken into account the manure and pesticide application.

3. DATA QUALITY IN AGRICULTURAL LCA STAGE

LCA methodologies were primarily designed for industrial applications, and therefore this tool shows some difficulties when is applied to agricultural systems. Table 1 shows different characteristics assessing industrial and agricultural systems and summarizes the main problems performing an agricultural LCA.

Characteristics (Table 1) determine that the choices made in the agricultural step of a vegetable lubricant life cycle have a major contribution to environmental impacts of the chain.

Table 1. Main characteristics of industrial and agricultural systems[Milà i Canals, 2003].

Characteristics	Industrial Systems	Agricultural Systems
Dependency from location	Highly independent	Highly dependent
System boundaries	Clearly defined	Unclear, both physically and temporally
Main source of impacts	Energy and materials consumption	Land use, energy and materials consumption, and field emissions
Degree of knowledge	High (simple and pre-designed processes)	Relatively low (complex, natural processes)
Functionality	One or few functions	Multifunctional

It is well known the importance of a methodical inventory data collection. However, agriculture process is characterised by a large variation of data sources depending on climate, soil and management systems. That means that agricultural inputs and outputs (yields, emissions to water, soil, and air) must be obtained with precise measurements methods or experts estimations and assumptions to assure their accuracy. Systematic errors leading to wrong issues have to be avoided, for example, the use of unrepresentative farm data, the use of inconsistent values of partial emissions or extrapolation from empirical models.

Consequently, all the assumptions and simplifications, which are made during the data collection, affect in different extend to each impact categories in the evaluation. Thus, it should be accompanied by a appropriately reliability analysis, which is frequently not considered.

Unfortunately, a lack of reliability is frequently observed because LCA performers tend to save time using data from inadequate time periods or sites.

Various methods have been proposed to manage data inaccuracy in LCA, such as analytical uncertainty propagation methods, calculation with intervals and fuzzy logic and stochastic modelling. In particular, stochastic modelling, which can be performed by Monte Carlo simulation, is widely recognized as a valid technique for making data inaccuracy in LCIs operational. The level of mathematics required to perform a Monte Carlo simulation is quite basic [Huijbregts, 1998; Maurice, 2000].

The Monte Carlo simulation basis is to select a value within the distributions assigned for each input variable and to compute the outputs. The process is repeated and the collective outputs from each interaction combine to form a probability distribution function [McCleese, 2002].

The input data are calculated as an average outcome, that is, the medium value of the data collection from different sources. Therefore, every input data has associated a gaussian distribution, with a corresponding confidence interval. Thus, the confidence interval for these input parameters is expressed as the coefficient of variation, in other words, a percentage of the standard deviation.

For this study, the basic scenario of the Monte Carlo simulation analysis is indicated in the Table 2.

Table 2. Basis scenario used in the Monte Carlo simulation analysis.

Operation	Input parameter	Basis scenario	Coef. variation
Machinery operations	*diesel*	109 l/ha	20%
Crop fertilization	*P*	46 kg/ha	39%
	K	60 kg/ha	56%
	N	129 kg/ha	27%

4. RESULTS

The software used in this study is GaBi 4, which applies Eco-Indicator 95 methodology. This software implements parameter variation and sensitivity analysis functions, allowing the modeller to examine how changes in input parameters or different scenarios affect the output results [Spatari, 2001].

Primarily, an analysis of sensitivity using parameter variation has been performed. It has been studied the influence of the following input variables: fuel consumption and fertilizer use in the farming operations.

As an example, Figure 3 shows the influence of variation of fuel consumption parameter on the Global Warming Potential (GWP) impact category. The scenario 1 involves a crop cultivation without pre-treatment soil operations and the scenario 2 includes the pre-treatment soil operations. This reveals an 5% increase in the GWP impact category from the scenario 1 to the scenario 2.

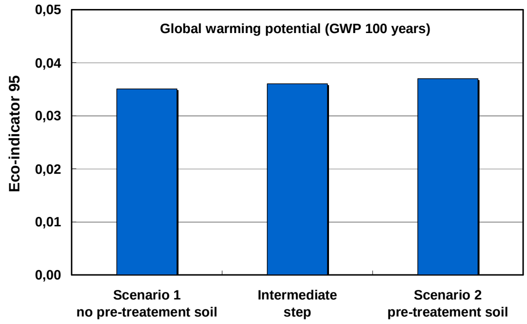


Figure 3. Fuel consumption parameter variation on the GWP impact category.

The parameter variation is more influential when analysing the fertilization practices. The Figure 4 shows how affects the variation of fertilizer parameter on the GWP impact category, where scenario 1 represents a conventional crop (using NPK fertilizers) and scenario 2 implies ecological cultivation (without fertilizers). There is a 15% decrease in the impact category from the first scenario to the second one.

This parameter variation analysis shows the strong repercussion of the considered data inventory assumptions on the study results. It can be deduced the importance to compile the closest field data to the real processes.

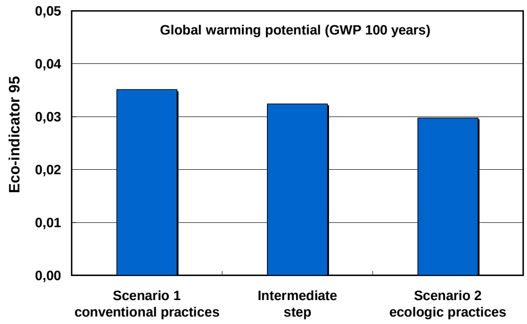


Figure 4. Fertilizer parameter variation on the GWP impact category.

On the other hand, a Monte Carlo simulation has been carried out to promote reliability to the output in the biodegradable hydraulic lubricant LCA.

The Monte Carlo analysis evaluates how the propagation of inputs variation is reflected in output values. The resulting output value corresponding to the impact category presents a mean value with a corresponding standard deviation. As a example, the shows graphically the Monte Carlo simulation outcomes for the GWP impact category.

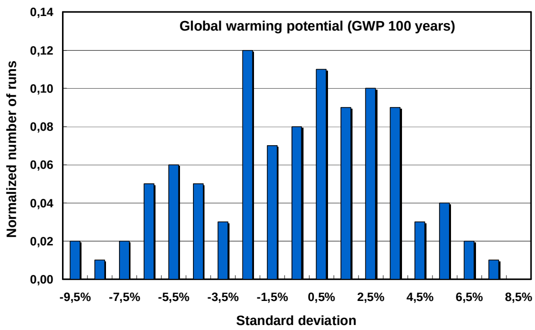


Figure 5. Results of a Monte Carlo Simulation for GWP impact category.

Numerically, Table 3 shows the coefficient of variation (CV) obtained for each impact category analyzed.

The Aquatic Ecotoxicity Potential impact category has not been considered because its contribution is not relevant for this process uncertainty analysis. The variation of the standard deviation is quite different, depending on the impact category studied.

Table 3. Result s for a Montecarlo Simulation in the different impact categories.

Impact Category	Unit/ Normalization	Mean Value	CV
Global Warming Potential, (GWP 100 years)	Eco-indicator 95	0,04	4%
Acidification Potential (AP)	Eco-indicator 95	0,04	4%
Eutrification Potential (EP)	Eco-indicator 95	0,04	14%

For example, Eutrification Potential category is the most influenced for the input parameters variation. It is a consequence of the fertilization step assumptions and considerations through applied fertilization rates. Consequently, it could be convenient a new estimation of the input parameters corresponding to fertilization stage.

5. DISCUSSION AND CONCLUSIONS

To sum up, uncertainty analysis techniques have been used to identify the important weak points in the biodegradable hydraulic lubricants Life Cycle Assessment study. The Monte Carlo simulation based screening methods are beneficial to the LCA results which give added value to environmental impacts obtained.

In other words, this kind of analysis allows to know which environmental impact categories support the major uncertainty. By means of the knowledge of input data which affects each environmental impact category, a revision of data inventory quality could be made improving the uncertainty in the final result.

The assumptions and simplifications made in the LCA study for biodegradable lubricant inventory (pre-treatment soil operations and fertilization rates) clearly influence the Eutrification Potential impact category in more than 10%, as obtained from Monte Carlo analysis.

The set-up of consistent models based on realistic input-output relations using detailed farm data from case studies, surveys, or detailed accounts statistics, makes the LCA to improve the results feasibility, as well as to gain a better knowledge of emission processes together with a check and adjustment of partial emissions of nutrients with balances at farm and enterprise level.

Suitable format databases with verified information concerning production on typical and representative farms and using a combination of detailed farm data, models and comprehensive accounts statistics, could be recommended.

6. ACKNOWLEDGEMENTS

This work has been partly funded by the EC (project contract G5RD-CT-2000-00410). We would like to thank all other partners involved in the project: Vito (Belgium), Volvo (Sweden), ICMM-CSIC (Spain), Doimak (Spain), CTD (Slovenia), Petrol (Slovenia) and Bekaert-Dymonics (Belgium).

7. REFERENCES

Audsley, A., Alber, S., Clift, R., Cowell, S., Crettaz, P., Gaillard, G., Hausheer, J., Jolliet, O., Kleijn, R., Mortensen, B., Pearce, D., Roger, E., Teulon, H., Weidema, B. and van Zeijts, H., Harmonisation of Environmental Life Cycle Assessment for Agriculture. *Final Report for Concerted Action AIR3-CT94-2028*, 1997.

Bouwman A.F., Direct emission of nitrous oxide from agricultural soils. *Nutrient Cycling in Agroecosystems*, 46, 53-70, 1996.

Brentrup, F., Küsters, J., Lammel, J., Kuhlmann, H., Methods to estimate on-field nitrogen emissions from crop production as an input to LCA studies in the Agricultural Sector, *The International Journal of Life Cycle Assessment*, 5(6) 349-357, 2000.

Chouquet, J., Bodin, O., Kohler, N., Dealing with sensitivity and uncertainty analysis in integrated building LCA model, *The International Conference on the Applications of Computer Science and Mathematics in Architecture and Civil Engineering*, Weimar (Germany), 2003.

Dalgaard D. og Halberg, N.; Metode og forudsætninger for livscyklusvurdering af landbrugsprodukter, 2003.

FAOstat, FAO statistical database, http://apps.fao.org/

Huijbregts M, Application of uncertainty and variability in LCA. Part I: a general framework for the analysis of uncertainty and variability in life cycle assessment, *The International Journal of Life Cycle Assessment,* 3 (5) 273–280, 1998.

ISO 14040 Environmental Management - Life cycle analysis - Principles and framework; ISO 14041 Goal and scope definition and inventory analysis; ISO 14042 life cycle impact assessment, ISO/CD 14043 Life cycle interpretation, 1998.

Lee J, Cho H, Choi B, Sung J, Lee S, Shin M, Lifecycle assessment of tractors, *The International Journal of Life Cycle Assessment,* 5 (4) 205-208, 2000.

Maurice B, Frischknecht R, Coehlo-Schwirtz V, Hungerbühler K, Uncertainty analysis in life cycle inventory. Application to the production of electricity with French coal power plants. *Journal of Cleaner Production* 8 (2) 95-108, 2000

McCleese, D.L. LaPuma, P.T. Using Monte Carlo Simulation in Life Cycle Assessment for Electric and Internal Combustion Vehicles, *The International Journal of Life Cycle Assessment*, 7(4) 230–236 (2002).

Milà i Canals, L., Contributions to LCA methodology for agricultural systems. Site-dependency and soil degradation impact assessment, PhD Thesis, Universitat Autonoma de Barcelona, Spain, 2003.

Spatari, S., Betz, M., Florin, H., Baitz, M., Faltenbacher, M., Using GaBi 3 to perform life cycle assessment and life cycle engineering, *The International Journal of Life Cycle Assessment*, 6(2), 81-84, 2001.

Weidema B.P., Meeusen M, Agricultural data for life cycle assessments, *The Second European Invitational Expert Seminar on Life Cycle Assessments of Food Products*, January, 1999.

An implemented approach for estimating uncertainties for toxicological impact characterisation

R. Rosenbaum[a], **D.W. Pennington**[b], **O. Jolliet**[a]

[a] *Industrial Ecology & Life Cycle Systems, Swiss Federal Institute of Technology Lausanne (EPFL), 1015 Lausanne, Switzerland*

[b] *Soil and Waste Unit, Institute for Environment and Sustainability (IES), Joint Research Centre (JRC), European Commission (EC)*

Abstract: One approach accounting for parameter and model uncertainty is implemented in the LCIA (life cycle impact assessment) method IMPACT 2002. The uncertainty is estimated for intermediate results from the chemical fate, human intake fraction, and two toxicological effect modules. Overall uncertainty estimates are then arithmetically calculated. Results are presented for impact contributions in the contexts of aquatic ecosystems and human health. The approach of Hofstetter (1998) was adapted for estimating the uncertainty related to chemical fate and human intake fractions. A fundamental problem when estimating uncertainties for 1000's of substances consists of the lack of uncertainty distributions for all of the input data and the need to have a practical approach to assign distributions to each chemical. Hofstetter (1998) proposed the use of fixed factors for clusters of substances. The choice of a factor is then dependent on the emission medium, exposure route, and the robustness of the model relative to the chemical being considered. The factors are initially determined for representative substances for each category using evaluation data, expert judgement, or approaches such as Monte Carlo. There is then no need to repeat the Monte Carlo calculations. Multiplying and dividing the geometric mean estimate by a factor provides an estimate of the upper and lower 95th percentile confidence interval bounds. The human health effect factor uncertainty is similarly defined and readily combined through addition with that of the intake fraction. Using expert judgement, three uncertainty classes were proposed to estimate uncertainty related to the human effects input data. These effects data account for both the risk of an effect, as well as the potential consequences of population-based exposures. The uncertainty for ecotoxicological effects is currently related to the number of species tested for aquatic species in the water column. The more species test results available, the more robust the estimate of the ecotoxicological factor is assumed to be. For estimating the ecotoxicological effect factor uncertainty, the combined use of two distinct approaches was suggested, – the higher uncertainty estimate being adopted. The combination of both guaranteed more robust results compared to applying either method – both being based on differing assumptions related to the sample versus the population distribution. The presented approach proved to be very transparent, robust but while reflecting our current level of knowledge, quick to use, and is easily applied in practice to combine the uncertainty of the emissions inventory with those of the impact assessment phase in a life cycle assessment study.

Keywords: Uncertainty; LCIA; Toxicity; Multimedia Modelling

1. INTRODUCTION

Accounting for uncertainty is vital in comparative assessments, although this has remained neglected in most life cycle assessment (LCA) studies and in related decision making. In LCA, uncertainty is, in part, due to the uncertainty associated with the model input data, at least those that are significant in the calculations of a characterisation factor (the impact per unit emission) for a particular chemical. This is the parameter, or data, uncertainty. Additionally, the uncertainty is related to the inherent uncertainty of the overall models and the underlying correlations themselves (model uncertainty). In combination, these model and parameter uncertainties could be considered the accuracy of the model.

IMPACT 2002 (Pennington et al., 2003) provides estimates for the uncertainty accompanying every characterisation factor (CF). By proposing a straightforward way of combining uncertainties of intermediate results to derive the final overall characterisation factor uncertainty, which can then be combined with LCA inventory uncertainties, the IMPACT 2002 methodology facilitates the calculation of LCA results with a related overall uncertainty estimate. In this way, judgement of LCA results is improved as their reliability can be taken into consideration by decision makers.

The general framework for calculating a human health effect characterisation factor is established as:

$$CF_h = FF \cdot XF \cdot EF_h = iF \cdot EF_h \quad (1)$$

where *FF* denotes the fate factor, *XF* the exposure factor, EF_h the human health effect factor, and *iF* the human intake fraction which is the product of *FF* and *XF*.

Assuming that *XF* = 1, in IMPACT 2002 the parallel concept for (aquatic) ecotoxicological effect characterisation is:

$$CF_{aqu} = FF \cdot EF_{aqu} \quad (2)$$

As an example we used the non-spatial version of IMPACT 2002 for an emission of 1,1,2,2-Tetrachloroethane into European surface waters. Based on the framework presented in Equations 1 and 2 we obtained the following illustrative results:

Table 1: Illustrative geometric mean estimates for an emission of 1,1,2,2-Tetrachloroethane into European surface waters.

(intermediate) Result	Value
Human health characterisation factor, CF_h [*number of cases* /$kg_{emitted}$]	1.66E-06
Ecotoxicological characterisation factor, CF_{aqu} [PAF $m^3 \cdot day^2/kg$]	2.93E+01
Human health effect factor, EF_h [*number of cases*/kg_{intake}]	8.74E-02
Ecotoxicological aquatic effect factor, EF_{aqu} [PAF $m^3 \cdot day/kg$]	2.4E+02
Human Intake Fraction, *iF* [$kg_{intake}/kg_{emitted}$]	1.90E-05
Human exposure factor, *XF* [1/*day*]	1.56E-04
Fate factor, *FF* [*day*]	1.22E-01

Based on these geometric mean estimates their 95th percentile confidence intervals are calculated in the following subsections.

2. UNCERTAINTY MODULES

2.1. Fate and exposure uncertainty

For estimating the overall uncertainty related to fate and exposure the approach of Hofstetter (1998) was adapted due to its straightforward and pragmatic, yet robust nature. By considering the emission medium, exposure route, and the appropriateness of the model relative to the chemical, this approach provides estimates for clusters of chemicals of the combined model, scenario, and parameter uncertainty – i.e. the likely overall accuracy of the results.

A fundamental problem when calculating the uncertainties of results for many substances is that not every input parameter has a “known” uncertainty distribution that can be used to estimate the final uncertainty. Hofstetter (1998) therefore proposed a set of fixed factors, based on representative distributions for clusters of chemicals with similar attributes. These factors are the square geometric standard deviations (SD_g^2) associated with log-normal distributions. The factors depend on the exposure route, emission medium, and certainty criteria classifying the robustness of the model in the context of the substance into one of three certainty classes (*h* – high, *m* – medium, *l* – low certainty).

The representative uncertainty distributions for each cluster of chemicals can be estimated using evaluation data for representative substances and expert judgement, or other approaches such as Monte Carlo. Caution is advocated when using approaches such as Monte Carlo as only input data uncertainty is often taken into account. The uncertainty of also the correlations in a model, etc., are often not addressed in practice and may be more significant. For example, estimates of contaminant uptake into meat and vegetation can be a major source of the uncertainty in human exposure estimates for most persistent organic chemicals that is often associated with the correlations adopted and not the data input into the model. The factors in the Hofstetter approach are the square geometric standard deviation values, SD_g^2. These factors are used to calculate the upper and the lower bound of the 95th percentile confidence interval by assuming a log-normal distribution of the uncertainty. For the intake fraction the calculations are:

$$iF_{97.5} = iF \cdot SD_g^2 \quad (3)$$

$$iF_{2.5} = \frac{iF}{SD_g^2} \quad (4)$$

Tables 2 and 3 list the values originally proposed by Hofstetter (1998) for emissions to air and water/soil, respectively. It should be noted that Hofstetter did not provide factors for soil emissions; hence the working proposal in Table 3 was adopted. Thus, the values for emissions to soil are the same as for water emissions. This might not be correct for soil emissions where exposure via drinking water or fish consumption are the main concern, but this may be rarely the case at a European population level for example (e.g. Bennett et al. 2002).

Table 2: Estimated fixed square geometric standard deviations for emissions to air (Hofstetter, 1998)

Certainty criteria	Certainty class	Exposure via: air	water	food
- >95% of intake is inhaled - chem. properties are well known - substance is well suited for model	*h*	2	4	8
- substance partitions - chem. properties are well known - substance is well suited for model	*m*	3	6	12
- substance properties are uncertain or unknown - substance is not well suited for model	*l*	20	40	80

Table 3: Estimated fixed geometric standard deviations for emissions to water and soil (Hofstetter, 1998)

Certainty criteria	Certainty class	Exposure via: air	water	food
- >95% of intake by drinking water - chem. properties are well known - substance is well suited for model	*h*	4	2	8
- substance partitions - chem. properties are well known - substance is well suited for model	*m*	6	3	12
- substance properties are uncertain or unknown - substance is not well suited for model	*l*	40	20	80

Hofstetter (1998) also provided a list of example substances that fit into each of the established clusters.

For the 1,1,2,2-Tetrachloroethane example introduced above we chose the certainty class *m* as inhalation is below 95% of intake, the properties are peer reviewed high quality data and the model was considered well suited for this organic non dissociating compound.

For an emission to water we find the SD_g^2 estimates for different exposure routes in Table 3. The detailed results suggest drinking water as the exposure pathway associated with the highest risk, hence SD_g^2 = 3. The 95th percentile confidence interval is calculated according to Equations 3 and 4, resulting in:

$iF_{97.5}$ = 5.70E-05

$iF_{2.5}$ = 6.33E-06

Hence, it is assumed that there is a 95% confidence that the true intake fraction is within a factor of 3 of the geometric mean estimate.

2.2. Effects uncertainty

The uncertainty of the effect factors is exclusively linked here to the likely reliability of the related input data, e.g. the Toxic Dose causing tumours in 50% of species (TD_{50}), the No Observed Adverse Effect Level (*NOAEL*), etc. used to derive the effect factor (*EF*). Only parameter uncertainty is addressed.

Different approaches and values are proposed to estimate the uncertainties for human health and for ecosystem effects. These methods have been chosen with consideration of the need to assess thousands of substances and their related input data, as consistently as possible, while being appropriate for use in comparative assessments. Further research and development is strongly encouraged, particularly in the context of estimating the low dose-responses that will be associated with many of the emissions and toxicological impacts assessed using tools such as LCA.

2.3. Human health effects uncertainty

The human health effect factor (effect per unit intake) uncertainty is currently estimated based on a rough orders-of-magnitude approach, pending completion of reviews of the various uncertainties and propositions for more robust values and rules.

Using expert judgement and building on classical approaches, three uncertainty classes were proposed to estimate uncertainty related to the human health effects input data (see Table 4). This is typical of current practice for regulatory chemical risk screening, except the uncertainty factors are retained as a measure of the uncertainty and, unlike in screening, were not included in the estimate of the effect factor. This helps to avoid bias amongst factors with similar geometric means in a relative comparison context, whilst retaining information on the different uncertainties.

Table 4: Expert judgement based uncertainty factors estimation of human health effects data uncertainty.

Uncertainty criteria	Uncertainty factor
Data from peer reviewed sources, such as the critical effect from US EPA's IRIS (IRIS, 2004) or Gold et al.'s carcinogenic data handbook (Gold & Zeiger, 1997) for example. These data should also be for chronic exposures.	10
Data from less peer reviewed sources that are based on chronic or sub-chronic exposure test results.	100
Data extrapolated from acute test results to humans (based on insights in e.g. Crettaz et al., 2002 and Pennington et al., 2002, as well as many other non-LCA sources).	1000

The calculation of the upper and the lower bound of the 95th percentile confidence interval of the effect factor (EF) is performed in a similar way to Equations 3 and 4.

For the 1,1,2,2-Tetrachloroethane example, peer reviewed high quality chronic effects data were used. This justifies an uncertainty factor of 10 (there is 95% confidence that the estimate is within a factor of 10 of the geometric mean). Hence, the upper and lower bounds of the effect factor can be estimated as:

$EF_{h,\,97.5}$ = 8.74E-01

$EF_{h,\,2.5}$ = 8.74E-03

2.4. Ecosystem effects uncertainty

Ecosystem effects uncertainty is currently related to the uncertainty of the underlying Hazardous Concentration data used to calculate the effect factor ($0.5/HC_{50}$) (Payet & Jolliet, 2004 Pennington et al. 2004). These HC_{50} data reflect the concentration at which 50% of species are likely to be affected.

In practice, the HC_{50} is estimated using the geometric mean of test results for different species in the relevant medium, currently in the water column for IMPACT 2002. The more species tested, the larger the sample size that is available to estimate the HC_{50} and the more robust it is considered to be. This robustness is expressed in terms of parameter uncertainty, the lower the sample size the higher the uncertainty.

As with the uncertainty calculations discussed above, the lower 95th percentile confidence interval HC_{50} is calculated by dividing by the square geometric standard deviation (similar to Equation 4) and the upper bound by multiplying by SD_g^2 (similar to Equation 3).

For estimating uncertainties related to the ecotoxicological effect factors, the combined use of two distinct statistical approaches was suggested (pending further investigation) – the higher uncertainty estimate being adopted:

Payet & Jolliet (2004) proposed adoption of the Student approach to calculate the 95%ile confidence interval from the sample distribution of the logarithmic values to estimate the 2.5th and the 97.5th percentiles (assuming a log-normal distribution). Secondly, a fixed-factor approach, dependent only on the sample size (*n*), was adopted based on a review of methods, distributions, and available data to determine the 95th percentile confidence interval for the HC_{50} (Pennington, 2003).

As shown in Table 5 for the fixed factors, the uncertainty significantly decreases with sample sizes higher than three.

Table 5: Median and upper 95th percentile confidence interval limits for extrapolations from the sample-based estimate to the population HC_{50} as a function of sample size – sample/population HC_{50} 97.5th percentile ratio (Pennington, 2003)

Sample size (*n*)	Log-logistic distributed	Log-triangular distributed	Log-normal distributed
1	-	-	-
2	231	266	231
3	134	242	168
4	53	88	68
5	44	87	69
6	27	46	38
7	25	45	36
8	18	29	26

The two approaches differ in terms of the assumed relationship between the distribution of the sample and of the actual population. The Student approach assumes that the distribution standard deviation is equivalent in the sample and in the population. Approaches considered in Pennington (2003) assumed that the standard distribution of the population was between likely limits observed from available toxicological insights, but not that it is equal to that of the sample.

So far, as both include necessary assumptions, the combination of the two estimates was considered to provide a more robust result compared to applying one or the other. In practice, this combination is applied by simply choosing the higher uncertainty estimate from the two approaches.

From Table 5, the choice of model is less important for the estimation of the geometric mean. The log-normal distribution is consistent with the approach adopted in IMPACT 2002 and reflects the middle estimate for the uncertainty for each sample size from these different common distributions. When compared with the Students approach, the estimate may be higher or lower for a given sample size depending on the distribution of the actual sample of test results (noting again that this is then assumed to be equivalent to the actual, or population, distribution).

In practice, the tendency is that the Student's approach typically estimates lower uncertainty values and therefore reliance is more on the values in Table 5. The estimates in Table 5 were based on the assumption of a maximum plausible standard deviation from observations. It could be argued that smaller plausible deviations could be adopted for certain chemicals based on their likely modes of action, although this is not without complexity and may not be justifiable in a practical context in LCA.

These calculations do not currently account for acute-to-chronic uncertainty, which is likely to be negligible according to both Payet & Jolliet (2004) and Pennington (2003), nor was additional uncertainty from using QSARs taken into account (where such data are adopted). Further research is required to account for the scenario and model uncertainty associated with these effect factor estimates, particularly in the context of estimating toxicological effects at likely low concentrations in the context of complex mixtures at regional scales that are relevant in LCA.

Applying Student's approach to the 1,1,2,2-Tetrachloroethane example results in a SD_g^2 estimate of 4.7 for the lower bound and 0.75 for the upper bound HC_{50}. As the HC_{50} is based on 10 species choosing the log-normal distribution from Table 5 yields an SD_g^2 of 26. A higher uncertainty is indicated by the latter approach. Applying the principle of Equations 3 and 4 to the HC_{50} results in the following interval:

$HC_{50\,97.5}$ = 8.01E-05

$HC_{50\,2.5}$ = 5.42E-02

This will be a slight overestimation since the HC_{50} is based on 10 species, while Table 5 only provided SD_g^2 estimates for up to 8 species. The effect factor (0.5/HC_{50}) is:

$EF_{aqu,\,97.5}$ = 6.24E+03

$EF_{aqu,\,2.5}$ = 9.23E+00

3. COMBINING TWO INDEPENDENT UNCERTAINTIES

Combining uncertainties is a vital step throughout the process of deriving a final overall uncertainty estimate for the model result (the characterisation factor in this case). It also provides a basis to quickly identify what is causing the highest uncertainty.

Uncertainty estimates for the human intake fractions for the fate and exposure part and the effect factors for the effects side, are readily combined when both uncertainties are log-normally distributed (without the need for further assumptions). Table 6 presents the results for the 1,1,2,2-Tetrachloroethane.

The geometric mean fate and exposure estimate is multiplied by the geometric mean effect factor, to estimate the overall characterisation factor (Equations 1 and 2). The two square geometric-standard deviations of the uncertainty distributions can be simply summed to calculate the overall uncertainty distribution's standard deviation, as outlined for example in the appendices of Hofstetter (1998). This overall uncertainty estimate can similarly be combined in a quantitative manner with those of the inventory data if they are presented in the form of log-normal uncertainty distributions.

To derive the overall uncertainty in the 1,1,2,2-Tetrachloroethane example the previously calculated uncertainties have to be combined. For the human health CF:

$$\begin{aligned} SD_g^2(CF) &= SD_g^2(iF) + SD_g^2(EF_h) \\ 13 &= 3 + 10 \end{aligned} \qquad (6)$$

In this particular case and using the values presented, the uncertainty of the effect factor is clearly dominating the overall uncertainty. Similar to Equations 3 and 4, the 95th percentile confidence interval bounds on the CF_h are then given by multiplying and dividing the geometric mean estimate by a factor of 13. There is 95% confidence that the CF_h is within a factor of 13 of the geometric mean):

$CF_{h,\,97.5}$ = 2.16E-05

$CF_{h,\,2.5}$ = 1.28E-07

The uncertainty of EF_{aqu} is dominating the overall CF_{aqu} uncertainty (as the combined fate and exposure uncertainty is a factor 3 compared to that of EF_{aqu} of 26).

Table 6: Illustrative IMPACT 2002 based geometric mean estimates accompanied by upper and lower bounds of the 95th percentile confidence interval for an emission of 1,1,2,2-Tetrachloroethane into European surface water.

(intermediate) Result	Lower bound	Geometric mean	Upper bound
Human health CF_h [*number of cases* /$kg_{emitted}$]	1.28E-07	1.66E-06	2.16E-05
Ecotoxicological CF_{aqu} [PAF $m^3 \cdot day^2/kg$]	1.13E+00	2.93E+01	7.62E+02
Human health eff. fact., EF_h [*number of cases*/kg_{intake}]	8.74E-03	8.74E-02	8.74E-01
Ecotox. aquatic eff. f., EF_{aqu} [PAF $m^3 \cdot day/kg$]	9.23E+00	2.4E+02	6.24E+03
Human Intake Fraction, *iF* [$kg_{intake}/kg_{emitted}$]	6.33E-06	1.90E-05	5.70E-05
Human exposure factor, *XF* [1/*day*]		1.56E-04	
Fate factor, *FF* [*day*]		1.22E-01	

While some insights have been provided in the literature related to uncertainty in LCA (Hofstetter, 1998; Hertwich et al., 1999; Hertwich et al., 2000; Huijbregts et al., 2000) and a working framework has been presented here for use in a relative comparison context, further research and development remain necessary for the quantification, but also to help take these uncertainties better into account in decision making.

4. REFERENCES

Bennett D.H. Margni M.D., McKone T.E., Jolliet O. (2002): Intake Fraction for Multimedia Pollutants: A Tool for Life Cycle Analysis and Comparative Risk Assessment. Risk Analysis, vol. 22 (5), 903-916.

Crettaz P., Pennington D.W., Rhomberg L., Brand K., Jolliet O. (2002): Assessing Human Health Response in Life Cycle Assessment using ED10s and DALYs - Part 1: Cancer Effects. Risk Analysis, vol. 22 (5), 931-946.

Gold L.S., Zeiger E. (1997): Handbook of carcinogenic potency and genotoxicity databases; ISBN 0-8493-2684-2, CRC Press.

Hertwich E.G., McKone T.E., Pease W.S. (2000): A systematic uncertainty analysis of an evaluative fate and exposure model. Risk Analysis 20: 439-454.

Hertwich E.G., McKone T.E., Pease W.S. (1999): Parameter uncertainty and variability in evaluative fate and exposure models. Risk Analysis 19: 1193-1204.

Hofstetter P. (1998): Perspectives in life cycle impact assessment: A structured approach to combine models of the technosphere, ecosphere and valuesphere. Kluwer Academic Publishers, Boston.

Huijbregts M.A.J., Thissen U., Jager T., van de Meent D., Ragas A.M.J. (2000): Priority assessment of toxic substances in life cycle assessment. Part II: assessing parameter uncertainty and human variability in the calculation of toxicity potentials. Chemosphere 41: 575-588.

IRIS (2004): US EPA's Integrated Risk Information System, available online at http://www.epa.gov/iris/.

Payet J. & Jolliet O. (2004): Comparative Assessment of the Toxic Impact of metals on aquatic ecosystems: the AMI method. In Life Cycle Assessment of Metals: Issues and research directions, A. Dubreuil Ed., SETAC Press, in Press.

Pennington, D.W.; Payet, J.; Hauschild, M. (2004): Aquatic ecotoxicological indicators in life cycle assessment. Environmental Toxicology & Chemistry, 23, in press.

Pennington D.W., Margni M., Payet J., Charles R., Jolliet O. (2003): Estimating cumulative toxicological risks and potential impacts for human health and ecosystems in LCA, submitted to Environmental Toxicology & Chemistry.

Pennington D.W. (2003): Extrapolating ecotoxicological measures from small data sets. Ecotoxicology and Environmental Safety, 56, 238-250.

Pennington D.W., Crettaz P., Tauxe A., Rhomberg L., Brand K., Jolliet O. (2002): Assessing Human Health Response in Life Cycle Assessment using ED10s and DALYs - Part 2: Non-Cancer Effects. Risk Analysis, vol. 22 (5), 947-963.

US EPA (1997): Exposure Factors Handbook, Vol. I-III. EPA/600/P-95/002Fa. Office of Research and Development, U.S. Environmental Agency, Washington.

An Integrated Approach for the Management of Uncertainty in Decision Making Supported by LCA-Based Environmental Performance Information

L. Basson[a,b] **and J.G. Petrie**[a,b]

[a] *Department of Chemical Engineering, University of Cape Town, Rondebosch, 7700, South Africa*

[b] *Department of Chemical Engineering, University of Sydney, NSW 2006, Australia*

basson@chemeng.uct.ac.za

Abstract: This paper presents an approach for the integrated consideration of both technical and valuation uncertainties during decision making supported by LCA-type environmental performance information. Key elements of this approach include "distinguishability analysis" to determine whether the uncertainty in the performance information is likely to make it impossible to distinguish between the activities under consideration, and the use of a multivariate statistical analysis approach, called principal components analysis (PCA), which facilitates the rapid analysis of large numbers of parallel sets of results, and enables the identification of choices that lead to similar and/or opposite evaluations of activities. The integrated approach for the management of uncertainty is demonstrated for a technology selection decision for the recommissioning of a coal-based power station. Distinguishability analysis showed that it was not possible to obtain a conclusive answer with regard to the preferred technology due to the extensive uncertainty in the LCA-based environmental performance information. PCA of the ranking of the design scenarios demonstrated that valuation uncertainties associated with choices made during intra- and inter-criterion preference modelling had a more significant effect on the ranking of the design scenarios than the inclusion/exclusion of environmental indicators reflecting local concerns or the choice of the position of the LCIA impact indicators in the cause-effect network. The results suggest that stakeholder involvement in intra- and inter-criterion preference modelling is important, and that the "encoding" of value judgements and preferences into LCA environmental performance information is to be avoided. As a whole, the paper supports a call for diversity in LCA methodology rather than one for greater standardisation, and provides a foundation for the consideration of the implications of such methodological diversity as part of an overall approach to promote effective decision making based on LCA environmental performance information.

Keywords: Decision Making; Valuation; Uncertainty; Multivariate Statistical Analysis

1. INTRODUCTION

Life Cycle Assessment (LCA) has gained recognition as a tool that can provide environmental performance information to support both decision making in the private and public sectors. The consideration and management of uncertainty is regarded as an essential part of good decision making practice. A distinction can be made between uncertainties that affect the estimation of the potential consequences of the activities under consideration and those that pertain to uncertainties in variables which are used for the evaluation of these consequences [Bonano, 1995]. It is suggested that these uncertainties be described as "technical uncertainties" and "valuation uncertainties", respectively. To date, methodological development for LCA has focused on the former, with less consideration being given to valuation uncertainties.

The evaluation of the (environmental) consequences of activities requires the modelling of value judgements, which inform both model structure and the choice of specific parameters in LCA models and in any decision models created to facilitate the comparative evaluation of the activities under consideration. A large number of choices of model structure and model parameters are made in both LCA and decision models. The consideration of the effects of all valuation uncertainties, in conjunction with technical uncertainties, can thus become an unmanageably large task. This paper presents a comprehensive and efficient approach for the integrated consideration of both technical and valuation uncertainties during decision making supported by LCA-type environmental performance information. The approach is illustrated in the context of a technology selection decision for the recommissioning of a coal-based power station.

2. AN INTEGRATED APPROACH FOR THE CONSIDERATION OF UNCERTAINTY IN DECISION MAKING

2.1 A Foundation in Multiple Criteria Decision Analysis (MCDA)

The integrated approach for the consideration of uncertainties in decision making has been developed to be used in conjunction with an approach to decision support based on multiple criteria decision analysis (MCDA). MCDA [e.g. Belton and Stewart, 2002] is a well-established approach to aid or support decision makers faced with complex decisions. It has been demonstrated that the structure of LCA has parallels with a decision analysis approach to decision making [e.g. Miettinen and Hämäläinen, 1997; Seppälä, 1999; Hertwich and Hammitt, 2001; Seppälä, et al., 2001]. However, more significantly from a decision making point of view, has been the use of MCDA to enable the consideration of trade-offs between environmental performance reflected by LCA-type indicators and performance pertaining to at least one other concern (e.g. financial, technical and/or social issues) and/or to provide an overall indication of the performance of a set of alternatives to facilitate decision making [e.g. Azapagic, 1996; Alexander, et al., 2000]. For discrete decision problems (i.e. ones in which a finite set of alternatives are considered), the use of MCDA methods based on multiple attribute value theory (MAVT) (also called "value function methods ") have been demonstrated to be of particular value to assists in the evaluation of a set of alternatives based on LCA-type environmental performance information [e.g. Miettinen and Hämäläinen, 1997; Seppälä, 1999].

2.2 Key Elements of the Approach

2.2.1 Placing Bounds on Particular Aspects

Uncertainties that are introduced during the problem structuring stage of the decision analysis process result from choices that define the scope of the decision process. These choices include the definition of the decision problem, the choice of stakeholders, the choice of temporal and spatial system boundaries for modelling of alternatives, and the choice of alternatives for consideration. The suggested approach for the management of these uncertainties is to attempt to achieve consensus amongst those involved in the decision process that appropriate bounds have been placed around the decision problem under consideration. Those who manage to decision process thus need to ensure that the necessary discussion takes place to enable the selection of appropriate bounds, and it is recommended that additional steps that call for the explicit verification of the choices made be introduced into the decision process.

2.2.2 Distinguishability Analysis

There would be little value in commencing with detailed preference modelling for the evaluation of the alternatives unless some of the alternatives are distinguishable from others despite the uncertainty in the performance scores. To ensure such "distinguishability", it is suggested that a step called "distinguishability analysis" be incorporated into the decision analysis process prior to any evaluation of the performance of the alternatives. Two approaches to distinguishability analysis have been identified, namely principal components analysis (PCA) [e.g. Jackson, 1991] and the "distinguishability index" (DI) approach developed for this purpose by Basson [2004]. In the latter, the potential ranges of the performances of the alternatives are compared in a pairwise manner and an assessment made of the number of aspects in which the performance ranges overlap due to the uncertainty in the performance information compared to the number of aspects in which the performance ranges do not overlap. In this manner it is possible to obtain an overall indication of whether any alternatives are completely distinguishable from any others, and, if not, in which aspects the alternatives are / are not distinguishable from one another.

2.2.3 Sensitivity Analysis

The implications of uncertainties introduced during the problem analysis elements of the decision analysis process, such as the choice of multiple criteria evaluation methods, associated preference models and parameters used in these models, need to be examined via sensitivity analysis. However, as a first pass, this effect of several of the model form- and model parameter- uncertainties may be investigated in a parametric manner to obtain an overall impression of their relative significance. (See Morgan and Henrion [1990] for parametric approaches for the consideration of model form uncertainties.) Finally, PCA can be applied to samples of the range of potential ranks of the alternatives to enable the graphical comparison the effect of the uncertainties examined through sensitivity analysis.

3. ILLUSTRATION OF THE OVERALL STRATEGY IN THE CONTEXT OF A TECHNOLOGY SELECTION DECISION PROBLEM

3.1 Problem Definition

A decommissioned power station belonging to an electricity supply company is to be recommissioned to meet a projected electricity demand. At issue for the recommissioning of the power station is whether it would be preferable to repower a station with fluidised bed combustion (FBC) boilers burning discard coal rather than to refurbish the existing pulverised fuel (PF) boilers which would burn run-of-mine coal?

The decision situation was originally considered by Notten [2001] in the context of a thesis which focused on data quality and the management of uncertainty in Life Cycle Inventory (LCI) information in resource-based industries. The case study decision presented here draws from Notten's work for the definition of alternatives for consideration and information on the environmental performance of these alternatives in terms of both LCI and LCIA indicators. However, the work extends that of Notten by framing the decision problem explicitly based on MCDA, which enables the consideration of financial and social aspects in addition to environmental aspects.

3.2 The Alternatives for Consideration

Notten proposed that the PF and FBC systems for consideration be defined by selecting particular combination of design variable choices and operating conditions, rather than attempting to evaluate all possible designs and operating conditions, or defining a discrete set of alternatives for consideration based on variations of a base case design. Notten defined three design scenarios per technology type in such a manner that they spanned extremes in environmental performance. These corresponded roughly to particular levels of refurbishment of the decommissioned power station. (See Notten [2001] for details regarding the definition of the design scenarios.)

3.3 The Evaluation Criteria and Performance Indicators

The aspects of concern to the company and to stakeholders that would inform the choice of preferred technology (i.e. the evaluation criteria) could be described broadly as financial, technical, environmental and social. The alternatives were designed to meet all technical requirements and indicators typically used by the company when considering the financial and social performance of designs were selected for the evaluation of the design scenarios. To allow for the comprehensive consideration of the environmental impacts of the technologies, Eco-Indicator 99 Hierarchist Perspective Life Cycle Impact Assessment (LCIA) indicators considered relevant to the decision context were selected (see Notten [2001] for details). Furthermore, to reflect specific concerns of relevance to the particular decision situation, two additional indicators were selected. These were an indicator for water consumption (considered important in the water scarce location of the decommissioned power station) and the affected land footprint (ALF) indicator developed by Hansen [2004], which provides a more comprehensive approach to the consideration of the impacts of solid wastes than the standard LCIA approach. The choice of evaluation criteria and the choice of specific performance indicators introduce valuation uncertainties into the decision process. The implications of such valuation uncertainties for decision making needed to be examined through sensitivity analysis (see section 3.7 later).

3.4 Data Collection and Enumeration of the Environmental Performance Scores

The environmental performance information was obtained using life cycle models (described in detail by Notten) and included information on the uncertainty in the inventory information based on the systematic evaluation of uncertainty associated with empirical parameters, model parameters and model structure. The values of the Eco-Indicator 99 LCIA indicators included the uncertainty in the equivalency factors used to calculate the impact potentials, which were available for most of the selected impact categories. Where such information was not present, estimates for the distributions of the characterisation factors were based on the work of Meier [1997]. To assist in the interpretation and preference modelling, the environmental performance information was normalised relative to average company performance for coal-based power generation. Uncertainty due to the variation in installed technologies and uncertainties associated with modelling the company's operation were included in the estimate of the average performance.

3.5 Distinguishability Analysis

The normalised performance scores were subjected to distinguishability analysis via both PCA and the DI approach. The PCA results can be summarised by way of PCA biplots. This is illustrated in Figure 1 for the criteria set where the environmental aspects were assessed in terms of the Eco-Indicator 99 impact level environmental criteria and the additional criteria reflecting local concerns.

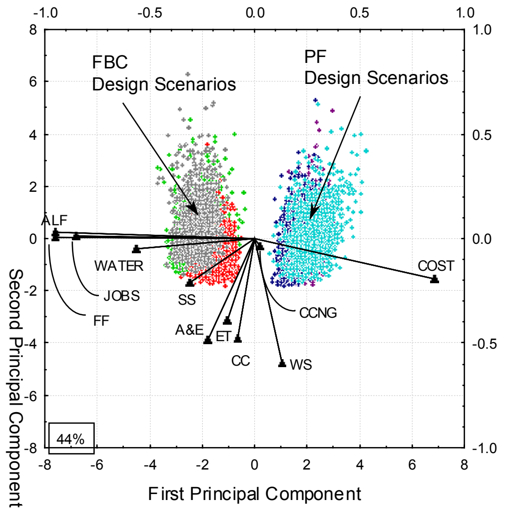


Figure 1 PCA biplot for the normalised performance scores

Codes: COST = Generating cost), JOBS =number of permanent jobs created, CCNG = carcinogenic effects on humans, SS = respiratory effects on humans caused by organic substances (summer smog), WS = respiratory effects on humans caused by inorganic substances (winter smog), CC = climate change, ET = ecotoxic emissions, A&E = combined effect of acidification and eutrophication, LO = land occupation, FF = extraction of fossil fuels, WATER = water use, ALF = affected land footprint

In Figure 1, the PF and FBC design scenarios form separate clusters, but the individual design scenarios overlap. This suggests that the technologies, but not the individual design scenarios, may be distinguishable from one another in at least some of the criteria. The arrows for the "principal factor loadings" associated with generating cost (COST), job creation (JOBS), fossil fuel consumption (FF) and the affected land footprint (ALF) are the longest and show the greatest correspondence to the axis representing the first principal component. This suggests that the two technology types differ most in these aspects. Furthermore, the principal factor loadings associated with generating cost and those associated with other three criteria lie almost diametrically opposite. The placement of the principal component scores for the two technologies suggests that the PF design scenarios perform better with regard to generating costs, while the FBC design scenarios perform better in terms of resource use and job creation. The key trade-offs to be considered for technology selection are expected to be between performance in generating cost and performance with regard to fossil fuel consumption, impacts on land (as represented by the AFL indicator) and job creation.

To verify the extent of distinguishability between the design scenarios, the DI approach was applied (results not shown). The DI results confirmed that the two technology types were distinguishable from one another with regard to generating cost, fossil fuel consumption, the ALF and job creation, but showed that the performance scores in the other criteria overlapped due to the uncertainty in the performance information. In general, it would thus have been recommended that an attempt be made to reduce the uncertainty in the performance information before commencing with the detailed modelling of preferences with the view to selecting a preferred technology type. However, in this particular case the uncertainty in the performance information was considered to be associated with empirical parameters (see Notten for details), and could thus be described in terms of probability distributions. The uncertainty in the performance information could therefore be propagated through the analysis and some indication obtained of the relative ranking of the design scenarios in full cognisance of the implications of the uncertainties in the performance information.

3.6 Evaluation of Design Scenarios Using Value Functions Analysis

The performance of the design scenarios considering all the financial, environmental and social performance indicators was compared using value function analysis. A weighted additive aggregation function was used to create an overall value score to reflect the relative preference for the different design scenarios. Valuation uncertainties were thus introduced during this stage of the decision process through the choice of model forms for the intra-criterion preference models (i.e. shapes of value functions) and the choice of model parameters for the inter-criterion preference models, (i.e. values for the weights).

3.7 Sensitivity Analyses

To determine the sensitivity of the ranking of the design scenarios to the valuation uncertainties introduced during the decision process (highlighted in sections 3.3 and 3.6), it was necessary to conduct several parallel evaluations of the design scenarios. Furthermore, by propagating the technical uncertainties (i.e. uncertainties associated with the performance information) through these parallel evaluations using a random sampling technique, it was possible to consider the implications for decision making of technical and valuation uncertainties simultaneously. The integrated approach for the consideration of technical and valuation uncertainties is demonstrated here by considering the evaluation of the design scenarios

- with and without the environmental criteria reflecting local concerns (sensitivity analysis for choice of evaluation criteria)
- using environmental indicators at inventory, impact (midpoint) and damage (endpoint) levels (sensitivity analysis for choice of performance indicators)
- using a combination of value function shapes and weights that place emphasis on good performance in environmental aspects and another combination that places emphasis on good performance in financial aspects (sensitivity analysis for intra- and inter-criterion preference models).

3.8 Results and Discussion

The overall ranking of the design scenarios, including the uncertainty associated these rankings, could be observed by plotting cumulative distribution functions for the overall value scores. This is illustrated in Figure 2 and Figure 3 for the value scores obtained using the criteria set illustrated in section 3.5 (Figure 1). Note that in Figure 2 and Figure 3 design scenarios with higher overall value scores are preferred to ones with lower value scores. The codes ER and SR correspond to "expected level" and "significant level" of refurbishment, respectively.

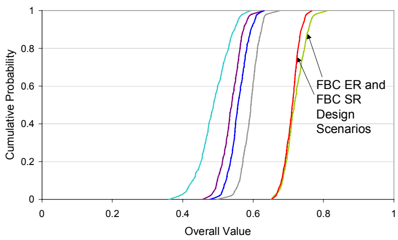


Figure 2 Overall value scores where emphasis is placed on good environmental performance

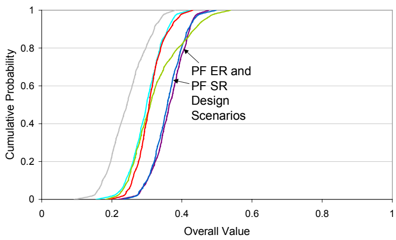


Figure 3 Overall value scores where emphasis is placed on good financial performance

Figure 2 and Figure 3 show that there is substantial overlap in the overall value scores of the design options. This is due to the uncertainty in the performance scores propagated through the analysis. However, despite this, it can be observed that when greater emphasis is placed on performance in environmental criteria, the FBC ER and FBC SR design scenarios are likely to be the preferred design scenarios (see Figure 2). If greater emphasis is placed on performance in the economic criterion, then the PF ER and PF SR design scenarios tend to perform better than the FBC design scenarios, but clear preference can not be established due to the uncertainty in the performance information (see Figure 3).

The rankings of the design scenarios are compared in Figure 4 and Figure 5, which present the principal factor loadings for the ranks of the design scenarios. Figure 4 compares the ranking of the alternatives with and without the additional environmental indicators, while Figure 5 considers the rankings for evaluations where the environmental criteria are represented by the Eco-Indicator 99 inventory (INV), impact (IMP) and damage (DAM) level indicators.

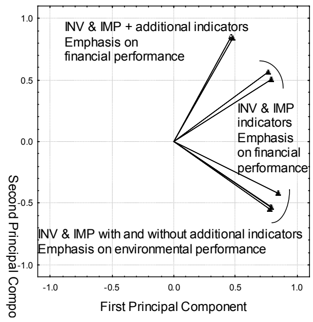


Figure 4 Principal factor loadings for the ranking of the design scenarios with and without the additional environmental indicators

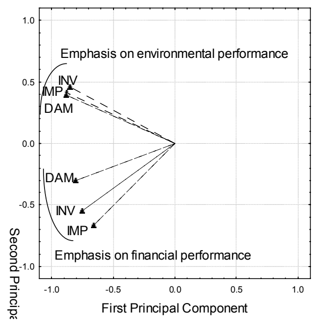


Figure 5 Principal factor loadings for ranking of the alternatives (environmental criteria represented by Eco-Indicator 99 inventory (INV), impact (IMP) and damage (DAM) level indicators)

The relative positions of the arrows representing the principal factor loadings of the ranks of the design scenarios for the parallel evaluations indicate that the relative preference of the design scenarios was sensitive to the valuation uncertainties associated with the choice of criteria (i.e. inclusion and exclusion of environmental criteria reflecting local concerns, Figure 4) and to the choice of the position of the indicators in the cause-effect network (assessment of environmental performance using inventory, impact and damage level indicators, Figure 5). However, in both cases, choices made during intra- and inter-criterion preference modelling had a more significant effect on the ranking of the design scenarios. For the case study decision, the results suggested that stakeholder involvement in intra- and inter-criterion preference modelling was essential.

4. GENERAL CONCLUSIONS

When making decisions it is necessary to ensure that the alternatives selected for further consideration/implementation are consistent with the value systems and preferences of the stakeholders. This requires the explicit modelling of stakeholder values and preferences rather than "encoding" of value judgements and preferences in the information used to support decision making. The latter may well be the effect of a call for greater standardisation of LCA methodology and a drive towards endpoint modelling which denies stakeholder an opportunity to explore their values and preferences and examined the effects of these on the decisions to be made. As a whole, this provides support for a call for diversity in LCA methodology rather than one for greater standardisation. The approach for the consideration of uncertainty presented in this paper provides a foundation for the consideration of the implications of such methodological diversity as part of an overall approach to promote effective decision making based on LCA environmental performance information.

5. ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial and practical support of Eskom, South Africa.

6. REFERENCES

Alexander, B., G. Barton, J. G. Petrie and J. A. Romagnoli, Process synthesis and optimisation tools for environmental design: methodology and structure, *Computers and Chemical Engineering*, 24 (2-7), 1195-1200, 2000.

Azapagic, A., *Environmental System Analysis: The Application of Linear Programming to Life Cycle Assessment*, PhD Thesis, Centre for Environmental Strategy, University of Surrey, Guildford, UK, 1996.

Basson, L., *Context, Compensation and Uncertainty in Environmental Decision Making*, Ph. D Thesis, Department of Chemical Engineering, University of Sydney, Australia, Sydney, 2004.

Belton, V. and T. J. Stewart, *Multiple Criteria Decision Analysis: An Integrated Approach*, Kluwer Academic Publishers, Dordrecht, 2002.

Bonano, E. J., Decision Making under Uncertainty and Risk Assessment, *Cost-Effective Risk Assessment for Process Design*, R. L. Deshotels and R. D. Zimmerman, Eds., McGraw-Hill, New York, 1995.

Hansen, Y., *Assessing the environmental impact of solid waste management in the primary Industries - a new approach*, PhD Thesis, Department of Chemical Engineering, University of Sydney, Sydney, 2004.

Hertwich, E. G. and J. K. Hammitt, A Decision-Analytic Framework for Impact Assessment. Part 1: LCA and Decision Analysis, *International Journal of Life Cycle Assessment*, 6 (1), 5-12, 2001.

Jackson, J. E., *A User's Guide to Principal Components*, John Wiley and Sons, New York, 1991.

Meier, M. A., *Eco-efficiency Evaluation of Waste Gas Purification Systems in the Chemical Industry*, Eco-Informa Press, 271, Bayreuth, Germany, 1997.

Miettinen, P. and R. P. Hämäläinen, How to benefit from decision analysis in environmental life cycle assessment (LCA), *European Journal of Operational Research*, 102 279-294, 1997.

Morgan, M. G. and M. Henrion, *A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, New York, 1990.

Notten, P. J., *The Management of Uncertainty in Life Cycle Inventories for Major Resource-Based Industries*, PhD Thesis, Department of Chemical Engineering, University of Cape Town, Cape Town, 2001.

Seppälä, J., *Decision Analysis as a Tool for Life Cycle Impact Assessment*, LCA Documents. Volume 4. Eco-Informa Press, 180, Bayreuth, Germany, 1999.

Seppälä, J., L. Basson and G. A. Norris, Decision Analysis Frameworks for Life-Cycle Impact Assessment, *Journal of Industrial Ecology*, 5 (4), 45-68, 2001.

Implications of Uncertainty and Variability in the Life Cycle Assessment of Pig Farming Systems

C. Basset-Mens, **H.M.G. van der Werf, P. Durand, Ph. Leterme**

INRA, UMR Sol Agronomie Spatialisation de Rennes-Quimper, 65, rue de Saint Brieuc CS 84215, 35 042 Rennes Cedex, France. Tel. +33 2 23 48 52 22; fax +33 2 23 48 54 30; email : basset@roazhon.inra.fr

Abstract: In a case study of pig production systems we propose a simple quantification of the uncertainty of LCA results (intra-system variability) and we explored inter-system variability in order to produce more robust LCA outcomes. Our quantification of the uncertainty took into account the variability of technical performance (crop yield, feed efficiency) and of emission factors (for NH_3, N_2O and NO_3) and the influence of the functional unit (FU) (kg of pig versus hectare used). For farming systems, the inter-system variability was investigated through differentiating by the production mode (conventional, quality label, organic (AB)), and farmer practices (Good Agricultural Practice (GAP) versus Current practice (CP)) while for natural systems, variability due to physical and climatic characteristics of catchments expected to modify nitrate fate was explored. For the eutrophication impact category, the variability of field emissions contributed more to uncertainty than the variability of building emissions, crop yield and feed efficiency. The influence of the FU on eutrophication results was very important when comparing systems with different degrees of intensification such as GAP and AB. Concerning inter-system variability, differences in farmer practices had a larger effect on eutrophication than differences between production modes. Finally, the physical characteristics of the catchment and the climate strongly affected the eutrophication result.

Keywords: uncertainty, variability, Life Cycle Assessment, pig production, eutrophication.

1. INTRODUCTION

The uncertainty of LCA outcomes stems from our attempt to convert the variability of the real world into LCA results through parameters, models and choices (Huijbregts, 1998). Treating uncertainty of LCA results represents a challenge at different levels. Currently, the analysis of the uncertainty in LCA studies, even if crucial, is rarely done (Ross, 2002), because of the lack of simple methods allowing its quantification. However, when data and knowledge are available, integrating a major part of the variability of both human and natural systems can reduce the uncertainty of LCA results. Concerning natural systems variability, the choice of the model used to convert emissions into regional impact such as aquatic eutrophication will strongly affect the result of assessments of the environmental performance of farming systems.

In a case study of pig production, we proposed an approach based on one reference LCA result, (i) to produce a simple quantification of its uncertainty, (ii) to explore the variability due to production systems (farmer practice and production mode) and due to nitrate transfer in the natural system (physical and climatic characteristics of catchments).

2. MATERIALS AND METHODS

2.1. Production systems

2.1.1. Production modes

This study dealt with the processes up to and including the production of pigs on the farm. Three contrasting production systems were defined. The Good Agricultural Practice (GAP) scenario corresponds to current intensive production (or "conventional" production), optimised in particular with respect to fertilisation practices, as specified in the French "Agriculture Raisonnée" standards (Rosenberg and Gallot, 2002). In the GAP scenario, pigs are raised at high density in a slatted-floor confinement building. The Agriculture Biologique (AB) scenario corresponds to organic agriculture according to the French version of the European rules for organic animal production (Ministère de l'Agriculture et de la Pêche, 2000) and the European rules for organic crop production (CEE, 1991). The Label Rouge

(LR) scenario corresponds to the Porc Fermier Label Rouge quality label (Groupements des fermiers d'Argoat, 2000). In the AB and LR scenarios pigs are born and raised outdoors until weaning, and in an open-front straw-litter building at low animal density after weaning.

Table 1. Characteristics of the animal production stage for the Good Agricultural Practice (GAP), Current Practice (CP), Label Rouge (LR) and Agriculture Biologique (AB) scenarios

	GAP/CP	LR	AB
Piglet production			
Weaned piglet/ sow/year	25.5	22.6	20.3
Weaning age, days	25.7	28	42
Feed per sow (boar included), kg/year	1313	1490	1695
Weaning to slaughtering			
Surface per pig, m^2	0.85	2.6	2.3
Feed to gain ratio	2.7	2.9	3.2
Slaughter age, days	175	190	195
Slaughter weight, kg	113	115	120
Feed consumed, kg	275	312	340

Data concerning resource use and emissions associated with the production and delivery of inputs for crop production (fertilisers, pesticides, tractor fuel and machines) were derived according to Nemecek and Heil (2001). Data for energy carriers for road and sea transport were from the BUWAL 250 database (BUWAL, 1996). Data concerning resource use and emissions associated with buildings (production and delivery of materials, construction) were from Kanyarushoki (2001). Data on crop production, transport distances, feed composition and system performance were based on statistics, estimates from experts and data from producers' associations.

For all crops, production corresponded to good agricultural practice, i.e. fertilisation according to anticipated crop needs and integrated pest management for GAP and LR. For the three scenarios, we assumed that pig manure (liquid manure for GAP, solid manure for LR, composted solid manure for AB) was used to fertilise Brittany-grown crops used as feed ingredients. For LR and GAP, yield levels were averages for 1996 – 2000 (AGRESTE, 2001; FAO, 2002). The yield levels of AB crops were according to experts from the region of production. For the processes concerning the transformation of crop products into feed ingredients and the production of feed, the inventory of resources used and emissions to the environment was limited to resources and emissions associated with the use of non-renewable energy. For ingredients resulting from processes yielding more than one product (e.g. soy cake, wheat gluten), resource use and emissions were allocated according to the economic value. Data for feed production (involving, amongst others: grinding, heating, mixing, pelleting) were from Sanders (2000).

For GAP, data on technical performance of the animal production stages (Table 1) were according to published statistics (ITP, 2001). For LR, data concerning piglet production (PP) were from ITP (2001), data concerning weaning to slaughtering production (WS) were averages supplied by the LR producers' association. For AB, data on technical performance were based on an optimised model of organic pig production (Berger, 2000) adjusted according to expert judgement. For GAP and LR, manure was stored, while for AB, manure was composted. Overall, GAP was more intensive than AB: higher feed efficiency, younger age at slaughter and less surface per pig. LR was intermediate between GAP and AB.

Ammonia emissions due to the application of ammonium nitrate fertiliser were estimated according to ECETOC (1994) and ammonia emissions following application of slurry were according to Morvan and Leterme (2001). Ammonia and nitrous oxide emissions from slurry in pig buildings were from IPCC (1996) and UNECE (1999). Methane emissions due to enteric fermentation and housing type were from IPCC (1996). For LR and AB, data on the production of excreta, emissions from buildings, during storage, during composting and from crops and paddocks, were chiefly obtained with the support of an expert panel from the Institut National de la Recherche Agronomique. The panel comprised: J. Y. Dourmad, Th. Morvan, J.M. Paillat, P. Robin and F. Vertès. The panel based its expertise on their experiments, simulation models and on their interpretation of the available literature.

For the four production scenarios, the contributions of the major emissions to eutrophication, acidification and climate change are summarised in Table 2.

Table 2. Contributions of the major emissions to eutrophication, acidification and climate change for GAP, CP, LR and AB scenarios and corresponding characterisation factors. Emissions are expressed in the unit indicated for each impact category.

Impact category (unit)	GAP	CP	LR	AB	Characterisation factors
Eutrophication (g PO_4 -eq /kg pig)					
NO_3	11	19.1	11.4	12.5	0.1
NH_3	8.3	8.5	3.6	6.1	0.35
NO_x (as NO_2)	0.95	0.94	0.95	1.71	0.13
PO_4	0.42	0.48	0.41	1.11	1
Total	20.8	29.3	16.6	21.6	
Acidification (g SO_2 -eq /kg pig)					
NH_3	37.8	38.7	16.4	27.9	1.6
NO_x (as NO_2)	3.65	3.61	3.67	6.57	0.5
SO_2	1.47	1.66	1.61	2.13	1.2
Total	43.5	44.9	22.6	37.2	
Climate change (g CO_2 -eq /kg pig)					
N_2O	964	1440	2150	2320	310
CO_2	882	950	1120	1390	1
CH_4	458	460	187	256	21
Total	2300	2850	3460	3970	

2.1.2. Farmer practice

In order to explore the influence of the farmer practice on the final result, a Current Practice scenario (CP) was defined with current fertilisation practice: fertilisation exceeded crop needs (Houben and Plet, 1997) for four of the major crops used as feed ingredients, leading to a three to four-fold increase of nitrate losses for these crops.

2.1.3. Uncertainty analysis

In order to explore the robustness of the GAP results, an uncertainty analysis was conducted. Crop yields, WS feed to gain ratio, field emissions (NH_3, N_2O and NO_3) and emissions of NH_3 and N_2O from buildings and manure storage were identified as important issues for the variability of results. For the parameters concerning these issues, a high and a low value reflecting what we coined "realistic" rather than overall variability were defined in addition to the default reference value. The "realistic" uncertainty interval thus defined contains about two thirds of the overall variability for the parameter concerned. In order to assess the relative importance of each of the four issues, we constructed for each issue favourable and unfavourable variants by combining on the one hand all favourable values and on the other hand all unfavourable values. The summation of the uncertainty sources quantified is proposed as an indicator of the uncertainty of the GAP results. Finally, in order to assess the influence of the choice of functional unit, impacts were expressed by two functional units corresponding to the two main functions of agricultural production systems. Kg of pig produced (live weight at slaughter) reflects its function as a producer of market goods, whereas ha of land used reflects its function as a producer of non-market goods (e.g. environmental services).

2.2. Natural systems

The natural context was considered for the transfer of nitrate in catchments through hydrological behaviour and rainfall. Three contrasting catchment types were selected (Table 3).

Table 3. Characteristics of the catchment types

	Geology	Wetlands (% of total surface)	Seasonal cycle
Type P	Granite	High (25%)	Reversed*
Type K	Schist	Middle (10%)	Normal*
Type S	Granite	Low (5%)	Normal*

*: Normal cycles present high nitrate concentration in winter and low in summer and conversely for reversed cycles.

Three levels of effective rainfall (resulting runoff of the rain falling on a catchment) were selected : 300, 435 and 700 mm. Thus, nine catchment scenarios were obtained by combining the three catchment types and the three levels of effective rainfall.

2.3. Evaluation methodology

2.3.1. Current LCA

The current LCA methodology was applied for seven impact categories. Only eutrophication, climate change and acidification are presented in this article. As recommended by Guinée et al. (2002), Eutrophication Potential (EP) was calculated using the generic EP factors in kg PO_4-eq., Global Warming Potential for a 100 year time horizon (GWP_{100}) was calculated according to the GWP_{100} factors by IPCC (Houghton et al., 1996) in kg CO_2-eq. and Acidification Potential (AP) was calculated using the average European AP factors by Huijbregts (1999) in kg SO_2-eq. (Table 2).

2.3.2. Eutrophication

Assessing the fate factor for nitrate in catchments (ratio between annual fluxes of N export from the catchment in the river and annual fluxes of N input in the catchment, namely leachable nitrate) require the quantification of the N retention capacity of the catchment, which is generally thought to be due to heterotrophic denitrification in the upper horizon of bottom land (Sebilo et al., 2003). This was done using the hydrology and biogeochemistry model INCA (Integrated Nitrogen in Catchments) (Whitehead et al., 1998). INCA is a semi-distributed and process-based model simulating the nitrogen fate through terrestrial systems and rivers. The model was calibrated against flow and chemistry data from the selected catchments. For the simulation of nitrate transport and the calculation of fate factors, estimated values for leachable nitrate based on historical fertilisation data were used. Leachable nitrate increased from 3.7 kg/ha (1965) to 93 kg/ha (2003), and it remained stable afterwards. This stabilisation of the nitrogen load allowed to reach an equilibrium state and to estimate the real fate factor for nitrate, eliminating the long term storage or release effect. The N retention capacity of the catchments was a function primarily of the percent of wetlands and secondly of the effective rainfall. The fate factors obtained ranged thus from 0.9 for the scenario crossing catchment type S and highest effective rainfall to 0.35 for the scenario crossing catchment type P and lowest effective rainfall. These fate factors were used to assess the eutrophication impact of the GAP scenario by multiplying with the generic EP factors.

2.4. Reference result

All the results were referred to one reference LCA result obtained by combining one production mode (conventional production mode), one level of farmer practice (good agricultural practice = GAP scenario), with average values for key parameters and the standard evaluation methodology, included aquatic eutrophication (fate factor for nitrate = 1).

3. RESULTS

Per kg of pig, the eutrophication result for GAP was: 0.0208 kg PO_4-eq while per ha it was 38.3 kg PO_4-eq. Both per kg and per ha, uncertainty was large (around ± 50%) and was mainly due to field emissions (around ± 35%) (Fig. 1). Both per kg and per ha, eutrophication was lower for LR (-20% and -30%, respectively). For AB, the result was very dependent on the choice of the FU: eutrophication was similar per kg but 40% less per ha. Eutrophication was 40% higher for CP than for GAP. Finally, when the fate factor was 0.9 or 0.35 instead of 1, the eutrophication result was reduced by 5 to 35% (Fig. 1).

The reference result for climate change was 2.30 kg CO_2-eq per kg of pig and 4236 kg CO_2-eq per ha. Both per kg and per ha, uncertainty intervals were very large and were mainly due to field emissions. Both per kg and per ha, climate change was higher for LR (+ 50% and +30%, respectively.). As for eutrophication, the climate change result for AB was very dependent on the FU: climate change was 70% higher per kg but was similar per ha. Finally, CP resulted in an increase of more than 20% of the climate change impact relative to the reference result. Per kg of pig, the reference result for acidification was 0.0435 kg SO_2-eq, while per ha it was 80.1 kg SO_2-eq. Both on a per kg and a per ha basis, uncertainty intervals for acidification were smaller than for eutrophication and climate change (± 30% and ± 20%, respectively).

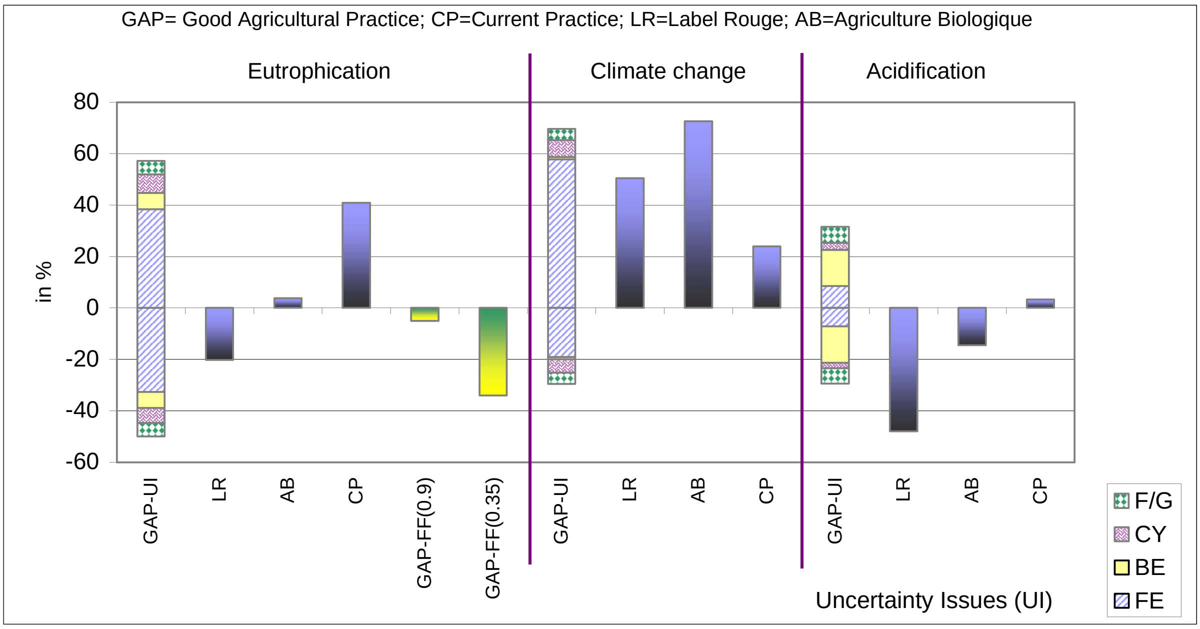


Fig 1. Uncertainty of LCA results (per kg of pig) for eutrophication, climate change and acidification. GAP-UI: Contribution of feed to gain ratio (F/G), crop yield (CY), building and manure emissions (BE) and field emissions (FE) to uncertainty for the GAP reference scenario. Other bars indicate differences for the LR, AB and CP scenarios and for 0.9 and 0.35 fate factors relative to the GAP reference scenario.

Uncertainty of the reference result was mainly due to building emissions and secondarily to field emissions. Both per kg and per ha, acidification was much smaller for LR: 48% less per kg and 55% less per ha. The difference between the reference result and AB depended on the FU once more: acidification results were close when expressed per kg (15% less for AB) (Fig. 1), while when expressed per ha, acidification was 53% less for AB. Finally, the CP scenario was similar to the reference result for acidification.

4. DISCUSSION AND CONCLUSION

Uncertainty for GAP (the reference) was large and originated primarily from the estimation of emission factors at the inventory stage. For eutrophication and climate change, field emissions were the main source of uncertainty while for acidification it was building emissions. The uncertainty of the results can reflect the real variability of the processes involved: for instance nitrate leaching is a function of soil characteristics and climate, but uncertainty can also arise from a lack of knowledge about these processes. Namely, for the emission of N_2O in the field due to nitrogen input we used the emission factor and the uncertainty interval proposed by Mosier et al. (1998), which are based on a literature review of field studies conducted in temperate regions of the world, with different fertiliser types, soils and climates. The large uncertainty range reflects the contrasting background conditions of the measurements. An approach is required, for instance the use of a suitable simulation model, which allows a more reliable estimation of emission factors (for N_2O, but also for NO_3 and NH_3) by assigning this variation to its controlling variables in order to produce an estimate that takes into account both environmental conditions (climate, soil…), farmer practices and technology used.

A practical methodology was defined to analyse the uncertainty of the reference scenario results. It could also be interesting to perform this uncertainty analysis using Monte Carlo simulation.

We explored the inter-system variability. The difference between LR and CP on the one hand and GAP on the other did not depend much on the FU used, because these systems present similar degrees of intensification whereas the difference between AB and GAP was very dependent on the FU.

For eutrophication, the difference between GAP and CP was larger than the difference between GAP and the other two systems (LR and AB). This result illustrates that farmer practices may affect the final result more than production modes.

LR and AB scenarios did better than or similar to GAP for eutrophication and acidification but they did worse for climate change. Eutrophication and acidification are considered as hot spots for the GAP scenario and even more for the CP scenario. However, this study reveals climate change as a hot spot for LR and AB. Basset-Mens and van der Werf (submitted) have demonstrated that the straw litter housing system was the main responsible for this hot spot, but also that this production stage seems to present important margins of improvement in this respect.

For eutrophication, the consequence of integrating the different catchment types in the analysis has been considered. Contrasting fate factors for nitrate were obtained by simulating nitrate transfer in nine catchment scenarios with the INCA simulation model. These fate factors ranged from 0.35 to 0.9 revealing potentially diverse and important N retention capacities for the catchments depending on their hydrology, the effective rainfall and the wetland surface. The reference result was reduced of 5% with a fate factor of 0.9 up to 35% with a fate factor of 0.35. These first simulations illustrate the importance of taking into account the environment where the pollutants are transfered in assessing aquatic eutrophication. These results complete the work of Huijbregts and Seppälä (2000; 2001) and demonstrate the need for further research on the simulation of the fate of pollutants in LCA models.

5. ACKNOWLEDGEMENT

This work is part of the research program “Porcherie verte” (“Green piggery”) and was financially supported by ADEME and OFIVAL.

6. REFERENCE

AGRESTE, Tableaux de l’Agriculture Bretonne 2001. Direction Régionale de l’Agriculture et de la Forêt, Service Régional de Statistique Agricole, Rennes, France, 2001.

Basset-Mens, C., van der Werf, H.M.G., Scenario-based environmental assessment of farming systems : the case of pig production in France, submitted to Agriculture, Ecosystems and Environment.

Berger, F., Les défis du porc biologique. Chambre régionale d’agriculture des Pays de Loire. Angers, France, 2000.

BUWAL, Ökoinventare für Verpackungen. Schriftenreihe Umwelt Nr. 250/1+2, Bundesamt für Umwelt, Wald und Landschaft, Bern, Switserland, 1996.

CEE, Règlement du conseil (CEE) n°2092/91 concernant le mode de production biologique de produits agricoles et sa présentation sur les produits agricoles et les denrées alimentaires. Journal Officiel de la communauté européenne L198 (22 juillet 1991). 1, 1991.

ECETOC, Ammonia emissions to air in western Europe. Technical report no. 62. European Chemical Industry Ecology & Toxicology Centre, Brussels, Belgium, 1994.

FAO, FAOSTAT Agriculture Data. Available at : http://apps.fao.org/, 2002.

Groupement des fermiers d’Argoat, Cahier des charges éleveur – porcs fermiers élevés en plein air. Homologation n°20-88. Saint-Brieuc, France, 2000.

Guinée, J.B., Gorrée, M., Heijungs, R., Huppes, G., Kleijn, R., de Koning, A., van Oers, L., Wegener Sleeswijk, A., Suh, S., Udo de Haes, H.A., de Bruijn, H., van Duin, R., Huijbregts, M.A.J., Life cycle assessment. An operational guide to the ISO standards. Centre of Environmental Science, Leiden University, Leiden, The Netherlands, 2002.

Houben, V., Plet, P., Elaboration des programmes d'action de la directive nitrate en Bretagne. Eléments de diagnostic et de préconisations. Volume 4 : La gestion de l’azote par système d’exploitation. Chambre d’Agriculture de Bretagne, Rennes, France, 1997.

Houghton, J.T., Meira Filho, L.G., Callander, B.A., Harris, N., Kattenberg, A., Maskell, K., Climate change 1995: the science of climate change. Cambridge University Press, Cambridge, United Kingdom, 1996.

Huijbregts M.A.J., Application of Uncertainty and Variability in LCA. Part 1 : A General Framework for the analysis of Uncertainty and Variability in Life Cycle Assessment. Int. J. LCA 3 (5), 273 – 280, 1998.

Huijbregts, M.A.J., Life-cycle impact assessment of acidifying and eutrophying air pollutants. Calculation of characterisation factors with RAINS-LCA. Interfaculty Department of Environmental Science, Faculty of Environmental Science, University of Amsterdam, Amsterdam, The Netherlands, 1999.

Huijbregts, M.A.J., Seppälä, J., Life Cycle Impact Assessment of Pollutants Causing Aquatic Eutrophication. Int. J. LCA 6(6), 339-343, 2001.

Huijbregts, M.A.J., Seppälä, J., Towards Region-Specific, European Fate Factors for Airborne Nitrogen Compounds Causing Aquatic Eutrophication. Int. J. LCA 5(2), 65-67, 2000.

IPCC, Intergovernmental Panel on Climate Change. Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories: Reference Manual (3). Available at: http://www.ipcc-nggip.iges.or.jp/public/gl/invs6.htm, 1996.

ITP, Résultats Gestion Technico-Economique. ITP, Paris, France, 2001.

Kanyarushoki, C., Evaluation de l'impact environnemental des bâtiments porcins – Application de la méthode d'analyse du cycle de vie. Mémoire de fin d'études DESS. Université Rennes 1, INRA, Rennes, France, 2001.

Ministère de l'Agriculture et de la Pêche, Cahier des charges concernant le mode de production et de préparation biologique des animaux et des produits animaux définissant les modalités d'application du règlement CEE n° 2092/91 modifié du conseil et/ou complétant les dispositions du règlement CEE n° 2092/91. Homologué par l'arrêté interministériel du 28 août 2000. Journal Officiel de la république française (30 Août 2000). Paris, France, 2000.

Morvan, Th., Leterme, Ph., Vers une prévision opérationnelle des flux d'azote résultant de l'épandage de lisier : paramétrage d'un modèle dynamique de simulation des transformations de l'azote des lisiers (STAL). Ingénieries 26, 17 – 26, 2001.

Mosier, A., Kroeze, C., Nevison, C., Oenema, O., Seitzinger, S., van Cleemput, O., Closing the global N2O budget: nitrous oxide emissions through the agricultural nitrogen cycle. Nutrient Cycling in Agroecosystems 52, 225-248, 1998.

Nemecek, T., Heil, A., SALCA – Swiss Agricultural Life Cycle Assessment Database. Version 012, December 2001. FAL, Swiss Federal Research Station for Agroecology and Agriculture, Zuerich, Switserland, 2001.

Ross, S., Evans, D., and Webber M., How LCA Studies Deal with Uncertainty. Int. J. LCA 7 (1) 47 – 52, 2002.

Rosenberg, P.E., Gallot, J., Référentiel de l'agriculture raisonnée. Arrêté du 30 avril 2002 relatif au référentiel de l'agriculture raisonnée. Journal Officiel de la République Française 104 (4 mai 2002), 8519, 2002.

Sanders, J., L'evaluation des systèmes de production animale: l'Analyse de Cycle de Vie peut-elle servir à construire des indicateurs environnementaux pour l'alimentation porcine ? INRA, UMR SAS, Rennes, France, 2000.

Sebilo, M., Billen, G., Grably, M., Mariotti, A. Isotopic composition of nitrate-nitrogen as a marker of riparian and benthic denitrification at the scale of the whole Seine River system. Biogeochemistry, 63(1), 35-51, 2003.

UNECE, Control options/techniques for preventing and abating emissions of reduced nitrogen compounds. EB.AIR/WG.5/1999/8United Nations Economic and Social Council, Economic Commission for Europe. Geneva, Switserland, 1999.

Whitehead, P.G., Wilson, E.J., Butterfield, D., A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA) : Part I – model structure and process equations. The Science of the Total Environment 201/211, 547-558, 1998.

Enhanced Presentation and Analysis of Uncertain LCA Results with Principal Component Analysis

P.J. Notten[a] **and J.G. Petrie**[b]

[a] *Dept. of Chemical Engineering, University of Cape Town, South Africa; now at: 2.-0 LCA consultants, Amagertorv 3,2, 1160 Copenhagen K, Denmark, pin@lca-net.com*

[b] *Dept. of Chemical Engineering, University of Sydney, Australia*

Abstract: A significant challenge of an uncertainty assessment is the presentation of the results, since a quantitative uncertainty analysis dramatically increases the already considerable amount of data that needs to be communicated in an LCA study. This paper investigates three graphical options to interpret output samples from quantitative uncertainty analyses. The output samples are from case studies within the coal-fired power generation sector, and include an assessment of empirical uncertainty from a stochastic uncertainty assessment and an assessment of uncertainty in decision variables from a parametric sensitivity analysis. Two commonly used representations of probabilistic samples are demonstrated, namely "box and whisker" plots and plots of the cumulative probability density function, as well as the multivariate geometric technique, principal component analysis (PCA). Cumulative probability plots are useful representations of uncertainty where a quantitative estimate of the relative uncertainty between options is required, but they become extremely tedious (many pair-wise combinations) and difficult to interpret when a large number of options are compared over many criteria. In such cases, PCA can be used to provide a valuable overview of the results, where it is able to clearly present any trade-offs that have to be made between selection criteria, and the "spread" of the options under consideration over the decision space. Box and whisker plots are good at representing the relative importance of empirical parameter uncertainty and the uncertainty arising from the choice of decision variables, and show the degree of shifting between the options as well as the full range over which the options potentially act. The three representations of uncertainty are found to complement each other, as each enhances different aspects of the results. The most appropriate graphical presentation method is found to depend on the particular decision context and the particular stage of the analysis.

Keywords: Stochastic results; Uncertain LCA results; Principal Component Analysis; Presentation

1. INTRODUCTION

A quantitative analysis of the uncertainty is becoming an increasingly accepted component of a life cycle assessment (LCA) study. The use of stochastic models and the presentation of results in ranges or as confidence intervals have been shown by a number of authors to enhance the decision support capabilities of an LCA study (e.g. Meier [1997]; Maurice et al. [2000]). Nonetheless, considerable challenges remain with regard to incorporating quantitative uncertainty analyses into life cycle assessment [Notten, 2002], notably with characterising the uncertainty of the input parameters, modelling correlated inputs, and analysing the considerable volume of data resulting from the analysis. This paper focuses only on the latter aspect of the uncertainty analysis.

Presenting and analysing the large data sets resulting from an LCA study is already a demanding process. These demands increase considerably when the results are extended to include a consideration of uncertainty, with each data point replaced by an output sample, as well as an increase in the number of scenarios requiring consideration. This paper therefore looks at ways of analysing and communicating uncertain results, including a novel technique using principal component analysis (PCA). The paper demonstrates the use of the methods with respect to a case study looking at technology options in coal-fired power generation, and draws conclusions as to the relative strengths of the different methods.

2. GRAPHICAL REPRESENTATIONS OF UNCERTAINTY

2. 1 Commonly used statistical methods

Uncertain data are often defined using common statistical measures, such as the variance, confidence intervals etc. Whilst such statistics are efficient at summarising the uncertainty of the data sample, graphical methods are typically the most effective in communicating insights into uncertain data sets. Three basic ways of presenting probabilistic results are:

- The probability density function (PDF),
- The integral of the PDF, the cumulative density function (CDF), and
- Displaying selected fractiles, as in box and whisker plots.

Examples of these commonly used graphical representations of uncertainty can be found in the following case study section. Each of the representations emphasise different aspects of the probability distribution of the output sample. PDFs give the relative probabilities of the different values and show the shape of the distribution. Box and whisker plots emphasise confidence intervals and means, and are thus a simple way to represent uncertain results when the range and not the distribution shape is primarily of interest.

The CDF also provides little information on the shape of the distribution, but is the best option if information on the fractiles of the distribution is required (i.e. the probability that the actual value of the variable is less than a particular value). The CDF is often preferred to the PDF when presenting stochastic output samples, because it looks a lot less noisy with the equivalent sample size. Comparative studies often share a number of sub-processes for which identical data have been used, with the resultant correlation between the output samples removed by basing the analysis on the normalised difference between the output samples [Coulon et al., 1997; Meier, 1997]. This has the added benefit of producing an easy to interpret CDF plot, in that the y-intercept of a CDF plot of the difference between two options shows the degree of confidence that can be held that the one option always performs better than the other.

A significant draw-back of CDF plots is that they are limited in the number of dimensions they can display. Conclusions have to be drawn across a large number of single plots if many options are to be assessed across many different environmental indicators. It is therefore difficult to get an overview of the results. A potential solution is to use the multivariate data analysis technique principal component analysis (PCA), which is able to reduce the dimensionality of the data set by producing a planar view of the data. The use of PCA for the presentation and analysis of LCA results has been demonstrated by Le Teno [1999], and its use explored in a number of complex case study situations in Notten [2002]. The following section discusses the main features of PCA with respect to its use in LCA.

2. 2 Principal component analysis

The goal of PCA is to represent the variation present in many variables in a small number of factors (or principal components), which are found via a mathematical manipulation of the data matrix. A new space in which to view the data is constructed by redefining the axes using these factors, instead of the original variables. The new axes allow the analyst to view the true multivariate nature of the data in a relatively small number of dimensions, allowing the identification of structures in the data that were previously obscured. The principal components are the eigenvalues of a correlation or covariance matrix of the input data, whilst the co-ordinates of the transformed variables on the principal component plane are given by the eigenvectors. The theory of principal component analysis can be found in most multivariate data analysis textbooks, e.g. Murtagh and Heck [1987].

The results of a PCA are best analysed graphically (see Figure 1). Stochastic output samples plot as clouds of points, which can be interpreted as "zones of confidence" [Le Téno, 1999] (see Figure 1). In this figure it is the structures or patterns in the data that are of interest, where the distances between the clouds of points determine similarities (or differences) between the options, and the overlap between the clouds visually identifies the significance of the rankings between the options. The axes do not have any physical meaning, and are merely measures of proximity that are interpreted as similarity. It is thus only the relative distance between the clouds of points, and their size and degree of overlap that are of note.

To interpret the principal component plot it is necessary to look at the factor-variable correlations given by the eigenvectors. These provide a measure of each variable's contribution to the principal components, and indicate which variables are best at discriminating between the options under investigation. In Figure 1, the eigenvectors are represented by the lines emanating from the origin, and their length and orientation indicate which variables have the greatest influence in “pulling” the data apart to create the spatial arrangement of the clouds of points. The eigenvectors also provide insights into the data structure. Highly correlated variables plot close together, thereby pointing to redundant selection criteria, whilst the relative lengths of the lines provide a measure of the relative ability of the variables to discriminate between the options (e.g. impact categories showing no significant differences between the options plot with short lines). Thus an analysis of the eigenvectors identifies the minimum set of impact categories useful for distinguishing between the options.

PCA is based on correlations between the data points, and thus finds where the greatest variations are occurring between the options, not where the largest absolute changes in indicator scores occur. It therefore does not provide information on the relative importance of the potential impacts. This also means that there is no need to correct for the problem of common data elements by basing the analysis on a difference between samples. However, it may be necessary to normalise the data so that the analysis is not skewed by variables operating on very different scales.

The following section demonstrates these features of PCA in a case study, and shows how it complements an analysis using other common representations of uncertain output samples.

3. CASE STUDY

The following case study is an excerpt from a larger study looking at technology options to refurbish coal-fired power plants in South Africa [Notten, 2002]. In particular, possibilities to utilise discard coal, a waste product from coal beneficiation, are investigated. An average year’s performance is chosen as the functional basis for comparison, because of the need to capture those environmental interventions (notably those from solid waste dumps), that only become evident after the plant has been operating for a number of years. The study includes all major processes in the coal-electricity supply chain (mining, coal preparation, coal combustion, flue gas cleaning and solid waste disposal), as well as the production and transport of major ancillary materials (liquid fuels, treatment chemicals etc.).

A rigorous assessment of uncertainty is undertaken in the study, including a parametric sensitivity analysis that systematically investigates model parameter uncertainty (i.e. the choice of operating conditions for the technology options), as well as a probabilistic assessment of empirical uncertainty. An iterative method based on successively refining the distributions of the most influential input parameters is followed, where the probability distributions take into account the certainty and completeness of the data source used to define the parameter, as well as an assessment of the adequacy of the data source in the particular context of the case study. Full details of the uncertainty assessment framework developed in the study can be found in Notten [2002].

The primary aim of this paper is to demonstrate methods for the presentation and analysis of uncertain results. Thus, in order to clearly present the extensive results of this case study in a suitable format (i.e. small, black and white graphics), it has been necessary to reduce the number of options considered, as well as to reduce the scatter of the uncertainty samples (although conclusions are still based on the full results). The results presented here are therefore generated from a random re-sampling of the interquartile range of the output distributions. Also, the results are normalised with respect to a "base case" option (the technology currently in operation), and presented on a relative scale on which the "base case" is assigned a score of 100 for each impact category considered (i.e. a score of less than 100 represents an improvement relative to the currently employed technology, whilst a score greater than 100 represents a higher contribution to that particular impact potential).

Figure 1 presents the results of a principal component analysis on the uncertainty samples of the option sets summarised in Table 1. The principal component "loadings" (the eigenvectors) and the principal component "scores" (the transformed data points) are calculated using standard statistical algorithms (available in most statistical software packages, see Notten [2002] for a summary of the underlying calculations).

Table 1. Primary decision variables causing the observed differences between the options (four plant configurations with fluidised bed combustion (FBC), and one with pulverised fuel (PF) combustion).

Option	PF	FBC A	FBC B	FBC C	FBC D
Boiler	PF	FBC	FBC	FBC	FBC
Cooling	wet	wet	wet	dry	wet
Fuel	coal	discard	discard	discard	blend
Sorbent	-	lime	dolomite	lime	dolomite
SO_2 removal	-	90%	40%	90%	40%

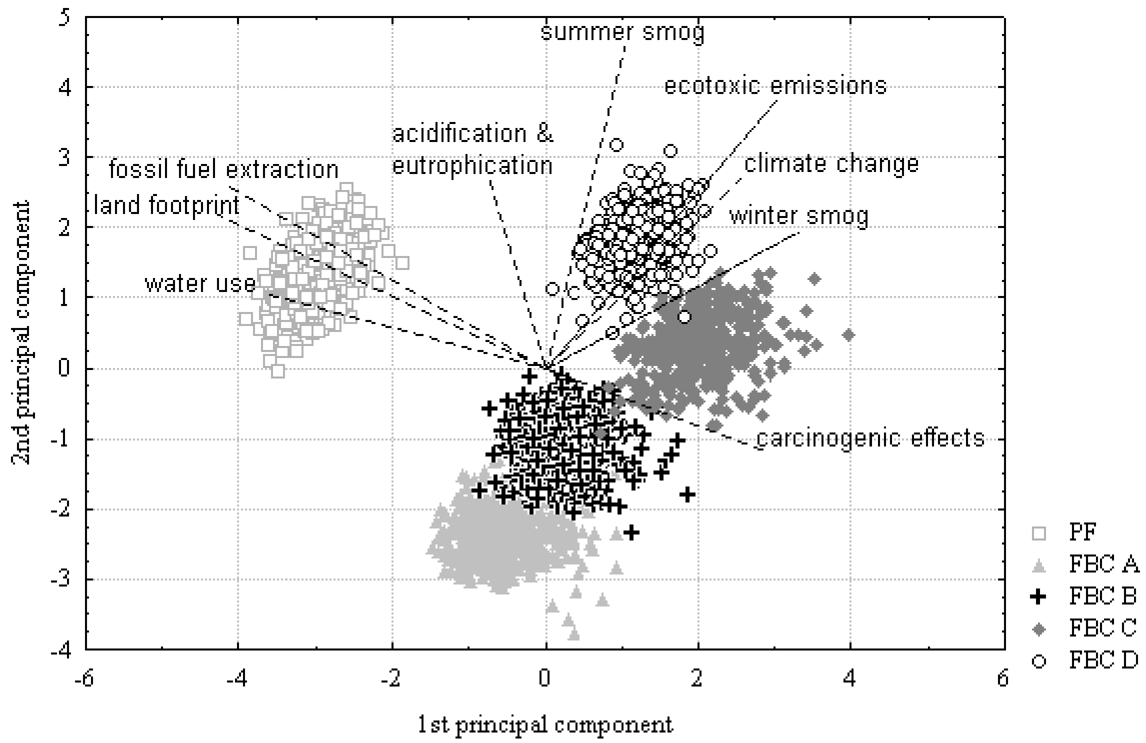


Figure 1. Normalised output samples transformed onto the 1st and 2nd principal component plane (points), with the eigenvectors or principal component "loadings" superimposed (labelled dashed lines).

In Figure 1 the output samples are shown transformed onto the principal component plane (i.e. placed in the new multivariate space defined by the principal components instead of the original variables), together with the relative contributions of each variable to the 1st and 2nd principal components (the lines emanating from the origin). If the lines are thought of as arrows, sample values plotting strongly in the direction of the arrow indicate a poor performance against that criterion, and the reverse for options plotting away from the direction of the arrow (for all indicators considered, a lower contribution means better performance). The relative length of the line indicates the strength of the observed difference between the options. For example, Figure 1 shows there is a certain and substantially higher contribution to fossil fuel extraction by the pulverised fuel (PF) option than the options with fluidised bed combustion (FBC). Whilst among the FBC options, there is a less certain and smaller observed difference between option D and the other FBC options (indicating its slightly higher contribution to fossil fuel extraction).

PCA captures the major sources of variability between the options, and shows which impact indicators are responsible for the observed variations. Although PCA efficiently summarises all sources of variability, Figure 1 only shows the first two principal components. Thus when interpreting Figure 1 it is important to bear in mind the sources of variance not captured by the first two principal components. From Table 2 it can be seen that the first two principal components, accounting for 63% of the variability, capture the large differences exhibited between the PF option and the three FBC options, and the greatest sources of variability between the FBC options (contributions to summer smog and ecotoxicity).

Table 2. Percentage contributions of the first four principal components to the overall variance, and the percentage contributions from the variables to these four principal components.

	PC 1	PC 2	PC 3	PC 4
% contribution to overall variance	35	28	19	7.5
Carcinogenic effects	10	2.1	2.9	85
Summer smog	1.2	31	1.2	0.7
Winter smog	13	5.6	20	3.0
Climate change	7.5	11	15	0.1
Ecotoxic emissions	11	21	1.1	0.1
Acidification and eutrophication	0.6	11	35	0.1
Impacted land footprint	21	7.0	4.5	5.6
Fossil fuel extraction	20	10	1.5	4.9
Water use	15	1.6	19	0.6

The difference in SO_2 removal efficiency between the options causes a different ranking between the FBC options to that found by the 1st principal component. This is responsible for the relatively high contribution to the overall variance by the 3rd principal component, whilst the contribution by the 4th principal component is a result of the very high uncertainty in predicting carcinogenic emissions (i.e. variability within rather than between the output samples).

A fairly extensive overlap between the output distributions is evident in Figure 2. Thus a different representation is required to determine the degree of confidence that can be held in the observed differences. For example, before selecting option A, a decision maker may wish to know how certain he/she can be that option A does indeed have a lower contribution to winter smog than option C. A CDF plot of the difference between their output samples provides this (see Figure 3), where the y-intercept gives the degree of confidence that can be held in the observed difference (shown by the arrow in Figure 3). In this case, there is a high degree of confidence in the relative performance of the options, with 94% of the sample predicting option C to have a higher contribution to winter smog than option A (i.e. their difference is less than zero for only 6% of the sample values). The precise representation of uncertainty in Figure 3 is extremely useful, but since it requires a pairwise comparison of a single indicator at a time, it was first necessary to narrow the option set under consideration.

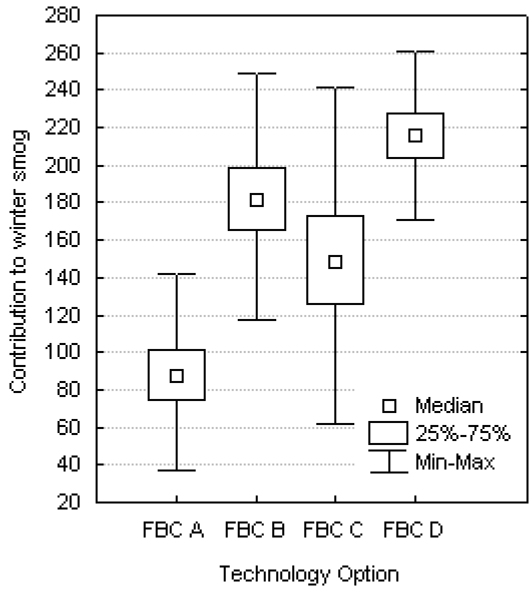


Figure 2. Contributions of the four fluidised bed combustion options to winter smog, relative to a "base case" option with a contribution of 100.

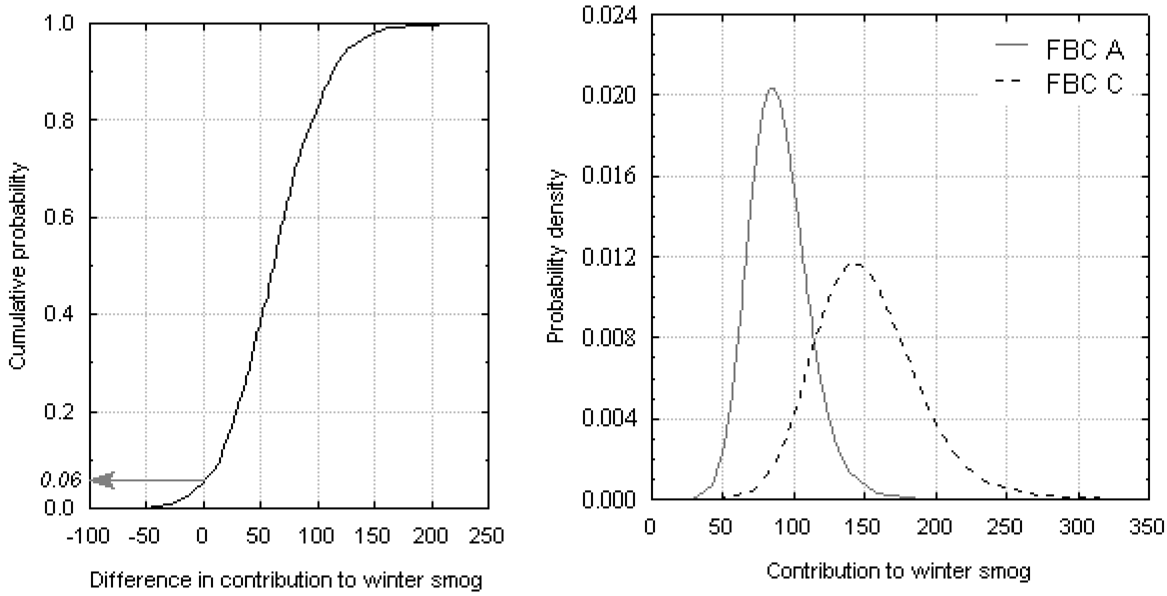


Figure 3. Cumulative probability function (CDF) of the difference between options A and C, and the probability density functions of these two options. The y-intercept of the CDF (shown by the arrow) gives the degree of confidence that option C performs better than option A.

4. CONCLUSIONS

Principal component analysis provides a valuable overview of LCA results, capable of highlighting where the most significant differences between options are occurring, and the impact categories responsible for the differences. This is extremely useful when a large number of scenarios need to be considered, giving a quick indication of the significance of the scenarios with respect to each other. PCA is thus most useful in a strategic type assessment, where a large number of decision variables are likely to require assessment as part of the uncertainty analysis. Even in studies considering a smaller option set, PCA still provides valuable assistance by making explicit the tradeoffs between the impact potentials. Furthermore, PCA provides useful information on the underlying structure of the result sample, particularly with respect to the strength and independence of the criteria chosen to evaluate the systems. However, PCA is best used in combination with other graphical representations of probabilistic samples. The three methods explored in the case study (PCA, box and whisker plots and cumulative probability plots) are found to complement each other, as each is able to enhance different aspects of the result sample.

The principal component representation allows a visualization of the trade-offs that have to be made between impacts, and the spread of the options over the operating space. Box and whisker plots are good at representing the relative importance of decision variable and empirical parameter uncertainty, whilst CDF plots are useful when a quantitative estimate of the relative uncertainty between options is required (i.e. the degree of confidence in the observed difference between the options). Whilst giving the most precise information, CDFs become extremely tedious and difficult to interpret when a large number of options are involved (many pair-wise combinations). In this case, PCA is invaluable, as it enables the full data set to be displayed on a single plot (provided a sufficiently high percentage of the overall variance is displayed by the first two principal components). Box and whisker plots provide a level of information between these two methods. The three representations of uncertainty are therefore most useful when used successively to narrow the option set.

REFERENCES

Coulon R., V. Camobreco, H. Teulon, and J. Besnainou, Data Quality and Uncertainty in LCI, International Journal of Life Cycle Assessment, 2(3), 178-182, 1997.

Le Téno J-F., Visual Data Analysis and Decision Support Methods for Non-Deterministic LCA. *International Journal of Life Cycle Assessment*, 1(4), 41-47, 1999.

Maurice, B., R. Frischknecht, V. Coelho-Schwirtz, and K. Hungerbühler, Uncertainty Analysis in Life Cycle Inventory. Application to the Production of Electricity with French Coal Power Plants, *Journal of Cleaner Production*, 8, 95-108, 2000.

Meier M.A., Eco-Efficiency Evaluation of Waste Gas Purification Systems in the Chemical Industry, LCA Documents, Vol. 2, Eco-Informa Press, Bayreuth, 1997.

Murtagh, F., and A. Heck, Multivariate Data Analysis, Reidel Publishing Company, Dordrecht, 1987.

Notten, P., Life Cycle Inventory Uncertainty in Resource Based Industries - A Focus on Coal-Based Power Generation, Department of Chemical Engineering, University of Cape Town, 2002.

A Review of Approaches to Treat Uncertainty in LCA

Reinout Heijungs[a] **and Mark A.J. Huijbregts**[b]

[a] *Institute of Environmental Sciences, Leiden University, Leiden, The Netherlands*

e-mail: heijungs@cml.leidenuniv.nl

b *Department of Environmental Studies, University of Nijmegen, Nijmegen, The Netherlands*

Abstract: It is important to know to what extent the outcome of an LCA is affected by various types of uncertainty, such as parameter, scenario and model uncertainty. These types may occur in the goal and scope definition, the inventory analysis and the impact assessment of an LCA. Information on the uncertainty of the model outcomes provides useful information to assess the reliability of LCA-based decisions and to guide future research towards reducing uncertainty. This paper reviews several approaches to treat different types of uncertainty in LCA. It will discuss the typology of uncertainty that may be encountered in LCA, the qualitative and quantitative techniques that are available to address these uncertainties, the inclusion of these techniques in LCA software tools, the (graphical) possibilities to show uncertainty in LCA outcomes, ways to simplify the uncertainty analysis, the inclusion of uncertainty analyses in case studies and (the difficulties in) the interpretation of uncertainty information.

Keywords*:* Uncertainty, Life cycle assessment; LCA

1. INTRODUCTION

If LCA is supposed to play a role in environmental decision-making, the quality of the decision-support should be made clear. It is natural that there is an interest of decision-makers and LCA-experts in the credibility of the results of LCA. In fact, it is amazing that this interest has not been natural since the development of LCA and the rise of its use. Although concerns about the quality of LCA-results have been raised at an early stage of LCA-development, assessment of this quality is still not a standard feature, and a systematic and comprehensive treatment is still lacking in most guidebooks, databases and software for LCA. Witness of this is the several guidebooks for LCA where the assessment of uncertainty is postponed and presented as some sort of additional feature. One still sees LCA case studies with bar charts showing that product A is 0.3% better than product B, without any indication of the significance or robustness of this difference. A decision-maker then has to figure out whether or not the difference of 0.3% is in any sense significant. The situation is even more complex due to the fact that writers of LCA-reports sometimes use the technical term “significant” when they have the intuitive term “large” in mind; in fact, some writers seem to think that “significant” is the scientific-sounding equivalent of “large”. The idea of statistical analysis is not always part of the standard vocabulary of practitioners and users of LCA.

There is, however, also good news. There have been quite a few initiatives and developments towards including uncertainty in LCA. Statistical uncertainty information is to an increasing extent percolating into methods, databases and software, and are increasingly being applied in case studies. Decision-makers increasingly recognize that uncertainties are important and should be made explicit.

The importance of including uncertainty in LCA has been long recognized. Already in 1992, a SETAC-workshop focussed exclusively on this topic (Fava *et al.*, 1993). During these years, the discussion was largely restricted to acknowledging the possible prohibitive effects of uncertainty and the setting-up of schemes for data quality indicators for LCI data. Approaches towards analyzing the uncertainty in final results were published (Heijungs, 1992; 1994), but remained unused for two main reasons: lack of knowledge of input uncertainties, and lack of appropriate software. For almost a decade, the two lacks seemed to be trapped in a vicious circle: as long as there is no software that deals with uncertain data, there is no need to collect uncertainty information for the data, and as long as there is no uncertainty information for the data, there is no need to develop software that deals with uncertain data. But the last few years, software and data providers are freeing themselves from this trap. Software that in one form or another supports Monte Carlo analysis is becoming standard, and one of the most widely used data sources, the Swiss ecoinvent, has started to include information on distribution and data quality indicators.

From the theoretical side, we mention the publication of a number of PhD theses in which uncertainty in LCA played a dominant role (Roš, 1998; Pohl, 1999; Huijbregts, 2001; Ciroth, 2001). Other important contributions are summarized by Heijungs & Suh (2002). Besides these structural developments, developments in case studies have shown some ingenuity in dealing with uncertainty issues. More and more case studies have used statistical methods to address uncertainties.

The increasing recognition of the role of uncertainty in LCA has also some darker sides. On the one hand, it may easily lead to pessimism or even cynicism. Results of LCA would be meaningless, as the uncertainties associated with these results would overshadow the results themselves. And carrying out LCA would become much more complicated, due to the additional data collection efforts and the more intricate calculations. Finally, interpretation of LCA-results would be more cumbersome, and involve a much more technical jargon: confidence intervals, significance levels, etc.

A practical problem with dealing with uncertainty is that the information is scattered and that terminology is confusingly non-standardized. This already applies to the definition of uncertainty. What is uncertainty, what is variability, what is sensitivity? And there are also rumors. Is Monte Carlo analysis the only possible method? Is it needed to take correlated variates into account?

This paper aims to provide an overview of the various aspects of uncertainty in relation to LCA, and of the practical approaches that have been proposed or employed. It partly builds on and supplements the survey of Björklund (2002). It starts with a theoretical part on the types of uncertainty and the techniques that are available to addresses these uncertainties. It then proceeds to give a survey of concrete proposals and implementation in guidebooks, databases, software and case studies. Finally, some proposals are made to arrive at a more uniform terminology of uncertainty issues in LCA.

2. TYPOLOGIES OF UNCERTAINTY

When speaking on uncertainties, one of the first things that could be defined is the very notion of uncertainty itself. Although a fully satisfying definition may be difficult to agree upon, we will here rely on a mere reference to the problem of using information that is unavailable, wrong, unreliable, or that show a certain degree of variability. The wording above suggests a division into three types:

- data for which no value is available;
- data for which an inappropriate value is available;
- data for which more than one value is available.

On top of that, one should acknowledge that LCA – and indeed any model – contains data, relationships and choices, so that the same division into three may be applied for relationships and choices as well, e.g., relationships for which no equation is available, or choices for which more than one option is available.

Before proceeding to study uncertainty in more detail, a contrast with variability should be made (US-EPA, 1989). Uncertainty relates to a lack of knowledge: no data is available, or the data that is available is wrong or ambiguous. Variability, in contrast, is a quality of data that is essentiality of a heterogeneous nature. The number of passengers in a specific train may be subject to uncertainty, while the number of passengers in a typical train may be subject to variability, because it differs from case to case. Likewise, the molecular weight of phenol may be uncertain, while the half-life time may be uncertain, because it depends on – variable – ambient conditions. Despite the different meaning and source of uncertainty and variability, the approaches for dealing with the two show a large overlap.

There are many ways of classifying uncertainty. Without going into the details of defining these categories, Table 1 lists a few typologies. Reviewing all these typologies, one might ask oneself whether a typology of uncertainties is useful at all. It appears that, no matter how you classify uncertainties, all uncertainties should be dealt with in the appropriate way. We believe, with Funtowicz & Ravetz (1990), that a typology is useful provided a distinction is made between sources and sorts of uncertainty. Moreover, we believe that it is the sorts of uncertainty that should be emphasised, because it ought to steer the approach taken to deal with uncertainty.

Another, perhaps underemphasized, aspect of uncertainty is that there are levels of uncertainty, relating to the role of the person that experiences the uncertainty. Thus, a scientist may feel uncertain on the value of a certain parameter, while a decision-maker may feel uncertain on the decision to be taken. This distinction may be of critical importance in the choice of methods to deal with uncertainty. For instance, an ISO-standard may settle the uncertainty problem for the decision-maker, but not for the scientist, who will like to do more research or to specify statistical distributions.

Table 1: Classification of uncertainties according to several authors.

Bevington & Robinson (1992)	Morgan & Henrion (1990) Hofstetter (1998)	Huijbregts (2001)
systematic errors random errors	statistical variation subjective judgment linguistic imprecision variability inherent randomness disagreement approximation	parameter uncertainty model uncertainty uncertainty due to choices spatial variability temporal variability variability between sources and objects
Funtowicz & Ravetz (1990)	Bedford & Cooke (2001)	US-EPA (1989)
data uncertainty model uncertainty completeness uncertainty	aleatory uncertainty epistemic uncertainty parameter uncertainty data uncertainty model uncertainty ambiguity volitional uncertainty	scenario uncertainty parameter uncertainty model uncertainty

3. TECHNIQUES AND TOOLS TO ADDRESS UNCERTAINTY

Approaches to deal with uncertainty exist in many kinds. Consider a concrete example: an LCA-practitioner runs across an uncertain data item, say the characterization factor for human toxicity of zinc. How could one proceed? Main lines are:

- the scientific approach (doing more research, like setting out laboratory tests to find out LC50s and other relevant parameters in the characterization model);
- the constructivist approach (involving stakeholders, discussing and finally deciding on or voting for a consensus characterization factor);
- the legal approach (relying on what authoritative bodies, like ISO or US-EPA, have decreed as the truth);
- the statistical approach (using methods from statistics, like Monte Carlo analysis or fuzzy set theory, to determine confidence intervals and other indicators of robustness).

It should be noted that the first three of these approaches aim to reduce uncertainty, while the last approach aims to explicitly incorporate it. Reducing uncertainty – although in itself a noble aim – will not further be discussed here; we merely refer to Von Bahr & Steen (2004) as a recent example in LCA. We will restrict the discussion to approaches to incorporate uncertainty. In doing so, we will be close to practicing post-normal science (Funtowicz & Ravetz, 1993), a form of applied science which claims to deal with policy issues in cases of large uncertainties and high decision stakes. Nonetheless, the statistical approach is in other respects more alien to post-normal scientists. In general, post-normal science prefers constructivist approaches to statistical approaches, thereby eventually doing away with the uncertainty. One should note that even statisticians eventually do away with the uncertainty, namely in their process of null-hypothesis significance testing, where all uncertainty information finally condenses into a yes-no decision.

In dealing with uncertainty, one is faced with problems at three places:

- the input side: where are the uncertainties, and how large are they?
- the processing side: how do we translate input uncertainties into output uncertainties?
- the output side: how can we visualize and communicate uncertain results?

Of course, answers within the three areas are highly dependent. For instance, if we choose to use Monte Carlo analysis for the processing side, our possibilities for the input and output sides are immediately restricted to a few options. One might be tempted to confine the picture of input, processing and output to those sources of uncertainty that pertain to parameter uncertainty or data uncertainty, the truly input-oriented elements of a model. However, some of the other sources of uncertainty can be captured in these terms as well. The choice between competing models, or the choice among the elements to be modeled can also be regarded as comprising an input uncertainty, for instance.

3.1 Processing uncertainties

As the method for processing are the pivot in this scheme, we start to discuss this aspect. Within the statistical approach, there are many possibilities. Some of these are:

- parameter variation/scenario analysis;
- sampling methods;
- analytical methods;
- non-traditional methods, such as the use of fuzzy set theory.

Each of these methods for processing uncertainties requires different forms of input uncertainties, and delivers different forms of output uncertainties. Below, we will discuss these methods and some applications in LCA.

With parameter variation/scenario analysis, a few different data sets and/or models and/or choices are investigated as to their consequences for the model results. For instance, the results are calculated for a data set with high emission values and a data set with low emission values. Good illustration of the use of this method within LCA are provided by Copius Peereboom *et al.* (1999) and Huijbregts (1998).

A sampling method is a method that employs the power of a computer for repeating calculations many times. If the input data for each parameter is drawn from some distribution, the results will differ from run to run and gradually give rise to a sample of results, of which the statistical properties may be investigated. Monte Carlo analysis, where a distribution of outcomes is calculated by running the model a number of times with randomly selected parameter representations, is the most well-known form, but there are more sophisticated ways, like Latin hypercube sampling (Morgan & Henrion, 1990). Monte Carlo analyses have been applied in LCA by a couple of authors, see, e.g., Meier (1997), Huijbregts (1998), Maurice *et al.* (2000) and Sonneman *et al.* (2003).

Sampling methods can also be used to address scenarios. In that case, the sample consists of combinations of different decision scenarios and model formulations, with a subjective probability reflecting the preference of the decision-maker or the faith of the modeler in a particular model formulation for an alternative (Efron & Tibshirani, 1991). According to Huijbregts *et al.* (2003), the resulting output distribution reflects the uncertainty of the decision-maker regarding the normative choices involved (scenario uncertainty) or the uncertainty of the modeler regarding the alternative model formulations (model uncertainty).

Analytical methods are based upon explicit mathematical expressions for the distributions of the model results. Their use is based on a first-order approximation of the Taylor expansion of the underlying model (Heijungs & Suh, 2002; Heijungs, 2002). Distribution-free variances of input parameters can then be used to calculate variances of output variables. Their use in LCA has been limited so far, probably because the mathematics was too complicated to be implemented in software. Results have, however, been reported by Heijungs *et al.* (in press).

Under non-traditional methods, we will capture all methods of dealing with uncertainty that are not part of the traditional statistics curriculum. It comprises a variety of methods, for instance:

- fuzzy set methods;
- Bayesian methods;
- non-parametric statistics;
- robust statistics;
- neural networks and other methods from artificial intelligence.

Methods for uncertainty analysis based on fuzzy sets have been introduced into LCA by several authors, see, e.g., Weckenmann & Schwan (2001), Chevalier & Le Téno (1996), Rong *et al.* (1998), Roš (1998). Bayesian statistics has hardly been mentioned in the context of LCA, although Shipworth (2002) provides an exception. The other mentioned methods are even less used within LCA, although the sign test and the Kruskall-Wallis test are briefly touched by Heijungs & Kleijn (2001). It should be noted that some of these methods are sometimes mentioned within LCA (see, e.g., Sangle *et al.*, 1999), but that the emphasis is in those cases not so much on the processing side of uncertain information, but on the approaches that have been developed within decision theory for dealing with unclear preferences.

3.2 Input uncertainties

Parameter variation requires that a number of different values is available for one or more parameters. Treating all parameters individually may lead to an exceedingly large number of scenarios. Therefore, it is usual to vary one parameter and keep all other parameters fixed at some "most probable value", and to repeat this procedure for all parameters in separate analysis. This type of analysis can be found in Copius Peereboom *et al.* (1999). An alternative is to define a limited number of scenarios with specific but consistent realizations of each parameter. Hofstetter (1998) and

Goedkoop & Spriensma (2000) employ these types of "perspectives" in their analyses. A more systematic treatment of the use of scenarios of the future in LCA is given by Pesonen *et al.* (2000) and Fukushima & Hirao (2002), and applied by Contadini *et al.* (2002).

Sampling methods are based on the random variation of uncertain parameters. They require the specification of a statistical distribution of every stochastic parameter. For instance, an emission may be specified as following a normal distribution with a mean of 12 kg and a standard deviation of 1 kg. Frequently encountered distributions are:

- the normal distribution;
- the lognormal distribution;
- the uniform distribution;
- the triangular distribution.

These distributions may or may not be correlated across parameters. In principle, correlations between parameters may be expressed by a correlation matrix or a covariance matrix (Heijungs & Suh, 2002). Huijbregts *et al.* (2003) showed how correlations between input parameters can be included in Monte Carlo analyses. Apart from correlations between input parameters, correlations between model outputs should be accounted for in comparative LCAs (Huijbregts *et al.*, 2001; 2003). This can be done in the form of a comparison index (Huijbregts, 1998) for the case of two alternatives, or in a more general discernibility analysis (Heijungs & Kleijn, 2001; Heijungs & Suh, 2002).

Analytical methods are based on the estimation of the moments of the distributions (Morgan & Henrion, 1990). In particular the second moment, the variance, is used in a first order Taylor approximation. Thus, not the distribution, but only the variance (or standard deviation) of the parameter is needed here. Thus, less information is needed for analytical methods than for sampling methods. Like for Monte Carlo analysis, correlations between variates can in principle be included, although this is seldom seen in practice. Inclusion of correlations in the analytical case implies a broadening of the scope to second-order Taylor approximations (Heijungs & Suh, 2002).

Because methods for processing uncertainties on the basis of non-traditional methods have hardly been applied in LCA, it is not clear which types of input information would be needed.

3.3 Output uncertainties

At the output side there are fewer differences. In combination of parameter variation, one often sees the consecutive presentation of tables and/or graphs for the different sets of parameters or scenarios; see e.g. Copius Peereboom *et al.* (1999), Huijbregts (1998).

Results of sampling methods can be presented in different forms. Sampled probability density plots, so-called histograms, are a typical example; see, e.g., Huijbregts (1998) and Sonneman *et al.* (2003). An alternative is the graphical representation of an average value with two boundary values. These boundary values may indicate the smallest and largest value obtained, or a more robust measure such as the 5 and 95 percentile values (Huijbregts *et al.*, 2003).

Analytical methods do not provide a distribution of outcomes. Instead, they provide moments of the distributions, such as the standard deviation. These can be used to calculate and visualise 95% confidence intervals. As analytical methods have hardly been applied in LCA, we cannot give an example of its use.

This holds even more true for the non-traditional methods, like fuzzy sets methods and Bayesian methods.

4. PROSPECTS FOR INCLUSION OF UNCERTAINTY IN LCA

Reviewing the developments that have taken place the last few years, it seems likely that discussion and inclusion of uncertainty issues in LCA will no longer be restricted to academic exploratory work, like PhD-theses, and will no longer be regarded as a curiosity in real practical work. Rather, we expect that inclusion of uncertainties will become a standard feature of case studies. The three requirements for becoming a standard procedure, availability of data of input uncertainties, availability of methods and software for processing uncertainties, and availability of methods for interpreting and visualizing output uncertainties, start to be satisfied.

There is perhaps one more aspect that can be seen as a requirement: standardization (*cf.* Björklund, 2002). The ISO-standards for LCA have canonized parts of the terminology used. On top of that, the format by SPOLD has provided a standard for data exchange. But especially for uncertainty, clear standards are lacking.

As to terminology, there is first of all the confusion between uncertainty and variability, and within uncertainty all the sorts and sources of uncertainty. It may be difficult (and unnecessary) to single out one single terminology. Then, there is a large number of types of approaches that are used interchangeably, or at least in a non-standardized way. We mention just a few:

- uncertainty analysis;
- sensitivity analysis;
- perturbation analysis;
- scenario analysis;
- error analysis;
- discernibility analysis.

It is a disturbing (or perhaps: consoling) fact that even outside LCA, within the uncertainty community itself, meanings and nomenclature give rise to disagreement. For instance, sensitivity analysis means to US-EPA (1989) the systematic changing of one parameter while keeping the other parameters constant, whereas to Saltelli *et al.* (2000) it means the apportioning of an output uncertainty to the various contributing input uncertainties, which is in turn referred to as key issue analysis by Heijungs (1996) and as uncertainty importance analysis by Björklund (2002).

But at least the way uncertain data is described can be standardized. In fact, this should be part of the normal data exchange process. In a small study, Heijungs & Frischknecht (in prep.) discussed the differences in representing a basic entity like the uniform distribution in just one database (ecoinvent), one LCA-program (CMLCA) and mathematical statistics. Clearly, one may choose in representing a uniform distribution between giving:

- the mean value and the width;
- the mean value and the half-width;
- the lowest and highest value;

The required transformations in going from one representation to another one is quite simple. For more complicated distributions, like the lognormal distribution, the expressions are much more involved. As long as these different options are not clearly defined and distinguished, one is likely to confuse a standard deviation with a variance, a width with a half-width, or worse.

Obviously, those aspects of uncertainty that pertain to parameter uncertainty or data uncertainty have received most attention so far, at least in practical cases. The model uncertainty is much less addressed. And the more profound forms of uncertainty, for instance epistemic uncertainty may fundamentally be difficult to deal with. In this, we agree with parts of the analysis of Funtowicz & Ravetz (1993), who promote the development of non-traditional modes of research (i.e. post-normal science) to deal with intrinsically uncertain policy questions. Their NUSAP-scheme (Funtowicz & Ravetz, 1990) has been brought into LCA by Weidema & Wesnæs (1996). We think that the separation with which this paper started, into data for which more than one value is available, data for which an inappropriate value is available, and data for which no value is available, can be connected to the SAP-part of the NUSAP-scheme:

- spread, for data for which more than one value is available;
- assessment, for data for which an inappropriate value is available;
- pedigree, for data for which no value is available.

It is especially the S-part which can be processed with sampling or analytical techniques, and the A-part by parameter variation/scenario analysis techniques. The P-part is supposed to reflect our "ignorance of ignorance", for which the use of precision-suggesting numbers is by definition inappropriate.

5. REFERENCES

Bahr, B. von and B. Steen. Reducing epistemological uncertainty in life cycle inventory. *Journal of Cleaner Production* 12 (2004), 369-388.

Bedford, T. and R. Cooke. *Probabilistic risk analysis. Foundations and methods.* Cambridge University Press, Cambridge, 2001.

Bevington, P.R. and D.K. Robinson. *Data reduction and error analysis for the physical sciences. Second edition.* McGraw-Hill, Inc., New York, 1992.

Björklund, A.E. Survey of Approaches to Improve Reliability in LCA. *International Journal of Life Cycle Assessment* 7: 2 (2002), 64-72.

Contadini, J.F., R.M. Moore and P.L. Mokhtarian. Life Cycle Assessment of Fuel Cell Vehicles: A Methodology Example of Input Data Treatment for Future Technologies. International. *Journal of Life Cycle Assessment* 7:2 (2002), 73-84.

Copius Peereboom, E., R. Kleijn, S. Lemkowitz and S. Lundie. Influence of inventory data sets on life cycle assessment results. A case study on PVC. *Journal of Industrial Ecology* 2:3 (1999), 109-130.

Chevalier, J.-L. and J.-F. Le Téno. Life Cycle Analysis with Ill-Defined Data and its Application to Building Products. *International Journal of Life Cycle Assessment* 1:2 (1996), 90-96.

Ciroth, A. *Fehlerrechnung in Ökobilanzen.* PhD thesis, Technischen Universität Berlin, Berlin, 2001.

Efron, B. and R. Tibshirani. Statistical analysis in the computer age. *Science* 1991, 253, 390-395.

Fava, J.A., A.A. Jensen, L. Lindfors, S. Pomper, B, de Smet, J. Warren and B. Vigon. *Life-cycle assessment data quality. A conceptual framework.* SETAC, Pensacola, 1994.

Fukushima, Y. and M. Hirao. A Structured Framework and Language for Scenario-Based

Life Cycle Assessment. *International Journal of Life Cycle Assessment* 7:6 (2002), 317-329.
Funtowicz, S.O. and J.R. Ravetz. *Uncertainty and quality in science for policy*. Kluwer Academic Publishers, Dordrecht, 1990.
Funtowicz, S.O. and J.R. Ravetz. Science for the post-normal age. *Futures* 25 (1993), 739-755.
Goedkoop & Spriensma, 2000. *The Eco-indicator 99, a damage oriented method for Life Cycle Impact assessment, Methodology Report.* Pré Consultants, Amersfoort, The Netherlands
Heijungs, R. and R. Frischknecht. Statistical distributions for uncertain parameters in LCA. Relationships between mathematical forms, their representation in ecoinvent, and their representation in CMLCA. In preparation.
Heijungs, R., J.B. Guinée, G. Huppes, R.M. Lankreijer, H.A. Udo de Haes, A. Wegener Sleeswijk, A.M.M. Ansems, P.G. Eggels, R. van Duin and H.P. de Goede. *Environmental life cycle assessment of products. I: Guide – October 1992. II: Backgrounds – October 1992.* CML, Leiden, 1992.
Heijungs, R. On the identification of key issues for further investigation in life-cycle screening. The use of mathematical tools and statistics for sensitivity analyses. *Journal of Cleaner Production* 4:3/4 (1996), 159-166.
Heijungs, R. and S. Suh. *The computational structure of life cycle assessment.* Kluwer Academic Publishers, Dordrecht, 2002.
Heijungs, R. *The use of matrix perturbation theory for addressing sensitivity and uncertainty issues in LCA.* In: *Proceedings of The Fifth International Conference on EcoBalance - Practical tools and thoughtful principles for sustainability.* The Society of Non-Traditional Technology, 2002, Tokyo, Japan.
Heijungs, R. and R. Kleijn. Numerical approaches towards life cycle interpretation. Five examples. *International Journal of Life Cycle Assessment* 6:3 (2001), 141-148.
Heijungs, R., S. Suh & R. Kleijn. Numerical approaches towards life cycle interpretation. The case of the Ecoinvent'96 database. *International Journal of Life Cycle Assessment*, in press.
Heijungs, R. A generic method for the identification of options for cleaner products. *Ecological Economics* 10:1 (1994), 69-81.
Hofstetter, P. *Perspectives in life cycle impact assessment. A structured approach to combine models of the technosphere, ecosphere and valuesphere.* Kluwer Academic Publishers, Dordrecht, 1998.
Huijbregts, M.A.J. *Uncertainty and variability in environmental life-cycle assessment.* PhD thesis, Universiteit van Amsterdam, Amsterdam, 2001.
Huijbregts M.A.J., G. Norris, R. Bretz, A. Ciroth, B. Maurice, B. von Bahr, B. Weidema and A.S.H. de Beaufort. Framework for modelling data uncertainty in life cycle inventories. *International Journal of Life Cycle Assessment* 6:3 (2001), 127-132.
Huijbregts M.A.J., W. Gilijamse, A.M.J Ragas, and L. Reijnders, Evaluating uncertainty in environmental Life-Cycle Assessment, *Environmental Science and Technology*, 37 (2003), 2600-2608.
Maurice, B., R. Frischknecht, V. Coelho-Schwirtz and K. Hungerbühler. Uncertainty analysis in life cycle inventory. Application to the production of electricity with French coal power plants. *Journal of Cleaner Production* 8 (2000), 95-108.
Meier, M.A. Eco-efficiency evaluation of waste gas purification systems in the chemical industry. Ecoinforma Press, Bayreuth, 1997.
Morgan, M.G. and M. Henrion. *Uncertainty. A guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge University Press, Cambridge, 1990.
Pesonen, H.-L., T. Ekvall, G. Fleischer, G. Huppes, C. Jahn, Z. Klos, G. Rebitzer, G. Sonnemann, A. Tintinelli, B.P. Weidema and H. Wenzel, Henrik. Framework for Scenario Development in LCA. *International Journal of Life Cycle Assessment* 5:1 (2000), 21-30.
Pohl, C., M. Roš, B. Waldeck and F. Dinkel. *Imprecision and uncertainty in LCA.* In: S. Schaltegger (Ed.). *Life cycle assessment (LCA). Quo vadis?* Birkhäuser Verlag, Basel, 1996.
Pohl, C.E. *Auch zu präzis ist ungenau! Unsicherheitsanalyse in Ökobilanzen und alternativen zu "use many methods."* PhD thesis, Eidgenössischen Technischen Hochschule Zürich, Zürich, 1999.
Rong, M., Q. Guanjun, and J. Zhihao. *An application of fuzzy mathematics in life cycle assessment (LCA).* In: *Proceedings of The Third International Conference on EcoBalance.* The Society of Non-Traditional Technology, 1998, Tokyo, Japan. p. 696-701.
Roš, M. *Unsicherheit und Fuzziness in ökologischen Bewertungen. Orientierung zu einer robusten Praxis der Ökobilanzierung.* PhD thesis, Eidgenössischen Technischen Hochschule Zürich, Zürich, 1998.
Sangle, S., P.R. Babu and P. Khanna, P. Evaluation of Life Cycle Impacts: Identification of Societal Weights of Environmental Issues. *International Journal of Life Cycle Assessment* 4:4 (1999), 221-228.
Shipworth, D. A Bayesian Hybrid LCA Model of Stochastic Embodied Greenhouse Gas Emis-

sions in Construction Materials. DOI: http://dx.doi.org/10.1065/ehs2002.04.037.1.

Sonneman, G.W., M. Schuhmacher, and F. Castells. Uncertainty assessment by a Monte Carlo simulation in a life cycle inventory of electricity produced by a waste incinerator. *Journal of Cleaner Production* 11 (2003), 279-292.

US-EPA. *Exposure factors handbook*. EPA, Washington, 1989.

Weckenmann, A. and A. Schwan. Environmental Life Cycle Assessment with Support of Fuzzy-Sets. *International Journal of Life Cycle Assessment* 6:1 (2001), 13-18.

Weidema, B.P. and M.S. Wesnæs. Data quality management for life cycle inventories. An example of using data quality indicators. *Journal of Cleaner Production* 4:3/4 (1996), 167-174.

Integration of Multi-Scale Dynamic Spatial Models of Socio-Economic and Physical Processes for River Basin Management

Hedwig van Delden, **Patrick Luja and Guy Engelen**

Research Institute for Knowledge Systems bv, Maastricht, The Netherlands
E-mail: hvdelden@riks.nl

Abstract: Policy-makers have the difficult task to intervene in a complex human-natural system. Therefore it is not enough to focus merely on the individual processes affecting a region, rather it is necessary to look at the system as a complex integral whole. Many authors have stressed the importance of integrated assessment and model integration, but only few have proposed concrete solutions for the integration of different sub-models and the use of the resulting integrated models in practice. In this paper an attempt is made to cover both topics by means of an example: the MedAction Policy Support System (PSS). This system, implemented with the GEONAMICA® application framework, incorporates both socio-economic and physical processes, as well as strong interactions between them. It is intended to be used by technicians supporting policy-makers and planners in the fields of land degradation & desertification, water management and sustainable farming. This paper gives insights into the individual models, the model integration and the translation of model output into policy relevant information, but it also contends that more than the technical and scientific aspects of policy support systems are important for the actual use of such a system. It therefore provides six decisive criteria determining the actual use of PSS by their intended end users and assesses the MedAction PSS accordingly. The paper concludes with some lessons learned during the development and use of the MedAction PSS and similar Decision Support Systems.

Keywords: model integration; Policy Support System; use of PSS; indicators, linking science to policy relevant information

1. INTRODUCTION

Policy-makers have the difficult task to intervene in a complex human-natural system. This requires addressing the system as an integral entity rather than focussing on individual processes affecting a region. Policy Support Systems (PSS) and integrated models can support technicians, planners and policy-makers confronted with problems requiring integrated approaches.

The MedAction PSS is a system developed for this purpose. It aims to apply scientific research to the support of policy-making through the construction of an integrated model based on scientific knowledge from previous EU-projects, policy measures and policy relevant indicators.
This paper will provide some insights into the individual models and their interaction, as well as into the process of translating scientific knowledge into policy relevant information. However, it will also contend that there is more to the development of a PSS than addressing the right problems, representing the right processes and visualising the results in a transparent manner.

1.1 The MedAction PSS

Land degradation and desertification are important issues in arid, semi-arid and dry sub-humid regions. Climate change in combination with socio-economic processes acting on the land may induce a reduction of resource potential. With a view to mitigate these problems the EC is funding among others the research project MedAction (see Acknowledgements), in which a PSS is developed with the aim of providing a support tool for policy makers confronted with land degradation and desertification in Mediterranean watersheds and regions.

The MedAction PSS addresses three policy themes regarding regional development: land degradation & desertification, sustainable farming and water resources. For each of these themes the main problems, goals, policy options and policy indicators have been collected, structured, and translated into a conceptual framework. A summary of the themes in relation to the policy options and indicators is presented in Table 1. Based on this framework an integral quantitative PSS is designed and developed.

By using policy themes as a starting point for model integration and by linking the model to policy relevant indicators, the PSS is intended to support policy-makers and planners at the regional level by providing assistance in:

1. understanding the important processes in the region as well as their interactions,
2. identification of current or future problems in the river basin,
3. impact assessment of possible policy measures to mitigate the problems, and
4. evaluation of the different alternatives.

The user interface of the PSS features a system diagram (Figure 1) representing the different interacting sub-models and processes. It enables interactive access to the individual sub-models so that the user may change parameter values, visualise sub-model output and access the on-line help system which clarifies the underlying assumptions and formal definitions of the models and the data used.

The MedAction Policy Support System is generic for Mediterranean regions. A previous version of the system has been applied to the Marina Baixa (Spain) and the Argolidas (Greece). The current version is applied to the Guadalentín river basin in

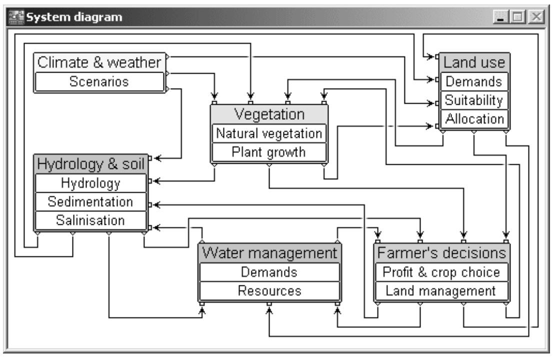


Figure 1: System diagram MedAction

Spain. It is implemented with the GEONAMICA® application framework, which enables a user to develop and run fully coupled multi-scale dynamic spatial models featuring complex feedback loops. Moreover, it provides a user-friendly interface with tools for visualisation and analysis of the outputs generated.

2. THE INTEGRATED MODEL

The core of the MedAction PSS consists of a number of sub-models, integrated into a single model that simulates the developments in the region for 30 years into the future.

Driving forces in MedAction are demographic and economic growth as well as climate change. The socio-economic growth is translated into demands for the different land use functions in the region. The demands are allocated dynamically in the *Land use* sub-models based on the spatial dynamics caused by the detailed interactions among land use functions, their dependencies on access to infrastructure, the zoning regulations restricting or facilitating particular land uses, and the physical suitability of the land to sustain the functions. For two of the land use functions, agriculture and natural vegetation, a dynamic suitability map is used, which changes as a result of the combined physical processes represented in

Themes	**Policy measures**	**Indicators**
Sustainable farming Long term profits Sustainable land use	▪ Subsidies, taxes ▪ Water price ▪ Water availability	▪ Profit ▪ Crop type ▪ Number and location of abandoned cells ▪ Dynamic suitability maps ▪ Irrigation water used from different sources ▪ Amount and cost of irrigation water
Water resources Availability and price of water	▪ Water availability and price ▪ Amount of water from outside the region ▪ Construction of desalinisation plants	▪ Change in aquifer and reservoir budget ▪ Natural water input (runoff and recharge) ▪ Costs and amount of water used
Land degradation & desertification		
Erosion	▪ Afforestation ▪ Grazing regulations ▪ Construction of check dams ▪ Dredging	▪ Fertile soil depth ▪ Erosion rates ▪ Change in storage capacity of reservoir ▪ Total cost of dredging
Preservation of nature and forests	▪ Afforestation ▪ Zoning	▪ Forested area ▪ Changes in natural vegetation type groups ▪ Dynamic suitability maps
Salinisation	▪ Maximum amount of water available from aquifer and desalinised water ▪ Maximum allowable percentage of salt in water from desalinised	▪ Soil salinity ▪ Salt concentration in the aquifer ▪ Restricted factor for plant growth (yes/no)
Sustainable land use in region	▪ Zoning ▪ Construction of infrastructure (dams, roads, channels)	▪ Land use map ▪ Dynamic suitability maps ▪ ESA

Table 1: Linking policy themes, measures and indicators

the system: soil moisture, soil salinity, fertile soil depth, slope, and temperature. These aspects are in turn calculated in the *Climate & weather*, *Hydrology & soil* and *Vegetation* modules and are influenced by social processes such as water use and land management practices. Dynamic suitability maps are also used in one of the *Farmers' decisions* sub-models in which crop choices are calculated based on physical characteristics, crop price, water price and use, and social aspects influencing farmers' decision-making. Availability of water is calculated in the *Water management* module and can be experimented with by the user of the system.

2.1 Individual models

Decisions relative to the selection and integration of individual models are based on the availability of existing models, and the degree to which they cover the policy themes and provide support to solve problems. To a large extent use is made of models incorporated in MODULUS, the predecessor of the MedAction PSS. Selection criteria used in MODULUS as well as the modifications made to the original models are described extensively in Engelen [2003] and Oxley et al. [in press]. In the development of the MedAction PSS existing models are re-used in their original form or modified to better suit the purpose of policy-making. New models are developed to fill the gaps and missing links in the integrated system representation. For the technical integration all original models have been recoded in C++. This turned out to be more efficient in terms of resource use and computational performance than re-using the existing code.

The individual models are designed, modified and implemented to represent the actual processes as realistically as possible. As real world processes can be very different in their nature, the models describing them will be so too. An example of this can be found in the two-step approach for calculating the detailed land use. In a first step, general land use functions (urban residential, rural residential, industry & commercial areas, tourism, agriculture, forest reserves and natural vegetation) are allocated by means of a cellular automaton based land use model taking into consideration zoning regulations, accessibility, and suitability for the different functions. Figure 2a shows the land use map with these general land use classes. This type of model represents very well the land use dynamics of socio-economic functions featuring characteristic locational preferences and spatial interactions vis-à-vis one another [White and Engelen, 1997], but is less suited for modelling transitions between different crop and natural vegetation types. These are therefore handled in a second step in the *Profit & crop choice* sub-model and the *Natural vegetation* sub-model, respectively. The *Profit & crop choice* sub-model represents choices made by individual farmers relative to the use of their land. For each hectare the specific crop type is calculated based on a stochastic utility function considering the profit for the different crop types as well as social aspects influencing the choices made by farmers. Figure 2b shows the agricultural area with the detailed crop distribution. The *Natural vegetation* sub-model represents the transitions of the different natural vegetation types toward a climax vegetation. These transitions are partly affected by humans through herding and protection against forest fires, but are mainly based on the height and vegetation cover of the natural vegetation, which in turn are the result of physical properties like temperature, soil moisture, soil salinity and fertile soil depth. Figure 2c shows the location of the detailed natural vegetation types. In Figure 2d the results of both steps are combined to produce a detailed map of the land uses in the region.

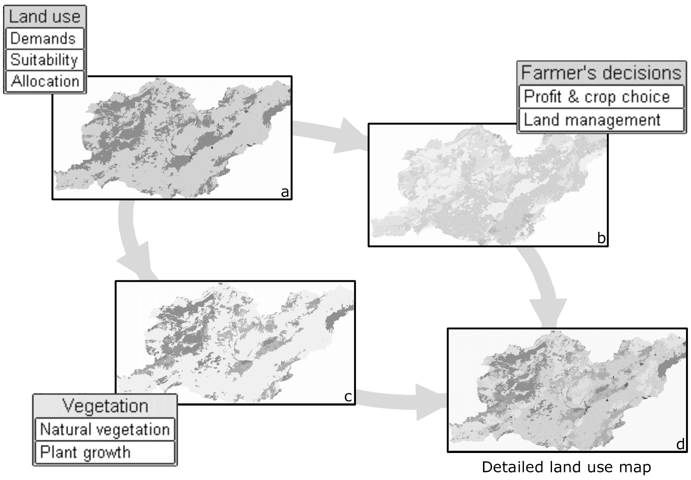


Figure 2: Two-step land use allocation

Besides capturing the right types of processes the PSS represents each process at the appropriate temporal and spatial resolution. For example the *Hydrology* sub-model runs on a variable time step, expressed in minutes, which is required to capture the short intense rainfall events causing erosion, while the farmers' crop choice model runs on the yearly time step representing the true timing of the farmers' decision making process. The spatial resolution of the integrated models is essentially the 1 ha grid. Some minor processes are modelled at the level of the whole region.

2.2 Model integration

After selecting the right individual models, choices had to be made relative to their integration. The MedAction PSS has strong linkages among all its sub-models. For example the soil moisture calculated in the *Hydrology* sub-model is used in the *Plant growth* sub-model for the calculation of biomass. But biomass is in turn used for the calculation of soil moisture. There are also strong couplings between the physical and socio-economic models. For example, the yield obtained from the *Plant growth* sub-model is used to calculate the crop choice in the *Profit & crop choice* sub-model, and, in the *Water management* module replenishment obtained from the *Hydrology* sub-model is combined with demands from the socio-economic sub-models.

growth and the physical variables, are applied to transform the model output into suitability values ranging between 0 and 1 (1 being very suitable, resulting in a maximal possible growth).

This is schematised in detail in Figure 3. The first column shows the maps of the important factors for plant growth calculated in the different sub-models: soil moisture, soil salinity, fertile soil depth, and temperature. Since all of these sub-models run on a time step smaller than one year, decisions had to be made relative to the aggregation of the model output. For the soil salinity and the fertile soil depth, the latest values are chosen, since they represent the current state of these factors. For the other two factors, soil moisture and temperature, the average is taken over the representative period for the growth of the specific plant; e.g. given that for plant type Y the temperature in April and May is crucial, only the average of these two months is taken into account. This results in a set of factor maps for each plant type. The second column shows the plant-specific response curves for all factors. They represent the suitability of a cell for the growth of this plant type (y-axis) as a function of its factor value (x-axis). The third column shows the result of applying the second to the first. Note that areas with low soil moisture (light coloured) are unsuitable for growth of plant type Y (light coloured). Areas with a higher soil moisture are more suitable, but areas that are too wet (dark coloured) result in a lower

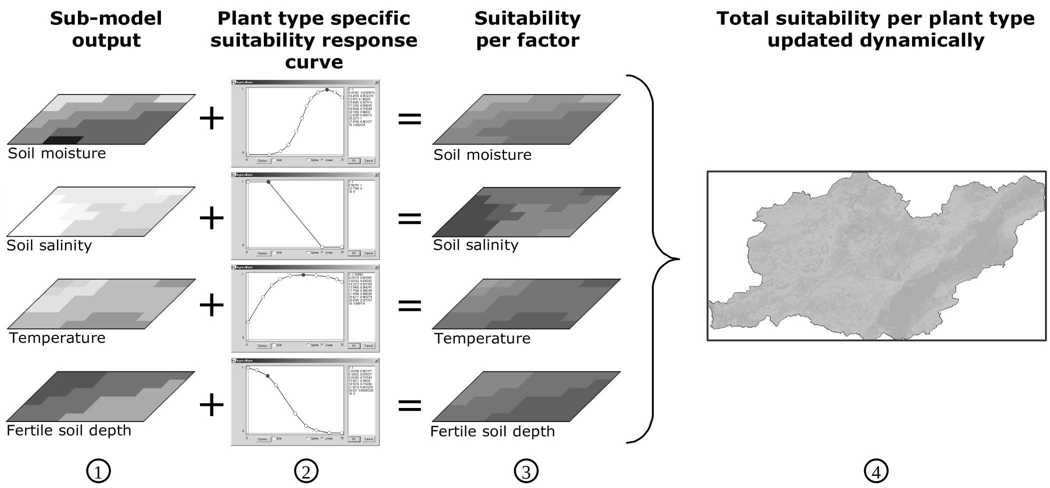


Figure 3: Dynamic suitability maps

Another important linkage between the physical and socio-economic models is realised through dynamic suitability maps. They represent the physical suitability of the land for the growth of different plant types and are an input in the *Land use allocation* and *Profit & crop choice* sub-models. They are recalculated on a yearly basis using up to date model output from the physical sub-models. Response curves for the different plant types, describing the relation between plant

suitability value. Finally, the suitability maps per factor are combined into a single map using a minimisation function, thus assuming that the factor with the lowest suitability restrains plant growth.

3. POLICY RELEVANT OUTPUT

One of the main drawbacks of integrated models and their encompassing software systems is that they produce types and amounts of output that require an expert in all domains to be able to derive the relevant results. To assist policy-makers in overseeing and interpreting all the model output, three main aggregated indicators are developed that describe the dimensions in which the system can evolve. Each indicator covers one of the three themes of interest in MedAction:

- The Environmentally Sensitive Areas (ESA) indicator describes the desertification status of an area on the basis of some 24 aggregated physical and socio-economic variables [Kosmas et al., 1999]. In the PSS the ESA indicator is calculated dynamically and presented as a map as well as an overall figure for the whole region;
- The Water Shortage indicator describes the water shortages the socio-economic activities experience individually and as a whole, and is also presented in a map for the agricultural land-uses;
- The Farmers' Profit indicator describes the profit made by individual farmers, per crop type and for the agricultural sector as a whole.

For representing the dynamic path of the modelled system, GEONAMICA® features a three-dimensional state space diagram consisting of an equilateral triangle. On each side of the triangle a different variable can be selected for representation. In the present case, the three indicators described above are represented. Initially the indicators are assumed to be in balance, therefore the system is represented as a point in the centre of the triangle (see Figure 4, t=0). During a simulation exercise the indicators are recalculated yearly. Consecutive states of the system are represented by a series of new points in the state space. Figure 4, t=1, shows a system evolving towards more severely degraded areas with decreasing water shortages and a very profitable agricultural sector. This could be the result of a warming climate with reduced precipitation in combination with expanding irrigated agriculture at the expense of dryland agriculture driven by high market prices and supported by sufficient water availability from new sources, partly from outside the region. Because of the changing climate and the resulting decrease in soil moisture, non-irrigated areas will become more degraded. Irrigated areas will become more polluted because of increased salt concentration caused by the new irrigation possibilities (assuming the salt concentration of the irrigation water will stay the same), which will lead to more environmentally sensitive areas and more severely degraded areas. Figure 4, t=2, shows the situation a few years later when the level of land degradation causes a noticeable impact on agriculture, resulting in decreasing profits.

These two examples demonstrate the purpose and visualisation power of the state space diagram in the context of the MedAction PSS. It goes without saying that the dynamic path of the system as the result of autonomous developments, external driving forces and policy options can also be displayed, analysed and interpreted by means of this tool.

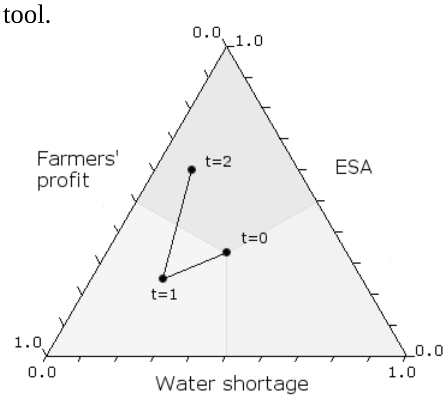


Figure 4: State space diagram

4. DISCUSSION

In order to understand why a PSS will or will not actually be used by the intended users in their decision-making processes, criteria are formulated in the following six questions:

- The strategy question: What is the usefulness of the system? What are the possibilities for application and what is its added value? What are the intended functions of the system (analysis, communication, knowledge management…)?
- The availability question: Is the availability of data, knowledge and models sufficient, in quality as well as quantity, to offer support? How difficult is it to obtain the necessary data, knowledge and models?
- The credibility question: Is there consensus on the models and the underlying assumptions? Can the output of the system be trusted?
- The language question: Does the system provide output that relates to the information needed by the intended end-users? Can information available to end-users be used as input into the system?
- The culture question: Are the intended end-users willing to adopt the PSS and willing and able to adjust their decision-making process?

- The structure question: Where can the instrument be placed in the organisation? Who is going to work with the instrument and what are the actual and practical changes that have to be made?

Based on the answers to these questions, the chances for success or failure of the PSS in a particular organisation can be determined, as well as the amount of time and money that would be necessary to resolve all questions positively.

The MedAction system was developed in a research context and first had to go through phases in which the technical and scientific challenges of integration, calibration and validation could be resolved, before it could be brought to the end-users. Even though during both the MODULUS and MedAction projects there has been contact with the potential end-users, this has not been sufficient to answer all of the questions mentioned above positively. Currently the PSS scores relatively well on the more technical and scientific elements such as those concerning scientifically sound models, model integration, and visualisation tools, and it is developing into a system that can be used by policy-makers in Mediterranean regions to help answer strategic questions, assess the impacts of policy options, and translate the output of scientific models into policy relevant information. But it still needs to be brought closer to the end-user to deal with the remaining, mostly 'softer', questions. This may become possible in the near future in the form of an actual policy exercise with policy makers in the Guadalentín river basin.

5. CONCLUSIONS AND RECOMMENDATIONS

Developing a Policy Support System is a difficult and time consuming task. Thorough knowledge of the individual models and model integration is crucial in the development of the core of such a system—an integrated model. This paper has presented two different techniques that can be useful in model integration: the two-step land use allocation procedure, and the concept of dynamic suitability maps to link physical and socio-economic models.

However, developing a PSS is more than developing an integrated model. The system needs to be able to support policy questions and provide policy relevant information. In the MedAction PSS this has been realised by discussing the policy themes, options and indicators with the intended end-users as a starting point for the development, so linking science to policy making.

Furthermore it is important to remember that the actual use of a policy support system depends not only on the contents of the system but also on the perceptions of the intended end-users regarding its usefulness and practicality, as well as their actual willingness to use the system. For them it is not only important that the system connects to the *context* of their problems but also that it relates to their decision-making *process*. Moreover they need to believe in the added value as well as the credibility of the system.

The MedAction PSS is developing in the right direction, but closer interaction with the intended end-users is necessary to bring it to the next level.

6. ACKNOWLEDGEMENTS

Research project 'MedAction: Policies to combat desertification in the Northern Mediterranean region', supported by the European Commission under contract EVK2-2000-22032: www.riks.nl/projects/medaction.

Research project 'MODULUS, A spatial Modelling Tool for Integrated Environmental Decision-Making' supported by the European Commission under contract ENV4-CT97-0685: www.riks.nl/projects/modulus.

7. REFERENCES

Engelen, G., Development of a Decision Support System for the integrated assessment of policies related to desertification and land degradation in the Mediterranean, p. 159-195 in: Giupponi, C. and M. Shechter (eds.), Climate Change in the Mediterranean: Socio-economic Perspectives of Impacts, Vulnerability and Adaptation, Edward Elgar, Cheltenham, 2003.

Kosmas, C., M. Kirkby, and N. Geeson (eds), The Medalus project Mediterranean desertification and land use: Manual on key indicators of desertification and mapping environmentally sensitive areas to desertification, European Commission, Directorate General XII, Contract ENV4-CT95-0119, Brussels, 1999.

Oxley, T., B. McIntosh, N. Winder, M. Mulligan, and G. Engelen, Integrated modelling and decision-support tools: a Mediterranean example, *Environmental Modelling & Software*, in press.

White, R., and G. Engelen, Cellular automata as the basis of integrated dynamic regional modelling, *Environment and Planning B: Planning and Design*, *24*, 235-246, 1997.

Towards an Open Modelling Interface (OpenMI) The HarmonIT project

M. Blind[a] and **J. B. Gregersen**[b]

[a] *RIZA Institute for Inland Water Management and Waste Water Treatment, The Netherlands, m.blind@riza.rws.minvenw.nl*

[b] *DHI Water & Environment, Denmark*

Abstract: The Water Framework Directive (WFD) poses an immense challenge to water management in Europe. Aiming at a "good ecological status" of surface waters in 2015, integrated river basin management plans need to be in place by 2009, and broadly supported by stakeholders. Information & Communication Technology (ICT) tools, such as computational models, are very helpful in designing river basin management plans (rbmp-s). However, many scientists believe that a single integrated modelling system to support the WFD cannot be developed, and integrated systems need to be quite tailored to the local situation and evolve during a collaborative planning process. As a consequence there is an urgent need to increase the flexibility of modelling systems, such that dedicated model systems can be developed from available building blocks. In the recent past a number of initiatives have been started to develop an IT framework for modelling to meet the required flexibility.
In Europe the international project HarmonIT, which is sponsored by the European Commission, is developing and implementing a standard interface for modelling components and other relevant tools: The Open Modelling Interface (OpenMI). This paper describes the HarmonIT project and objectives in general. The current progress is described. It describes the roles for different types of stakeholders in modelling, varying from software coders to non-specialized users of decision support systems. It will provide insight in the requirements imposed when using the OpenMI.

Keywords: Water Framework directive, Model linking, IT Frameworks

1. INTRODUCTION

In 2000 the European Parliament and Council passed the ambitious directive 2000/60/EC establishing a framework for Community action in the field of water policy, known as the Water Framework Directive (WFD). The key objective of this law is to achieve 'good ecological status of Europe's water resources by 2015. A key aspect of the WFD is integration. The WFD aims at integrating, amongst others: i) environmental objectives - combining quality, ecological and quantity objectives; ii) all water resources - combining fresh surface water and groundwater bodies, wetlands, coastal water resources at the river basin scale; iii) all water uses, functions and values into a common policy framework; iv) disciplines, analyses and expertise, combining hydrology, hydraulics, ecology, chemistry, soil sciences, technology engineering and economics; v) stakeholders and the civil society in decision making, etc [1].
To achieve the WFD objective a number of activities need to be carried out, leading to an Integrated River Basin Management Plan (RBMP) in 2009. Such a plan will include programmes of measures, leading to the desired state of the water resources. Since the task of integration is very complex, it is believed that models and other ICT based tools are very valuable to structure and analyse all information and estimate effects of measures. It is also clear that there will not be one single instrument that can support all activities across Europe, different (spatial and temporal) scales within river basins, objectives, stakeholders, data-availabilities, (etc.). Furthermore, at different phases of the implementation different levels of detail will be required.

As a consequence there is an urgent need to

increase the flexibility of modelling systems, such that dedicated model systems can be developed from available building blocks. In Europe, the international project HarmonIT, which is sponsored by the European Commission, is developing and implementing a standard interface for modelling components and other relevant tools: The Open Modelling Interface (OpenMI). In this project a number of otherwise competing software vendors have joined forces. This is considered a key advantage to achieve the objective of setting a standard.
This paper describes the HarmonIT project and objectives in general. The current progress is described. The paper focuses on the implications for different types of stakeholders in modelling, varying from system administrators to non-specialized users of decision support systems.

2. THE HARMONIT OBJECTIVES AND BENEFITS

The aim of the HarmonIT project is to develop an 'Open Modelling Interface and Environment (OpenMI)'. The purpose of this environment is to facilitate the linking of models. This allows catchment process interactions to be represented in the formulation and selection of sustainable policies for catchment management [2]. Such a modelling interface and environment should resolve or improve a number of complicated linkage issues, such as for example: difference in spatial and temporal scales, feedback loops, differences in spatial and temporal concepts (distributed vs. lumped, steady state vs. dynamic), different units and naming of variables, distributed computing, etc.

Much knowledge and value is already implemented in existing (integrated) modelling systems. A key requirement of the OpenMI is to facilitate, with limited effort, the use of such valuable legacy software within the OpenMI environment.

Though focussing on data exchange between models in a runtime environment, the OpenMI standard is not limiting: linkages to databases, (real-time) monitoring devices are obvious uses of OpenMI. It can also be used for linking to end-user dedicated user interfaces, and be the base of gaming tools, assessment software (etc.) as well.

For the implementation of the WFD the OpenMI standard and environment will provide the means to speed up the development of both simple Decision Support Systems for non-specialist end-users, and linking complex models to increase insight and improve results. Improving the possibilities to swap models allows changing and tailoring integrated systems to emerging needs. From a scientific perspective the OpenMI will lead to an improved ability to model process interactions, the ability to use appropriate model combinations and the ability to swap in and out different models of the same process and hence facilitate sensitivity analyses and benchmarking. The OpenMI will further lead to increased choice for model users, an increased market for suppliers and increased opportunities for the creation of small-medium enterprises (SME's). Finally the OpenMI will allow better reuse of generic parts such as visualisation tools, thus a higher return on investment.

3. STATE OF PLAY & OPENMI ARCHITECTURE

In 2002 an overview of existing frameworks [4] and a HarmonIT requirements report [5] have been published. Based on this a framework architecture has been designed, which has been reviewed by an international panel of experts. Given some minor revisions the architecture is considered stable. The draft can be found for registered HarmonIT users on www.HarmonIT.com.

The core problem of the design was how to facilitate data exchange between models, databases and tools in a run-time environment. The solutions to this problem lie in two principles: component based design and the 'GetValue' mechanism.

Component based design implicates, that all models, databases and tools must be software components. For existing models this is achieved by 'wrapping' the model. All components have at least the same minimum set of properties, methods and events.

The main mechanism for (run-time) data exchange is the 'GetValue'-mechanism (figure 1). This mechanism is pull driven, meaning that a model that requires input asks a providing model for a valueset for a given time at (a set of) location(s). Three types of valuesets are defined: scalarsets (doubles), vectorsets or stringsets. The providing model calculates the required values and returns them. The GetValue mechanism is also used when getting data during e.g. initialisation of a model or retrieving data from databases or monitoring stations. Generic tools, such as for example a graph tool utilise a 'listener' mechanism: models signal when values are available and based on

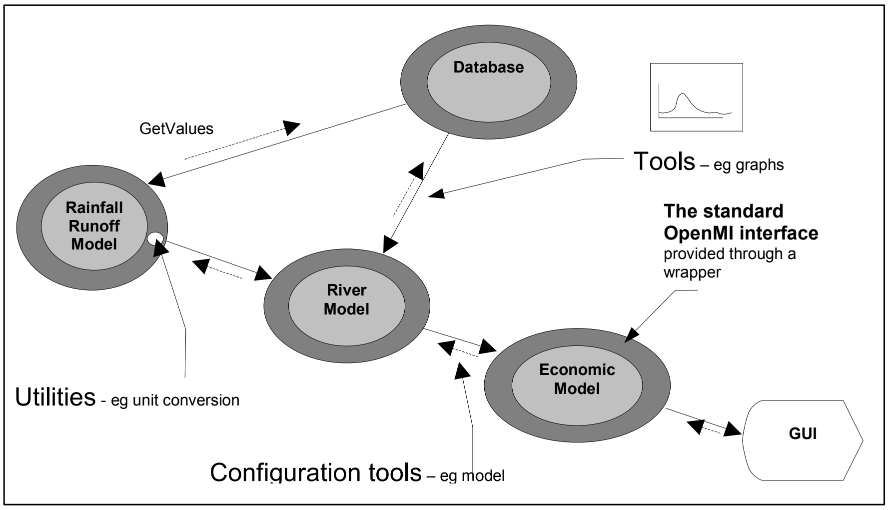


Figure 1: Facilities for wrapping, linking and running models [3].

such signals generic tools use the GetValue mechanism to retrieve the required data.

The overall architectural design is layered, consisting of the standard itself (system layer), and three supporting layers: utilities, configuration and tools:

- The system layer consists of a description of the mechanisms, all interface definitions and definitions to describe the models and exchange variables. In principle, this is all you need to work with the standard.
- Configuration files for the integrated model system can be created based on the system layer alone. To ease this process the configuration layer has been designed, which provides the tools that facilitates linking.
- The utilities layer is independent of the architecture as well, and may or may not be used by software developers. The layer does however contain a huge amount of functionality that is especially useful when existing (legacy) software needs to be adapted to work within an OpenMI environment. Adapting legacy software to the OpenMI is called migration, and typically the existing code will be 'wrapped' with some OpenMI compliant code. To keep the work on the legacy code to a minimum, the utilities layer provides a SmartWrapper, which contains helpful functionalities such as bookkeeping of links, unit conversions, buffering, time and space interpolations/extrapolations, etc. (see also chapter 5)
- The tools layer consists of a number of useful tools, such as visualisation tools, logging tools, optimisation controllers, iteration controllers etc. Here again, the OpenMI user is free to use any (OpenMI-compliant) tools.

In 2003 detailed designs of all layers have been developed, and core principles of the design have been prototyped. The first two tests concerned the functioning of bi-directional links in hydrological modelling. The more advanced of these tests consisted of splitting a river model in an upstream and a downstream model, which is visualized in figure 2. At the same time step the downstream model requires inflow from the upstream model, whereas the upstream model requires water level from the downstream model for calculations. Obviously, linking the models leads to the situation that both need input from the other model at the same time. To avoid a deadlock a model can provide an initial estimate allowing computations to start, and through iterations the system should approach the required computational and hydrological stability. The test showed, that in this particular case the difference between the original model and the linked parts were negligible, even without iterations. We find this result very promising: bi-directional will become increasingly important in integrated systems: Links between surface and groundwater models, river bedding plant growth model and water level river model, sediment model and river model are other examples of illustrating the use of bi-directional

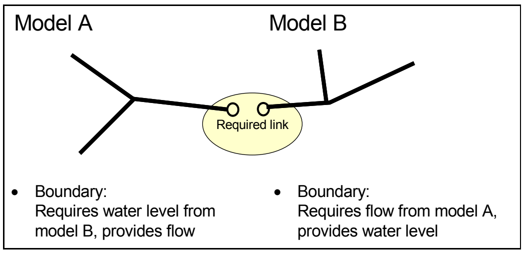


Figure 2: Visualization of the bi-directional link in the case of a split 1 dimensional flow model

links.

A third test involved running a distributed linked system over four computers, consisting of three models (each from a different partner) and a generic graph tool. This test has been successful, too, though the bi-directional link described above could not fully be tested due to technical limitations of the developer environment (.Net). We consider bi-directional links over multiple machines to be of less importance for practical situations. Furthermore, the problem encountered is a technical implementation problem, and independent of the OpenMI standard.

In this year (2004) the OpenMI is implemented, including utilities such as the SmartWrapper, configuration tools and generic tools. A number of models are migrated. Much work is carried out to provide guidance on how to use OpenMI. This guidance will be tested from December 2004 onwards, when a number of project partners not involved in the current implementation phase, will migrate a number of their models and tools. We believe that within the project duration partners will migrate about 30 models to the OpenMI environment, covering (aspects of) hydrology (channel flow, groundwater, etc.), water quality, ecology, economy, sewers, real-time control, etc.

4. HOW HARMONIT WILL WORK IN THE FUTURE

The way integrated modelling systems are developed and tailored to the user's needs in the future is visualized in figure 2. Starting from the bottom of the figure, models, tools and databases with an OpenMI interface will be developed by model and tool coders. This development will frequently be 'migration', which means that existing code is adapted, such that it meets the OpenMI Interface specifications. For the purpose of linking models it is only required to provide the following through the OpenMI interface: required inputs, outputs and some parameters for tuning integrated systems. What the responsibilities are of the model code developer is described in section 5. Modelling specialists will populate and tune individual models, such as for example groundwater models. For existing models, the already available user interface for populating models will be used, since, (as stated above) not all variables may be accessible through the OpenMI interface.

A model integrator, who specializes in establishing scientifically sound links between models, tools and databases, will link the populated models. Since an integrated system of validated models does not imply that the integrated system is valid, the model integrator will tune the system in co-operation with the specialist on the individual models.

On top of the integrated modelling system dedicated user interfaces for both specialist and non-specialist end-users will be build by the "end-user application builder". For non-specialist end-users this means that s/he may be able to change some settings, develop scenarios etc, and is not burdened with the working and parameters of all different models and tools in the background. We

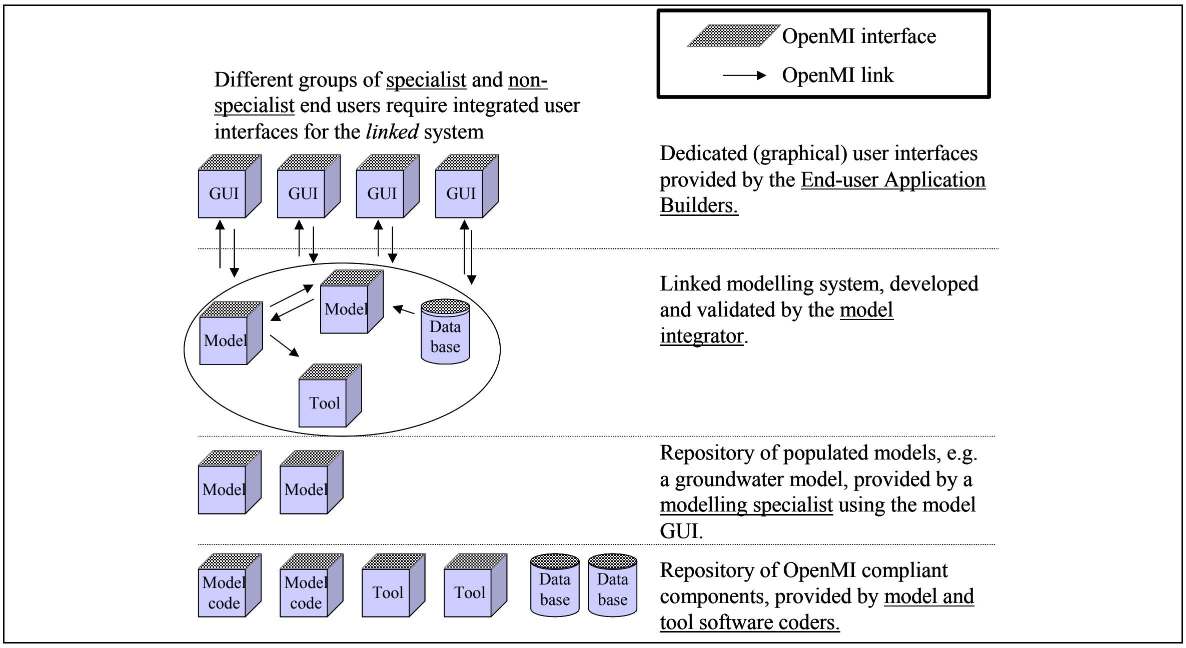


Figure 2: Visualization of the use of OpenMI in the future.

believe that the work of the end-user application builder will become increasingly important.

Not represented in figure 2 is the role of an organisation's computer (network) systems administrator. Especially in the case of distributed systems, the OpenMI environment needs to be able to cope with a variety of networks and their security properties.

5. THE RESPONSIBILITIES OF THE MODEL CODE DEVELOPERS

The model code developer's foremost responsibility is to ensure the compliance with the OpenMI standard. But, complying with the technical OpenMI is not the only task of the code developer. We assume, that the model code developer is also the expert who knows best how an existing system needs to be migrated. First of all s/he needs to decide if individual functional components or the entire model as such is migrated. Conceptually, and for parsimony reasons small functional units are preferred, but economic, commercial and practical reasons may lead to larger components (which internally can again be OpenMI compliant). Second, s/he needs to decide which data need to be exposed to the outside world both from the viewpoint of providing and accepting values. This work will usually also involve the modelling specialist. Which data need to be exposed is very dependent on the use of the model component. In many cases it will be sufficient to expose data that need to be exchanged by models. But if scenarios need to be calculated it may be necessary that underlying schematisations can be altered using the OpenMI interface. For example, widening a river's bedding and changing its overgrowth is a strategy to reduce flood risk. An end-user decision support system should thus allow calculating the effects of such measures. This means, that the schematisations of individual models need to be adaptable, e.g. the cross-sections and riverbed roughness parameters. There are different strategies to accomplish this and we believe that more and more features of models will become accessible through the OpenMI interface. It remains the responsibility of the code developers to decide if and how such data can be altered in an OpenMI environment.

Though the OpenMI design allows all models to calculate at their own time stepping and spatial representation, the GetValue mechanism requires that output can be delivered at any given time and place (reasonably within the models overall temporal and spatial boundaries). The model coder needs to make sure that interpolation and extrapolation mechanisms are in place: The HarmonIT project will provide a library of functions, which can be extended: it is the responsibility of the model coder to decide which functions interpolation and extrapolation functions are appropriate.
To achieve this the model needs to know at what time step it is. Based on this knowledge the model component may

- Start calculations, if the required time step has not been reached but can be computed.
- Extrapolate in time if the required time step cannot be reached.
- Search its buffer (if implemented) if the requested time step has passed, otherwise calculate required values)
- Interpolate between different time steps if the requested time step does not match the internal time stepping.
- (Dis-) aggregate in space and time, if the requested spatial and temporal representation does not match the internal representations.

To support the model code developer the SmartWrapper utility will provide a library containing some basic functionality to support the actions listed above. It is up to the model code developer to use the library or implement desired functionality.

In advanced integrated systems iterations and optimisations may be required. To facilitate efficient computing it is recommended that models are able to revert to an earlier state in the calculations. Again, it is the model code developer's responsibility to implement such functionality.

6. DISCUSSION

The HarmonIT team feels that it is progressing quickly in its effort to develop a standard for exchanging data in integrated water management. The HarmonIT initiative is all but alone in its effort. Dozens of initiatives are ongoing to build modelling frameworks and integrated systems. See for reviews [4,6]. None of the existing frameworks appears to be successful in achieving the status of a de facto standard. Why do we believe that the HarmonIT initiative is not just another attempt?

1) First of all, three commercial, otherwise competing organisations in hydrological modelling have joined forces and invested quite large amounts of their own R&D funds.

Together these organisations have a significant share in the market of hydrological and environmental modelling. The three organisations and many of the other project participants have significant expertise in integrated modelling, integrated modelling systems and IT frameworks. It is in their interest to use the standard, given the investments.

2) Secondly, the value of existing software at the participating organisations is extremely high, and all share the problem of increasing (customer) needs for flexibility, rising maintenance costs and the need to collaborate. As a result, the effects of the OpenMI design on existing software and on future requirements and development opportunities are continuously points of attention. A tangible result of this attention is the development of the SmartWrapper, which will support migration towards the OpenMI standard, making it both time and cost efficient.

3) The HarmonIT team spends much time and effort in documenting the developments from the perspective of different interest groups, making the OpenMI as transparent as possible and providing tailored guidance.

4) The HarmonIT team is actively preparing for an organisational structure to support OpenMI beyond the project's time line.

5) The designs are open to input from the outside world at early stages.

6) A significant number of well-known models will comply with the OpenMI at the project's end.

7) If required, the OpenMI system may be part of a larger system, and vice versa, existing (larger) frameworks and integrated modelling components may be included in an OpenMI environment as a single component.

8) Last, but certainly not least, the project is supported and reviewed by a number of leading experts in the field from all over the world.

We feel that the OpenMI architecture and associated tools and utilities will be an attractive way for model providers and model users to create systems for integrated catchment management.

We believe that on success OpenMI will be opening ways for advanced environmental research. The research community may benefit strongly: the water management community may adopt new scientific achievements much faster. Many aspects of integrated systems, such as the effects of integrating different spatial and temporal scales or modelling concepts can be researched in-depth. Helpful tools will make a researchers work more and more productive.

7. ACKNOWLEDGEMENT

The authors would like to thank their fellow partners in the HarmonIT project, originating from 14 different organisations, for their help and support in preparing this paper.
For more information please access the project's pages on www.HarmonIT.com.

7. REFERENCES

[1] CIS (Common Strategy on the Implementation of the Water Framework Directive), 2003, Best practices in river basin planning Work Package 2: Guidance on the planning process.

[2] Moore, R.V., 2001, Description of work for the HarmonIT project agreed during contract negotiation, IT Frameworks (HarmonIT), Proposal number: EVK1_2001-00037 (www.HarmonIT.com).

[3] Moore, R., Tindall, I. and D. Fortune, Update on the HarmonIT project – The OpenMI Standard for model linking, paper submitted to the 6th International Conference on HydroInformatics, 2004.

[4] Hutchings, Ch., J. Struve, S. Westen, K. Millard, and D. Fortune, 2002, State of the Art Review, Work Package 1, IT Frameworks (HarmonIT) Contract EVK1-CT-2001-00090, (http://www.harmonit.org/docs/repu_01_09_soa_review_approved.pdf)

[5] Gregersen, J, A. Malmgren Hansen, L. Michael, H. G. Enggrob, Requirements report, Work Package 2, IT Frameworks (HarmonIT) Contract EVK1-CT-2001-00090, http://www.harmonit.org/docs/harmonit_wp2_final_report_v2_approved.pdf

[6] Whelan, G. and Th. Nicholson, Proceedings of the Environmental Software Systems Compatibility and Linkage Workshop, NRC Report Number: NUREG/CP- 0177 (http://www.nrc.gov/reading-rm/doc-collections/nuregs/conference/cp0177/0177.pdf)

DIVA: An Iterative Method for Building Modular Integrated Models

J. Hinkel [a]

[a]Potsdam Institute for Climate Impact Research
GPO Box 601203
14412 Potsdam, Germany

Abstract: Integrated modelling of global environmental change impacts faces the challenge that knowledge from the domains of Natural and Social Science must be integrated. This is complicated by often incompatible terminology and the fact that the interactions between subsystems are usually not fully understood at the start of the project. While a modular modelling approach is necessary to address these challenges, it is not sufficient. The remaining question is how the modelled system shall be cut down into modules. While no generic answer can be given to this question, communication tools can be provided to support the process of modularisation and integration. Along those lines of thought a method for building modular integrated models was developed within the EU project DINAS-COAST and applied to construct a first model, which assesses the vulnerability of the world's coasts to climate change and sea-level-rise. The method focuses on the development of a common language and provides domain experts with an intuitive interface to code their knowledge in form of modules. However, instead of rigorously defining interfaces between the subsystems at the project's beginning, an iterative model development process is defined and tools to facilitate communication and collaboration are provided. This flexible approach has the advantage that increased understanding about subsystem interactions, gained during the project's lifetime, can immediately be reflected in the model.

Keywords: model integration; modelling framework; modular modelling

1 INTRODUCTION

This paper presents a method for building modular integrated models. The method was developed and first applied within the EU project DINAS-COAST (Dynamic and Interactive Assessment of National, Regional, and Global Vulnerability of Coastal Zones to Climate Change and Sea-Level Rise, www.dinas-coast.net). The aim of the three-year project is to develop a dynamic, interactive, and flexible CD-ROM-based tool that will enable its users to quantitatively assess coastal vulnerability to sea-level rise and explore possible adaptation strategies. Underlying are various climatic and socio-economic scenarios and adaptation policies on national, regional, and global scales covering all coastal nations. This tool is called DIVA, Dynamic and Interactive Vulnerability Assessment, and is centred around an integrated model.

DINAS-COAST was motivated by apparent limitations of previous global vulnerability assessments (Hoozemans et al. [1993], Baarse [1995]), including: the obsolescence of underlying data sources and the static, one-scenario approach. To overcome these limitations DINAS-COAST combines data, scenarios, and assessment models into an integrated tool, and makes it available to a broad community of end-users on a CD-ROM.

For the development of such a tool expert knowledge from the domains of Natural and Social Science must be integrated, calling for a modular approach to model development. Individual partners independently develop modules representing coastal sub-systems which are then "plugged" together to form one integrated model. While modularity is certainly necessary, it is not sufficient. The development of an integrated model is an organisationally challenging task. Efficient means of communication, methods to harmonise concepts, and a rigorously organised development process are essential for success.

Facing these challenges the DIVA method for integrated modelling was created. The method organises the development process and facilitates communication and cooperation. The actual DIVA tool is currently being built using this method. While the DIVA tool is specific to DINAS-COAST, the DIVA method is not and can be reused in other contexts with similar requirements.

This paper first analysis the DINAS-COAST requirements as perceived from the perspective of model integration and software development (section 2), then explicates some concepts of the modelling process needed for the following discussions (section 3). Section 4 explores the space of solutions to the requirements and section 5 presents the DIVA method as a possible answer. Finally sections 6 and 7 list some limitations and conclusions, respectively.

2 Requirement Analysis

The development of an integrated model faces several challenges. Knowledge from the domains of Natural and Social Science must be integrated. This is complicated by the often incompatible terminology, differing model types and modelling styles, and also by the fact that domain experts are distributed over various institutes worldwide. Frequent project meetings are not possible. Most of the model development must be coordinated via email, web-sites, and telephone calls.

While the requirements listed above are common to integrated modelling, some special challenges needed to be addressed in DINAS-COAST. Due to lack of an appropriate data source the model had to be developed simultaneously with its proper worldwide database (see Vafeidis et al. [2003]). The interactions between sub-systems were not fully understood at the start of the project; instead, such understanding is a major result of the project itself. Both circumstances necessitated a flexible model design that accounts for the incorporation of new knowledge in form of data, algorithms, or sub-system interactions at any stage during the development process.

3 Modelling Process

This section explicates four concepts involved in the modelling process that are needed for the following discussions.

1. **Ontology:** The modelling process starts with some concepts that we have at our disposal to perceive the world. It is good practise, especially in integrated modelling, to make this basic conceptualisation explicit. An explicit specification of a conceptualisation shall be called ontology.
2. **Mathematical Problem:** Based on the ontology a mathematical problem is formulated. For example one might have a system of differential equations and be interested in knowing the state of that system at a given time in the future.
3. **Algorithm:** Since in most cases the mathematical problem cannot be solved analytically, it's solution must be approximated by applying numerical methods. The result of this step is the numerical solution or the algorithm.
4. **Computer Model:** The last step considered here is the implementation of the algorithm in a programming language. This step yields the executable computer model.

4 Modular Integrated Modelling

An integrated model is composed of various submodels. It is evident that such a model, like other complex software, should be built in a modular rather than monolithic fashion: all contributers provide their knowledge about sub-systems in form of self-contained components (modules).

While modularity is a necessary answer for integrated modelling, it is not sufficient. Among others, four questions need to be addressed:

1. At which stage of the modelling process shall the integration take place?
2. What are the modules' interfaces or how shall the system be decomposed into sub-systems?
3. Which technology or software shall be used?
4. How shall the process of model integration be organised?

The following subsections explore possible answers to these questions and motivate the decisions taken in the case of DINAS-COAST.

4.1 Integration Level

The first question which arises in integrated modelling is at which stage of the modelling process the integration shall take place. Clearly, model integration has to start with a common ontology. Any attempt without a common conceptualisation of the system to be modelled is likely to fail. The remaining question is whether to integrate mathematical problems, algorithms, or executable computer models.

From an idealistic point of view models should be integrated at the level of the mathematical problems. Having a complete mathematical formulation of the system allows for careful selection of appropriate numerical methods and leads to stable and efficient algorithms. In praxis this route is seldom taken. Reasons for that are: the existence of legacy computer models; the need for a lot of cooperation at an early stage of the project; unclear linkages between sub-systems and that it is uncommon to “think” about integrated modelling in terms of mathematical problem specifications rather than algorithms and computer programs.

From a pragmatic point of view it makes sense to integrate or couple existing computer models. Legacy models, in which a lot of development time was invested, can then be reused. The flip-side of the coin is that the coupling of computer models involves a lot of technical issues, due to the heterogeneity in platforms, computer languages, compilers and data structures involved. A further disadvantage of this approach is that due to the absence of a complete specification of the mathematical problem it often remains unclear whether the numerics of the coupled computer models adequately represent the problem.

In the case of DINAS-COAST an intermediate approach was taken: the models were integrated at the level of the algorithms. Thus the project partners were free to solve their mathematical problem individually, but then had to implement the algorithms as modules in a common programming language. This route could be taken, because there were no legacy models to include.

4.2 Module Interfaces

An elementary question of any modular approach to integrated modelling is how the modelled system shall be de-composed into sub-systems or, phrased differently, what the modules’ interfaces are.

An efficient way of developing an integrated model would be to define specialised interfaces between the modules. However, a distinguishing feature of interdisciplinary research is that interactions between subsystems are usually not fully understood at the start of the project. Thus, general interfaces that leave the developers of modules with more freedom to define subsystem interactions are required. The flexibility offered by general interfaces is essential for taking advantage of the interdisciplinary learning process during the project’s lifetime. General interfaces also have implications on the development process. While specialised interfaces would not require extensive collaboration between partners developing the modules, general interfaces do, asking for a rigorously defined process of model development .

4.3 Technology

A wealth of methods and technologies from software engineering, like for example object-oriented programming or component technologies, are based on the concept of modularity. The necessity to build complex and integrated models has brought these techniques to the modelling communities and triggered the development of modelling frameworks.

Frameworks provide a conceptual frame, that is an abstract ontology for certain classes of the problems. Frames often support one (or several) modelling paradigms. For example an object-oriented framework for agent-based modelling might provide classes for agents, organisations, and environments. Models implemented in a framework use its basic concepts and specialise them further to their own needs.

An up-to-date overview of modelling frameworks developed within the environmental modelling community is given by Argent [2003]. Most approaches, just like the one presented here, tackle model integration at the algorithmic level of the modelling process. Consequently, sub-models must be implemented in a framework-specific language. The route to integrate existing computer models implemented in different languages or on different platforms is taken by Leimbach and Jaeger [2004]. Few approaches support model integration at the stage of the mathematical problem specification. A popular example is the M software environment (Jos de Bruin [1996]). Other mathematical approaches can be found within the the Decision Support Community. See Dolk [1993] for an introduction.

In the case of DINAS-COAST it was decided to develop a new framework. This was motivated by the will to provide the project partners with a very simple and efficient interface for expressing their knowledge. To this end the framework has to provide the “right” framing. If it frames too little a lot of coding needs to be done to express the specific problem. If it frames too much some aspects of the problem cannot be represented in the frame. A second motivation for developing something new was the aim to tightly couple the framework to tools supporting the actual (social) process of integrated model development.

4.4 Organisation

Model integration is an organisationally challenging and communication intensive process. While there is a wealth of modelling frameworks framing model design, there is little framing communication and the process of model development. Model documentation and meta-data are first steps in the right direction (see for example Rahman et al. [2003]).

The DIVA-Method emphasis and structures the process of integrated modelling. This necessity arose specifically from the requirement, that the model must be flexible to account for changes in interfaces, algorithms, and data-structures at any stage of model development.

5 THE DIVA METHOD

The DIVA Method consists of a conceptual frame (section 5.1), an iterative development process (section 5.2), a generic model (section 5.3), and a build and documentation tool (section 5.4). The DIVA Method was designed to be generic and can be applied to problems with similar requirements as DINAS-COAST.

5.1 The Frame

The DIVA-method provides, just like any other modelling framework, a conceptual frame for modelling. For modelling dynamical systems concepts for expressing static information about the system (data model) as well as concepts for representing the system’s dynamics are needed.

The statics of the system is represented by a relational-data model consisting of geographic features, properties, and relations. The geographic features represent the real-world entities, like rivers or countries. Properties capture the quantitative information about the features; e.g. a country might have the property area or a river the property length. Finally, relations describe how the features are structured; for example the feature region might contain several country features.

The dynamics of the system is represented by first-order difference equations: the state of the system is a function of the state at the last time-step and the drivers. All properties of the features must be classified according to the role they play in the system’s dynamics into the four categories: driver, state variable, diagnostic variable and parameter. For example the country’s area would most likely be static, that is a parameter, while its population might be driving the model.

5.2 The Development Process

The first step of the development process consists in defining the model’s ontology: Given the abstract frame the specific features, properties and relations which constitute the modelled system must be specified. The result of this step is the *list of system properties* containing the property names, the features they belong to, their type (that is whether they are drivers, parameters, or variables) and some meta-information. The compilation of this list is a joint responsibility of the project consortium.

The list of system properties is then automatically translated into Java source code. For each feature a class containing public member fields for the feature’s properties is generated. Relations between the features are represented by class composition. For example the source code of a feature country might look like this:

```
public class Country implements
   Feature {
      public float area;
      public int population;
      public Region region;
      ...
```

The class has three public member fields: the first two are for the feature’s properties (area and population) and the last one points to the region the country belongs to.

In the next step the project partners code the algorithms. They express their knowledge about the dynamics of the system in form of difference equations written in Java and using the generated features classes. Since now the model’s ontology is hard-coded in Java an algorithm will only compile

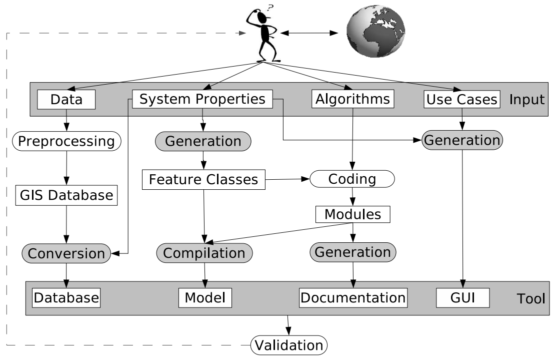


Figure 1: The DIVA Development Process. Boxes denote deliverables, ovals denote processes, and shaded ovals denote processes.

if it is consistent with the ontology. Related algorithms are grouped into modules. Before a module is submitted for inclusion into the integrated model it is run and validated individually.

The last step of the development process consists in the analysis of the modules, their linkages, and the validation of the complete model. Whenever a new or updated module is submitted the project's website is automatically updated offering documentations and the new model for down-load (see 5.4).

Figure 1 illustrates the work-flow of the development process. Knowledge about the modelled system enters the process via four categories: the list of system properties, that is the model's ontology; the modules, which express the functional relationships between the system properties; the data, expressing the actual state of the system and possible futures in form of scenarios; the use-cases, which specify the end-user requirements. Those four categories are interrelated: new data may create the need to change existing algorithms or develop new ones with the consequent need to update the list of system properties. Once the knowledge has entered the development cycle most of the subsequent processes are automated. The development process can be iterated as many times as needed. At any stage a complete model is available. This approach allows for rapid-prototyping of new models and their incremental refinement until a satisfactory result is produced.

5.3 Model

The integrated model consists of a generic kernel and a number of problem specific modules that are developed as described in the last section. The kernel is responsible for data input, data output, and the time-loop. It dynamically creates the data structures according to the list of system properties and the data files, sets the parameters, initialises the state variables, and reads the drivers. The kernel loads the modules at run-time and invokes them sequentially for each time-step. The modules to be loaded and their order of invocation are given in a configuration file. In the case of DINAS-COAST all modules operate on the same time-scale. The model, however, could be easily extended to support multiple time-scales.

Figure 2 shows the modules and the data-flow of the DINAS-COAST model. The model is driven by sea-level rise and socio-economic scenarios. The first modules invoked compute the geo-dynamic effects of sea-level rise, including direct coastal erosion, erosion within tidal basins, changes in wetlands, and the increase of the backwater effect in river basins. This is followed by an assessment of socio-economic impacts. The last module implements adaptation measures based on user-defined decision rules. These adaptation measures then influence the calculations of the geo-dynamic effects and socio-economic impacts of the next time step.

5.4 Build and Documentation Tool

A tool for building, testing, and documenting the model accompanies the development process. It takes the Java modules and the list of system properties as inputs and generates a web-site offering documents in various human- and computer-readably formats (HTML, XML, CSV and PDF). The documents include meta-information about the modules, the model, and the system properties, as well as documents used for the generation of the graphical user interface and the production of the model input data files. The tool also generates a diagram that shows the data flow through the system of modules (figure 2). The whole build and documentation process is fully automated: all documents are always consistent with the current model development status and available on the web.

6 LIMITATIONS

The flexibility of the iterative model development process comes at a price. The danger is that model development doesn't come to an end and not enough project time remains for model validation and application. Another drawback of this approach is that no complete mathematical problem specification is formulated. This is common to all approaches which

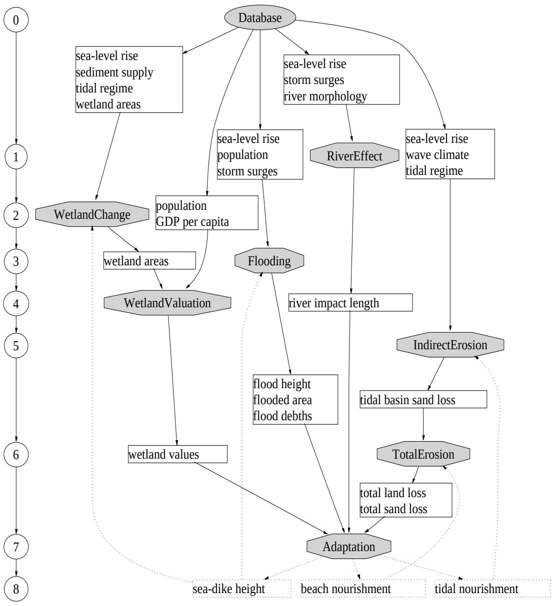


Figure 2: Module linkages in the DINAS-COAST model. Ovals represent the modules, boxes represent data, the drawn through arrows represent the flow of data during one time step, and the dotted arrows represent the data fed into the next time step.

integrate models at the algorithmic or computational level. Unintended model dynamics can result from that and more efficient numerical solutions cannot be found.

7 CONCLUSIONS AND OUTLOOK

The DIVA method is an innovative method for building integrated models by distributed partners. Unlike other integrated modelling frameworks it emphasises communication and the organisation of the development process. It provides scientists from different backgrounds with a way to harmonise their conceptualisations of the system to be modelled and an intuitive interface to express their knowledge about it. The process of model development is well defined and automatically documented. As a result, the status quo is constantly available on the web, providing a basis for efficient communication between project partners.

Within the project DINAS-COAST the DIVA method has been applied to build a tool for assessing the coastal vulnerability to sea-level rise. Meanwhile, application and improvement of both the DIVA tool and the DIVA method can go hand in hand. The global scientific and policy relevance of DIVA have already been recognised and collaboration on a range of initiatives is anticipated, including the EU ICZM (Integrated Coastal Zone Management) Strategy, and the new LOICZ (Land Ocean Interactions in the Coastal Zone) Science Plan. Improvements on the current DIVA tool could include a module for coral reefs and atolls, refining the adaptation module and increasing the spatial resolution of the analysis, thus increasing DIVA's usefulness to coastal management. In addition, it is conceivable to develop regional versions of the DIVA tool, such as a DIVA-Europe or a DIVA-India.

REFERENCES

Argent, R. M. An overview of model integration for environmental applications - components, frameworks and semantics. *Environmental Modelling & Software*, pages 1–15, 2003.

Baarse, G. Development of an operational tool for global vulnerability assessment. *CZM Centre Publication*, No. 3, 1995.

Dolk, D. R. An introduction to model integration and integrated modeling environments. *Decision Support Systems*, pages 249–254, 1993.

Hoozemans, F., M. Marchand, and H. Pennekamp. *Sea Level Rise: A Global Vulnerability Assessment: Vulnerability Assessments for Population, Coastal Wetlands and Rice Production on a Global Scale. 2nd revised editioni*. Delft Hydraulics and Rijkswaterstaat, 1993.

Jos de Bruin, P. d. V. M - a visual simulation tool. *Simulation in the Medical Sciences*, 1996.

Leimbach, M. and C. Jaeger. A modular approach to integrated assessment modelling. *Environmental Assessment and Modelling, in press*, 2004.

Rahman, J. M., S. P. Seaton, and S. M. Cudy. Making frameworks more usable: using introspection and metadata to develop model processing tools. *Environmental Modelling & Software*, pages 1–10, 2003.

Vafeidis, A. T., R. J. Nicholls, and L. McFadden. Developing a database for global vulnerability analysis of coastal zones: The dinas-coast project and the diva tool. *Remote Sensing in Transition*, pages 333–341, 2003.

Geospatial Interoperability in Modeling Frameworks - The ‘GEOLEM’ Approach

O. David[a] **, R. J. Viger**[b]**, I. W. Schneider**[c]**, and L. Garcia**[a]

[a] *Colorado State University, Fort Collins, CO, U.S.A.*

[b] *US Geological Survey, Denver CO, U.S.A*

[c] *USDA Agricultural Research Service, Great Plains System Research, Fort Collins, CO, U.S.A*

Abstract: Environmental models and modeling frameworks (MF) typically do not represent geographic information in a way that enables the direct translation of this information between geographic information systems (GIS) and the model or modeling framework. Parameters of the characteristics of geographic features are processed as part of a model’s mathematical solution and are thus explicitly represented in the model. In addition to the lack of semantic definition of geographic information in the environmental modeling process, current modeling approaches suffer from a lack of interoperability. The GEOLEM (Geospatial Object Library for Environmental Modeling) project forms a interagency working group which implements GEOLEM as a middleware solution (i) for the definition, storage, and manipulation of geographic metadata and (ii) for the transformation of information from the form of one context into another based on metadata specification (e.g. from the spatial data formats of GIS into the parameter organization of an environmental model). The purpose of this system is to eliminate the need for GIS-specific knowledge in the modeling framework and model-specific knowledge in the GIS. More specifically, GEOLEM will result in Modeling Frameworks being able to specify methods conceptually for the (i) Delineation of geographic features, (ii) Parameterization of geographic features, (iii) Visualization of model and GIS data entities, and (iv) the Exploration of model and GIS data entities.

Keywords: Modeling Frameworks, GIS, Interoperability, Modeling

1. INTRODUCTION

Environmental models and modeling frameworks typically do not represent geographic information in a way that enables the direct translation of this information between geographic information systems (GIS) and the model or modeling framework. Parameters of the characteristics of geographic features, such as area and volume, are processed as part of a model’s mathematical solution and are thus explicitly represented in the model. Information about the specific geometry or location of the geographic features whose behavior is being simulated is rarely represented explicitly in the model. The specification of methods needed to generate parameters describing the geographic features, or to actually delineate those geographic features in the first place, are never represented within the model. Further, the semantic definition of geographic feature types and their parameters are not represented within models. While environmental models are not expected to actually derive input geographic information, they should provide information about how to derive that information.

In addition to the lack of semantic definition of geographic information in the environmental modeling process, current modeling approaches suffer from a lack of interoperability. In all cases where GIS data are used in an environmental modeling context, it is necessary to develop algorithms or methods that provide for the translation between the two representations of spatially relevant information. Translation algorithms developed to date have been tightly designed to the needs and characteristics of specific models and GISs. As a result, two increasingly important types of problems present themselves: (i) resource, (ii) technical. The resource problem occurs because the translation algorithms are not, by and large, reusable. This means that each connection between a specific pairing of an environmental model and a GIS requires a unique translation algorithm, which, in turn, requires new resources to repeatedly solve the same conceptual problem

The more significant problem resulting from incompatible translation algorithms is technical in nature. Because each translation algorithm is unique to a model/GIS combination, sharing and coupling of environmental models and GIS methods is hindered. Once an environmental model or modeling framework has been "wired" to a GIS, it should automatically gain access to a wide array of community-developed geoprocessing libraries and geographic visualization software. To achieve this, standard protocols should be established for translating spatial information between GIS, environmental models, and other tools common to environmental modeling tasks. In addition, a metadata nomenclature should be established for referencing the array of geographic feature types relevant to environmental modeling. With these standards in place, generic sets of information translation adapters can be developed such that once a modeler has "mapped" their unique nomenclature to the standard nomenclature, transfer of information to and from GIS can more readily be automated. In addition, whenever a new geo-processing or visualization tool is developed in conformance with these standards, the tool is immediately available to the larger community.

2. BACKGROUND / RELATED WORK

Despite the lack of semantic description about geographic information in simulation models, tighter integration of GIS and simulation models has been sought. One approach has been to put "the model in the GIS". Although GIS and computing resources have evolved to a point that implementation of an environmental model within a GIS is possible, programming within the confines of a GIS has not become the norm. This is largely due to the relatively poor computational efficiency of the programming languages of most GISs and the burden of interacting directly with GIS data structures. Developers typically prefer to design models that ingest the simplest form of spatial data possible and concentrate their development efforts on the simulation of environmental processes. Examples of this have ranged from using GIS as a map-based interface for the selection of pre-existing data and model execution (US Environmental Protection Agency) to the creation of model inputs and model execution (Robinson and Mackay 1995) to a full integration of environmental models into spatial decision support systems (Taylor, Walker et al. 1999).

Alternatively, putting "the GIS in the model" has largely been rejected because the complexity of implementing GIS within models or model frameworks is not cost-effective. This approach has been most clearly adopted in groundwater models, where the modeling response units are delineated according to relatively simplistic methodologies such as finite difference or finite element meshes (McDonald and Harbaugh 1988). Although these models do not normally generate the original maps of modeling response units, they do exploit the spatial topological connections between units.

As a result, a third approach has gained popularity. GIS is used as a standalone pre- or post-processor for a model, reducing spatial data to the simplistic descriptions expected by the model. The USGS GIS Weasel is an example of this (Viger, Markstrom et al. 1998). A GIS operator usually works with a modeler to manipulate and digest spatial data into a file or set of files that will eventually be read by the model. The knowledge that was used to apply the GIS appropriately usually resides in the mind of modeler and the GIS operator. This knowledge is not normally formalized or codified. In the case of well-established models, dedicated GIS software applications may be developed as pre- or post-processors. Although these applications do serve to codify the knowledge used to delineate geographic features and derive parameters of those features, they fail to enable the re-use of those geo-processing methodologies in newly created models.

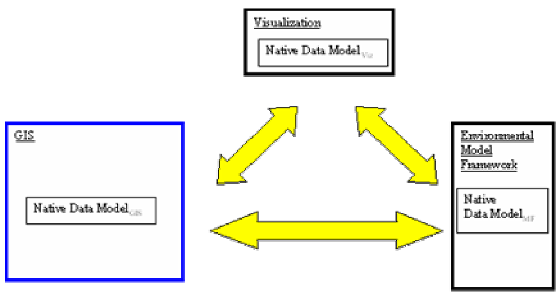


Figure 1. GIS, VIZ, and MF components with distinct internal "native data models" and custom integration directly mapping between specific native data models.

A fourth approach, depicted in Figure 1, seeks a looser coupling of GIS, MFs, and visualization software (Viz). This configuration relies on communication between discrete software components, rather than merging functionalities of disparate components into a monolithic piece of software (Leavesley, Grant et al. 1996).

This approach towards the integration of GIS and environmental models will be used as a starting point for the research proposed here.

3. GEOLEM DESIGN

In order to be able to allow software components to more readily interconnect in a generic way, Figure 2 shows a middleware architecture that allows the most effective data model for each software component to continue being used by each respective component, yet facilitate the movement of information across these contexts. This middleware is referred here as the Geographic Object. The authors seek to leverage the ideas of the OGC Geographic Object initiative (OpenGIS Consortium 2003), and participate, if feasible, in this effort. One way to describe the role of the Geographic Object is that it maps the relevant details of one context to those of another.

3.1 Objective

The objective of GEOLEM is to enable the systematic integration of MFs and GIS during (i) pre-run, (ii) run-time, and (iii) post-run phases of modeling. In addition, this effort seeks to eliminate the need for GIS-specific knowledge in the environmental model and environmental model-specific knowledge in the GIS. More specifically, this effort will result in MFs being able to specify methods for the

- Delineation of geographic features
- Parameterization of geographic features
- Visualization of model and GIS data entities
- Exploration of model and GIS data entities

Central to achieving these goals is the development of template metadata specifications for geographic information to be used by models developed in the MFs. The developer of an earth science model will be able to use previously specified types of geographic features in that model. These template specifications will not significantly alter the style of data model (e.g. arrays of parameters) typically found within traditional earth science modeling components.

3.2 Approach

This mapping is represented by chains of arrows that connect one external component to another. Consider the chain at the bottom of Figure 2, representing the MF requesting information that is generated in the GIS. The MF communicates with the Geographic Object, requesting information in a format suitable to the MF. The Geographic Object understands that the information requested by the MF corresponds to some GIS-based information. Based on this understanding, the Geographic Object requests the appropriate information from the GIS. Once this GIS information is returned from the GIS to the Geographic Object, the Geographic Object then uses its understanding of the MF-GIS correspondence to return a set of MF-appropriate information to the MF. The broad arrows represent component specific communication. The thin, black arrow represents the work that the middleware does to translate information from one context into another. In our example, this could be the reduction of shape files to arrays of parameters.

The benefit of this mapping is that the MF does not need to understand how to make low-level requests to the GIS, nor does it need to know how to extract what it needs from the alien data formats of the GIS. It needs only to know how to make request to the Geographic Object for the higher-level information that it needs.

To extend the sample explanation of the arrow chain started above, let the information requested by the MF be the elevation of hillslopes within a watershed. The Geographic Object will understand that the MF ultimately wants to receive an array of real numbers, because the name of the parameter within the MF, say `elev`, has an association within the Geographic Object middleware to a description which states the numerical format of elev. The Geographic Object will also have a specification that the method used to derive the real numbers in the array `elev` is to find the median value in the distribution of elevations within each hillslope. In addition, the Geographic Object middleware will know about a hillslope and how it should be derived. Although it is obvious to human users that in order for the median elevations of hillslopes to be derived, hillslopes must first be delineated, this information is not known to the MF which simply knows that it needs an array of real numbers. The Geographic Object will manage this relationship

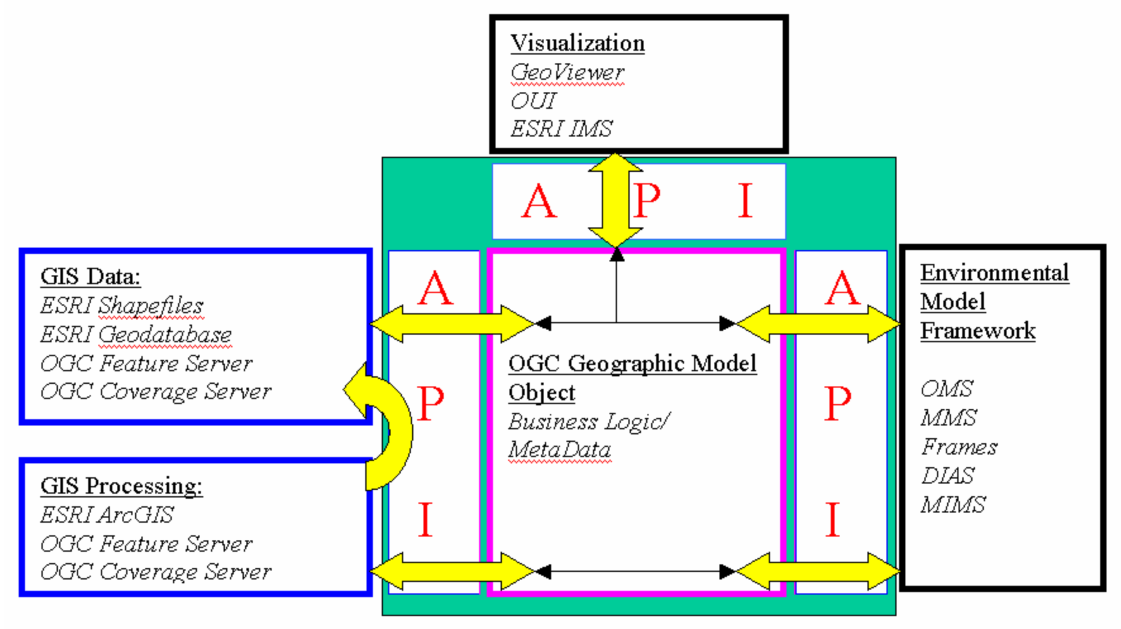


Figure 2. Integration of software components based on middleware (OGC Geographic Object) conceptual model.

4. IMPLEMENTATION

Figure 3 shows how an environmental simulation model can access GIS functionality. The model is represented on the left, in this case by PRMS (Precipitation Runoff Modeling System, Leavesley et al.,1983). The model is intended to make high-level calls to GEOLEM, based on the internal vocabulary of the model. The rationale is to avoid as much as possible changing the internal workings of the pre-existing model or modeling framework. Requests will typically be for parameters. GEOLEM will be able to take these high-level requests and translate their meaning into generic terms.

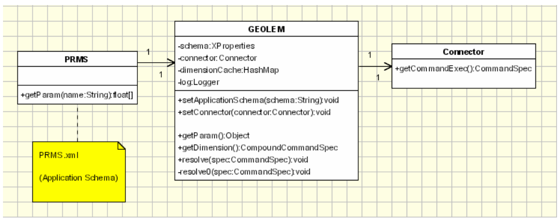


Figure 3. General structure of the connections between an environmental simulation model, GEOLEM, and a GIS server.

There are three different linguistic contexts in the scenario described above: (i) the environmental simulation model, (ii) GEOLEM, and (iii) the GIS server. In order to communicate across these contexts, two different translations are made. The first is from the language of the model into that of GEOLEM and the second is from the language of GEOLEM into that of the GIS server. The first translation relies on what is referred here as the conceptual schema for the environmental model. This is a metadata store that encodes the conceptual model, described in the previous chapter, for geographic information within the environmental simulation model. The conceptual schema is configured to relate the elements of conceptual model to analogs available within the GEOLEM library.

4.1 GEOLEM Functions

The GEOLEM library, which is not shown in Figure 3, contains several types of functions: commands, compound commands, and parameter providers. These functions are intended to provide a generic way to use basic GIS functions. These basic functions can be grouped to develop more complex geo-processing methodologies or they can be used to extract information from a GIS. As noted in Figure 3, GEOLEM is not a GIS itself and relies on the existence of a GIS server with which it can communicate.

4.1.1 Commands

Commands are used here to denote simple, single-step functions. Examples of some commands could be “calculate aspect” or “derive a watershed”. Figure 4 depicts the class hierarchy used to create such a command.

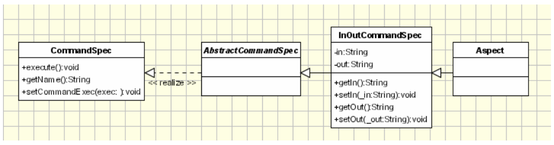


Figure 3. Inheritance Hierarchy of the simple GIS Command, Aspect.

The *CommandSpec* box on the left is an interface, meaning that it is merely a general definition that is not actually implemented. It serves to set a minimum level of functionality that all descendent functions must support. The *AbstractCommandSpec* is the basic implementation of the *CommandSpec* interface. Note that this implementation is denoted by the dashed arrow accompanied by the <<realize>> label. At the right of Figure 3, the Aspect class is what could be thought of as an actual GIS command. It extends the *InOutCommandSpec*, which in turn extends the *AbstractCommandSpec*. Extension, a basic principle of object oriented design, effectively allows new functionality and properties to be added to a general class. The new functionality and properties are encoded in a sub-class. The sub-class, by referencing the more general super-class, will gain all the functionality and properties that existed in that original class. *InOutCommandSpec* merely serves as a helper to add some functionality that is likely to be widely used by other actual GIS commands. Commands like Aspect, Slope, *FlowDirection*, etc. can simply extend *InOutCommandSpec* and avoid having to re-implement the exact same functionality.

This helper implements the setting of the names of inputs and outputs to the command.

- Commands as indivisible, atomic units of geoprocessing functionality
- These devices allow the GEOLEM compound commands to develop sophisticated methods of reasoning, sometimes described as business logic, without the constraints common to many of the languages associated with GIS.

4.1.2 CompoundCommands

Compound commands are intended to allow sequences of simple commands to be created. In addition, compound commands allow logic to be associated with these sequences. Implementations of this interface are intended to provide a way to encode high level representations corresponding to the semantics of a type of geographic feature. The *CompoundCommandSpec* box at the lower left of Figure X shows that this is an interface which extends the *CommandSpec*. Classes that implement *CompoundCommandSpec*, such as *HillslopeMethod*, are able to enumerate all of the *CommandSpec* objects that will be referenced within that class. Put another way, a compound command is able to reveal all of the simple GIS commands that it will use (with the *getCommandSpec*() method). The significance of this will be discussed below. In addition, the implementation of a *CompoundCommandSpec* interface is expected to extend the *AbstractCommandSpec*, thereby gaining access to standard methods such as *execute*().

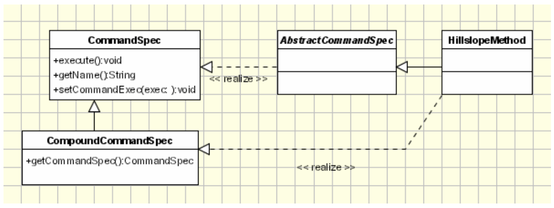


Figure 4. Inheritance Hierarchy of a Compound Command, HillslopeMethod

4.1.3 ParameterProvider

The third main type of function that GEOLEM exposes is the parameter provider. This concept is intended to be a generic way to derive new information, most likely parameters, based on an input map of geographic features and some methodology fixed within the implementation of the parameter provider. This idea is represented by the ParameterProviderSpec interface, shown at the upper left of Figure 5 below. The interface defines two significant methods. The first is getParam(). This method is explicitly designed to return a data object, some form of which will ultimately be returned to the environmental simulation model. Most commands and compound commands return only a character string indicating whether or not an operation has succeeded. The second method, setDimension(), associates the particular parameter provider with the input map of geographic features, alluded to above. The term dimension, introduced in the previous chapter, is used to refer to the map of geographic features that the methodology contained within the parameter provider will be applied to. Figure 5 shows an example of a parameter provider implementation designed to derive the median elevation for a set of geographic features (each feature is regarded as a zone in this context).

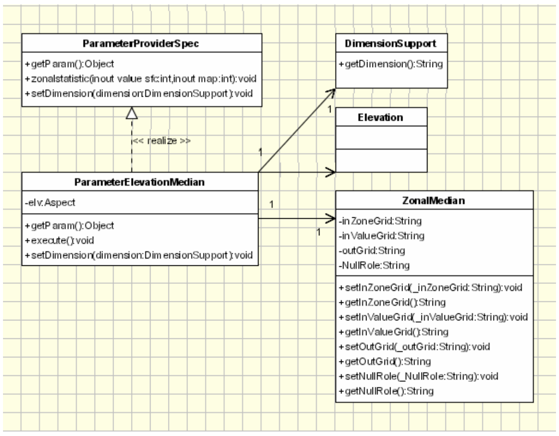


Figure 5. Inheritance Hierarchy of a Compound Command, HillslopeMethod

Besides these three concepts there is the Core GEOLEM class which maps the specifications of basic GIS commands to some GIS server. It uses XML configuration files, which describe how a generic GIS specification gets mapped into a GIS call. This can be understood by a real GIS. Up to now, there are GEOLEM prototype bindings to ARC GIS 9.0 beta via JNI/Python and COM.

5. APPLICATION PRMS

For the PRMS model a following scenario can be applied to derive the parameter hru_elev (the elevation value for each hydrological response unit) using GEOLEM:

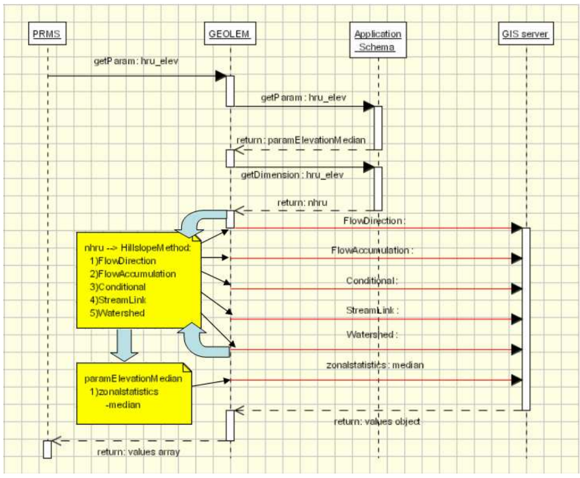


Figure 6. Scenario diagram depicting the PRMS usage of GEOLEM

The delineation *hru_elev* requires the application of *CompoundCommand* *ParameterElevationMedian*. The *HillslopeMethod* itself uses simple GIS commands such as *FlowDirection*, *FlowAccumulation*, etc. and a *ZonalStatistics* command to generate a value array, which can be consumed by the model. The sequence diagram shown in Figure 6 depicts the sequencing of interactions between several components in GEOLEM. The Application Schema object describes in XML the dependencies of the *hru_elev* parameter, its dimensions, data input, and optional unit conversion (Figure 7). Such ApplicationSchema represent the backbone of GEOLEM. They express the model/modeling framework requirements of model parameter data.

```
<xprop name="geolem">
 <xprop name="applicationschema">
  <xprop name="prms">

   <xprop name="hru_elev">
     <entry name="cmd">
   geolem.spec.gp.ParameterElevationMedian
      </entry>
     <entry name="dimension">
        nhru
      </entry>
     <entry name="type">
        double
      </entry>
     <entry name="units">
        feet
      </entry>
   </xprop>

   <xprop name="nhru">
     <entry name="cmd">
       geolem.spec.gp.HillslopeMethod
     </entry>
   </xprop>

  </xprop>
 </xprop>
</xprop>
```

Figure 7. GEOLEM ApplicationSchema fragment for PRMS

A detailed description of the XML syntax would exceed the length and scope of this paper, but can be found under http://oms.ars.usda.gov/geolem.

A model such as PRMS which is using the ApplicationScheme in Figure 7 is only required to add the code

```
double hru_elev[] =
   (double[])GEOLEM.getParam("prms/hru_elev");
```

to obtain the HRU elevation data as a array of doubles. This call causes GEOLEM to invoke the orchestration of the lookup scenario shown in Figure 6 to generate these values.

6. CONCLUSIONS

The GEOLEM effort represents a significant undertaking to simplify the usage of Geographical Information Systems for models and modeling frameworks. It is actually prototyped with frameworks such as the Object Modeling System (OMS) (David et.al 2002) amd PRMS and will be adapted to FRAMES in conjunction with ESRI GIS products. A prototype implementation using the ESRI ARC GIS 9.0 demonstrated the feasibility to delineate and derive parameter for PRMS in OMS.

The overall expected benefits of the GEOLEM project and its implementation can be summarized as follows:

- Easier exchange of scientific expertise due to improved interoperability of spatial modeling applications in modeling frameworks among the agencies and other institutions
- Leveraging and saving the investments being made in simulation model development/GIS adaptation, data-management and visualization for all project partners and avoid duplication of development efforts.
- Highly efficient integration of new spatial simulation models into existing GIS solutions and easier adaptation of models on new versions of COTS GIS packages without model reimplementation.
- Establishing an informal and formal collaboration platform by means of a model metadata standard and template reference library implementation, which is seen as the foundation for other interagency efforts like “Data Representation and Interchangeability”
- Offer model developer a path for “better” handling of meta information for future model application by providing “executable”, operational meta-info; meta-info becomes a part of the execution model, rather being optional documentation “sugar”.

7. ACKNOWLEDGEMENTS

This work is supported by the Interagency Steering Committee (MOU) on Multimedia Environmental Models, Workgroup 1 on ‘Software Systems Designs and Implementation for Environmental Modeling’. It is funded by the Environmental Protection Agency EPA. Thanks to Gerry Laniak (from EPA) for his visionary lead of the working group and the support of this effort.

8. REFERENCES

David O., S.L. Markstrom, K.W. Rojas, L.R. Ahuja, and I.W. Schneider (2002). The Object Modeling System, In: Agricultural System Models in Field Research and Technology Transfer, L. Ahuja, L. Ma, T.A. Howell, Eds., Lewis Publishers, CRC Press LLC, 2002: 317—331.

Leavesley, G. H., R. W. Lichty, et al. (1983). Precipitation-runoff modeling system--User's manual, U.S. Geological Survey: 207.

Leavesley, G. H., G. E. Grant, et al. (1996). A modular modeling approach to watershed analysis and ecosystem management. Watershed '96 A National Conference on Watershed Management, U.S. Government Printing Office.

McDonald, M. G. and A. W. Harbaugh (1988). A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model, U.S. Geological Survey

OpenGIS Consortium (2003). GO-1: Geographic Objects Initiative, OpenGIS Consortium. 2003.

Robinson, V. B. and D. S. Mackay (1995): "Semantic modeling for the integration of geographic information and regional hydroecological simulation management. Computers Environment and Urban Systems 19(5-6): 321-339.

Taylor, K., G. Walker, et al. (1999). "A framework for model integration in spatial decision support systems." International Journal of Geographical Information Science 13(6): 533-555.

Viger, R. J., S. L. Markstrom, et al. (1998). The GIS Weasel - An Interface for the Treatment of Spatial Information Used in Watershed Modeling and Water Resource Management. First Federal Interagency Hydrologic Modeling Conference, Las Vegas, Nevada.

Development of Multi-Framework Model Components

Robert M Argent[a] **and Andrea E Rizzoli**[b]

[a] Cooperative Research Centre for Catchment Hydrology, The University of Melbourne, 3010, Australia. R.Argent@unimelb.edu.au

[b] IDSIA, Manno, Switzerland, andrea@idsia.ch

Abstract: A number of environmental modelling frameworks have been developed recently, and plans for new frameworks are under way. Examples such as TIME, OpenMI, SME and OMS share an approach to environmental modelling based on model components, and offer improved model development and deployment. These approaches have methods for ensuring model component-linking compatibility using manual and machine processes either internal or external to the model component. Examples include matching output to input and checking data type compatibility. Semantic integration is also possible, such as with the OpenMI, where a component requests and receives particular data. However, each framework does model component checking in a different way and interoperability between model components of different frameworks is limited. To improve the use of model components it is necessary to consider the development of multi-framework model components (MFMC). Existing software standards enable communication at a low level, but many problems remain at high levels. This paper discusses development of an MFMC in each of TIME and the OMS, that can be accessed from the other framework. Additionally, the requirements for further framework compatibility, such as the OpenMI, are considered. Six main approaches are described, covering methods relevant to both between- and cross-platform compatibility, which range from re-implementation, through Web Services, to declarative modelling. Web services are suggested as a viable option for the problem considered here, although the other techniques warrant further investigation in particular cases.

Keywords: Component-based modelling; Modelling frameworks; Model development, Multi-framework model components.

1. INTRODUCTION

Development of new environmental modelling ideas has been taking place over recent years. Two primary ideas, of simplified modelling focussing on dominant processes and using parsimonious methods, are being combined with modular or component-based modelling ideas. In this approach components are created that represent discrete system functions, and these functions are then combined flexibly to address particular problems using a tailored modelling approach.

These developments have been associated with the production of various support tools, the most significant of which are the modelling frameworks, or development environments, within which developers can build and test components, and click components together, and together with data and visualisation components, to build multi-component models. One of the problems in this approach, and a major problem of the environmental modelling science world, is that a component build by and for one framework is not directly compatible with another framework.

The problems of making algorithms from one component into a useable form for other components can be addressed using a range of approaches. This paper explores these approaches for the development of model components for porting between three specific modelling frameworks – the OpenMI, OMS, and TIME. Furthermore, the semantic issues are explored further to identify some of the critical issues that need to be addressed before appropriate components for use in multiple frameworks can be identified and developed.

2. PORTING MODEL COMPONENTS

We have identified at least six methods to port a model component across modelling frameworks, but we must first distinguish between whether or not the modelling frameworks are implemented using the same software platform.

Where components are based on the same platform, with binary level compatibility, porting options are a) *interface extension*, and b) *on-demand data binding.* For different platforms the options are: c) *manual re-implementation*; d) *cross-compilation*; f) *web-servicing*, with declarative interfaces; and e) *declarative modelling.*

We now examine the advantages and the disadvantages of these approaches.

2.1 Binary Compatible Modelling Frameworks

This is the case where the modelling frameworks use the same software architecture (e.g. two frameworks are implemented using the .NET software architecture).

2.1.1 Interface Extension

This approach proposes development of a different model component interface for each framework. The model component code stays mostly unchanged when being used in a new framework, and an interface or wrapper has to be developed. This activity, to be done by hand, may take a lot or little time depending on the requirements and features of the frameworks.

2.1.2 On-Demand (XML) Data Binding

This approach involves the specification of the model component interface by means of a standard XML description. The description is then parsed by software tools that automatically generate the source code for the model component interface, targeted to the appropriate modelling framework. The source code is then compiled and linked into the modelling framework. This approach has been called XML data binding (McLaughlin, 2002)

This option, while possibly requiring more effort up-front in the definition of an XML schema to define the interface data formats, may be better in the long term, as new components can be readily created to be multi-framework compatible. It also has the advantage of being easy to implement both in J2EE, using JAXB (Java Architecture for Data Binding: http://java.sun.com/xml/jaxb) and in .NET, using `xsd`, the W3C Schema Definition Tool. A nice feature of this approach is that the same XML interface can be used to generate code for modelling frameworks based on both Java and C#.

2.2 Heterogeneous Modelling Frameworks

This is the case where the model components are designed and implemented to work under incompatible modelling frameworks (e.g. one modelling framework is based on the .NET architecture, and one on J2EE).

2.2.1 Manual Re-implementation

This is the oldest and most common approach to make a model component re-usable in another modelling framework: make public the algorithms, code and explanation of the science in the component. This model component can then be manually transformed for use in another framework. This transformation will be of varying complexity depending on the nature of the algorithms, the language/s used for the components and the features and capacities of the alternative frameworks. It is the most straightforward approach as it requires only a single unit block of work to produce a workable solution. However, over time this may become the most time consuming, as every model component needs to be broken open, re-written, de-bugged and tested.

2.2.2 Cross-Compilation and Translation

To limit the programming effort, model components written for one modelling framework can be ported to another using cross-compilation and translation tools. For instance, JNBridge (www.jnbridge.com), an implementation of the .NET Remoting wire-level protocol, and Remotesoft's Java.NET (www.remotesoft.com) allow translation of a model component written in Java into a corresponding C# component. However, in this approach the interfaces may be not compliant with the specifications issued by the target modelling framework, and therefore interface extension or XML data binding might be required.

2.2.3 Web-Servicing

Model components can be implemented as web-services, providing a published interface that is remotely callable. The use of a standard protocol, such as SOAP, to exchange data, and the fact that the model component resides on a remote server, enables interoperability among non binary-compatible modelling frameworks. This approach requires that the modelling framework knows about the interface of the remote model component. Thus, we fall back into one of the two approaches for binary compatible systems: interface extension or XML data binding. An example of the latter can be found in Rizzoli *et al.* (2001).

2.2.4 Declarative Modelling

This is the most generic approach to component porting, but is also the most difficult to bring into common practice and to implement.

The basic idea is to shift from a *procedural* approach to modelling to a *declarative* one. In the procedural approach, models are written as sets of instructions for simulating the model, written in the programming language of the modelling framework, for instance C# or Java. In the declarative approach, models are represented as a set of statements defining the structure of the model. These statements are written in a text file, using a standard and open format (again, an XML schema can be useful). Experiences of declarative modelling are found in the Simile modelling environment (Muetzelfeldt and Massheder, 2003) and in the Integrating Modelling Architecture (Villa, 2000). A declarative model can be processed and transformed automatically, by means of a model compiler, into a model component targeting any modelling framework.

The main disadvantage with this approach is that every modelling framework needs to adhere to a standard and common declarative modelling language, which currently exists only in some domain. A major advantage is the ability to link elements declared in the model with entities declared in distributed ontologies, thus reducing the risk of ambiguities and misuse of data, which are quite frequent in all other approaches

This overview provides a range of techniques for developing multi-framework model components, with the primary determinants being the platform and software structure used by the associated modelling frameworks. The following explores development of a simple component model in the TIME framework, and the requirements for operating this component in the OMS framework and making the component compatible with the OpenMI.

3. TIME – THE INVISIBLE MODELLING ENVIRONMENT

TIME is an environmental modelling framework constructed using .NET (Rahman *et al.*, 2003). Its primary features are a thin architecture and a strong capability to use model metadata. By adding metadata about parameters and variables, control of the model by the TIME system can be automated in many ways. One of these is to support automatic creation of user interfaces. If a model has a declared parameter with metadata regarding range and default value, then a user interface consisting of, say, a slider bar ranging between the extremes of the allowable range, can be readily generated. Also, if metadata information on data types relating to inputs and outputs is given, then the linking behaviour of two component models can also be controlled by an intelligent model management system, by means of XML data binding. Another feature is the use of component technology, in that components are designed to be multi-purpose, either being run with command line techniques, to support multiple runs for, say, stochastic modelling, run remotely, for Web Service applications, or have attached an automatic GUI, providing control over parameter values for scenario exploration.

3.1 The Component

The selected component was one of the more simple components that can be developed, and one which has often been used for examples in development of the TIME modelling environment (Rahman *et al.*, 2003). This is a rainfall runoff model that uses a runoff ratio to create runoff from rainfall (equation 1), which can be run once or for every time step in a temporally dynamic simulation, and which has spatial application to a point, polygon or any cell in a raster.

$$Ro = RR * (Rain - ET) \quad (1)$$

Where

Ro is runoff

RR is runoff ratio

Rain is precipitation

ET is evapotranspiration

In a Java version of the component in TIME, the component code is as follows:

```
package
TIME.Models.Examples.RainRunoffCoefft;

import TIME.Core.*;

/**
 * Rainfall runoff coefft test.
 * @author David Verrelli
 */
```

```
public   class   RunoffCoeff   extends
Model
{
  /** @attribute Input() */
  double Rainfall, PET;

  /** @attribute Parameter()
  *  @attribute Minimum(0.0)
  *  @attribute Maximum(1.0) */
  double Coeff;

  /** @attribute Output() */
  double Runoff;

  //Constructor
  public RunoffCoeff()
  {
   super();  //This is implicit
  }

  public void runTimeStep()
  //This runs at every timestep.
   {
   Runoff = Coeff * Math.max(Rainfall
   - PET, 0.0);
   }
 }
```

This kind of straightforward TIME routine has three basic sections:

- An initiation section, containing basic declarations of the package, any external library requirements, class declaration and, if required, parent class declaration
- declaration section for parameters and variables, and
- run section, wherein lies the core code for the algorithm

The component above extends the abstract parent class `Model`, which controls the timestepping of the model through a `runTimeStep()` abstract method, and also keeps track of the progress of the model run. `Model` has a `Subject` parent class that, together with an `Observer` class, provides a communication structure built on the Observer pattern (Gamma *et al.*, 1994). `Data` is also an abstract parent class, sharing the `Subject` parent class with `Model`. All child classes of `Data` implement the `setItem (in i: int, in val: double)` method and the `item( in i: int)` query to provide a common data interface (Rahman *et al.*, 2003).

A similar code structure to the above could be used to write the model component in other languages, such as C#, FORTRAN95, Eiffel, and Visual Basic, each of which would access the same parent classes and methods.

4. THE OBJECT MODELING SYSTEM - OMS

The OMS is a modelling framework written in Java, developed by members of the USGS, USDA and Friedrich-Schiller University. Modular modelling lies at the core of the OMS with a structure that clearly separates the system core, system extensions, and the user interface. The core provides the functionality for basic module operation, data handling, input-output, visualisation and remote access. Extensions cover features such as module development, application construction and the management of 'dictionaries' that covers, to some degree, the semantics of interaction between modules.

A model component, which performs a similar computation to the one examined above, can be implemented in the OMS modelling framework. The model component also contains three basic sections: an initialisation section, a run section and a clean-up or handover section, as shown below.

```
/**
 * RunOffCoeff.java
 * @author adapted from Sven Kralisch
 */

package de.unijena.jenamodel;
import org.omscentral.data.*;
import org.omscentral.model.*;
import java.io.ObjectInputStream;
import org.j2k_io.j2kBinFileHeader;

public class RunOffCoeff extends
OMSComponent {
  transient  OMSTimeInterval time;
  transient  OMSEntitySet es;

  /** @attribute Input() */
  transient public double Rainfall =
0;
  transient public double PET = 0;

  /** @attribute Parameter() */
  transient public double Coeff = 0.2;

  /** @attribute Output() */
  transient public double Runoff = 0;

  // Constructor
  public RunOffCoeff() {}
  public void register() {}
  public int init() {return 0;}
  private void initData() {
    OMSEntity currentEntity =
this.es.current;

  try {
```

```
  Coeff = Double.parseDouble((String)
currentEntity.getAttribute("Coeff"));
  Rainfall =
Double.parseDouble((String)
currentEntity.getAttribute("Rainfall"
));
  PET = Double.parseDouble((String)
currentEntity.getAttribute("PET"));

}
 catch
(org.omscentral.data.OMSEntity.NoSuch
AttributeException nsae)
{System.out.println("Attribute not
found");
}
}

  public int run() {
    initData();
    double Temp = 0.0;
    if(Rainfall> PET)
      Temp = Rainfall-PET;
    else
      Temp = 0.0;
    Runoff = Coeff * Temp;
cleanup();
return 0;
  }

public boolean cleanUp(){
  OMSEntity currentEntity =
this.es.current;
try {
currentEntity.setAttribute("Runoff",
new Double(Runoff));
}
 catch
(org.omscentral.data.OMSEntity.NoSuch
AttributeException nsae) {
System.out.println("Attribute not
found");
}
return true;
}
}
```

Comparison between this component and that built for TIME highlights some similarities and differences between the two. In terms of control structures and looping or stepping through time and space, both have a similar approach, with the 'run part' of the component, containing the operational algorithm, being separate from these structures. The differences arise from more framework-specific attributes, such as the type and arrangement of parent classes.

These differences influence the selection of an approach for accessing a TIME component from the OMS, and indicate that the web service approach offers the most efficient approach in this case. Undertaking this would be done through the core support in OMS for accessing remote components. In the opposite direction, that of operating an OMS component from within TIME, a web service option is also suggested. The .NET system, upon which TIME is based, has native support for remote component access, and implementation of these within TIME would be straightforward.

This straightforward use of a web service approach is a reasonably elegant approach to framework interaction, although it has the difficulties of remote operation. Despite these difficulties, web services are growing in use in areas outside of environmental modelling, and there is a considerable potential for providing modelling services in this way.

An alternative approach is to support a common component interface within each framework, and to create local component interfaces that conform to this standard.

5. THE OPEN MODELLING INTERFACE (OpenMI)

The OpenMI is an approach to components and models that focusses on the linking of models, rather than internal model or component operation or construction (Gijsbers *et al.*, 2002). The OpenMI has arisen from consideration of existing environmental modelling and the needs of the European Water Framework Directive, and consists of a set of interfaces and concrete classes that specify the requesting and exchange of data between models or components.

In the OpenMI a model component, identified as a *linkable* component, is populated with persistent data by an initialisation method (`Initialise()`), then data are obtained using a `GetValues()` run method invoked from a calling component. The main difference from the above two components lies in the pull mechanism, which allows a calling component to perform the computation and extract the results with one call to the `GetValues()` method.

Given this, the OpenMI approach could be used to support component interaction between different frameworks. To achieve this, each framework would separately contain the methods and classes necessary to use the OpenMI, and components, created as discrete objects with published interfaces, would be made compatible with OpenMI. This approach, or that of web services, provides a way of allowing components constructed in different frameworks to be used together in a confederated model. Additional difficulties arise, however, over the semantics of the data exchange between components.

6. SEMANTIC CONSIDERATIONS

In component-based modelling a common problem is that of the *meaning* of data, variable and parameter names that are used by and passed between components. When components are built by different people using different frameworks, and then offered for use by other people with other frameworks, misunderstandings can arise unless a clear meaning is given to data, parameters and variables. To do this a clear language needs to be established covering not only these, but also the modelled concepts.

This problem is well recognised within disciplines, and approaches such as formal ontologies have been proposed. As environmental management expands and becomes more multi-disciplinary, components are necessarily re-constructed to fit new scales or conceptual structures. This brings with it an increased problem on accurate information exchange between components. Technical solutions to this include a greater use of metadata and meta-information, supported by XML. In particular, the declarative modelling approach appears to be well suited to this purpose. The Semantic Web initiative of the W3C (http://www.semanticweb.org) also offers some hope, through improved definition and communication of the 'meaning' of web information.

7. CONCLUSION

The area of multi-framework model components is one of considerable challenge, although the range of technical solutions listed here, and the development styles of those working on environmental modelling frameworks offers an indication that the difficulties are not insurmountable. Future key areas for investigation and practice include testing of the approaches listed here, identification of the advantages and disadvantages of implementation, and extension of these ideas beyond the limited range of frameworks considered here.

7. REFERENCES

Gamma, E., Helm, R., Johnson, R., Vlissides, J., 1995. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley: Reading, MA.

Gijsbers, P. J. A., Moore, R. V., Tindall, C. I., 2002. HarmonIT: Towards OMI, an Open Modelling Interface and Environment to harmonise European developments in water related simulation software, Hydroinformatics 2002. Fifth International Conference on Hydroinformatics: Cardiff, UK, IAHR, p. 7pp.

McLaughlin, B. 2002, *Java & XML Data Binding*, O'Reilly.

Muetzfeldt, R., Massheder, J., 2003. The Simile visual modelling environment. *European Journal of Agronomy*, 18, 345-358.

Rahman, J. M., Seaton, S. P., Perraud, J.-M., Hotham, H., Verrelli, D. I., Coleman, J. R., 2003. It's TIME for a new environmental modelling framework. In: Post, D. A., (Ed.), MODSIM 2003 International Congress on Modelling and Simulation: Townsville, Modelling and Simulation Society of Australia and New Zealand Inc., p. 1727-1732.

Rizzoli, A.E., Argent, R.M., Manglaviti, M., Mutti, M. 2001. Encapsulating Environmental Models And Data: A Self-Consistent Approach. In: Ghassemi, F. et al.. (eds). Proceedings of *MODSIM 2001, International Congress on Modelling and Simulation*, 10-13 December 2001, The Australian National University, Canberra, Australia, 1649-1654.

Villa, F., 2001. Integrating modelling architecture: a declarative framework for multi-paradigm, multi-scale ecological modelling. *Ecological Modelling*, 137 (1), 23-42.

MOIRA DSS – Architecture, Model Integration and User Interface Design

D. Hofman

Studsvik RadWaste AB, 611 82 Nyköping, Sweden

Abstract: The acronym MOIRA stands for "A Model-Based Computerised System For Management Support To Identify Optimal Remedial Strategies For Restoring Radionuclide Contaminated Aquatic Ecosystems And Drainage Areas" (EC projects MOIRA, COMETES, EVANET-HYDRA). MOIRA system is an integrated friendly tool for the users with different level of responsibility in decision-making and different experience in computers, environmental modelling and radioecology. With MOIRA GUI the decision-maker works with the custom things – maps, tables, graphs, reports. User can be guided through all the steps of the decision-making or can start from the direct request of the data of interest (it could be for example "ranking of the alternative strategies" for the top-level decision maker or "concentration of Sr-90 in the water" – for expert in radioecology). MOIRA system is developed using component architecture by integration of the components such as GIS (MapInfo), models (.exe, Powersim®-based, MapInfo-based) and Muli-Attribute Analysis Module (.exe) with the MOIRA Software Framework that provides kernel and GUI functionality of the system. To support quick development of the system by distributed international team same models can be used both in the frame of MOIRA and standalone. Exchanging of data sets between system kernel and the models as the text files supports this flexibility. The model's developer decides format of the file and kernel adapts itself to it using data set description. Each model can be supplied also with the files (in the same format as input ones) containing default, minimum and maximum values for its input parameters. These files compose the contents of the Reference Data Base (RefDB) used to validate user's input and to help inexperienced users with the default values for the unknown parameters.

Keywords: Decision support system; Model integration; Object-oriented language; User interface;

1. INTRODUCTION

The acronym MOIRA stands for "A Model-Based Computerised System For Management Support To Identify Optimal Remedial Strategies For Restoring Radionuclide Contaminated Aquatic Ecosystems And Drainage Areas" [Monte et al. 2000, Gallego et al. 2000, Monte et al. 2002b]. The decision support process with MOIRA is based [Appelgren et al. 1996] on evaluation via modelling of social, environmental and economic consequences of implementation of alternative countermeasures and ranking the effectiveness of the countermeasures using Multi-Attribute Analysis techniques. Necessary site-specific parameters for models are received from GIS-based data, database or provided by user.

MOIRA DSS contains ten predictive environmental models, dose model, economic model and Multi-Attribute Analysis (MAA) Module [Monte et al. 2000, Håkanson et al. 2000, Gallego at. al. 2002]

Efficient management of these tools requires knowledge in nature of models, their input and output parameters and file formats, knowledge in GIS and data base as well as efforts to keep results of evaluation of different countermeasure strategies consistent with the changes in input data. To make MOIRA DSS user-friendly, accessible, flexible and practical system, which can be used by users – not necessary experts in environmental modelling [Appelgren et al. 1996, Monte, Brittain 1998], the MOIRA Software Framework had been developed [Hofman 1998a, Hofman et al. 2000]. MOIRA Software Framework features have been constantly developed during preparation and issue of updated versions of MOIRA DSS [Gallego et al.2002b, Hofman moiradss.topcities.com 2003, 2004]

Due to the development of the MOIRA DSS by the distributed international team the following considerations was taking into account during MOIRA architecture design and MOIRA Software Framework design and development:

- It must be possible to simultaneously develop models, “system” part and user interface of the DSS
- It must be easy to integrate existing models into the system with minimal changes in the models code. Developers of the model should not concern about technical (programming) aspects of model integration and GUI.
- User interface of the system need to be easy updateable in response to the changes in models input and output parameters.

2. ARCHITECTURE OF THE MOIRA DSS

The component architecture where models are independent applications communicating with the system kernel by exchange of the data had been selected as most suitable for the MOIRA DSS (Fig. 1).

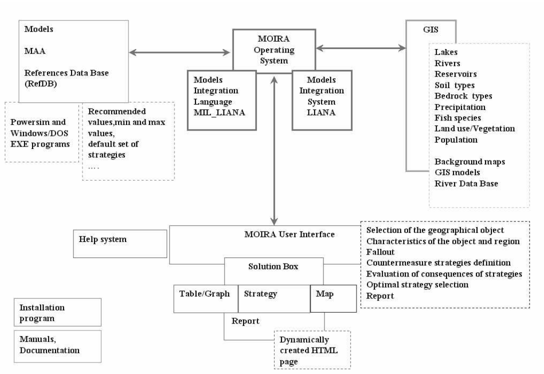


Fig. 1. Architecture of the MOIRA DSS

Similar architecture showed it reliability and usefulness in the RODOS system [Erhardt, Shershakov 1996]. There is one important difference between data exchange in RODOS and MOIRA. RODOS models are exchanging with the kernel via shared memory [Bentz et al. 2001]. This simplifies description of the I/O data for the but requires using of special tool (RODOS/2) to convert input and output files to and from shared memory for the model standalone run during the testing. MOIRA models exchange data with kernel by text files. It is possible to use exactly the same model both in the frame of MOIRA DSS and as stand-alone tool. This helps to support development, testing and update of the models with minimal interaction between developers of models and system. In the same time it requires to provide not only description of input and output data types, arrays length etc. (as in RODOS) but also description of corresponding I/O files formats.

Components of the MOIRA system can be subdivided to:

- Modelling components - models, MAA module and MOIRA GIS
- MOIRA Software Framework

Most of the models was realised in the environment of Powersim® 2.5 package [www.powersim.com]. MAA module and economic model was realised as stand-alone Windows and DOS applications respectively.

Specific modelling component is the MOIRA GIS. During the MOIRA project collection of environmental data for 30 Swedish lakes and for river Tevere (Italy) as well as European-wide population, land-use, soil types and precipitation data had been made in the environment of MapInfo GIS [www.mapinfo.com]. Several programs was developed using MapBasic® to help user in querying of available environmental data and in estimation of number of person affected by the contamination of aquatic body. Results of these models are automatically available for the MOIRA users after user’s selection of the geographical object via the MOIRA GUI.

MOIRA Software Framework consists of MOIRA Operating System and MOIRA User Interface and “covers” all the models- and system- specific functionality of the DSS giving user the possibility to work with “native” things – geographical maps, data tables, graphs, report**.** In the same time it supports easy extending of the system with the new or updated models. Detailed explanation of Software Framework architecture and functionality is given in the next chapter.

3. MOIRA SOFTWARE FRAMEWORK

3.1 Software Framework classes and objects.

MOIRA Software Framework is Windows application internally developed (using Visual

C++ language) as based on Document/View architecture (supported by Microsoft Foundation Class library (MFC)) framework of classes. The simplified schema of the key MOIRA Software Framework classes and their associations is shown on figure 2.

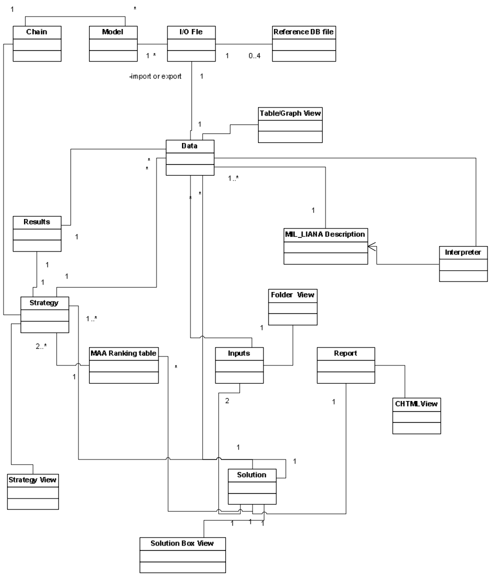


Figure 2. Simplified class diagram of MOIRA Software Framework

The objects of type Data containing values (scalar or time-dependent) for one or more model parameters or results. Content and properties of each data object are kept in persistent form as file (further in this article referred as “data set”). First time properties of particular Data object are set by MOIRA Operating System using correspondent description file written on MIL_LIANA language. Data object properties describe Data object purpose (“input”, “result”, ”countermeasure”), type of information in the object (“scalar”, “time-dependent”), templates for object information importing or exporting from/to text file and format of the table presenting object data in GUI.

Each data object is normally related to certain text file which is either directly input or output file of one of the models integrated to MOIRA or file created by the MOIRA Framework. Before run of the model content of the respective Data objects exported to its input files. When model finish execution its output files is imported in other Data objects. As. mentioned above MIL_LIANA description contains template for such exporting or importing operations.

Objects of classes Inputs, Strategy, Results, Solution are collections keeping references to name of each object, its data set and BaseTemplate id. BaseTemplate is the basic reference to the place in array containing Document/View/Frame templates and to array of icons. Object may have several “states” (for example – does not exists, exists but contain undefined values, exists and do not contain undefined values). Depending of “state” corresponding offset will be added to the BaseTemplate. Collection may contain references not only to Data object but also to objects of other classes (for example Solution contains references to all Strategy collections). A content of each type of collection is described by Framework’s configuration files. Software Framework always keeps consistency between data referred in Inputs, Strategy and Results collections

User interacts with the MOIRA User Interface “browsing” Solution and “activating” data set referred in collections by clicking of mouse on its icon. After activation system will create corresponding Document/View/Frame objects using information provided by MFC CDocTemplate mechanism. For Data objects if data set does not exist it will be first created using MIL_LIANA description[1] and (in the case of activation of “results” Data object) the chain of models will be invoked (via Chain object) in order to receive object’s data.

.

3.2 MOIRA Operating System.

3.2.1 LIANA Model Integration System.

LIANA Model Integration System [Hofman 1998b, Hofman 1999] is the general-purpose framework of classes and functions for the construction of “shell” for model-based DSS. Most of MOIRA Software Framework classes are either directly classes of LIANA system or inherited from classes provided by LIANA or use its low-level API functions.

LIANA Model Integration System consists of the following main parts:

- Classes and low-level functions for management of Data objects and their collections

[1] To save time system use techniques based on idea similar to just-in-time compiling. When certain MIL_LIANA description is used first time system creates on the base of it empty data set, keep it in system directory and then just make a copy of it when object data set need to created.

- Prototype classes for Model and Chain framework classes, realising the basic functionality for running of the individual models as well as for construction and running of the chain of the models
- Interpreters for LIANA and MIL_LIANA languages
- API managing interaction with the Reference Data Base

3.2.2 Integration of models with the MOIRA OS

LIANA system requirement is that model has to be Windows or DOS standalone application (.exe file, .bat file). Specially designed (.exe) application performing link with Powersim® package allow also integrating Powersim® models. MOIRA OS integrates MapBasic® models via connection with MapInfo package based on OLE Automation

Model has to receive inputs and produce outputs as one or more text files. LIANA adapts itself to formats of these files at run-time using description of each input/output file. This description has to be written on the MIL_LIANA language and contain:

- Type of the data set
- Types, names and (optionally) initial values of variables of data set
- Format of the table pesenting data set in MOIRA GUI
- Template for saving data set as the input file of the certain model; or for obtaining the data set information from the output file of the model

MIL_LIANA description can be relatively easy prepared by editing of model's input and output files. All I/O files of MOIRA models were described using MIL_LIANA language. Complete freedom in selection of file format is available with using of LIANA language (see 3.2.4)

During the integration the references to data objects related to model's input and output files must be included in the configuration files describing Framework's collections. In addition each model integrated with the MOIRA Frameworks is described in data set Times by:

- Unique Id number
- Model file (for example neweco.exe or lake1.sim)
- Simulation time-step (if applicable)
- Place in the model chain

During the simulation LIANA system prepares all the input files required for the model, starts model, checking until model will be finished and collect model output information. If model writes data about time passed in special file then this information is transferred to the user during model run-time.

Limited requirements of the MOIRA Software Framework to the model help to integrate model in the form it received from developers. For the model there is no difference either to run in the frame of MOIRA DSS or standalone. This gives the reach possibility for the models testing and development.

3.2.3 Construction and running of the chain of the models

MOIRA models running as a "chain". User has possibility to run complete chain of the models or only part of the chain (if user activates the data set related to the output results of the model situated in the "middle" of the chain).

MOIRA has two "main" chains related to the lake and river scenarios and selects one of them depending on the current scenario. As the next step the models related to the evaluation of concentration of Cs-137 and Sr-90 may be excluded from the chain if user did not provide fallout for one of these radionuclides. Model for estimation of chemical properties of lake water depending on the environmental characteristics of the region can be excluded from the chain if user directly provides the lake water properties. For Framework objects (of Chain class) chain is given as an array containing Id-s of the models on the certain positions. Excluding of the model simply mean placing of 0 instead corresponding model Id.

While present method for excluding model from the chain depending on aquatic object type and provided radionuclide fallout is suitable for the present MOIRA conditions (two types of aquatic objects and two radionuclides) the further development of the system may require automatic construction of the chain using user-defined conditions. This will be done by using LIANA language (see 3.2.4).

During the run of the chain before run of each individual model system verifies if all output data sets of the model are already present for the corresponding strategy (as described earlier due to the data consistency if some of the data in strategy or scenario have been changed then system will delete corresponding results). If all data sets are present then it is no need to run the corresponding model and system simple convert the output data set to text files and place them to the working directory.

3.2.3Reference Data Base

Each model integrated with the MOIRA system may be optionally supplied with

- Files with the default values for certain parameters
- Files with the min values
- Files with the max values
- Files corresponding to the "no countermeasure"

These files have the same format as corresponding input file of the model and compose the Reference Data Base (RefDB). Default data are available for the user during working with the corresponding data tables. Range values are used to check user's input.

3.2. 4 LIANA object-oriented language

LIANA object-oriented language [Hofman, 1999] is the build-in programming language used by LIANA Model Integration System. Schematic functionality of LIANA Model Integration System with using of this language is shown on Figure 3. LIANA language was used in prototype versions of MOIRA [Hofman 1998], while MIL_LIANA

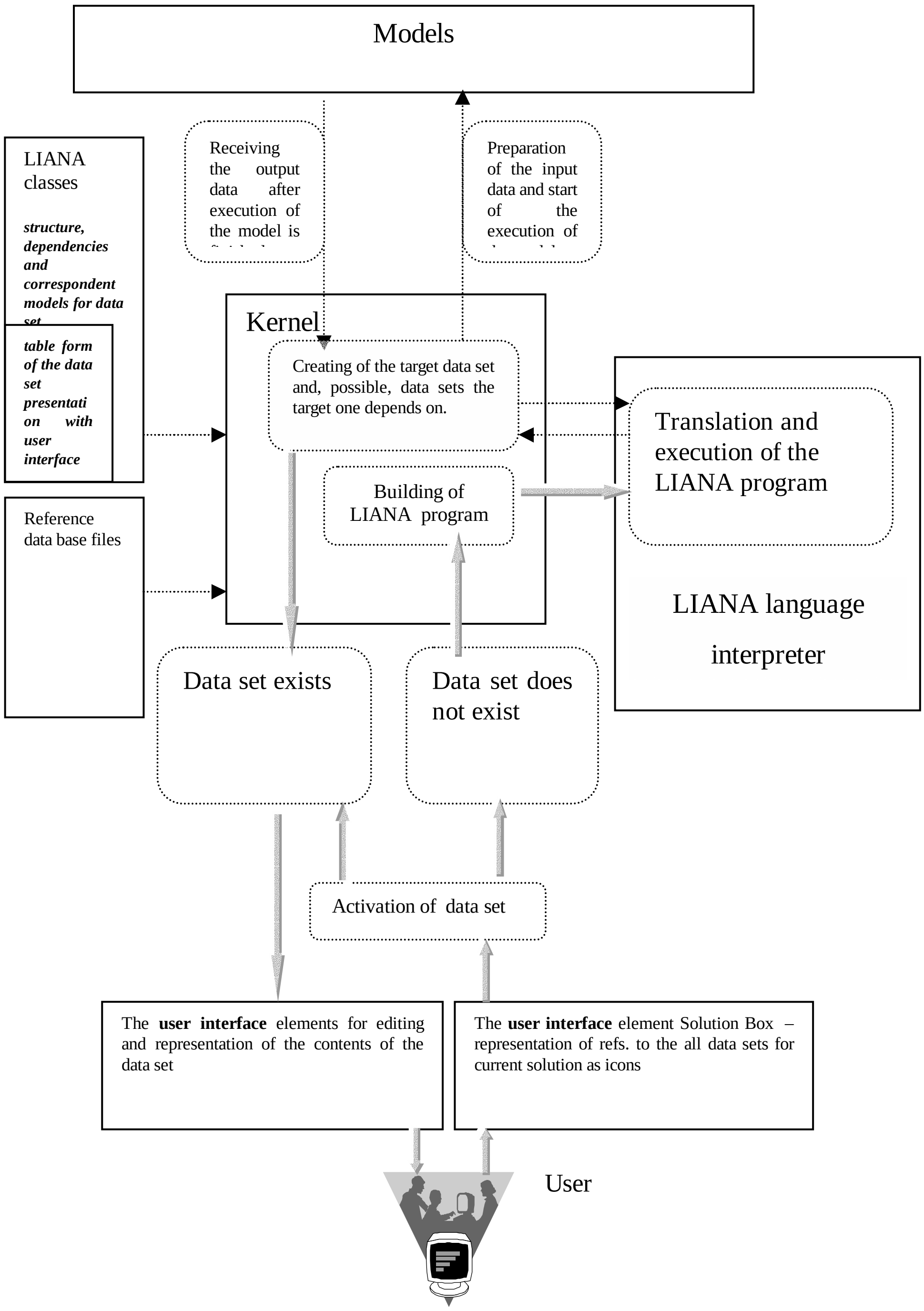


Figure 3. Schematic representation of functionality of LIANA Model Integration System

have been used in end-user versions [Gallego et. al 2002, Hofman moiradss.topcities.com 2003, 2004]. The version of MOIRA currently in preparation for release (v. 2.4) will allow using both MIL_LIANA and LIANA language.
In LIANA language types of Data objects are described using C++-like class definition, which may contain the following additional sections:

- **Produces** – class members available for access (normally read-only) from other objects. These members only will be saved in the data set.
- **Private** – class members not available for access
- **Needs** – other objects, which must be prepared before execution of the command described in Realization. Conditional creating of the certain data objects can be specified by using "if" statement.
  In order to make description simple, for each model it can be created one class with the Needs section describing model inputs. The classes describing model outputs can refer to object of this class in the Needs section.
- **Realization** – normally the command starting external model, which provides data defined in section Produces (in the form RUN ModelId, for example RUN 3). Alternatively can be a message to the user or SAVE command for the data sets created just by query and pre-processing of information from other data sets. The possibility to specify call to Framework's internal function is in plans.
- **Represented** – LIANA statements for reading of model's output or writing the model's input file. The possibility to specify call to Framework's internal function is in plans.
- **Table -** description of the format of the table presenting given object

The STORAGE, GLOBAL, LOCAL, UNIQUE keywords could be used in LIANA class declaration. The STORAGE identifies that there is a data set corresponding to the object of this class. The GLOBAL, LOCAL, UNIQUE keywords help to identify the location (from the current "solution" directory) and name of the data sets.

LIANA language can help to integrate the models with the complicated or changing structure of the I/O data. System can adapt itself to the I/O files with arbitrary structure selected by developers of models. Language gives also possibility to automate completely model chain construction. It easy to see that creating of object described by LIANA language class with the non-empty Needs section will automatically constructs the necessary chain of the models. In addition LIANA language can help users themselves write (if necessary) complex procedures manipulating with Data objects and in this way automate the system

Differences in the structure of MOIRA Software Framework classes (in comparison with Fig. 2) in the case of using LIANA language are shown on Fig. 4. There are class LianaInterpreter, link between LianaInterpreter and Chain object and possibility to recursively load and parse required LIANA class descriptions (bidirectional association between LianaInterpreter and LIANA class).

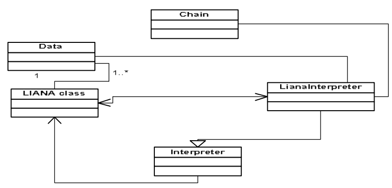


Figure 4. Changes in the MOIRA Software framework class diagram in the case of using LIANA language for the data objects description.

3.3 MOIRA User Interface

MOIRA User Interface (Fig. 5) is the part of MOIRA Software Framework and contains the classes derived from MFC CView and CMDIChildWnd related to browsing, previewing and editing of Software Framework objects as well as following of progress during model chain execution. Browsing of the data sets connected to the geographical location is provided by using of Map view available in MOIRA GUI by making OLE Automation connection with MapInfo (this technique called in MapInfo "Integrated Mapping").

The interesting feature of MOIRA GUI is that model developers indirectly participate in the design of GUI final view. After including in the MOIRA configuration files of the reference to Data object related to I/O file of certain model the corresponding icon will appear in one of the windows of MOIRA User Interface. Now the edition of input or preview of output data is available in table format through Table/Graph Tool after clicking on this icon. This gives model developers the possibility to participate in the DSS user interface construction by data separation in the different files and by description of these files with MIL_LIANA or LIANA language. Such participation helps to construct user interface quickly and utilise broad experience of the model developers in scientific subject and their experience with the communication with users of particular model.

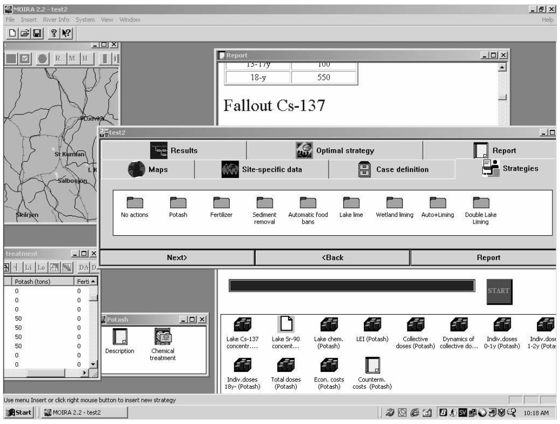


Figure 5. MOIRA Graphical User Interface

4. CONCLUSIONS

The architecture, model integration techniques and user interface design implemented in the MOIRA helped quickly create the reliable user-friendly decision support system based on the integration of the predictive environmental, dose and economic models and MAA.

MOIRA software framework could be used as the basis for the creation of the wide range of the user-friendly and easy-to-learn decision support systems.

5. SOFTWARE

The MOIRA software is working on PC under Windows 95/98/NT/2000/XP. The recent version of MOIRA software is the MOIRA 2.3.1 released in October 2003. MOIRA software and documentation web site is http://moiradss.topcities.com

6. ACKNOWLEDGEMENTS

The MOIRA system is the result of collective efforts undertaken along the MOIRA [Monte et al. 2000] and COMETES [Monte et al, .2002a] projects, as well as during ongoing EVANET-HYDRA thematic network [http://info.casaccia.enea.it/evanet-hydra].

Design and development of the MOIRA Software Framework would be impossible without its constant assessment by A. Appelgren (Studsvik, Sweden) J. Brittain (University Oslo, Norway), E. Gallego (UPM, Spain), R.Helling (NRG, the Netherlands), L.Håkansson (Uppsala University, Sweden), S. Nordlinder (Studsvik, Sweden) and L. Monte (ENEA, Italy).

L.Papush (Linköping University, Sweden) had participated on voluntary basis in the design of elements of MOIRA GUI.

7. REFERENCES

Appelgren, A., Bergström, U., Brittain, J., Gallego, E., Håkansson, L., Heing, R., Monte, L., 1996. An outline of a model-based expert system to identify optimal remedial strategies for restoring contaminated aquatic ecosystems: the project “MOIRA”. ENEA Technical Report, RT/AMB/96, ISSN/1120-5556, Rome, Italy,1996

Erhardt J., Shershakov V.M. .Real-time on-line decision support systems (RODOS) for off-site emergency management following a nuclear accident. Final Report, EUR 16533, 1996

Monte, L., Brittain, J.(Eds.) Principles for the development and implementation of the MOIRA computerised system. ENEA Technical Report RT/ABM/98/4, ISSN/1120-5555, Rome,Italy, 1998

Hofman, D.,. Design and development of the MOIRA software framework. In: Principles for the development of the MOIRA computerised system (ref. Monte, Brittain 1998), pp.37-57.

Hofman, D., Usage of software system LIANA for the integration of application tasks, GIS and data bases into the decision support systems based on models. Mathematical machines and systems N1, Kiev, Ukraine, p.75-88, (on russian)., 1998

Hofman D..Software framework for integration of the models into the decision support system on ecological safety, , Ph D dissertation (in Russian) , Kiev, 1999

Gallego E., Brittain J., Håkanson L., Heling R., Hofman D. and Monte L.. MOIRA: A Computerised Decision Support System for the Restoration of Radionuclide Contaminated Freshwater Ecosystems, Proc. of IRPA 10 congress, Hiroshima, 2000 (http://www.irpa.net/irpa10/cdrom/00393.pdf)

Hofman D., Appelgren A., Bergström U., Karlsson S., Mathiasson L., Nordlinder S. The MOIRA software framework, geographical information system and integration of models. In MOIRA Final Report (ref. Monte 2000)

Håkanson L., Gallego E., Ríos-Insúa S. "The application of the lake ecosystem index in multi-attribute decision analysis in radioecology", pp. 319-344, Journal of Environmental Radioactivity, 49-3, 2000

Monte L., Van der Steen, J., Bergström, U., Gallego, E., Håkanson, L., Brittain, J. The project MOIRA: a model-based computerised system for management support to identify optimal remedial strategies for restoring radionuclide contaminated aquatic ecosystems and drainage areas. Final report. ENEA Report. RT/AMB/2000/13, 2000

Bentz G., Lorentz A., Rafat M., O. Schüle. Integration of external programs in RODOS, RODOS-TN(1)9601, 2001

Monte L. , Brittain J., Håkansson L., Gallego E., Zheleznyak M., Voitsehovich O., Kryshev I., Marinov Petrov K. . "Implementing Computerised Methodologies to Evaluate the Effectiveness of Countermeasures for Restoring Radionuclide Contaminated Fresh Water Ecosystems. COMETES Project. Final Report". Ente per le Nuove Tecnologie, l'Energia e l'Ambiente (ENEA), RT/AMB/2001/28, Roma (Italia), 2002

Monte, L., Van der Steen, J., Bergström, U., Gallego, E., Håkanson, L., Brittain, J.. A model-based computerised system for management support to identify optimal remedial strategies for restoring radionuclide contaminated aquatic ecosystems and drainage areas (MOIRA). In: Schulte, E-H, Kelly, G.N. & Jackson C.A. (Eds), Decision support for emergency management and environmental restoration, EURATOM, EUR 19793, Belgium, pp. 26-27, 2002

Gallego E., Jiménez A., Mateos A., Sazykina T., Rios-Insua S., Windengård M., Application of Multiattribute analysis (MAA) methodologies to the evaluation of the effectiveness of remedial strategies with the Moira system, pp. 11-120 in COMETES Final Report (Ref. Monte et al. 2002a)

Gallego E., Jiménez A., Hofman D. , The version 2 of MOIRA system. A direct result of COMETES project. pp. 179 – 184 in COMETES Final Report (Ref. Monte et al. 2002a)

Continuous Simulation in Material Flow Networks

A. Moeller[a]

[a] *Faculty of Environmental Sciences, University of Lueneburg, Germany*

Abstract: Software systems supporting industrial ecology, corporate material flow analysis and life cycle assessment, are normally used as environmental accounting systems. Using these software systems it is easy to model material and energy flow networks with hundreds or even thousands of processes. Environmental accounting systems are not only utilised to analyse and control existing material and energy flow systems, they are also applied to assess future scenarios in steady-state models. But these systems do not support dynamic modelling. Even though dynamic behaviour is not the main focus of environmental accounting systems, sometimes simulation models are required to estimate material and energy flows depending on specific decision criteria (stock-keeping policies, water circulation design options in manufacturing processes, etc.). This article describes a way to integrate continuous simulation approaches into an environmental accounting system such as the material flow network approach.

The integration of continuous simulation concepts into material flow networks is only one example. It demonstrates how to integrate advanced concepts into an industrial environmental information system (EnvIS). For several other concepts such a way is possible (e.g. Coloured Petri-Nets to support discrete simulation). As a result the environmental information system acts as an infrastructure and consists of several modelling components. The accounting system of the EnvIS plays the role of a conceptual framework.

Keywords: Continuous Simulation, Dynamic Modelling, Environmental Information System, Material Flow Network, System Dynamics

1. FOUNDATIONS OF MATERIAL FLOW NETWORKS

The representation and interpretation of material and energy flows in a particular environmental or economic system has become one of the most important tasks of environmental management in organisations [Schaltegger, Burritt, 2000]. To provide accurate data for managerial uses it makes sense to base the material and energy flow analysis on business accounting methods [Wohlgemuth et al., 1997], which - similar to environmental management - also form the basis for rational decision making.

A well-established accounting method in the private sector is double-entry bookkeeping. The purpose of double-entry bookkeeping is to represent the financial flows and stocks of an organisation. This approach distinguishes between two types of accounts - asset accounts and nominal accounts. Asset accounts describe the capital and asset stocks of a company. Financial flows into the asset account increase the assets, outflows reduce them. By contrast, nominal accounts describe the profit and loss account. The accounting entries record the financial flows between source and destination accounts. This uniform method makes it possible to carry out two different evaluations - the profit and loss calculation and the inventory or asset calculation with the aim of achieving an inventory balance. Together they constitute the annual financial statement of the company.

This type of accounting system is essentially period oriented. It provides information on what financial flows have occurred in a period under review (fiscal year) and how the opening inventory has changed after this period.

The material flow network approach is such an accounting system. Here, material and energy flows are considered, instead of financial flows. This directly results in the necessity for including energy and material stocks in a company into the calculation. The objective of material flow networks is to trace the material and energy flows and stocks within a company or between different companies within a value chain [Möller, 2000, Schmidt et al., 1997].

Based on the concept of Petri Nets [Reisig, 1985] material flow networks consist of three different types of elements (see figure 1): The nodes in the network may be either transitions (visualised as squares) or places (represented by circles). Between these nodes there are arrows as linking elements. Transitions are those locations in a material flow network, where material and energy transformation processes occur. Each transition can, on the one hand, be connected to places from which it is supplied with materials and energy (input), and on the other hand be connected to places to which it delivers materials (output). Places represent storages in which time conversions take place [Wohlgemuth et al., 1997, Möller, 2000].

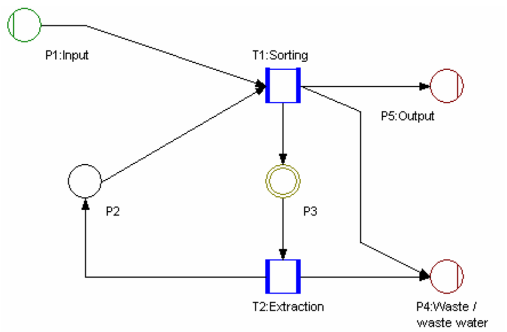


Figure 1. Simple material flow network (process water cycle)

Material and energy flow models are often very large and complex. With hierarchical networks the processes in the value chain can be displayed with regard to the different responsibilities and the necessary degree of detail. A transition can contain a whole subnet which is displayed by another network diagram so that any number of network levels can be modelled. Thus the degree of detail can be comfortably increased step by step at different levels of the network.

As a result a period-oriented material and energy flow model is obtained. The evaluation of such models result in different reports – that is, providing data for inventories, ecobalances, life cycle assessments, Sankey diagrams, and eco-efficiency indicators.

2. CALCULATIONS IN MATERIAL FLOW NETWORKS

A basic task of software systems based on material flow networks is to calculate the stocks at the end of the period and to aggregate the material and energy flows in order to show them in balance sheets. Therefore the algorithm utilises the initial inventory and the material and energy flow data.

A major challenge of software systems based on material flow networks is data collection. How can we estimate all the material and energy flow in our network? One solution is to connect the software system to enterprise resource planning systems (ERP systems) and to utilise the business accounting components. Unfortunately, these components contain only a small part of the data needed. Normally they don't include environmental data like carbon dioxide emissions or water consumption.

To draw the complete picture of the material and energy flows in the network, it is necessary to calculate unknown flows using models of sub-systems in the network. In material flow networks transitions are used to integrate models of sub-systems.

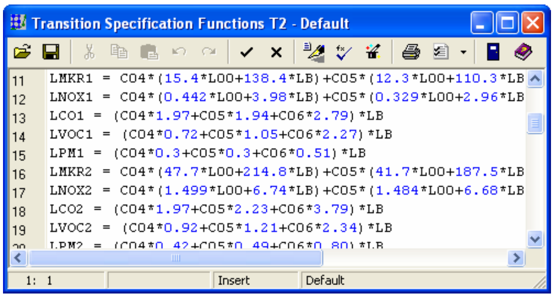


Figure 2. Typical transition specification (transport process)

The aim of these models is to estimate material and energy flows based on known flows. In a simple case the model consists of coefficients. These coefficients specify the linear relationship between different material and energy inputs and outputs. One advantage of the period oriented material flow networks is the possibility to integrate more complex models using non-linear functions (see figure 2).

The main purpose of calculation algorithms in software systems based on material flow networks is to support data collection and estimation. As an accounting system the material flow network approach is not focussed on dynamic behaviour of systems.

3. DYNAMIC MODELLING

Even though material flow networks are widely used in industry to control and to forecast material and energy flows in organisations, it is often difficult to specify the input data of the models. Sometimes it is easier to describe the dynamic behaviour of the system in a simulation model. In this case the simulation model "imitates" a real system during a given time period.

If the number of state changes in the simulation model is finite, we call these models discrete event simulation models. Discrete event simulation models are used to analyse for example warehouses, production chains or call centres [Page, 1991]. In some systems the states change continuously. To deal with these dynamic systems continuous simulation models are more appropriate.

One approach to integrate simulation models into material flow networks is to interpret simulation models as single transitions. In that case the result of the simulation should be a collection of input and output flow data of the transition. Further information regarding discrete simulation models as transition specifications in material flow networks is available in Wohlgemuth et al. [2000] and Möller et al. [2001].

4. SYSTEM DYNAMICS

A more ambitious approach to integrate simulation models into material flow networks is to utilise the structure of the network. To support discrete event simulations we can apply high-level Petri Nets like Coloured Petri Nets [Jensen, 1992].

Another well-known modelling technique is System Dynamics [Forrester, 1961, Hannon, Ruth, 1994]. System Dynamics supports continuous simulation. Therefore System Dynamics models contain a number of stocks (shown in System Dynamics diagrams as rectangles, see figure 3) and flows (shown in System Dynamics diagrams as double-lined arrows).

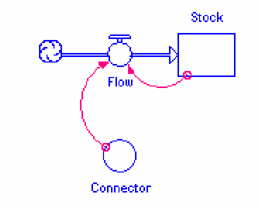


Figure 3. Elements of System Dynamics diagrams

Any flow directed to the stock increases its level, and the flow going out of the stock decreases its level. The amount of flow in and out is regulated by rates (visualised as "water valves"). Connectors (visualised as circles) are used as helper elements to specify user-defined functions and parameters. They are linked to other nodes in the diagram ("information flows").

As in material flow networks, System Dynamics diagrams contain nodes that represent storages. The water valves can be interpreted as special transitions. However, a visualisation of the information flows for calculation purposes is not provided in material flow networks (compare figure 4 and figure 5).

System Dynamics software tools use classical integration procedures like Euler-Cauchy or Runge-Kutta to perform the simulation [Cellier, 1991]. Particularly, Euler-Cauchy algorithms work similar to period-oriented calculations in material flow networks. Schmidt [1995] has imitated the Euler-Cauchy algorithm within material flow networks using a sequence of very short periods and "multi-period" calculations. In multi-period calculations the periods are linked so that the stocks are transferred between the periods. That's exactly how the Euler-Cauchy algorithm works.

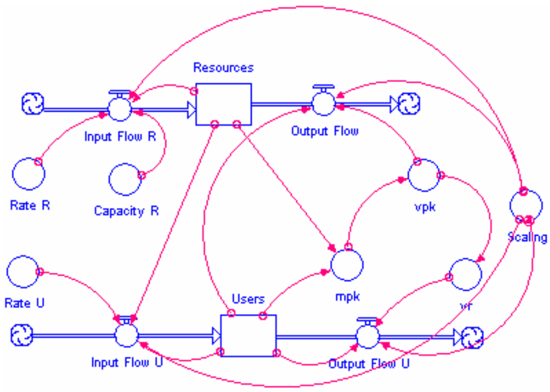


Figure 4. Tragedy of the commons
(System Dynamics diagram)

From here it is a small step to integrate numerical integration procedures seamlessly into material flow networks: The time periods of the network become the simulation period of the System Dynamics model, and the periods of the multi-period calculations are substituted by the time steps of the numerical integration procedure.

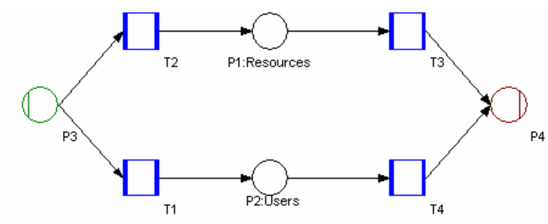


Figure 5. Tragedy of the commons
(material flow network)

Figure 4 shows a simple example (tragedy of commons) in the System Dynamics version; figure 5 the equivalent material flow network model. The tragedy of commons model is a typical System

Dynamics reference model. Several other reference models are available [Bossel 1994, Hannon, Ruth, 1994].

The absence of information flows is obvious in figure 5. From the System Dynamics perspective this is a disadvantage. On the other hand the material flow networks are targeted on the visualisation of material and energy flows. The usage of the same diagram elements avoids breaks in the user interface (see figure 8) and makes it possible to apply the same graphical presentations, particularly Sankey diagrams.

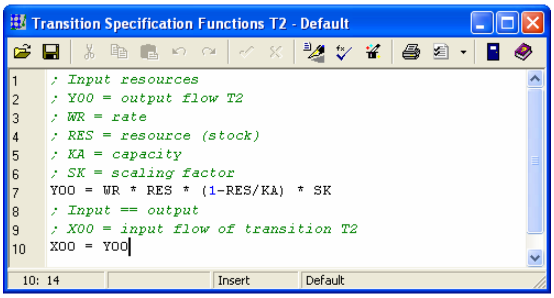


Figure 6. Transition specification functions of T2 (logistic growth)

Whereas in the System Dynamics models the water valves control exactly one flow, in transition specifications several input and output flows can be incorporated and mappings between inputs and outputs can be specified. So the water valve can be considered as a special case of transition specification (see figure 6, line 10). Regarding material and energy flow analysis this is a practical enhancement of the System Dynamics approach.

5. THE LINK BETWEEN SYSTEM DYNAMICS MODELS AND MATERIAL FLOW NETWORKS

As described above the simulation models basically serve as a data source for the environmental accounting system. The question arises how to transform data from the simulation models into a period oriented accounting system. In fact, the transfer of stocks is quite simple, because the simulation periods comply with the periods in the accounting system. The stocks in the simulation models at the end of the simulation period correspond to the stocks at the end of the accordant period in the accounting system. Regarding the flow data it is necessary to calculate the average value for each period (see figure 7). This requires minor extensions of the numerical integration procedures. The aggregated values are transferred into the accounting system.

On the one hand this results in a substantial extension of the material flow network approach: The extensions facilitate System Dynamics on the level of material and energy flows and stocks. On the other hand apart from calculation time conventional material flow networks yield the same results as before. Indeed, the integration of numerical integration methods into the calculation algorithm of material flow networks constitutes a substantial extension of this modelling approach.

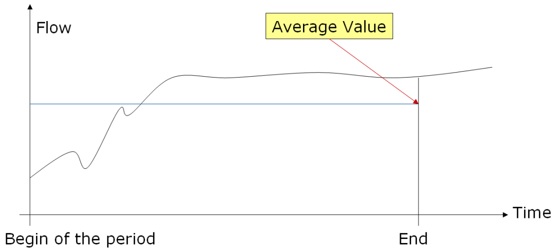


Figure 7. Calculation of period oriented flow data based on a simulation model

Normally it isn't necessary that a simulation model covers the whole network. In hierarchical material flow networks the simulation can be restricted to a subnet so that the simulation model comprises only those system elements which are essential for the simulation model (see figure 8). As material flow network models normally contain hundreds or even thousands of transitions and places [e.g. Skrzypek, Wohlgemuth, 2000], embedding simulation models into material flow networks in such a manner allows combining huge material flow network models and complex simulation models.

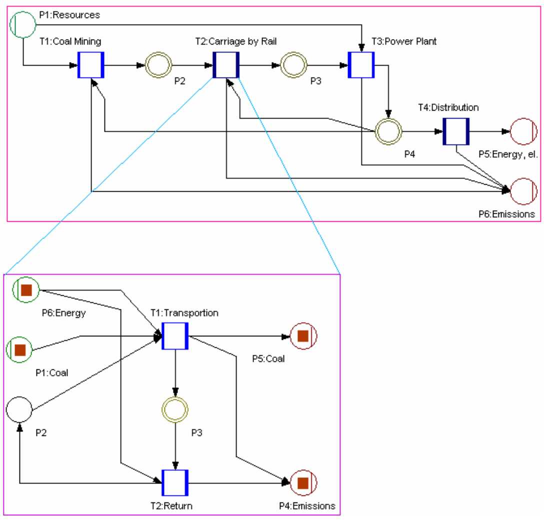


Figure 8. Simulation model as a subnet of transition T2

Subnets of hierarchical material flow networks can be exported and imported and saved separately so simulation models can be incorporated in a so-called process library. Process libraries contain a large number of process specifications that can be used in material flow models and life cycle assessment respectively [e.g. Frischknecht, 2001]. Process libraries supporting material flow networks not only contain linear specifications, it is possible to store non-linear specifications as well as encapsulated subnet models in such databases [Mampel, 1997]. Systems Dynamics reference models like tradegy of commons (see figure 4 and 5) constitute the basis of a new type of models in the process library [Bossel, 1994]. Indeed some of the available System Dynamics reference models do not comply with the requirements of representing material and energy flows and stocks in companies and value chains. It will be necessary to develop more reference models that cover the requirements.

Making such process libraries accessible to a number of users also allows to separate work tasks and to assign them to different users according to their role and level of know-how:

Modelling experts can design and validate generic simulation models (reference models), while other users download and customise these generic models.

6. ENVIRONMENTAL INFORMATION SYSTEMS

The integration of System Dynamics is only one example of integrating modelling techniques into an environmental accounting system such as the material flow network approach. The accounting system becomes an overall concept or rather a conceptual framework of the environmental information system [Rautenstrauch, 1999, Page, Rautenstrauch, 2001]. Modelling techniques used by System Dynamics, high level Petri Nets or discrete event simulation can be integrated in the framework as components.

This approach allows using component-oriented programming [Möller, Rolf, 2003]. The framework has to provide several interfaces to support of

- registering and un-registering components to integrate calculation algorithms as well as specification components,
- accessing the database of the information system,
- using the other components of the system on a standardised way,
- making the components accessible by script languages of the system.

New software technologies like CORBA, .NET, J2EE, XML or Web Services facilitate component-oriented software development. The programming languages of the application developers can be referred to as a “glue language” [Hammond, Robinson, 2000]. They are utilised to glue the components together.

7. CONCLUSIONS

As shown above, material flow networks can be characterised as an environmental accounting system. Appropriate calculation algorithms provide assistance for data collection, in particular by evaluating transition specifications.

Compared to other material flow analysis approaches, material flow networks consist of two different types of nodes. Although places are irrelevant in life cycle assessment and material flow based cost accounting, they become more important in simulation models. In contrast to life cycle assessment dynamic modelling is focussed primarily on stocks and their changes over time. Because of the places continuous simulation techniques like System Dynamics can be integrated seamlessly into the overall concept.

Environmental information systems require an overall concept and an environmental accounting system respectively. This concept has to provide interfaces to different modelling approaches so that component-oriented programming can be applied. In fact, the integration of continuous simulation approaches into material flow networks can be interpreted as an experiment to survey the coverage of the material flow network approach in this respect.

8. REFERENCES

Bossel, H., Modellbildung und Simulation, 2. Edition, Braunschweig, Wiesbaden, 1994

Cellier, F. E., Continuous System Modeling, New York, Berlin, Heidelberg, 1991

Forrester, J. W., Industrial Dynamics, Cambridge, MA, 1961

Frischknecht, R., Life cycle inventory modelling in the Swiss national database ECOINVENT 2000, in: Hilty, L.M., Gilgen, P.W. (Eds.), Sustainability in the Information Society, Marburg, 2001

Hammond, M., Robinson, A., Python Programming on Win32, Beijing et al., 2000

Hannon, B., Ruth, M., Dynamic Modeling, Berlin, Heidelberg, New York, 1994

Jensen, K., Coloured Petri Nets – Basic Concepts, Analysis Methods and Practical Use, Volume 1, Berlin, Heidelberg, New York, 1992

Mampel, U., Die Prozeßbibliothek in Umberto, in: Schmidt, M., Häuslein, A., Ökobilanzierung mit Computerunterstützung, Berlin, Heidelberg, New York, 1997

Möller, A., Grundlagen stoffstrombasierter Betrieblicher Umweltinformationssysteme, Bochum, 2000

Möller, A., Page, B., Rolf, A., Wohlgemuth, V., Foundations and Applications of Computer based Material Flow Networks for Environmental Management, in: Rautenstrauch, C., Patig, S. (Eds.), Environmental Information Systems in Industry and Public Administration, Hershey, London, 2001

Möller, A., Rolf, A., Informationsversorgung eines erweiterten Umweltcontrollings, *uwf UmweltWirtschaftsForum*, 11. Jg, H. 2, 2003

Page, B., Diskrete Simulation, Berlin, Heidelberg, New York, 1991

Page, B., Rautenstrauch, C., Environmental Informatics – Methods, Tools and Applications in Environmental Information Processing, in: Rautenstrauch, C., Patig, S. (Eds.), Environmental Information Systems in Industry and Public Administration, Hershey, London, 2001

Rautenstrauch, C., Betriebliche Umweltinformationssysteme, Berlin, Heidelberg, New York, 1999

Reisig, W., Petri nets: an introduction, Berlin, Heidelberg, New York

Schaltegger, S., Burritt, R., Contemporary Environmental Accounting, Sheffield, UK, 2000

Schmidt, M., Modellierung von Stoffrekursionen in Ökobilanzen, in: Schmidt, M., Schorb, A. (Eds.), Stoffstromanalysen in Ökobilanzen und Öko-Audits, Berlin, Heidelberg, New York, 1995

Schmidt, M. et al., Environmental Material Flow Analysis by Network Approach, in: Geiger, W. et al. (Eds.), Umweltinformatik'97, Marburg, 1997

Skrzypek, V., Wohlgemuth, V., Anwendung und Integration von Umberto® in die betriebliche IT-Struktur am Beispiel der Bertelsmann Großdruckerei MOHN Media Mohndruck GmbH, in: Bullinger, H.-J., Beucker, S. (Eds.), Stoffstrommanagement – Erfolsfaktor für den betrieblichen Umweltschutz, Stuttgart, 2000

Wohlgemuth, V. et al., Computer based support for LCAs and company ecobalances using material flow networks, in: Alef, K. et al. (Eds.), Information and Communication in Environmental and Health Issues, Eco-Informa-97, Vol. 12, 1997

Wohlgemuth, V., Page, B., Einbettung von Transportmodellen und diskreten Simulationsmodellen in Stoffstromnetze, in: Cremers, A. B., Greve, K. (Eds.), Umweltinformatik `00, Marburg, 2000

Structured Modeling Technology: A Modeling Environment for Integrated Environmental Assessment

W. Purwasih[a] and **Y. Nakamori**[a]

[a] *Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Tatsunokuchi,Japan,*

Abstract: Sound environmental decision-making and assessment require integration of modeling resources such as models and complex tools, large volumes of often disparate and multi-format data from numerous sources and effective methods of disseminating and presenting results that allow broad, interactive participation in the assessment and decision-making process. This paper discusses how data sets, models and modeling tools are incorporated into the design of a structured modeling environment. In this paper, we describe the background and implementation of structured modeling environment, which was developed to facilitate and integrate many aspects of the modeling process. The conceptual framework underpinning the structured modeling technology identifies the major aspects of integration between database management system, models, tools, object oriented paradigm, XML and Web technology to provide some of the elements to effectively support Integrated Environmental Assessment (IEA). It helps lay the foundation for the development of implemented prototype using AIM/Trend model to meet the requirements of modeling activities for supporting IEA in Asia Pacific region.

***Keywords**:* Modeling environment; Structured modeling; Structured modeling technology; IEA

1. INTRODUCTION

Integrated Environmental Assessment (IEA) is increasingly recognized as an important technique for managing the environmental impacts of human actions. Therefore, producing accurate assessments of both the state of environment and the consequences of environmental policies becomes critical for effective decision-making. Integrated environmental assessment requires knowledge and understanding of nature and trends in characteristics of complex environmental processes. Because the pertaining phenomena in environmental problems are typically multidimensional, it may require drawing causal inferences from knowledge and data that are combined from different disciplines. Hence, efforts to develop responses to the complex of global environmental change have created the need for access to communication networks, individual and organizational expertise, and types of computerized modeling resources such as data sets, models, solvers, and tools.

Modeling has the potential to provide decision makers and others involved in environmental assessment with tools that enhance understanding of their natural system and the interactions between system components, and allow 'what-if' scenario testing of environmental policies based on strategies and environmental conditions [Dolk, 1993; Makowski, 2004; Rizolli et al., 1998; Wierzbicki et al., 2000]. Due to the interdisciplinary nature and complexity of environmental problems, IEA requires the use of a variety of modeling approaches that can be coupled together to adequately address all elements of a given problem. This needs to include both individual system components and broad system interactions. An integrated model that will allow comparison of the effects of suggested actions, and various combinations of actions, provides a tool not only for option analysis, but also for redirecting debate from a finger-pointing approach to constructive dialogue, thereby facilitating progress towards policy consensus. The Club of Rome has initiated the use of integrated modeling and later various models explicitly addressed environmental issues emerged in the late 1970s.

While the advantages of modeling have been articulated for some time, the development of integrated environmental models that are actually used by assessment is no simple task. Attempts in the development of integrated models have been hampered by a number of problems. 'Integration' will require the diverse range of modeling paradigms to enhance modeling process (e.g. specification, analysis and documentation). Furthermore, the process of model formulation and subsequent integration of model with data in a system is a complex and ill-structured process [Parker et al., 2002]. Due to incompatible modeling representation and data handling used by each modeling paradigm, it is difficult to simply link the corresponding modeling resources. Another issue that has been emphasized by Parson [1995] and Parker et al. [2002] is the requirement of appropriate communication that could facilitate the involvement of stakeholders on model results and the detailed needed for developing alternative strategies.

It is noted in Geoffrion [1989] and Dolk [1993] that structured modeling, as a lingua franca, can be used widely in many different applications, and therefore simplify modeling resources integration. However, it should be extended to address the wider issues of model sharing and multiple users that allow wide participatory in modeling process. Learning from the previous work [Dolk, 1993; Kokkonen, 2003; Rizzoli, 1998], in this paper we attempt to report our ongoing work on incorporating modeling resources for environmental assessment using Structured Modeling Technology [Makowski, 2004]. In the work described here, we use Structured Modeling Technology (SMT) that adopt Geoffrion's structured modeling approach as the vehicle for model representation to provide some of the elements of an integrated modeling environment in such a manner that subcomponents of the system would be reusable and adoptable to others similar modeling process. This modeling technology combines database management system, object-oriented programming, XML and Web technology to promote integration and collaboration in solving complex environmental problems.

The rest of this paper is organized as follows. Section 2 gives the formal description of AIM model as an example of integrated models that was developed to meet the requirements of modeling activities for supporting Integrated Environmental Assessment in Asia Pacific region. Section 3 provides some background knowledge of structured modeling. The platform for integrated modeling environment using structured modeling is explained in Section 4. Section 5 presents implementation issues using structured modeling technology. Finally, some concluding remarks are briefly discussed in Section 6.

2. AIM: AN EXAMPLE OF INTEGRATED MODELS

Here we describe briefly the AIM model that is used in our work. The EIA sub-project of APEIS (Asia-Pacific Environmental Innovation Strategy Project) has provided a set of computer models and database that can be used to analyze the impacts of climate change and mitigation policies for the region. By using the AIM modules and databases, policy/decision makers in the region will be able to develop the most appropriate innovation strategies for specific socio-environmental problems in their countries, taking into account local social, economic and environmental conditions.

The Asia-Pacific Integrated Model (AIM) has been developed by National Institute of Environmental Studies Japan as tool for assessing policy options on sustainable development particularly in the Asia-Pacific region. For this purpose, six models have been developed based on the AIM's family, those are, AIM/Trend, AIM/CGE, AIM/Material, AIM/Energy, AIM/Ecosystem, and AIM/Water. AIM has been used by Intergovernmental Panel on Climate Change (IPCC) as one of models used to develop SRES scenarios [IPCC, 2000].

Aim/Trend, which was developed to predict the economic, energy and environmental trend in 42 countries in Asia-Pacific region, uses simple-econometric method and develops several scenarios for capacity building [Kainuma *et al.* 2003]. This model consists of two sub-models, Model A and Model B, and it has been developed based on the availability of energy balance's data in IEA. In Model A, five scenarios were prepared: Reference, Market First scenario, Policy First scenario, Security First scenario, and Sustainability First scenario. An overview of the structure of Model A is given in Figure 1.

3. STRUCTURED MODELING AS PLATFORM FOR INTEGRATED MODELING ENVIRONMENT

Structured Modeling was originally developed by Arthur M. Geoffrion [1987] in order to identify the basic components of models, the relationships among these components, and conditions under which a model may be termed "structured". It is applied in order to enhance and facilitate some essential functions like model prototyping, data and model access, and model integration.

There are three basic structures: (i) *elemental,* (ii) *generic* and (iii) *modular*. A "structured model" can be defined as an elemental structure which may be partitioned into genera and further aggregated into modules. Here we summarize some general definitions and for more detail explanation please refer to [Geoffrion, 1987; Geoffrion 1989].

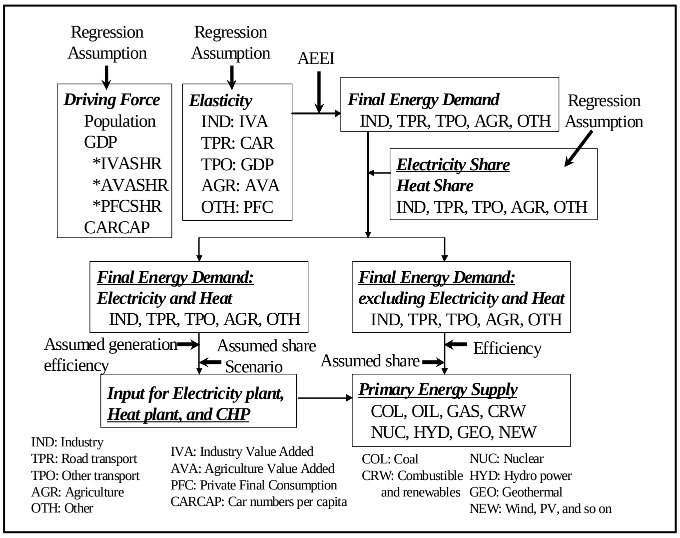


Figure 1. Calculation flow of Sub-model A [Kainuma, 2003]

Elemental structure consists of: primitive entity, compound entity, attribute, function and test. Primitive entities represent things or concepts and require no value. Compound entities represent other things or concepts and have no associated value. Attributes represent properties of things or concepts and have a constant value. Variables attributes are attributes whose values unspecified. Function elements represent calculable properties or mathematical equations and its value depends on a rule and the values of called elements. Test elements are function elements with a Boolean (True, False) value and it can be used to specify constraints in mathematical models. Elemental structure can be represented in terms of a directed graph of elements and "calls". A nonempty, closed, finite, and acyclic collection of elements must be satisfied by the elemental structured of a model.

A *generic structure* represents a partitioning of the elemental structure into genera, such that there is one genus for each element type. It must satisfy generic similarity in that every element in a genus calls elements in the same foreign genera. A tree defined on the generic structure all of whose leaves are genera, and all of whose non-terminal nodes are modules is defined as a *modular structure*. It should satisfy the qualification, which is called monotone. A model can be defined as a structured model if it consists of an elemental structured, a generic structure satisfying similarity, and a monotone modular structure.

Structured modeling does not deal with inputs and/or outputs explicitly. Input variables in structured modeling will usually be designated as fixed attribute (*a*) elements although it's conceivable that primitive entity (*pe*) elements could be inputs. Model outputs usually will be either variable attribute (*va*) or function (*f*) elements. Structured modeling extends semantic data models from the database world to capture the complexities of mathematical modeling. It allows the user to view models graphically, textually, or algebraically, and at different levels of abstraction. Hence, a new modeling environment based on structured modeling will be discussed in the next section.

4. THE REQUIREMENT OF MODELING ENVIRONMENT FOR IEA

Recently, a number of proposals have been made for achieving the desired level of integration between modeling resources. The notion of modeling paradigm integration arose from the issue of model representation coupled with data, tools and solver integration. Model integration, which characterized by linking various constituent modules, typically carries with it the requirement of solver integration as well since different models usually involve distinct solver technologies. A solver is defined as a method or program to manipulate a model and always paired with a model class, modeling paradigm or modeling tradition. Reliable assessments require not only models and modeling tools (e.g. visualization and various analytical tools) but also valid data that are required to populate a meaningful model instance. Environmental data sets are large and often complex. To maintain vast amount of computer-stored data, a database management system is commonly used. Therefore, the integration of databases and models that allows users to automatically retrieve and load input data for complex models is essential.

As computer-processing power has become more affordable, integrating (sub-) models, tools, solvers, and data sets into integrated modeling environment is a challenging task. A number of conventional modeling systems which successfully assimilate innovations like spreadsheets and relational databases still involve a good deal of inefficiency because it could not satisfy one or more the below requirements.

An integrated modeling environment has been rephrased [Geoffrion, 1989] to capture a broader notion of integration; not just integration of models and data but also integration of software tools (solvers, DBMS, GUI, etc.) as well. The required caharacteristics particularly could be categorized into: (i) support for models, (ii) support for solvers, and (iii) utilities (tools). Similarly, Brandmeyer [2000] defines that an overall modeling framework should provide functions and tools common to multiple models, while managing data and computing resources using a single GUI to the modeler and implementing shared data storage.

Therefore, the development of modeling environment for integrated environmental assessment should fulfill the requirement for:

- Providing context for integrated environmental assessment;
- Support for analyzing the cause-effect relationships of a specific environmental issue. It should cover a wide range of analytical tools;
- Providing automated support for modelers for activities performed during the *modeling life-cycle* phases in an integrated environmental assessment. These include: (i) problem identification; (ii) model specification in an algebraic statement; (iii) data collection; (iv) instance generation; (v) instance analysis; and (vi) model analysis [Marek, 2004|;
- Support for documentation, presentation and dissemination of the most important environmental concerns, trends, emerging issues and condition;
- Allowing participatory process that ensures involvement of stakeholders and experts from every discipline that relates to environment and development issues [Parker et al., 2002].

Based on the above requirement, the development of general framework should adopt the important principle: (i) each individual models that needs to be linked should satisfy the generic interface demands; (ii) non-model specific functionality is generally available as reusable tools that may be linked to models; (iii) models may easily use the same basic data. The user may interact with any model through the common Graphical User Interface (GUI). From the interface, users can start simulations, review state variables, and visualize the results. The used models are chosen based on the view points of the user needs, using data in database as input and the results as output of model is then stored in the database, which could be viewed and evaluated using tools and solvers by users based on their preferences. The solvers used to solve a problem might be a query processor or an equation solver for simulation and optimization.

5. IMPLEMENTATION ISSUES: STRUCTURED MODELING TECHNOLOGY

The main focus of development of this framework is to provide a platform for systems integration that adapted Structured Modeling framework. Structured Modeling Technology (SMT) was developed [Marek, 2004] to open client/server architecture, WWW technology, object-oriented paradigm and DBMS for extending the application focus of structured modeling towards the issues of sharing on modeling resources and multiple users. By taking an object-oriented view (C++) of Structured Modeling frameworks and its representation as a relational database or an object database, the implementation of model integration as a function of an object-relational DBMS (Oracle) becomes a feasible alternative and it can be accessed to using readily available Internet World Wide Web browser software.

<u>Object-oriented programming</u>. The applicability of object-oriented programming has been recognized by numerous researchers to model management and the associated benefits of models as objects, solvers bound to these models, and inheritance hierarchies. Object-oriented programming could provide techniques for managing enormous complexity and coupling data with the tasks that manipulate that data. Another advantage of this programming is its capability in achieving reuse of software component. Since all of features and functions of objects are encapsulated with the object themselves, systems can be structured in the terms of the relationships in a very problem oriented way, hiding the technically of implementation in the objects' declarations. Objects become the flexible building blocks of modular systems that can be more easily adapted and redesigned, and reused as requirements change.

Over the last decade C++ has been widely used due to its great advantage for creating types of objects that closely correspond to the real-world objects being modeled. We use C++ because it fully supports object-oriented programming, including the three pillars of object-oriented development: encapsulation, inheritance, and polymorphism as described above. Some description of the declaration of classes that supports encapsulation for SMT structures (using C++) can be seen in Makowski [2004].

Database Management System (DBMS). The DBMS is a vital tool needed for handling the complex data manipulation that earmarks large scale modeling. In this work, we use an object-relational DBMS, namely Oracle, as a way to meet the performance demand on the complex data.

The Internet and XML. The Internet has become the most significant new medium for communication between user and modeling resources that is essential to enhance the modeling process of an identified decision. The World Wide Web is used by the Internet to explore information. XML or eXtensible Markup Language is a widely used data format for structured and semi-structured text on the Web, for specifying the semantics of the model input and output [Rizzoli et al., 2001], linking databases to model inputs [Kokkonen, 2003] or documentations. Moreover, XML documents can be managed by DBMSs. An object-relational (object-based) mapping is used as the basis for software that transfers data between XML documents and relational databases. Replacing proprietary file formats with XML allows organizations to achieve much higher levels if reuse of semi-structured and unstructured data. For instance, information contained in an Excel spreadsheet is only accessible to the Excel program, or to program that uses Microsoft's COM APIs. The same information, stored in an XML document is accessible to any tool that can leverage the XML programming model.

6. CONCLUSIONS

Integrated Environmental Assessment requires modeling resources such as models, data sets, solvers, and tools that could support in producing accurate assessments of both the state of environment and the consequences of environmental policies. The lack of interoperability that permits integration of modeling resources remains a major constraint to accessing usable modeling resources. To do this effectively, they have to be integrated in a modeling environment to provide direct access to user community.

Recently, significant progress has been made towards the development and application of structured modeling technology on AIM/Trend. The most important aspects that could be offered by structured modeling technology are: *flexibility*, *versatility*, and *accessibility*. The approach described in this paper coupled with the SMT provides: (i) a framework and system design for management of modeling resources which are needed by IEA; (ii) communication that could facilitate the involvement of multiple users.

7. ACKNOWLEDGEMENTS

The authors wish to thank Marek Makowski (IIASA) and M. Kainuma (NIES Japan) for their kind collaboration and providing us with all available data and access to AIM/Trend.

8. REFERENCES

Brandmeyer, J.E., Karimi, H.A., Coupling methodologies for environmental models, *Environmental Modelling & Software 15*, 479-488, 2000.

Brouwer, F.M., *Integrated Environmental Modeling: Design and Tools*, Jubels B.V., 229 pp., Amsterdam, 1986.

Dolk, D.R., Kotteman, J.E., Model integration and a theory of models, *Decision Support System 9*, 51-63, 1993.

Geoffrion, A.M., An introduction to structured modeling, *Management Science 33*(5), 547-588, 1987.

Geoffrion, A.M., Integrated modeling systems, *Computer Science in Economics and Management 2,* 3-15, 1989.

Greenberg, H.J., Mathematical programming models for environmental quality control, *Operations Research 43 (4)*, 578-622, 1995.

IPCC, *Special Report on Emissions Scenarios,* Cambridge University Press, Cambridge, 2000.

Kainuma, M., Matsuoka, Y., Morita, T., *Climate Policy Assessment: Asia-Pacific Integrated Modeling,* Springer-Verlag, 402 pp., Tokyo, 2003.

Kokkonen, T., Jolma, A., Koivusalo, H., Interfacing environmental simulation models and databases using XML, *Environmental Modeling and Software 18,* 463-471, 2003.

Makowski, M., A structured modeling technology, *Feature Issues of EJOR on Advances in Modeling: Paradigms, Methods and Applications*, to be published in 2004.

Parker, P.,*et al*., Progress in integrated assessment and modeling, *Environmental Modeling & Software 17*, 209-217, 2002.

Parson, E.A., Integrated assessment and environmental policy making in pursuit of usefulness, *Energy Policy 23*, 463-476, 1995.

Rizzoli, A.E., Davis, J.R., Abel, D.J., Model and data integration and re-use in environmental decision support systems, *Decision Support Systems 24*, 127-144, 1998.

Wierzbicki, A.P., Makowski, M., Wessels, J., *Model-based Decision Support Methodology with Environmental Applications*, Kluwer Academic Publishers, 475 pp., Dordrecht, 2000.

Bhatt, N., and C.A. Cole, Impact of conservation on rates and operating costs, *Journal of et al.Water Resources, American Society of Civil Engineering, 11*(2), 192–206, 1985.

Spatial Modelling of Air Pollution in Urban Areas with GIS: A Case Study on Integrated Database Development

L. Matejicek

Institute for Environmental Studies, Charles University, Prague, 128 01, Czech Republic (lmatejic@mbox.cesnet.cz)

Abstract: A wide range of data collected by monitoring systems and by mathematical and physical modelling can be managed in the frame of spatial models developed in the GIS. In addition to data management and standard environmental analysis of air pollution, data from remote sensing (aerial and satellite images) can extend all the data sets. In spite of that simulation of air pollutant distribution is carried out by standalone computer systems, the spatial database in the frame of the GIS is used to support decision-making processes in a more efficient way. Mostly, data are included in the map layers as attributes. Other map layers are carried out by the methods of spatial interpolation, raster algebra and case oriented analysis. A series of extensions is built in the GIS to adapt its functionality. As examples, the spatial models of the flat urban area and the street canyon with extensive traffic polluted with NOx are constructed. Different scales of the spatial models require variant methods of construction, data management and spatial data sources. The measurement of NO_x and O_3 by the automatic monitoring system and data from the differential absorption LIDAR are used for investigation of air pollution. Spatial data contain digital maps of both the areas complemented by digital elevation models. Environmental analyses represent spatial interpolations of air pollution that are displayed in horizontal and vertical plains. Case oriented analyses are mostly focused on risk assessment methods. Finally, the LIDAR monitoring results and the results obtained by modelling and spatial analyses are discussed in the context of environmental management of the urban areas. The spatial models and their extensions are developed in the frame of the ESRI's ArcGIS and ArcView programming tools. Aerial and satellite images preprocessed by the ERDAS Imagine represent areas of Prague.

Keywords: spatial modelling; GIS; air pollution; Lidar

1. INTRODUCTION

The recent development of spatial data management in the frame of geographic information systems (GISs) has created the new era of environmental modelling. More powerful computers have made running air quality models at global and locale spatial scales possible. In order to understand the function of more complex models, the modelling system should consist of other subsystems (point and area sources of pollution, spatial description of terrain elevations, meteorological data, air quality monitoring networks). Obviously, the use of the GIS has become essential in providing boundary conditions to the air quality models. Certainly, the use of the GIS in air pollution modelling can be extended moreover to processing the surface data. Many models have been coupled with the GIS in the past decade to simulate various environmental processes as described in the book written by Longley et al. [2001]. Due to the four-dimensional nature of distribution of atmospheric pollutants, the concept of the GIS should be extended to include temporal variations of three-dimensional spatial data. Considering to a huge volume of numerical calculations, two-dimensional interpolations into the horizontal layers are used to interpolate three-dimensional atmospheric data onto a model grid system. The interpolations, integrations of land cover surface data and the GIS analyses focused on small scale spatial models carried out in the kilometer grid are discussed by Lee in the book published by Goodchild [1996]. In case of large scale air quality modelling, more detailed spatial data are needed to include the impact of buildings and other man-made barriers on distribution of air pollutants, [Janour, 1999; Civis 2001]. Apart from this approach, the statistical theory is also used to indicate spatio-temporal interactions as described by Briggs et al. [2000].

2. METHODS OF INTEGRATION AIR QUALITY MODELS INTO THE GIS

A few scenarios can be established to integrate air quality models into the GIS. The basic level is represented by the standalone software application for simulation of air quality models (ISCST3, ISC-PRIME), which is accompanied by data inputs and outputs. All data can be used independently by other software systems (GIS, RDBMS, Surfer, WWW-presentations). The individual programs form heterogeneous data structures that require the transport of data into various data formats. Figure 1 illustrates an example of steps carried out during the simulation of air quality models.

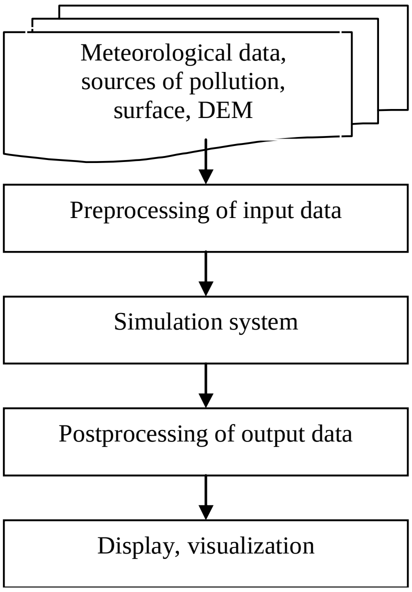


Figure 1. The standalone simulation of air quality models, which is extended by preprocessing and postprocessing software systems.

On the other side, a number of computer programs has been developed to integrated particular functions of the GIS, air quality modelling and graphic systems. Mostly, they are determined to carry out specific calculation without links to other software applications. The GIS based software applications are mostly based on spatial software libraries. The missing functions (air quality modelling, visualisation tools) can be complemented or shared through the dynamic–link libraries. The integrated emission evaluation systems, which offer alternative ways of using the emission models together with selected functionality of the GISs, are described by Rebolj [1999]. A number of software applications is focused on design of relational databases and their interconnection together with standard air quality modelling systems. The structure of the programs developed with spatial software libraries shows figure 2.

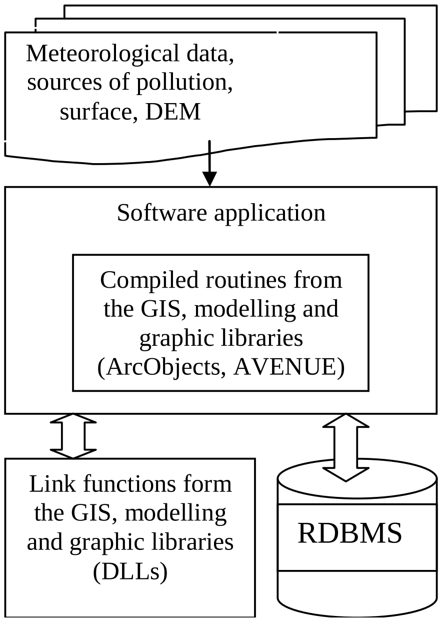


Figure 2. The standalone software application for integrated evaluation of air quality.

2.1 GIS data management and functionality

Considering to the both described scenarios of integration, the scope and scale of urban areas problems make the GIS a powerful tool for management of spatial and temporal data, complex analyses and visualization, [Matejicek, 2002]. Due to the ability to manage a number of spatial and temporal data formats, data structures created in the frame of the GISs open the ways to building air quality information systems that synthesize geospatial and temporal air quality data to support spatio-temporal analysis and dynamic modelling. There is also a growing amount of the digital maps in the GIS community, which are used to support decision-making processes of the urban authorities (data sets for land cover and climatic variables, digital elevation models, which are extended by blocks of buildings and trees, air pollution sources and monitoring networks, soil and hydrologic properties, road and railway networks). While much progress has been made with the mapping of environmental data and the creation of national, regional and local data sets, many challenges remain. For example, air quality models are not used to be included into the GIS. As standalone software applications, they use various data formats, which can usually operate independently with their own GIS database. Similarly, air quality management agencies are creating GIS data sets to support their operations without any data standards that can support spatio-temporal analysis and dynamic modelling. The common theme among these challenges is the need for integration of different spatial and air quality data, integration of data and modelling, integration across spatial scales. The requirements for the integrated spatial modelling of air quality in the frame of the GIS represent a common geospatial coordinate system, vector themes (points, lines and areas) for description of surface objects (buildings, bridges, vegetation) supported by raster and TIN surface data (digital elevation models), vector themes for representation of air pollution inputs (local point, line and area sources of pollution, long-distance transport of air pollution). The key parts of the projects represent data of air quality measured by monitoring networks, terrain measurements (LIDAR) and simulation results of air quality models.

2.2 GIS data models

Nowadays, all the mentioned properties can accomplish a few of the GISs. In the presented study, the ArcGIS distributed by the Environmental Systems Research Institute (ESRI) has been used for the proposed operations. The ArcGIS, a descendant of the widely used ArcInfo, can manage spatial data in a few levels as shapefiles, coverages and geodatabases. Moreover the expansion of the ArcGIS functionality by the COM technology, the Visual Basic is the standard interface language, just as Microsoft uses the Visual Basic as the interface language for other software applications. The ArcGIS can be customized for particular applications of the GIS using specially designed data models. Currently, a number of data models have been published in hydrology [Maidment, 2002], biodiversity, forestry, etc. Air quality modelling can be accomplished by exchanging data between ArcGIS and the independent air quality simulation system, by constructing a simulation tools attached to a project in the ArcGIS, or by customizing the behavior of the ArcGIS objects. The choice depends on the model complexity and calculation requirements in the frame of the various ArcGIS levels. All data are stored in the relational database, which can be represented on the basic level by the personal geodatabase (Microsoft Access), or by the RDBMS (Oracle, Microsoft SQL Server). So, the data transfer among other standalone software applications can be realized directly through the implemented database connections. In case of the ArcGIS's geodatabase, all the data are loaded into the relational database, so that the geospatial coordinate data of the GIS data layers are stored in the relational data tables. Since the relational database supports relationships between its tables, feature-to-feature spatial connections can be set up among the GIS data layers together with linking and joining of external data tables.

2.3 SPATIAL MODELS FOR AIR QUALITY ASSESSMENT EXTENDED BY THE LIDAR MEASUREMENTS

The data required for spatial models to serve air quality modelling can be grouped into a few classes. Figure 3 shows spatial data included into map layers in the frame of a GIS project.

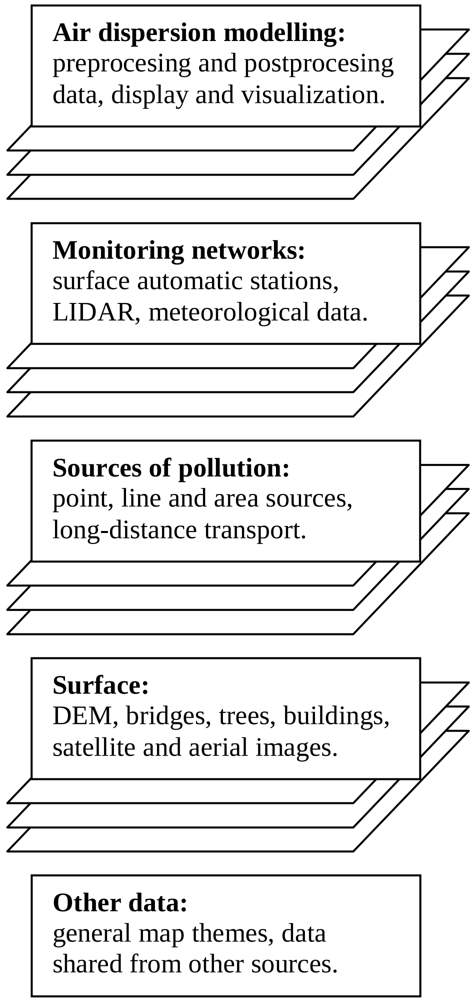


Figure 3. Data included into GIS map layers.

It is impossible to completely enumerate all the spatial and non-spatial data needed, since the more that is known, the better. However, the accuracy of the model results does not depend on the data alone. Choice of the appropriate modelling tools and their setting represents other key parts of air quality modelling. So, if the models do not require or are not capable of evaluating some detailed information, there is little benefit in putting that data in a GIS project. To examine the functionality of the spatial modelling system, the Industrial Source Complex-Short Term (ISCST3), its version with Plume Rise Enhancements (ISC-PRIME), and the AMS/EPA Regulatory Models (AERMOD/AERMOD-PRIME) have been included into the projects. The ISC-AERMOD View with its preprocessing and postprocessing modules has been used as the unified interface of the air dispersion models.

The spatial surface data (digital elevation model-DEM, buildings) make up the input into the preprocessing modules (Import of the Digital terrain data in ISC-AERMOD, Building Profile Input Program-BPIP). Other surface data (bridges, trees, satellite and aerial images) complement spatial information for display and visualization. The layers with sources of pollution contain (in addition to the coordinates and shapes) the attributes, which describe emission properties. The surface data and data about sources of pollution have to be transferred into appropriate input formats to run the air quality dispersion models. The primary storage in the GIS spatial database serves furthermore for spatial analysis, display and visualization. Likewise the previous data, meteorological data are also preprocessed from the database storage into the input formats for air quality dispersion modelling. The map layers, which represent monitoring networks and LIDAR measurements, serve for a comparison of the measured data with the predicted air pollution data calculated by the models.

The mentioned air quality models are steady-state Gaussian plume models used to assess pollutant concentrations from a wide variety of sources mostly associated with an industrial complex. The steady state values of variables are transferred and incorporated into the GIS database, which can be handy to manage data time series. To accommodate large data and many variables such as air quality data, climatic data, properties of sources of pollution, the data repository containing all types of time series data for all features and for all times is proposed. So, time series information can be depicted in 3-D space. The three coordinate axes mark space (S-identification code of a spatial future), time (T-discrete time) and variable being measured (V- identification code of a variable). The data value indexed by the space, the time and the variable can be defined as D(S,T,V). Thus, each stored value is represented by a point in three-dimensional space with its corresponding coordinates, figure 4. In order to extract time series, the space and variable coordinate have to be specified in a query. The result is represented by selected records that match up to the condition in a query. Due to spatial properties of the GIS, space coordinates can be derived from a spatial query in the frame of GIS functionality. The associations between data repository and spatial objects in the ArcGIS geodatabase are specified by relationships, which are stored into the relationship classes.

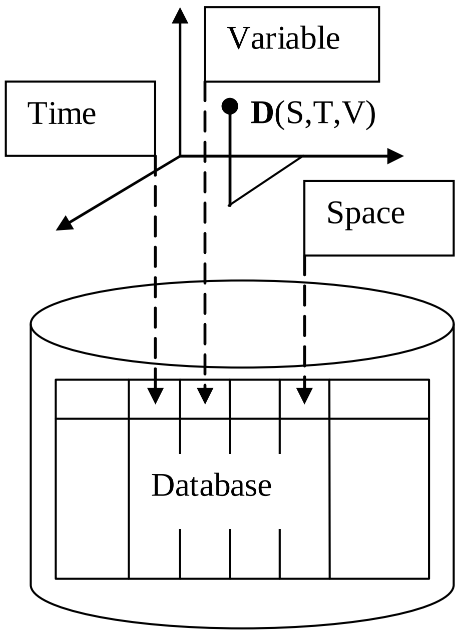


Figure 4. Data repository with 3-D space indexes.

4. CASE STUDIES OF THE URBAN AREAS

The various data sets (digital maps, aerial and satellite images, spatio-temporal data in the 3-D database, data outputs from simulation systems) have been linked together to make up projects for different spatial scales. The GIS, originally design to display 2-D digital maps, has been extended into 3-D mapping and data management in the frame of the ArcGIS. As examples, two urban areas of Prague have been used to demonstrate abilities of the spatial modelling.

4.1 Spatial modelling of the flat urban area

The inputs of spatial data represent digital elevation model, which can be used for air pollution modelling, and aerial or satellite images, which can serve for classification of the surface into classes to set the surface graininess and temperature. The sources of air pollution are mapped into a few categories according to a volume of pollution. Their locations and shapes (in case of the line and area sources) together with the attributes are stored in separate themes. Influential sources of pollution among others are represented by NOx (mostly traffic-related air pollution mapped as the line sources) and SO_2 (mostly stationary air pollution registered as the point sources). In addition to data from automatic monitoring system, the LIDAR [Zelinger, 2003], has been used to complete the data sets. The map composition, which contains the aerial images complemented by the layers with sources of air pollution and 3-D LIDAR data (O_3 concentration labelled by the elevation), is illustrated in figure 5.

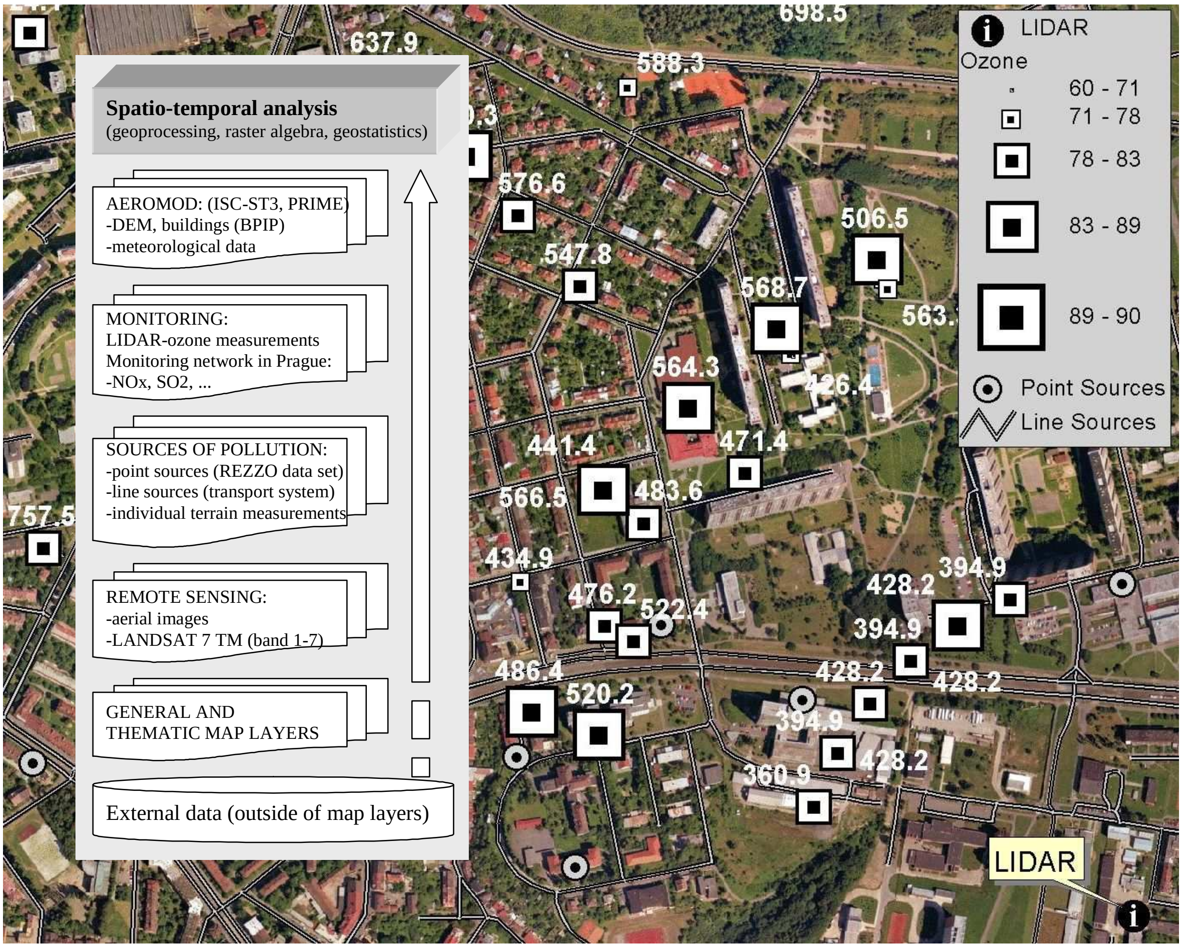


Figure 5. Map layers of the flat urban area.

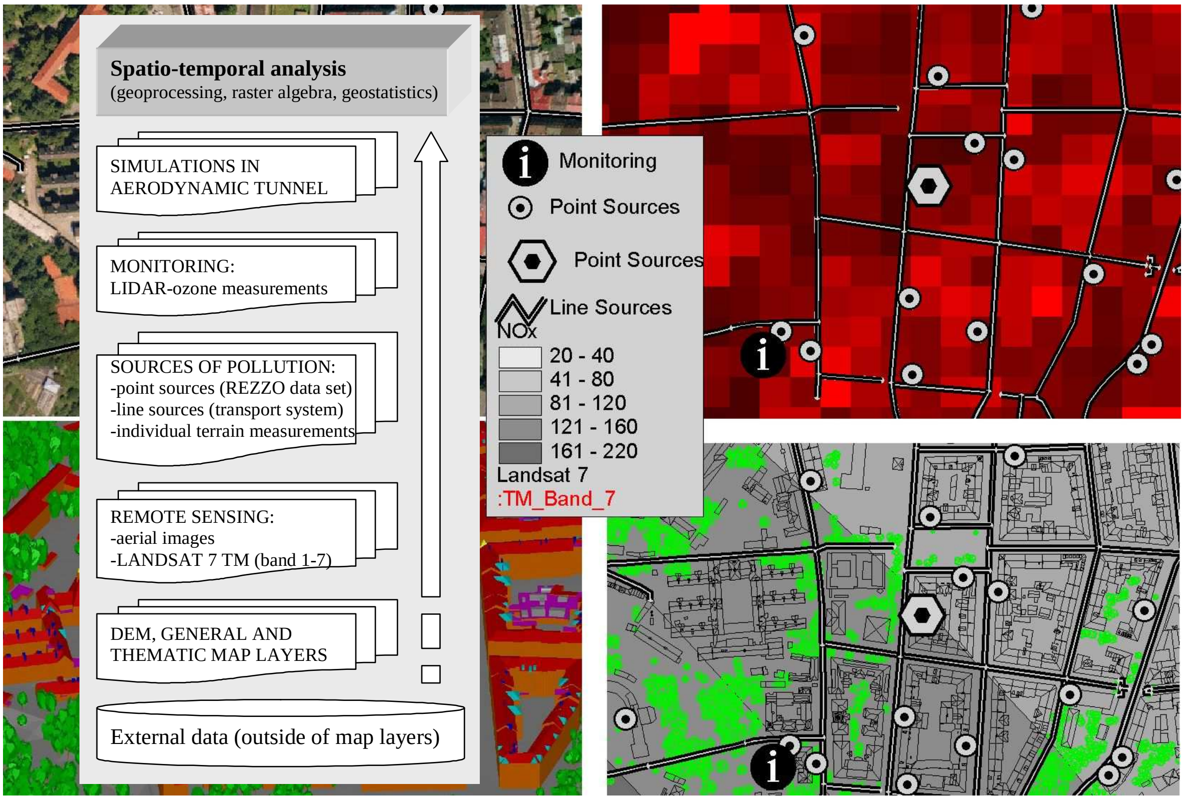


Figure 6. Map layers of the street canyons.

4.2 Spatial modelling of the street canyons

The streets surrounded by high buildings in the urban areas polluted with the traffic-related sources are spatially modelled as the street canyons. Accumulation of air pollution (mostly from cars) results in high concentrations of organic and inorganic compounds in the street canyons. Distribution and local accumulation of pollutants can be solved by mathematical and physical modelling. In the first stage, the digital terrain model complemented by buildings and other terrain objects is needed to support air quality modelling. Consequently, complex analysis of all spatio-temporal data has to be carried out. Spatial modelling in the frame of the GIS can help to accomplish nearly all these tasks. To demonstrate GIS suitability, a case study of spatial modelling of air quality in urban streets illustrates figure 6. Map compositions contain various sets of themes. The first part shows the aerial images of the studied local area complemented by the layers with sources of air pollution and one point of the monitoring network. Other map compositions contain the same area complemented by the satellite image from Landsat 7 (the 7th band, which refers to temperature of the surface), the digital terrain model with buildings and trees, and a sample of the spatial interpolation of air pollution in the area. Again in additions to standard analysis, the LIDAR system and the results of physical modelling in the scaled down models (simulations in wind tunnels) can be used to complete the data sets.

5. CONCLUSIONS

Spatial modelling of air quality in this paper is mainly focused on integration of a wide range of data in the frame of the GIS spatial database. This way of data management and analysis is also promoted by the LIDAR data, which represent measurements of compounds above the surface located by 3-D coordinates. Despite of complex spatial data management, analysis and visualization, modelling of air pollution has to be solved independently in the frame of standalone computer systems (mathematical modelling or physical scaled models). So, the GISs serve as the data stores, which can manage all the data together with model outputs to carry out risk assessment analysis and map compositions. The spatial modelling of street canyons in the frame of the larger urban area complemented by the 3-D LIDAR measurements requires more detailed three dimensional mapping that can generate extensive volume of data. So, the spatial modelling of air pollution extended by air dispersion models under the united interface can be used in case of the adequate hardware, software and data support.

6. ACKNOWLEDGEMENTS

The paper was carried out in the frame of the project GACR 205020898, which is generally focused on measurement and modelling of air pollution in urban areas and flat landscape. The GIS projects were realized in the GIS Laboratory supported by the Ministry of Education, Youth and Sports of the Czech Republic in the frame of the project MSM 113100007 of the Faculty of Natural Science, Charles University in Prague. The digital maps used in the case studies are administrated by the Institute of Municipal Informatics of Prague.

7. REFERENCES

Briggs, D.J., A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments, The science of the total environment 253, 151-167, 2000.

Civis, S, Z. Zelinger, M. Strizik, Z. Janour, Simulation of Air Pollution in a Wind Tunnel. In: *Spectroscopy from Space.* (Demaison, J., Ed.), Kluwer Academic, Dordrecht, 275-299, 2001.

Goodchild, M.F., L.T. Steyaert, B.O. Parks, C.A. Johnston, D.R. Maidment, M.P. Crane, and S. Glendinning, editors, *GIS and Environmental Modeling: Progress and Research Issues.* Fort Collins, CO: GIS World Books, 239-242, 1996.

Janour, Z., Z. Zelinger, S. Civis, Laser photoacoustic spectometry and its application for simulation of air pollution in a wind tunnel, Analyst, 124, 1205-1208, 1999.

Longley, P.A., M.F. Goodchild, D.J. Maguire, D.W. Rhind, Geographic Information Systems and Science, John Wiley & Sons, 27-58, New York, 2001.

Maidment, D.R., Arc Hydro: GIS for water resources, ESRI, California, 2002.

Matejicek, L., Environmental Modelling in Urban Areas with GIS, Integrated Assessment and Decision Support, Proceedings of the 1st biennial meeting of iEMSs, Lugano, 60-65, 2002.

Rebolj, D., P.J. Sturm, A GIS based component-oriented integrated system for estimation, visualization and analysis of road traffic air pollution, Transport and air pollution-8th international symposium, 15-21, 1999.

Zelinger, Z., M. Strizik, Z. Janour, P. Berger, A. Cerny, Comparison of model and in-situ measurements of distribution of atmospheric pollutants, Proceedings of PHYSMOD 2003: International Workshop on Physical Modelling of Flow and Dispersion Phenomena, Italy, 6 pp., 2003.

How tight are the limits to land and water use? – Combined impacts of food demand and climate change

Hermann Lotze-Campen[a], **Christoph Müller**[a,b], **Alberte Bondeau**[a], **Pascalle Smith**[c], **Wolfgang Lucht**[a]

[a]*Potsdam Institute for Climate Impact Research (PIK), 14412 Potsdam, Germany*

[b]*Max Planck Institute for Meteorology, 20146 Hamburg, Germany*

[c]*Laboratoire des Sciences du Climat et l'Environnement (LSCE), Gif-sur-Yvette, France*

Abstract: In the coming decades, world agricultural systems will face serious transitions. Population growth, income and lifestyle changes will lead to considerable increases in food demand. Moreover, a rising demand for renewable energy and biodiversity protection may restrict the area available for food production. On the other hand, global climate change will affect production conditions, for better or worse depending on regional conditions. In order to simulate these combined effects consistently and in a spatially explicit way, we have linked the Lund-Potsdam-Jena Dynamic Global Vegetation Model (LPJ) with a "Management model of Agricultural Production and its Impact on the Environment" (MAgPIE). LPJ represents the global biosphere with a spatial resolution of 0.5 degree. MAgPIE covers the most important agricultural crop and livestock production types. A prototype has been developed for one sample region. In the next stage this will be expanded to several economically relevant regions on a global scale, including international trade. The two models are coupled through a layer of productivity zones. In the paper we present the modelling approach, develop first joint scenarios and discuss selected results from the coupled modelling system.

Keywords: Integrated modelling; agricultural production; land use; water use; climate change

1. AGRICULTURE AS A CRUCIAL LINK BETWEEN SOCIETY AND NATURE

Agricultural production interacts more closely than most other human activities with socio-economic and environmental conditions. In economic terms the agricultural sector is losing importance over time. The share of agriculture in GDP and labour force is now below five percent in most industrialised countries. These trends occur despite wide-ranging government interventions to achieve the contrary. From an environmental point of view, however, agriculture is of key importance in rich and poor countries alike. On a global scale, agricultural production accounts for about 40 percent of total land use, it uses about 70 percent of all freshwater withdrawals, affects nutrient cycles and the climate, and it is the most important driver of biodiversity loss. At the same time, environmental changes strongly influence agricultural productivity [Kindall and Pimentel, 1994]. Understanding the links between food consumption, agricultural production and related environmental impacts is challenging, as socio-economic and environmental driving forces and impacts occur at different spatial, temporal and thematic scales. We present a coupled modelling framework as a tool for an integrated environmental-economic analysis of the food system across different scales. While our scope is global in principle, for the purpose of demonstrating the viability of our concept we chose Germany as a pilot sample region.

2. AGRICULTURAL CHALLENGES IN THE 21ST CENTURY

Whether food production will keep pace with the demand for improved diets for a rapidly growing world population is still under debate. Optimists note relatively low average crop yields, inefficiencies throughout the food production and consumption chain, and the ample reserves of potential arable land in many developing countries. Sounder government policies, wider application of green revolution technology, reduced inefficiencies, upgraded rural infrastructure, and greater investments in human resources and research will make much larger harvests possible and no insurmountable environmental constraints are foreseen [Alexandratos, 1999]. Pessimists point at many signs of environmental stress and increasing difficulties encountered in expanding agricultural land, water supply and crop yields, and in controlling pests. Large expansion of agricultural output may not be feasible and it seems doubtful that current levels of crop production can be sustained in a number of countries. With global warming the prospects for increased food production would become even less favourable than they are at present [Kindall and Pimentel, 1994].

World population will probably grow to about 10 billion people by the year 2100, with a median projection at 8.8 billion for the year 2050 [Lutz et al., 2001]. As income rises, people tend to consume more calories in total, and the share of animal calories increases, especially the consumption of animal fats (Table 1). Global meat consumption can be expected to rise by up to 3 percent annually over the next decades, due to a combination of population growth, growth in per-capita income and a high income elasticity of meat demand. Average global meat consumption per capita could increase from 32.6 kg/year now to 44-54 kg/year in 2030 [Keyzer et al., 2001].

Table 1: Major sources of food energy in industrialised and developing countries [Bender and Smith, 1997]

Product group	*Industrialised countries (Percent share)*	*Developing countries (Percent share)*
Cereals	31	56
Meat and dairy products	28	12
Sweeteners & vegetable oils	23	17
Roots and tubers	4	5
Others	14	10

The amount of land necessary for animal production depends very much on specific production systems. Different animals have different feed requirements and feed conversion rates (Table 2).

Table 2: Conversion rates of grain to animal products [Bender, 1997]

Animal product	*Kg of feed / kg of output*	*Kcal of feed / kcal of output*
Beef	7.0	9.8
Pork	6.5	7.1
Poultry	2.7	5.7
Milk	1.0	4.9

Even at conservatively reduced growth rates in crop yields, global food supply may still outpace demand up to 2020 and real prices for agricultural commodities are likely to continue to fall [Rosegrant and Ringler, 1997]. However, the assumption of exponential growth paths instead of logistic curves has been questioned for projections in the very long run [Harris and Kennedy, 1999]. The potential of biotechnology and genetic engineering for accelerating agricultural productivity growth remains unclear and subject to a strong public debate. Some initial trials show positive effects, but environmental consequences have to be further investigated and widespread social acceptance remains questionable [Qaim and Zilberman, 2003].

The total amount of land available for agriculture not only depends on biophysical conditions, but also on the demand for land for other economic and environmental purposes. Infrastructure development and urbanisation may reduce agricultural areas around the major population centres. In the course of a major energy transition a significant demand for bio-fuel production may arise, either from fast growing forests or from agricultural crops. Moreover, a certain share of land may have to be set aside for biodiversity protection [Sands and Leimbach, 2003].

More intensive production systems may lead to land degradation. This is a very important issue in some geographic regions and could become a serious threat to global food supply [Döös, 2002]. In order to assure sufficient nutrient supply for more intensive production on a global scale, the demand for fertilizer will rise. Especially nitrogen requirements will strongly increase up to 50 percent above current consumption by 2050. The consequences for sensitive environmental systems and the nitrogen cycle remain unclear [Gilland, 2002].

Water may pose the most serious limitation to future global food supplies. Irrigated areas account for nearly two-thirds of world rice and wheat production. Rising irrigation output per unit of land and water is essential to feed growing populations. Since the development of traditional irrigation and water supplies is increasingly expensive and new sources like desalination are not expected to play a major role soon, water savings at every level are absolutely necessary. Crop output per unit of evaporative loss has to be increased and water pollution has to be reduced. However, the size of potential water savings in agricultural irrigation systems is unclear. While specific water uses can be made more efficient through better technology, the potential overall savings in many river basins are probably much smaller, because much of the water currently lost from irrigation systems is reused elsewhere. Increasing water demand from households and industry will further exacerbate the challenge [Rosegrant and Cai, 2003]. The specific water requirements for various agricultural products differ widely, from less than 200 litres per kg output for potatoes, sugar beets or vegetables, to more than 1000 litres per kg output for wheat and rice [Barthélemy et al., 1993]. A typical diet with meat consumption at American levels requires about 5400 litres of water for crop evapotranspiration, while a comparable vegetarian diet requires only about half the amount. The future global challenge with respect to agriculture and water implies that over the next 25 years food production has to be increased by about 40 percent while reducing the renewable water resources used in agriculture by 10-20 percent [Rijsberman, 2001].

An additional constraint to agricultural production in the second half of the 21st century is global climate change. A rise in atmospheric CO2-levels and a corresponding rise in global temperatures will not only affect plant growth and yields, but also alter the regional patterns of precipitation and water availability as well as land erosion and fertility. So far, sensitivity studies of world agriculture to potential climate changes have indicated that global warming may have only a small overall impact on world food production because reduced production and yields in some areas are offset by increases in others. However, regional impacts vary quite significantly, with tropical regions especially suffering from droughts. Moreover, the combined effects of various changes are still highly uncertain [IPCC, 2001].

3. AN INTEGRATED ENVIRONMENTAL-ECONOMIC MODELLING FRAMEWORK

The impacts of agricultural production on natural conditions strongly depend on specific local conditions. Changes in water or nutrient cycles are related to soil conditions, terrain type and local climate conditions. Hence it is necessary to link economic conditions of agricultural production to the place-specific biophysical conditions, in order to better understand their interactions. The key challenge with respect to modelling is to link place-specific models of agricultural production and land use with models representing important elements of the biosphere and hydrology. Our starting point for improving the understanding of society-biosphere interactions is the extension of one of the most prominent models of the global biosphere – the Lund-Potsdam-Jena (LPJ) Dynamic Global Vegetation Model. We suggest a way to integrate human activities into LPJ and come up with a coupled climate-biosphere-economy modelling framework, including the water cycle. This is an important improvement on existing research, as LPJ endogenously models the dynamic linkages between climate and soil conditions, water availability and plant growth. It can be used to define plausible biophysical constraints to agricultural production and other human activities [Sitch et al., 2003]. The global version of LPJ has a spatial resolution of 0.5 degrees, which is equivalent to a grid size of about 50x50 km at the equator. The fundamental entity simulated in LPJ is the average individual of a plant-functional type (PFT). This concept provides a simple way for process acting at the level of the plant individual to be scaled up to the "population" over a grid cell. The grid cell is treated as a mosaic divided into fractional covers of PFTs and bare ground. LPJ simulates the global terrestrial carbon pool sizes and fluxes, and captures the biogeographical distribution of Earth's major biomes. In addition to the PFTs representing natural vegetation, 13 crop functional types (CFTs) have been implemented in order to simulate potential agricultural production. These CFTs represent 8 classes of agricultural crops, e.g. temperate cereals (wheat), tropical cereals (millet), rice, maize, pulses (lentil), oil crops (sunflower, soybean, groundnut, rapeseed), roots and tubers (sugar beet, maniok), and fodder crops (C3 and C4 grass). Input data required by LPJ are monthly fields of mean temperature, precipitation and cloud cover. Standard LPJ outputs include changes in net primary production and different fractions of biomass, changes in carbon pools and water balances. Under given climate conditions, soil type and water supply, the CFTs generate crop yields in terms of above-ground biomass as well as harvested organs. LPJ is written in C++ code.

The socio-economic part of the coupled modelling system, MAgPIE, is a linear-programming model with a focus on agricultural production, land and water use. The goal function is to produce a required amount of food energy at minimal costs. Food demand is defined for an exogenously given population in three energy categories (crops, meat, milk). Energy can be produced with 8 cropping activities (bread grain, feed grain, oil crops, sugar crops, roots/tubers/pulses, vegetables/fruits/nuts, rice, fodder crops) and 3 livestock activities (ruminant meat, non-ruminant meat, milk). Variable inputs of production are labour, chemicals, and other capital (measured in US$). Crop land, pasture and water are fixed inputs in limited supply, measured in physical units. MAgPIE is written in Python code. Currently the two models are coupled offline by exchanging text files. Later, we will use the typed data transfer (TDT) protocol.

Currently we only look at one pilot region without external trade. The regional demand for intermediate inputs like feed grain and green fodder has to be met by regional production. Water supply is currently defined purely by precipitation inflows. We abstract from groundwater reservoirs, lakes or other water storages. Water demand from production activities is calculated using fixed coefficients per unit of crop or livestock output. In order to keep the cropping mix within plausible bounds we introduce rotational constraints. Average production costs are based on data from FAO and the Global Trade Analysis Project. MAgPIE output includes the shares of different crops in total food energy production and land use, purchases of variable inputs, and shadow prices for fixed inputs and other constraints. The generation of shadow prices is especially useful, as it facilitates the assignment of internal use values to factors of production for which no proper markets and, hence, no observable prices exist.

Several challenges have to be overcome in coupling the two models. First, thematic scales have to be matched. CFTs in LPJ, based on plant-physiological properties, have to be matched with groups of crops which provide a similar type of output for human consumption. Oil crops, for instance, comprise a wide variety of plant species (e.g. rapeseed, groundnuts, sunflowers, oil palms etc.), but they all deliver similar types of oil, which are almost perfectly substitutable in the processing of agricultural products. Second, temporal scales have to be made consistent. While LPJ is usually run over a period up to the year 2100, most economic forecasts do not go beyond 2020, as changes in technology and input use are very hard to predict in the longer run. Third, we have to bridge the gap between the national or regional scale in MAgPIE and the 50x50km-grid scale in LPJ. While it is hardly possible to model economic activity on a 0.5-degree-grid for a larger region, it does not make much sense either to model environmental impacts on the aggregated national level. In order to bridge this gap, we grouped LPJ cells into a small number of "productivity zones", according to the normalised level of crop yields in each grid cell.

For our sample case of Germany we have 185 grid cells grouped in 6 different zones. Effectively this means, that MAgPIE can choose among 8 cropping activities and 3 livestock activities in 6 different zones, yielding in total 66 different production activities in the given region. We can distinguish between constraints to be fulfilled in each zone and constraints to be fulfilled at the regional level. This introduces aspects of trade between zones. For instance, feed grain produced in any zone is pooled across all zones and can be used in the whole region, as long as the overall balance is maintained. In contrast, green fodder has to be used locally and, hence, we impose a separate constraint for each zone. Land and water are also constrained in each zone. Having separate constraints for different zones implies that MAgPIE generates different land use shares and shadow prices for each zone.

4. SCENARIOS AND SELECTED MODEL RESULTS

Our reference scenario in this paper is the situation in Germany in the year 2000, defined by climate conditions, yields, the fraction of arable land and pasture in total area, and cost structures in agricultural production at this point in time. Then we look at 4 different drivers of agricultural production: a change in climate conditions as predicted for the year 2020 ("environmental change"); an increase in total food energy demand by 10 % ("income change"); a decrease in meat energy demand by 10 % ("lifestyle change"); a decrease in available crop land by 10 % ("demand for land from energy sector"). These different impacts can be analysed separately, but here we only present the combined results for environmental and socio-economic drivers.

In step 1 of our analysis we run LPJ separately with each CFT in order to define potential yields for each grid cell. Figure 1 shows yield distributions for "temperate cereals" in 2000. The map reveals significant variation in yields across the region. Currently, yields in LPJ strongly depend on precipitation and less on soil conditions. This is partly due to the rather crude soil classification in the FAO soil data set used in LPJ.

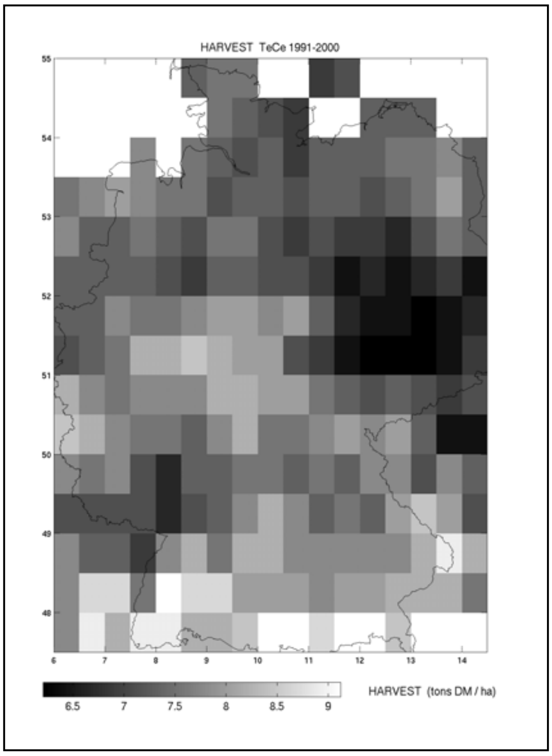


Figure 1: Regional distribution of potential cereal yields in Germany in 2000 (model results)

In step 2 we use normalised yields for cereals and maize as the most important crops in terms of land use share, in order to define 6 productivity zones. In step 3 of our analysis these characteristics of zones and yields are implemented in MAgPIE, and in step 4 agricultural production and resource use are optimised for the sample region. Total food energy demand for Germany is calculated by multiplying a population of 82 million by an average daily food *availability* of 3411 kcal or 14272 MJ (according to the FAO food balance sheets). More precise data on effective food *intake* are not available. The shares in total food energy consumption are 69 % for plant-based energy, 17 % for meat-based energy, and 14 % for milk-based energy. With the current specification of MAgPIE, in the reference situation total food demand in Germany can be met, the self-sufficiency ratio is about 110 %. Under these conditions the model leaves in the optimised solution about 10 % of crop land and 9 % of pasture unused (Table 3).

Table 3: Average land use shares for Germany under various scenarios (model calculations, %)

	Year 2000	Climate 2020	Demand increase	Reduced meat	Reduced crop land	Com-bined
Bread grain	16	11	11	21	11	13
Feed grain	50	53	55	45	55	52
Rapeseed	14	15	19	9	22	18
Sugar beet	0	0	3	0	1	0
Silage maize	10	11	12	10	11	12
Unused cropland	10	9	0	15	0	5
Unused pasture	9	1	4	9	10	1

The shares of rapeseed and sugar beet in total land use are currently underestimated in the model compared to observed data. Climate change, a demand increase and reduced crop land tend to shift land use towards more feed grain and high-energy crops like sugar beets and rapeseed. Reduced meat consumption leaves room for more bread grain production. In the combined scenario some of these effects eliminate each other. The resulting shadow prices for the combined scenario show considerable variation between zones, as e.g. crop land and pasture are scarce in some zones, but not in all (Table 4). Water is not a binding constraint in any zone, i.e. the shadow price is always zero. The rotational constraint on cereals is binding in all zones, except zone 2, which is rather small in this scenario.

Table 4: Regional and zone-specific shadow prices in combined scenario (model calculations)

Zone	*1*	*2*	*3*	*4*	*5*	*6*
Regional constraints						
Crop energy	17					
Meat energy	852					
Milk energy	526					
Feed grain balance	18					
Zone-specific constraints						
Fodder balance	14	127	12	14	11	13
Crop land	612	1288	319	213	106	0
Pasture	2800	0	2848	2819	2875	2841
Rotation cereals	291	0	319	310	301	272
Water	0	0	0	0	0	0

The fact that pasture has a higher shadow price than crop land is implausible and indicates that livestock production activities in MAgPIE have to be refined. In step 5 the land use patterns for each zone are implemented in LPJ and in step 6 the impacts on net primary production (NPP), carbon and water balances are calculated. Figure 2 shows the difference in NPP in the combined scenario compared to the reference situation in 2000. Changes in NPP, water and carbon balances are dominated by climate impacts in our current scenarios, as these affect both crops and forest. Forest and unused land account for about 40 percent of the total area in Germany. It will be subject to future research, whether different cropping patterns have a significant impact on environmental indicators in other regions.

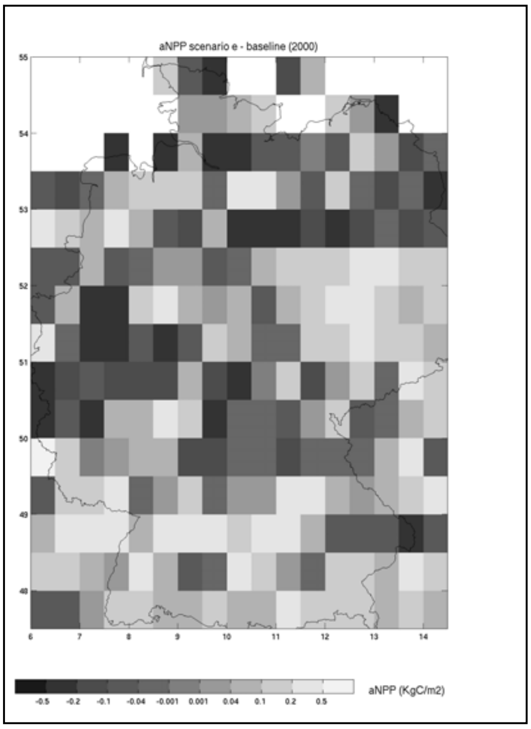


Figure 2: Regional changes in Net Primary Production (NPP), combined scenario compared to reference (model calculations)

5. CONCLUSIONS

With the preceding analysis we have shown how a grid-based dynamic global vegetation model and a non-spatial economic optimisation model can be coupled. The preliminary results show the viability of the concept. This modelling approach can in principle be run on the small scale of a single grid cell or even a single farm as well as on the global scale. Hence, it provides the opportunity for consistent spatial aggregation and dis-aggregation and nested modelling structures. It can also be coupled to a food demand model or an economy-wide model, in order to make markets and prices for outputs and inputs endogenous. Here we will build upon recent developments in the area of model coupling and meta-optimisation at PIK [Jaeger et al., 2002].

However, several caveats apply. The 0.5-degree-resolution of the current version of LPJ is appropriate on the global scale, but too coarse for the analysis of specific smaller regions. Crop yields and crop growth functions in LPJ have to be further evaluated. The specification of production activities in MAgPIE is rather preliminary, especially the linkages between livestock and crop production, and water requirements by crops have to be refined. The linear-programming technique is powerful, flexible, and computationally very efficient. However, a linear-programming model may not be robust under conditions of large structural breaks. Our current definition of productivity zones has to be reconsidered for global-scale applications.

Immediate further research steps include the definition of several economic regions and to allow for global trade in products among them. Activities of land conversion (e.g. deforestation, bio-fuel production) are also indispensable for modelling agricultural production on a global scale. A dynamic version of MAgPIE would be required to model perennial crops or forest management, and also to implement management of stocks of natural resources. The most challenging task will be the implementation of technological change, which is crucial in the very long run. Many aspects of water and nutrient cycles are only poorly monitored and not yet well understood, but they are strongly influenced by agricultural production technologies. The model results have to be validated using satellite remote sensing and agricultural statistics.

6. REFERENCES

Alexandratos, N., World food and agriculture: Outlook for the medium and longer term, *Proceedings of the National Academy of Sciences of the United States of America*, *96* (11), 5908-5914, 1999.

Barthélemy, F., D. Renault, and W. Wallender, Water for a sustainable human nutrition : inputs and resources analysis for arid areas, UC Davis Internal report, 1993.

Bender, W., and M. Smith, Population, food, and nutrition, *Population Bulletin*, *51* (4), 2-47, 1997.

Bender, W.H., How much food will we need in the 21st century?, *Environment*, *39* (2), 7-28, 1997.

Döös, B.R., Population growth and loss of arable land, *Global Environmental Change-Human and Policy Dimensions*, *12* (4), 303-311, 2002.

Gilland, B., World population and food supply - Can food production keep pace with population growth in the next half-century?, *Food Policy*, *27* (1), 47-63, 2002.

Harris, J.M., and S. Kennedy, Carrying capacity in agriculture: global and regional issues, *Ecological Economics*, *29* (3), 443-461, 1999.

Intergovernmental Panel on Climate Change (IPCC), Third Assessment Report - Climate Change 2001, 2001.

Jaeger, C.C., M. Leimbach, C. Carraro, K. Hasselmann, J.C. Hourcade, A. Keeler, and R. Klein, Integrated assessment modeling: modules for cooperation, FEEM Working Paper 53-2002, 2002.

Keyzer, M.A., M.D. Merbis, and I.F.P.W. Pavel, Can we feed the animals? Origins and Implications of rising meat demand, Centre for World Food Studies, Amsterdam, 2001.

Kindall, H.W., and D. Pimentel, Constraints on the expansion of the global food supply, *Ambio*, *23* (3), 1994.

Lutz, W., W. Sanderson, and S. Scherbov, The end of population growth, *Nature*, *412*, 543-545, 2001.

Qaim, M., and D. Zilberman, Yield effects of genetically modified crops in developing countries, *Science*, *299*, 900-902, 2003.

Rijsberman, F.R., Molden, D., Balancing water uses: water for food and water for nature, in *Thematic background paper to the International Conference on Freshwater*, Bonn, 2001.

Rosegrant, M.W., and X. Cai, Global water demand and supply projections. Part 2: results and prospects to 2025, *Water International*, *27* (2), 170-182, 2003.

Rosegrant, M.W., and C. Ringler, World food markets into the 21st century: environmental and resource constraints and policies, *Australian Journal of Agricultural and Resource Economics*, *41* (3), 401-428, 1997.

Sands, R.D., and M. Leimbach, Modeling agriculture and land use in an integrated assessment framework, *Climatic Change*, *56*, 185-210, 2003.

Sitch, S., B. Smith, I.C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. Kaplan, S. Levis, W. Lucht, M. Sykes, K. Thonicke, and S. Venevski, Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ Dynamic Vegetation Model, *Global Change Biology*, *9*, 161-185, 2003.

Using the Object Modeling System for hydrological model development and application

S. Kralisch[a], **P. Krause**[a], **and O. David**[b]

[a]*Friedrich-Schiller-University Jena, Institute for Geography, Jena, Germany (sven.kralisch@uni-jena.de)*

[b]*Colorado State University, Fort Collins, CO, U.S.A*

Abstract: State of the art challenges in sustainable management of water resources have created demand for integrated, flexible and easy to use hydrological models which are able to simulate the quantitative and qualitative aspects of the hydrological cycle with a sufficient degree of certainty. Existing models which have been developed to fit these needs are often constrained to specific scales or purposes and thus can not be easily adapted to meet different challenges. As a solution for flexible and modularised model development and application, the Object Modeling System (OMS) has been developed in a joint approach by the USGS (Denver, Col.), the USDA (Fort Collins, Col.) and the FSU (Jena, Germany). The OMS provides a modern modelling framework which allows the implementation of single process components to be compiled and applied as custom tailored model assemblies. This paper describes briefly the OMS and its main components and explains more in detail how the problems during coupling of models or model components are solved inside the system. It highlights the integration of different spatial and temporal scales by their representation as spatial modelling entities embedded into time compound components. As an example the implementation of the hydrological model J2000 is discussed.

Keywords: Object-oriented modeling system; OMS; modular model development

1 INTRODUCTION

With the implementation of the European Water Framework Directive (WFD) in December 2000 prognostic modelling for sustainable management of water resources has become even more important than it was before. The goals set up by the WFD require a stronger integrative and multidisciplinary approach than usually practiced in the last decades. In addition to quantitative and qualitative hydrological issues, socio-economic and legislative objectives must be considered to find the best solutions for the maintenance or improvement of the water quality in European water bodies.
Although such an interdisciplinary and holistic approach is the most promising way to reach the goals set up by the WFD, it also introduces new problems, which have to be solved in advance. The most obvious problem is that each discipline involved in the development of strategies for sustainable management of water resources is uses its own methods and tools for prognostic simulation and modelling of the single processes of the water cycle throughout Europe. This is true not only from the multidisciplinary standpoint but also from a regional point of view. Scientists from one discipline in one part of Europe often use different models for the same purpose than other scientists in other parts of the continent because the constraints and environmental circumstances are different.
The most significant differences between the single models or modelling systems applied in Europe and worldwide are the specific model cores which simulate the single processes. On the other hand, all models or modelling systems have systematic functionalities (e.g. data in- and output) which are principally common for all models even if they had been implemented in different ways.
For future proof model development and application, a modular approach which divides the systematic routines from the scientific parts is the most promising approach. Such an approach should provide the basic functionality for data in- and output, application and communication of the single components as well as an application programming interface (API) for the implementation

of the scientific methods in the form of encapsulated programme modules. The most relevant benefit of such a framework for model developers would be to enable them to concentrate on only the implementation of most suitable methods and always be confronted with a familiar interface and modelling environment.

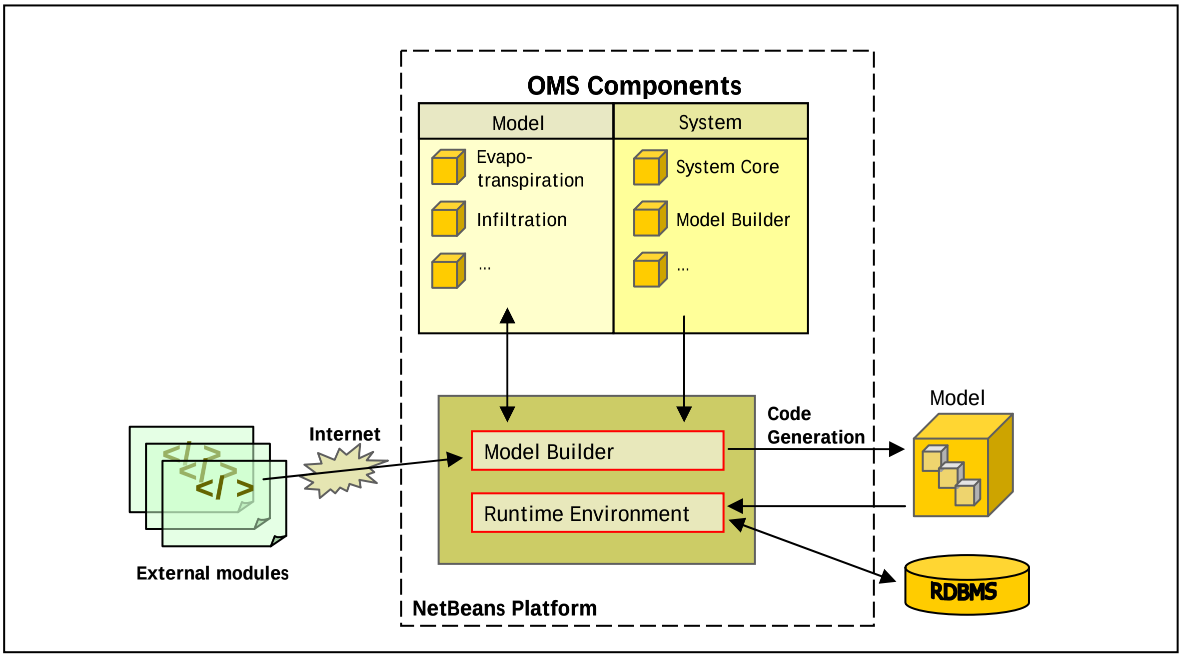


Figure 1. Principal layout of the Object Modeling System OMS

The Object Modeling System (OMS) (David et. al [2002]), developed at the Friedrich-Schiller-University in Jena, Germany and at the United States Department for Agriculture in Fort Collins, Colorado USA, is such a modelling framework. This paper deals with its introduction and the current development and implementation of suitable programme modules.

2 THE OBJECT MODELING SYSTEM OMS

The basic OMS concept is the representation of all its system and model components as independent modules. These modules are coupled by standardised software interfaces. In order to achieve maximum platform independence, OMS was implemented in JAVA on top on the NetBeans platform [http://www.netbeans.org]. NetBeans is an open source software which provides common desktop applications requirements like menus, document management, and settings for the user or developer. Extension of the OMS by new system components is guarantied through integration into the NetBeans context because of the flexible and generic interface of this platform. The principal layout of the Object Modelling System is shown in figure 1. On top the modules divided into system and scientific components are shown. The system components provide the Model Builder with single scientific modules that can be assembled to provide a running model. The model itself can then be executed inside the Runtime Environment.
The result is a modular system whose components can be divided into *system components* and *model components*.

2.1 OMS system components

All functions which are needed for the modelling system are implemented by the system components. System components handle all coupling and execution of model components. The following main system components can be identified.

- **System Core**

The *system core* provides basic functionality for all other components and forms the runtime environment for model as well as component development and execution. The system core implements all the data types that can be used by model components. Besides simple data objects a number of more complex objects and components are implemented. As an example, objects covering time issues (e.g. the OMSCalendar) implement specific time steps for model application and methods for time step management. Other objects, like the OMSComponents are used as basic classes for the implementation of model components. They can be understood as containers, which have to be filled by the model or module developers by the implementation of own methods or processes. The interaction and communication of such modules with others is done by standardised functions which hook the modules into their specific context. Such basic functionalities provided by these containers offers a solid basis for consistent and sound module development and allow the module developer to concentrate on the process implementation without thinking too much about the system's functions.

- **Model Builder**

The *model builder* supports the assemblage and configuration of complex models from single model component with an easy to use graphical user interface. This interface offers capabilities for the mapping of components output parameters to input parameters of subsequent components.

With the model builder different model configurations can be stored and managed. Once a model has been assembled and configured inside the model-builder it can be easily passed to other users or executed in other computing environments.

- **Update Center**

The *update center* is a standard NetBeans component and provides functionality for easy update and download of existing modules for integration into OMS. All developers of OMS model components can encapsulate their work within NetBeans modules. By offering them via the Internet they can make their work available to other OMS users who can then retrieve and install them through the use of the update center. OMS-tailored NetBeans modules include not only the model components themselves, but they can also provide additional data like parameter-sets and documentation.

All model components can be supplied with additional signature keys and licensing information to protect them against modification without permission and to secure the property rights of the developer.

- **User Interface components**

The OMS offers well designed *user interface (UI) components* which provide a number of visualization features for developers and users. As an example, the 2d plot component offers miscellaneous 2d representations for modelling results in form of graphs or xy-plots. Additional UI components can easily be added by implementing them as NetBeans modules and integrating them into the framework.

2.2 OMS model components

The OMS system components are complemented by the OMS model components which form the building blocks for all models created within the framework. For each model component the following properties are prototyped by the OMS and have to be implemented by the developer:

- The model component implements four common methods: *register(), init(), run()* and *cleanup()*. The register method comprises commands and functionality that needs to be executed once during the model initialization stage, like loading native libraries containing model functionality. The init method includes code to be executed at the first invocation of a module, mostly for presetting parameters to initial values. Code contained in the run method is executed at each module invocation and contains the real functionality, e.g. the Penman-Monteith equation for calculation of the potential evapotranspiration. At the end of model execution, the cleanup method of each model component is executed in order to free resources used by the component.
- The read and write access of each variable is supervised by the OMS model component. To secure consistency during runtime, each variable with read access must have been written by a preceding module during the model execution. With this information, the runtime system is able to synchronise variable values between the executions of two successive model components and thus guarantee a harmonised data flow.

Whereas the model components themselves are implemented as Java classes, they may include calls to functions from native libraries using the Java Native Interface (JNI). Thus, a large amount of pre-existing process implementations can be accessed from the OMS with minor reprogramming efforts.

In addition to the model components for the implementation and execution of process representations, OMS provides specialised compound components. These work as containers for other model components and can be used to represent hierarchical structures. Each compound component provides an internal iterator which controls if and how the contained elements are enumerated. By modifying this iterator, arbitrary control structures like conditional or iterated execution of contained components can be realised. In case of a conditional operator the representing compound component simply decides if or if not the contained components are executed - dependent on some external condition checked by the component. For example, if the compound component represents a while-operator, the compound component consecutively iterates over all contained components as long as a specific external condition is valid. With these compound components, the runtime behaviour of the models can be very effectively structured and implemented.

Because of the representation of temporal or spatial contexts in many models, the iterated execution of model components is of special interest. Therefore predefined compound components for representation of temporal contexts (TimeCompoundComponent) and spatial contexts (SpatialCompoundComponent) are already implemented in OMS. TimeCompoundComponents (TCC) own

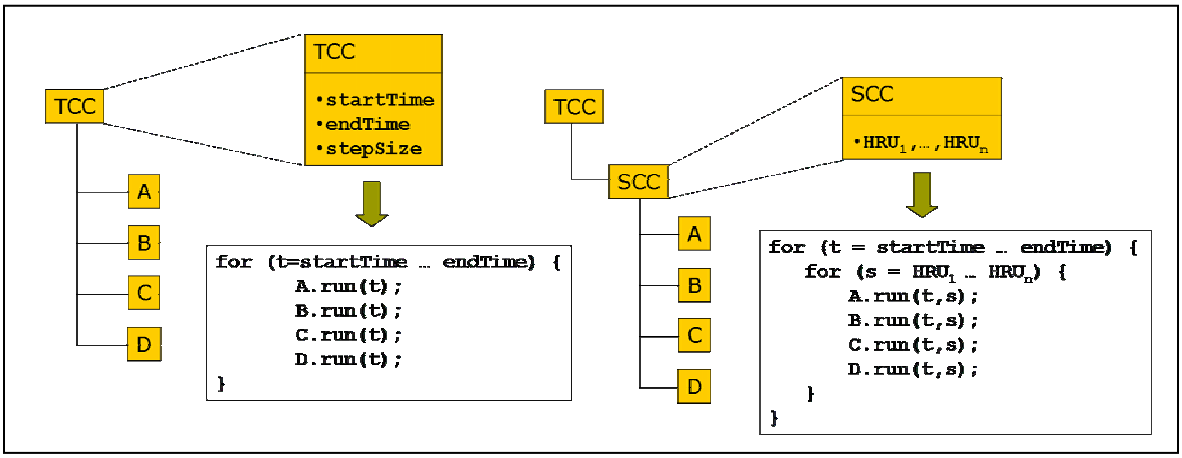


Figure 2. Application of compound components for representation of temporal and spatial contexts

a specific attribute that represents a user defined time interval and step size. With such data, TCC can create discrete points in time. On the other hand, SpatialCompoundComponents (SCC) represent discrete points in space by explicitly listing predefined spatial entities like Hydrological Response Units (HRU) or raster cells. Figure 2 shows an example of the execution of model components within a TCC on the left side. Here the process modules A to D are executed iteratively inside a time loop provided and controlled by the TCC. On the right side of figure 2, a SCC is integrated into the TCC. Again the modules A to D are executed iteratively but now across time and space. The right example also shows how compound components can be assembled in a hierarchical manner.

2.3 Representing space

In order to represent spatial model entities, the system core uses the data type OMSEntity. OMSEntities work as an abstract container for arbitrary attributes of spatial entities. These data which can vary for each OMSEntity object are stored in tables that map attribute names to their respective values. With this approach attribute sets of spatial model entities can easily be expanded if additional data are provided by new model components or external sources. Figure 3 shows an example of how spatial model entities (polygons or raster cells) can be obtained and represented as OMSEntity objects. The figure shows a basin with elevation and soil information. Each entity has a unique ID and a set of elevation and soil-type values. Specific getter and setter functions of the OMSEntity can be used to access the OMSEntity's attribute set. This attribute set is then used during model execution to provide the process modules with the required spatial distributed information.

When a model is assembled with the model builder, the single process modules can extend the OMSEntity set by process specific state variables

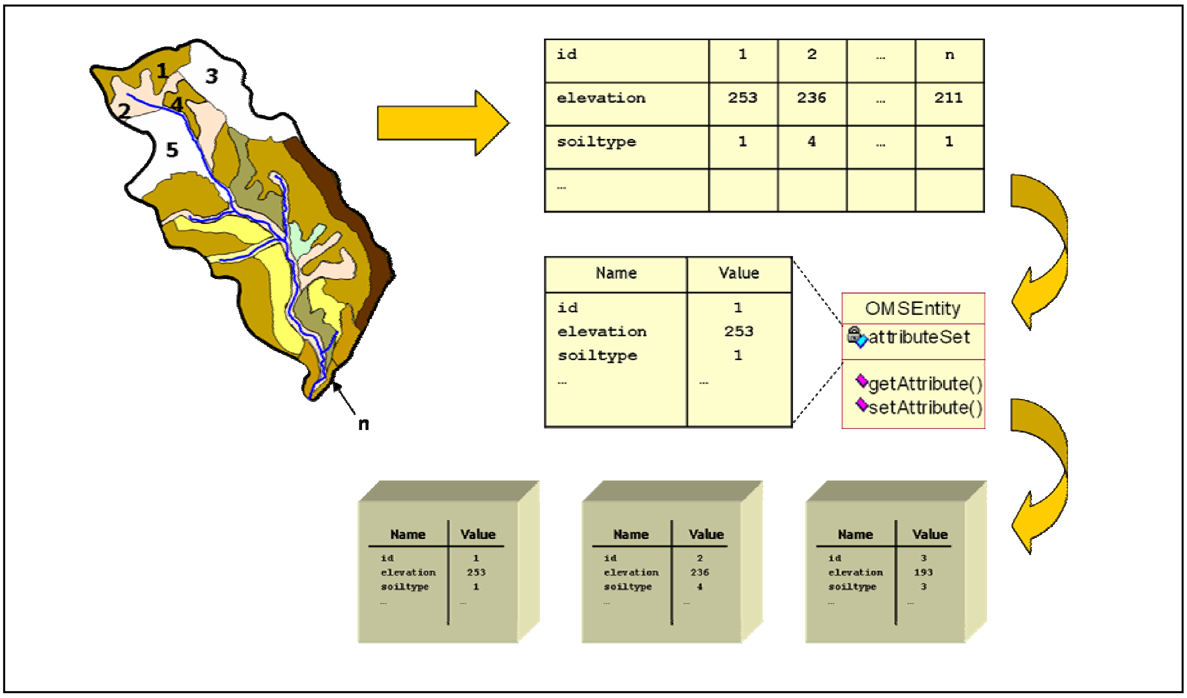


Figure 3. Representing spatial model entities as OMSEntity objects

and attributes. For example, a soil water process module with two different soil storages extends the attribute set of each OMSEntity object by two variables representing the storages during initialisation. Additionally, specific getter and setter routines are set up to retrieve or change the current state of these storages according to the running processes during the model execution.
The idea of representing spatial entities as abstract containers for arbitrary data opens the possibility of implementing model components which are not bound to a specific spatial discretisation (e.g., polygons or raster cells). Furthermore, they can work on OMSEntity data objects without consideration or knowledge of the underlying discretisation concept. This feature, of course, cannot solve problems of process validity and compatibility on specific spatial and temporal scales, which still have to be considered by the module or model developer.

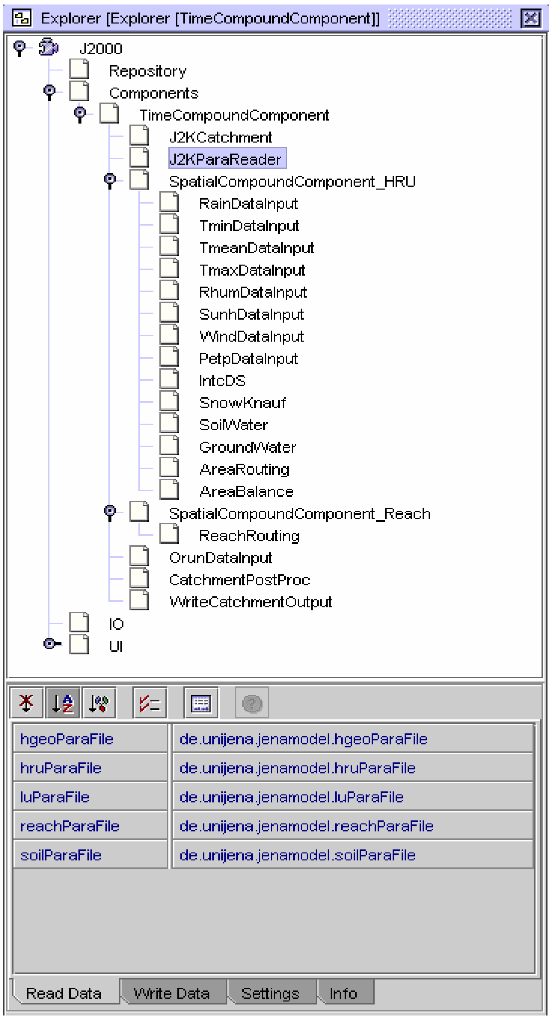


Figure 4. Composition of J2000 components with the model builder

3 APPLICATION OF THE OMS

In order to test the described system against an existing model, a number of model components have been implemented. The basis for these components was provided by the J2000 model (Krause 2001, Krause 2002) which is a conceptual fully distributive hydrological modelling system. The J2000 uses the topological HRU approach for catchment discretisation and implements the single processes of the hydrological cycle (ETP, snow, soil water, groundwater, lateral routing between the HRUs and channel routing) as encapsulated process modules. Therefore the J2000 already provides a number of cleanly implemented process descriptions written in Java. These were transferred into OMS model components with only minor adjustments in the initialisation and cleanup routines. Additionally, readers for transferring the J2000 parameter files (which describe the spatial entities `J2KParaReader`) and for the input data files (which contain the driving variables `xxxDataInput`) have been integrated into the OMS. Figure 4 shows the J2000 model setup inside the OMS Model Builder.
During initialisation within the OMS/J2000 model, the J2KParaReader component reads the J2000 two spatial model entity types (Hydrological Response Units and river reach units) together with their describing parameters from external data sources (files). The objects descriptions are then translated into two lists of OMSEntitiy objects, one for the HRUs and one for the river reaches. These lists can then be accessed by two different Spatial Compound Components which are used to create iterators over the lists. The attributes of the single objects in the OMSEntitiy lists are then updated by the process modules with additional attributes and state variables together with getter and setter routines to assess their content. Both SCC are then integrated into the temporal context of one TCC which iterates in daily time steps during model execution.
During model application for each iteration step of the TCC, the two SCCs execute a number of different process modules embedded in their context in a sequential order (figure 4). During the execution of the single process modules, the state of the spatial objects passed to them are changed according to the process implementation inside the modules and the updated states are given back to the system. An example of the run routine of a module implementing the interception process is shown in figure 5. In the beginning, the relevant HRU attributes are passed to local variables. Next, the process implementation follows and is terminated at the end by the return of the updated local variables back to the HRU object. Preliminary test runs of the OMS/J2000 model show that the system produces nearly identical results as the origi-

nal J2000 model. The differences are mostly determined by different treatment of the input variables and slightly differing parameter sets.

4 CONCLUSION AND OUTLOOK

This first implementation of a complete hydrological model into the OMS has shown the suitability of the system for model development and application. For the integration of the J2000 modules and its distribution, concept spatial entities have been developed and integrated into the OMS. Great care was taken to implement the OMSEntities as flexible and open as possible to guarantee their reusability for other process modules from other models. During the refactoring of the J2000 modules for the use inside OMS, only minor parts had to be adapted to ensure proper module installation and initialisation. The original process implementations were left untouched. Test runs of the OMS/J2000 implementation showed nearly identical results when compared to the original model results.

The comparison of the OMS/J2000 against the original implementation also showed that the OMS performance is much slower than the original implementation. This lack of performance is related to the flexibility gained by the use of dynamic attribute sets for model entities inside the OMS. To keep them updated and consistent during model execution, a lot of time consuming computing, like type casting, has to be performed. The optimisation of model performance and elimination of time consuming operations during model execution therefore forms the subject of ongoing research. A possible solution could be the utilization of alternative data structures for storing entity attributes. The GNU Trove library (http://trove4j.sourceforge.net/) e.g. provides a fast, lightweight implementation of the Java Collections API to realize this.

```
public class IntcDS extends OMSComponent {
 transient  OMSTimeInterval time;
 transient  OMSEntitySet es;

 public int run() {
  //pass hru variables to local variables
  OMSEntity currentEntity = this.es.current;
  int julday = (int) time.current.getDayInYear();
  double[] LAIArray     = (double[])
      currentEntity.getAttribute("LAI");
  double LAI            = LAIArray[julday-1];
  double area           =
      currentEntity.getDoubleAttribute("area");
  double dailyRain      =
      currentEntity.getDoubleAttribute("dailyRain");
  double dailyTmean     =
      currentEntity.getDoubleAttribute("dailyTmean");
  double dailyPetp      =
      currentEntity.getDoubleAttribute("dailyPetp");
  double dailyAetp      =
      currentEntity.getDoubleAttribute("dailyAetp");
  double deltaETP       = dailyPetp - dailyAetp;
  double actIntcStorage =
currentEntity.getDoubleAttribute("actIntcStorage");
  double throughfall    = 0;
  double dailyIntc      = 0;

  //calculate interception parameters
  double alpha = 0;
  if(dailyTmean < -2.0)
    alpha = 0.5;
  else
    alpha = 0.2;

  double maxIntcCap = (LAI * alpha) * area;

  if(actIntcStorage > maxIntcCap){
    throughfall = actIntcStorage - maxIntcCap;
    actIntcStorage = maxIntcCap;
  }

  double deltaIntc = maxIntcCap - actIntcStorage;

  if(deltaIntc > 0){
    double saveRain = dailyRain;
    if(dailyRain > deltaIntc){
      actIntcStorage = maxIntcCap;
      dailyRain = dailyRain - deltaIntc;
      deltaIntc = 0;
      dailyIntc = (saveRain - dailyRain);
    } else{
      actIntcStorage = (actIntcStorage + dailyRain);
      dailyIntc = dailyRain;
      dailyRain = 0;
    }
  }
  if(deltaETP > 0){
    if(actIntcStorage > deltaETP){
      actIntcStorage = actIntcStorage - deltaETP;
      dailyAetp = dailyAetp + deltaETP;
      deltaETP = 0;
    } else{
      deltaETP = deltaETP - actIntcStorage;
      dailyAetp = dailyAetp + (dailyPetp - deltaETP);
      actIntcStorage = 0;
    }
  }

  //return hru variables from local variables
  currentEntity.setDoubleAttribute("throughfall",
      throughfall);
  currentEntity.setDoubleAttribute("netPrecip",
      throughfall + this.dailyRain);
  currentEntity.setDoubleAttribute("dailyAetp",
      dailyAetp);
  currentEntity.setDoubleAttribute("actIntcStorage",
      actIntcStorage);
  currentEntity.setDoubleAttribute("dailyIntc",
      dailyIntc);

  return 0;
 }
```

Figure 5. Example of the run routine of a OMS/J2000 process module.

5 REFERENCES

David O., S.L. Markstrom, K.W. Rojas, L.R. Ahuja, and I.W. Schneider (2002). The Object Modeling System, In: *Agricultural System Models in Field Research and Technology Transfer*, L. Ahuja, L. Ma, T.A. Howell, Eds., Lewis Publishers, CRC Press LLC, 2002: 317—331.

Krause, P. (2001): Das hydrologische Modellsystem J2000 – Beschreibung und Anwendung in großen Flußgebieten (The hydrological modelling system J2000 – Documentation and application in large river basins); Schriften des Forschungszentrums Jülich, Reihe Umwelt/Environment, Band 29

Krause, P. (2002): Quantifying the impact of land use changes on the water balance of large catchments using the J2000 model; Physics and Chemistry of the Earth, 27, p. 663-673

Developing Natural Resource Models Using the Object Modeling System: Feasibility and Challenges

L.R. Ahuja [a], **O. David** [b], **and J.C. Ascough II** [c]

[a] *Soil Scientist, USDA-ARS, Great Plains Systems Research Unit (GPSRU), Fort Collins, CO 80526 USA*

[b] *Research Associate, Colorado State University, Fort Collins, CO 80523 USA (olaf.david@ars.usda.gov)*

[c] *Research Hydrologic Engineer, USDA-ARS, GPSRU, Fort Collins, Colorado 80526 USA*

Abstract: Current challenges in natural resource management have created demand for integrated, flexible, and easily parameterized hydrologic models. Most of these monolithic models are not modular, thus modifications (e.g., changes in process representation) require considerable time, effort, and expense. In this paper, the feasibility and challenges of using the Object Modeling System (OMS) for natural resource model development will be explored. The OMS is a Java-based modeling framework that facilitates simulation model development, evaluation and deployment. In general, OMS consists of a library of science, control, and database modules and a means to assemble the selected modules into an application-specific modeling package. The framework is supported by data dictionary, data retrieval, GIS, graphical visualization, and statistical analysis utility modules. Specific features of OMS that will be examined (with respect to natural resource/hydrologic modeling) include: 1) how to reduce duplication of effort in natural resource modeling; 2) how to make natural resource models easier to build, apply, and evaluate; 3) how to facilitate long-term maintainability of existing and new natural resource models; and 4) how to improve the quality of natural resource model code and ensure credibility of model implementations.

Keywords: Object Modeling System, Natural resource models; Object-oriented programming, Simulation framework.

1. INTRODUCTION

The problems facing both developers and users of natural resource models are becoming increasingly complex. Tremendous progress has been made in discovering basic principles in different scientific disciplines that created major advances in management and technology for natural resource systems. However, understanding natural resource management issues related to ecology (habitat restoration), hydrology (reservoir management), and farming practices (fertilizer and chemical application) become compounded when viewed within the physical, biological, chemical, and geological responses of the natural world. Computer simulations for prediction and management of watersheds, water supply areas, and agricultural fields and farms have also increased in complexity. The multidisciplinary nature of these problems usually requires accounting for a significant number of different models, data sources, management alternatives, and customers/stakeholders. This is particularly true in the case of agriculture where awareness of the general public requires careful management of agricultural systems in order to protect soil, water, and air quality. Simultaneously, market-based global competition in agricultural goods is challenging the economic feasibility of traditional agricultural production systems, and compels development of new and dynamic cropping and management strategies.

It can be argued that achieving the goal of sustainable natural resource management should involve consideration of whole system effects. Unfortunately, most natural resource systems involve highly complex interactions of soil-plant-weather-management components that are extremely difficult to quantitatively describe. Thus, state of the art challenges in optimal management of the natural resources have created demand for integrated, flexible and easy to use modeling tools which are able to simulate the quantitative and qualitative aspects of the system (e.g., the hydrological cycle) with a sufficient degree of certainty. Although a myriad of models are available, they are typically constrained to the specific scales and purposes they have been developed for and therefore are more robust in some areas than others (depending on the primary goal guiding their development). Furthermore, most of these monolithic models not modular; are very difficult to update, add to, or connect with other models; have diminishing technical support; and lack the flexibility to meet current needs for more integrated analysis of changing natural resource issues.

All of the above reasons indicate a need for a new framework of model development that can integrate existing (and future) natural resource models into a common, collaborative, and flexible system. Such a system will maintain modularity, reusability, and interoperabililty or compatibility of both science and auxiliary components. The system will also recognize the fact that different categories of applications may require different levels of scientific detail and comprehensiveness, as driven by problem objectives, scale of application, and data constraints. These functionalities of the system will be obtained by establishing standard libraries of inter-operable science and auxiliary components or modules that provide the building blocks for a number of similar applications. Module libraries have been successfully used in several domains, such as the manufacturing, transport, and other systems (Breunese et al., 1998; Praehofer, 1996). One of the earliest modular model developments was done for SHE, the European Hydrologic System Model (Abbot et al., 1986). Leavesley et al. (1996) reported the conversion of the Precipitation Runoff Modeling System (PRMS) to a Unix-based Modular Modeling System (MMS) for hydrologic modeling. Leavesley et al. (2002) presented some successful applications of this concept. To summarize, an approach for modeling natural resource systems is needed that will:

- Reduce duplication of effort;
- Improve the quality and currency of model code;
- Make natural resource models much easier to build, access, understand and use;
- Facilitate long-term maintainability of existing and new natural resource models;
- Lead to greater consistency of modeling for particular problems and scales;
- Enhance response and delivery times in scientific modeling projects;
- Ensure creditability and security of model implementations; and
- Function on any major computing platform.

The Object Modeling System (OMS) being developed by the USDA-ARS Great Plains Systems Research Unit (Fort Collins, CO) and the USGS (Denver, CO) meets the above criteria. The OMS provides a modular modeling framework which allows the implementation of single process modules which can be compiled and applied as custom-tailored model assemblies.

2. OBJECT MODELING SYSTEM (OMS) DESCRIPTION

2.1 Introduction

The OMS project was initiated in 1996 at the Friedrich Schiller University of Jena. In October 2000, OMS evolved into an interagency project between the USDA-ARS, USGS, and USDA-NRCS, with financial support from ARS and in-kind support from the partners. During the past 36 months, the OMS programming team has completed the development of most of the core components of OMS. The vision of OMS, described below for (initial) ARS implementation, is close to being realized (Ahuja et al., 2002):

"The OMS is a computer framework consisting of: 1) a library of science, control, and database modules; 2) a means to assemble the selected modules into a modeling package customized to the problem, data constraints, and scale of application; 3) an automatic generation of a friendly user interface; and 4) creation of a compiled, ready-to-run, version of the package. The framework is supported by utility modules such as data dictionary, data retrieval, GIS, graphical visualization, and statistical analysis. The framework employs the latest Java-based software technology for all its components. The science modules are also quickly updated or replaced as new knowledge and data become available. The OMS will be supported from a central server for use by all ARS scientists, NRCS specialists, USGS, and other collaborators."

OMS is built on top of the NetBeans platform. The NetBeans platform is a software framework for building desktop application software in the Java programming language. OMS leverages NetBeans features such as user interaction components (e.g., menu bars, tool bars, status displays, tabbed-window displays, properties, alerts and dialog boxes, printing controls, output consoles, wizards, setting persistence), storage access components, and help components (e.g., JavaHelp). A schematic of OMS implementation for natural resource modeling is presented in Figure 1.

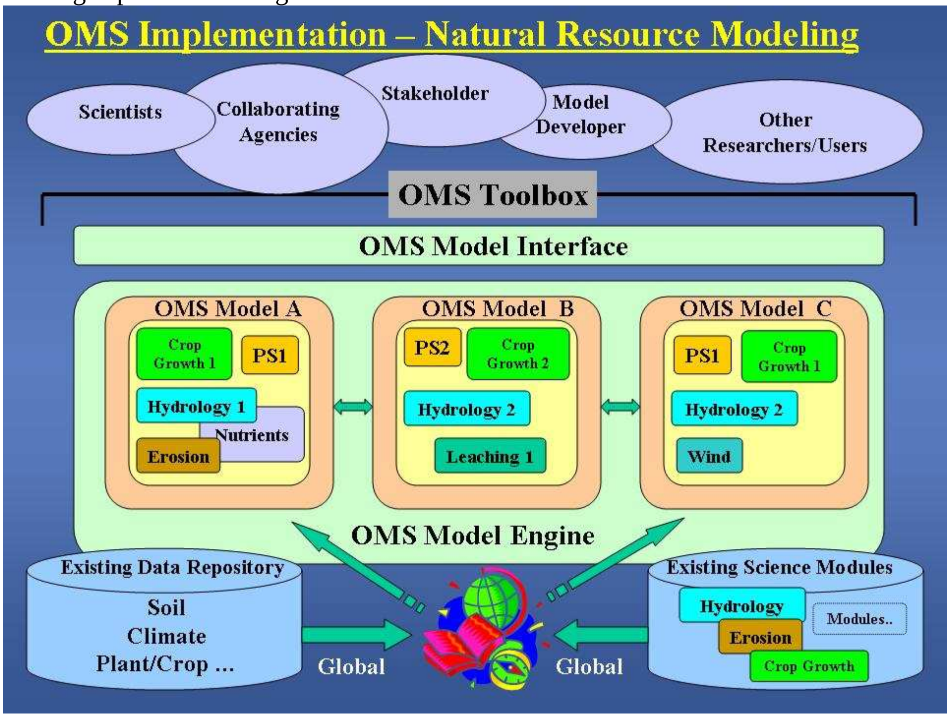


Figure 1. Schematic of OMS implementation for natural resource modeling.

2.2 Component Architecture

The general objectives of the OMS project included the development of generic software tools to extract modules from existing non-modular simulation models, and to incorporate them into the OMS framework with standard OMS descriptions. These tools have been developed, but need further testing and improvement. The OMS framework has the following functional components that are currently operational:

1. A module-building component that facilitates the integration of existing (legacy) code into the framework.
2. A module repository containing modules that can be readily utilized to assemble a working model (types of modules in the library will include science, control, utility, assessment, data access, and system modules.
3. A model builder that assembles modules from the module library into executable models and verifies data connectivity, and compatibility in scale and comprehensiveness.
4. A dictionary framework that manages extended modeling data type information and provides extended semantics checking for module connectivity verification.
5. An extensible user interface that facilitates an appropriate user interaction for general model development and application (it is supported by a number of contributing software packages for database management, visualization, and model deployment).

The components have the following architecture or characteristics:

1. OMS models are treated as hierarchical assembled components representing building blocks. Components are independent and reusable software units implementing processing objects for simulation models. They reside in a model library and are categorized into data access components, science components, control components, utility components, and system components.
2. OMS is able to integrate legacy code components. By an automated JAVA wrapper generation for legacy code, components written in languages such as Fortran or C can be embedded into OMS at the function level.
3. The “knowledge-backbone” of OMS is the dictionary framework. It enables OMS to verify state variables and parameters according to scientific nomenclatures during model development and application. Dictionaries are also used to specify parameter sets, model control information and the component connectivity. They are

implemented in the Extensible Markup Language (XML).

4. OMS is extensible. Extension packages exist for different aspects in model development and application. Extension packages are used for visual model assembly, model application, an interface to the dictionary framework, output visualization, and GIS integration.

5. OMS scales from a full-featured, stand-alone development system with tools for model assembly, visualization, and analysis to a runtime Web service environment.

For a more complete explanation of the OMS framework and architecture, the reader is referred to David et al. (2002).

3. NATURAL RESOURCE MODEL DEVELOPMENT WITH OMS

The following section lists the advantages (Ahuja et al., 2002) and disadvantages of developing and using OMS.

3.1 Advantages and Feasibility

1. *Efficient Transfer of Technology:* For the transfer of natural resource technology tools to stakeholders, researchers, and other users, the OMS will serve as a multidimensional platform for integration of various different software tools. The end users will then develop deployment links to only OMS, rather to develop links to each of the separate software tools as is current practice. This will result in a faster transfer of appropriate technology. For example, the USDA-NRCS, Information Technology Center (ITC) in Fort Collins, has been a partner in OMS development and is committed to using OMS as a means to provide technical tools to 2,500 field offices for natural resource conservation planning. The NRCS-ITC has already developed links to OMS for its automated runoff curve number approach. Finally, due to a common model building platform and a common user interface for all models, the OMS will result in reduced start-up time for model development and lower training costs for users.

2. *Cost Reduction in Upkeep, Maintenance, and Customer Support for Software Technology:* It is generally agreed that over the long-term, these items cost up to three times as much as for initial development of the software packages. At present, there are hundreds, if not thousands, of small to large natural resource software programs that need to be maintained and supported. A huge amount of time and labor is being spent on this process since monolithic models are becoming very expensive to use, difficult to update, have diminishing technical support, and lack the flexibility to meet today's needs for more integrated analysis of natural resource issues. These problems can be overcome if many of the existing packages were transferred to OMS and all new packages were developed within OMS. By using the existing Modular Modeling System (MMS) (Leavesley et al., 1996) for PRMS (Leavesley et al., 1983) model development and deployment, the USGS has realized a huge cost savings in upkeep, maintenance, and customer support. The web site (http://iscmem.org/Memorandum.htm) lists details of a MOU that eight U.S. federal agencies (NRC, EPA, DOE, COE, USGS, ARS, NRCS, NOAA) have endorsed for the development and deployment of common methods and techniques for platform integration across all agencies.

3. *Cost Reduction in Developing New Software Technology:* In the past, model development efforts have primarily consisted of large teams of scientists. For example, ARS has had a number of individuals and teams build software technology and simulation models, including the cotton and soybean models Gossym and Glycim; the erosion models WEPP and WEPS; and water quality models GLEAMS, RZWQM, AnnAGNPS, and SWAT. Each of these packages cost many millions of dollars to develop, including scientist and support time. Development costs were also inflated by significant duplication of work. Natural resource model developers can now leverage that investment by putting the science in those packages as modules in OMS to build new customized software packages at a small fraction of the cost. Preliminary results in the core OMS development phase on the modularization of existing hydrologic models RZWQM and PRMS showed a code reduction of OMS versions of these two by 20-33% while keeping the same modeling results. In this age of information technology, the demand for such software packages will increase tremendously. In the ARS alone, if we were to develop ten customized large new system packages to meet this demand over the next ten years, we will save at least $100 million dollars, assuming 80% of the science comes from existing modules already put in OMS and 20% is new code.

4. *Applying the Most Suitable Science for Specific Problems:* The OMS will allow the selection of the best evaluated and most appropriate science modules currently available depending upon the nature of the problem and required answers, availability of input data, and scale of application. The OMS library may have different modules for a research model versus a management decision tool. Similarly, a watershed-

scale management model may possibly require different (i.e., less complex) science modules than a field-scale model.

5. *Assure Reliability in Results From Software Tools for Similar Applications:* Natural resource software application users often report that different software tools or simulation models give vastly different results, say for predicting crop yield, because the tools used different science approaches in key process areas. The OMS will significantly reduce this problem by utilizing evaluated, documented, and standardized modules for the basic science components for a given category of applications.

6. *OMS Library as a Reference and Coordination Tool for Future Research and Development:* The OMS library will be a repository of current, quantitative knowledge in different areas of natural resource system science. Future scientists could look to this library to help determine where further research and development are needed.

7. *Integrated Analysis of Natural Resource System Production and Conservation Issues:* Effective analysis and management of natural resource systems and the environment requires integration of tools and data types that now exist in an array of individual disparate models. The OMS will provide customized, whole-system tools for the analysis of production and conservation issues (e.g., environmental quality and global climate change management) in natural resource systems.

8. *OMS Certification Mechanism for Approved "Science Building Blocks***":** The OMS is supporting the technical certification of library components based on X.509 Certificates and the validation of such certificates. This will allow the agency to certify approved modeling components and models.

9. *Enhance Productivity of Scientists and Researchers:* The customized, best-quality, software tools developed through OMS will help field scientists quantify their results and transfer them to other soils and climates very rapidly. The gaps identified in the process will make future research more focused. Overall, the productivity of scientists and the quality of science should increase as focus centers on science module implementation rather than Graphical User Interface (GUI) design, software deployment, packaging and maintenance.

10. *International Coordination in New Science Module Development and Publication:* Through the Internet, the OMS will serve a common platform for international scientists and researchers to contribute their findings as modules to the OMS library. A supervisory group or organization (e.g., IAHS) could coordinate this development, and provide a mechanism for peer review and quality control. The module contribution to the library will be considered a publication by scientists that could have world-wide impact.

3.2 Challenges

Challenges to using OMS for natural resource model development stem from the following problems:

- *Lack of motivation to share model code* – in order to fulfil the intended purpose of OMS, model developers must build a repository of modules through contribution to the OMS module library.
- *Acceptance of a modular coding structure* – model developers must spend more "up-front" time in module development in terms of module structure, I/O requirements, metadata description, etc.
- *Willingness to share data sets for a range of natural resource processes covering different climatic and physiographic regions across the world* – application of natural resource models developed under OMS will be difficult without data sets for comparison and evaluation.
- *Loss of model name recognition* – this can be overcome, however, by the development of a mechanism for peer review and quality control of individual modules.

4. SUMMARY AND CONCLUSIONS

A large number of complex natural resource system models are currently in use worldwide. Most of these models are based on sound science, but have specialized data requirements and significant duplication exists in many areas. The current OMS development tool will leverage these sizeable investments of developer time and money to: 1) facilitate an interdisciplinary effort extracting the best scientific routines of existing models; and 2) provide integration and interoperability of existing and new scientific modules and modern data resources. To summarize, principal advantages of the OMS include:

1. OMS will increase the probability of using the best science available in various combinations for the given conditions and problem.
2. OMS will be easier to maintain and update as new knowledge, data and technology become available. OMS will allow a "select, plug, and play" mechanism for modules consisting of

sub-models, equations, graphics, statistics, risk analysis, parameter estimation, standard data sources and various reporting formats.

3. New knowledge expressed in the form of modules will be relatively easy to verify and evaluate (and possibly lead to scientific peer review).
4. OMS will help to eliminate duplicate functionality across natural resource models. The OMS library of modules will serve as a reference and a coordination mechanism for future improvements. OMS will facilitate communication between model developers by providing a common standard for development and implementation.
5. OMS should significantly reduce the problem of different natural resource models giving different results by utilizing a library of evaluated, documented, and standardized modules and integrated output options.
6. OMS will provide a consistent interface for model creation and evaluation, and will reduce startup time for scientific users and developers.
7. OMS offers support-ready compliance with the distinct advantage of having a set of application packages (e.g., data input, parameterization, visualization) under a single user interface or a set of consistent user interfaces.
8. OMS will allow flexibility to choose scientific modules most appropriate for the scale or region of interest, or to respond to other unique influencing factors under consideration.

In conclusion, the object-oriented and modular approach of the OMS and the modules/models implemented in it will provide the basis for more efficient and collaborative model development in the future. This type of integrative and open-source approach is desperately needed in order to solve global challenges impacting natural resource systems such as sustainable management of natural resource systems and the impact of global climate change on natural resource systems. For more details on the OMS project mission, project documentation, or to download the entire application or individual modules, visit the OMS web site at http://oms.ars.usda.gov/.

5. ACKNOWLEDGEMENTS

The authors wish to thank Ian Schneider and Ken Rojas for their programming support on OMS core model development.

6. REFERENCES

Ahuja, L.R., O. David, W. Blackburn, R. Amerman, J. Werner, J. Carlson, R. Knighton, and G. Leavesley, The Object Modeling System (OMS): an advanced object-oriented, modular modeling computer technology for agricultural production systems, USDA-ARS Working White Paper, Vers. 1, 13 pp., 2002.

Abbot, M.B., J.C. Bathurst, J.A. Cunge, P.E. O'Connell, and J. Rasmussen, An introduction to the European Hydrological System-Systeme Hydrologique Euopeen, “SHE”, 2: structure of a physically-based, distributed modelling system, *J. Hydrol.*, 87, 61-77, 1986.

Breunese, A.P.J., J.L. Top, J.F. Broenink, J.M. Akkermans, Libraries of reusable submodels: theory and application, *Simulation*, 77(1) (July), 7-22, 1998.

David, O., S.L. Markstrom, K.W. Rojas, L.R. Ahuja, and I.W. Schneider, The Object Modeling System, p. 317-330, In: L.R. Ahuja, L. Ma, and T.A. Howell (Eds.), Agricultural System Models in Field Research and Technology Transfer, Lewis Publishers, 357 pp., New York, 2002.

Leavesley, G.H., R.W. Lichty, B.M. Troutman, and L.G. Saindon, Precipitation Runoff Modeling System: User’s Manual, Water Resources Investigations 83-4238, USGS, Denver, Colorado, 1983.

Leavesley, G.H., P.J. Restrepo, S.L. Markstrom, M. Dixon, and L.G. Stannard, The Modular Modeling System (MMS): User’s Manual, Open File Report 96-151, USGS, Denver, Colorado, 1996.

Leavesley, G.H., S.L. Markstrom, P.J. Restrepo, and R.J. Viger, A modular approach to addressing model design, scale and parameter estimation issues in distributed hydrological modelling, *Hydrol. Proc.*, 16, 173-187, 2002.

Praehofer, H., Object oriented, modular hierarchical simulation modeling: towards reuse of simulation code, *Simul. Pr. Theory*, 4(4), 221-234, 1996.

Incorporating Level Set Methods in Geographical Information Systems (GIS) for Land-Surface Process Modeling

D. Pullar

Geography Planning and Architecture, The University of Queensland, Brisbane QLD 4072, Australia

Abstract: Land-surface processes include a broad class of models that operate at a landscape scale. Current modelling approaches tend to be specialised towards one type of process, yet it is the interaction of processes that is increasing seen as important to obtain a more integrated approach to land management. This paper presents a technique and a tool that may be applied generically to landscape processes. The technique tracks moving interfaces across landscapes for processes such as water flow, biochemical diffusion, and plant dispersal. Its theoretical development applies a Lagrangian approach to motion over a Eulerian grid space by tracking quantities across a landscape as an evolving front. An algorithm for this technique, called level set method, is implemented in a geographical information system (GIS). It fits with a field data model in GIS and is implemented as operators in map algebra. The paper describes an implementation of the level set methods in a map algebra programming language, called MapScript, and gives example program scripts for applications in ecology and hydrology.

Keywords: Spatial dynamics; Map algebra; GIS; Modelling

1. INTRODUCTION

Over the past decade there has been an explosion in the application of models to solve environmental issues. Many of these models are specific to one physical process and often require expert knowledge to use. Increasingly generic modeling frameworks are being sought to provide analytical tools to examine and resolve complex environmental and natural resource problems. These systems consider a variety of land condition characteristics, interactions and driving physical processes. Variables accounted for include climate, topography, soils, geology, land cover, vegetation and hydro-geography [Moore et al., 1993]. Physical interactions include processes for climatology, hydrology, topographic landsurface/sub-surface fluxes and biological/ecological systems [Sklar and Costanza, 1991]. Providing a generic environmental modeling framework for practical environmental issues is challenging. It does not exist now despite an overwhelming demand because there are deep technical challenges to build integrated modeling frameworks in a scientifically rigorous manner. It is this challenge this research addresses.

1. 1 Background for Approach

The paper describes a generic environmental modeling language integrated with a Geographical Information System (GIS) which supports spatial-temporal operators to model physical interactions occurring in two ways. The trivial case where interactions are isolated to a location, and the more common and complex case where interactions propagate spatially across landscape surfaces. The programming language has a strong theoretical and algorithmic basis. Theoretically, it assumes a Eulerian representation of state space, but propagates quantities across landscapes using Lagrangian equations of motion. In physics, a Lagrangian view focuses on how a quantity (water volume or particle) moves through space, whereas an Eulerian view focuses on a local fixed area of space and accounts for quantities moving through it. The benefit of this approach is that an Eulerian perspective is eminently suited to representing the variation of environmental phenomena across space, but it is difficult to conceptualise solutions for the equations of motion and has computational drawbacks [Press et al, 1992]. On the other hand, the Lagrangian view is often not favoured because it requires a global solution that makes it difficult to account for local variations, but has the advantage of solving equations of motion in an intuitive and numerically direct way. The research will address this dilemma by adopting a novel approach from the image processing discipline that uses a Lagrangian approach over an Eulerian grid. The approach, called level set methods, provides an efficient algorithm for modeling a natural advancing front in a host of settings [Sethian, 1999]. The reason the method works well over other approaches is that the advancing front is described by equations of motion (Lagrangian view), but computationally the front propagates over a vector field (Eulerian view). Hence, we have a very generic way to describe the motion of quantities, but can explicitly solve their advancing properties locally as propagating zones. The research work will adapt this technique for modeling the motion of environmental variables across time and space. Specifically, it will add new data models and operators to a geographical information system (GIS) for environmental modeling. This is considered to be a significant research imperative in spatial information science and technology [Goodchild, 2001].

The main focus of this paper is to evaluate if the level set method (Sethian, 1999) can:

- provide a theoretically and empirically supportable methodology for modeling a range of integral landscape processes,
- provide an algorithmic solution that is not sensitive to process timing, is computationally stable and efficient as compared to conventional explicit solutions to diffusive processes models,
- be developed as part of a generic modelling language in GIS to express integrated models for natural resource and environmental problems?

The outline for the paper is as follow. The next section will describe the theory for spatial-temporal processing using level sets. Section 3 describes how this is implemented in a map algebra programming language. Two application examples are given – an ecological and a hydrological example – to demonstrate the use of operators for computing reactive-diffusive interactions in landscapes. Section 4 summarises the contribution of this research.

2. THEORY

2. 1 Introduction

Level set methods [Sethian, 1999] have been applied in a large collection of applications including, physics, chemistry, fluid dynamics, combustion, material science, fabrication of microelectronics, and computer vision. Level set methods compute an advancing interface using an Eulerian grid and the Lagrangian equations of motion. They are similar to cost distance modeling used in GIS [Burroughs and McDonnell, 1998] in that they compute the spread of a variable across space, but the motion is based upon partial differential equations related to the physical process. The advancement of the interface is computed through time along a spatial gradient, and it may expand or contract in its extent. See Figure 1.

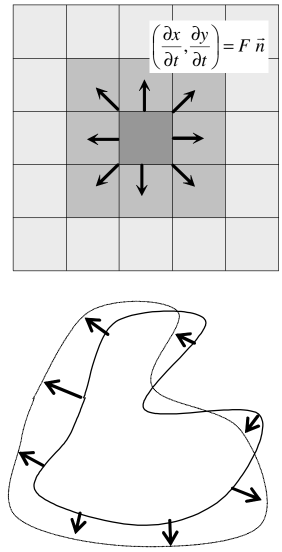


Figure 1. Shows a) a propagating interface parameterised by differential equations, b) interface fronts have variable intensity and may expand or contract based on field gradients and driving process.

2. 2 Theory

The advantage of the level set method is that it models motion along a state-space gradient. Level set methods start with the equation of motion, i.e. an advancing front with velocity F is characterised by an arrival surface $T(x,y)$. Note that F is a velocity field in a spatial sense. If F was constant this would result in an expanding series of circular fronts, but for different values in a velocity field the front will have a more contorted appearance as shown in Figure 1b. The motion of this interface is always normal to the interface boundary, and its progress is regulated by several factors:

$$F = f(L, G, I)$$

where L = local properties that determine the shape of advancing front, G = global properties related to governing forces for its motion, I = independent properties that regulate and influence the motion.

If the advancing front is modeled strictly in terms of the movement of entity particles, then a straightforward velocity equation describes its motion:

$$| \nabla T| F = 1 \qquad \text{given } T_0 = 0$$

where the arrival function $T(x,y)$ is a travel cost surface, and T_0 is the initial position of the interface. Instead we use level sets to describe the interface as a complex function. The level set function ϕ is an evolving front consistent with the underlying viscosity solution defined by partial differential equations. This is expressed by the equation:

$$\phi_t + F| \nabla\phi| = 0 \quad \text{given } \phi(x,y,t=0)$$

where ϕ_t is a complex interface function over time period $0..n$, i.e. $\phi(x,y,\ t= t_0..t_n,\ \nabla\phi$ is the spatial and temporal derivatives for viscosity equations. The Eulerian view over a spatial domain imposes a discretisation of space, i.e. the raster grid, which records changes in value z. Hence, the level set function becomes $\phi(x,y,z,t)$ to describe an evolving surface over time. Further details are given in Sethian [1999] along with efficient algorithms. The next section describes the integration of the level set methods with GIS.

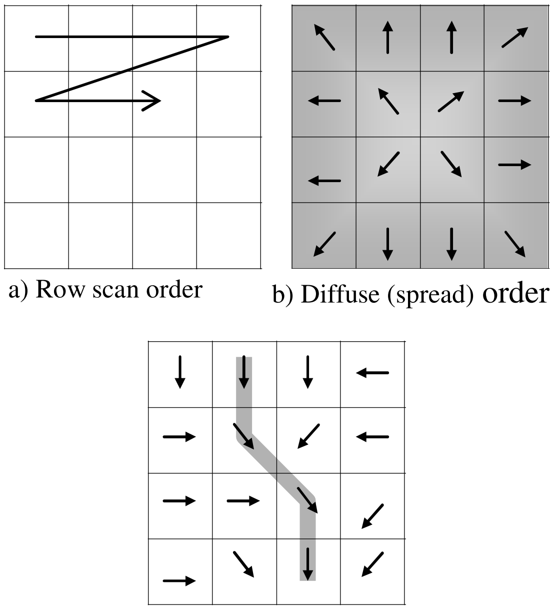
3. MAP ALGEBRA MODELLING

3. 1 Map Algebra

Spatial models are written in a map algebra programming language. Map algebra is a function-oriented language that operates on four implicit spatial data types: point, neighbourhood, zonal and whole landscape surfaces. Surfaces are typically represented as a discrete raster where a point is a cell, a neighbourhood is a kernel centred on a cell, and zones are groups of cells. Common examples of raster data include terrain models, categorical land cover maps, and scalar temperature surfaces. Map algebra is used to program many types of landscape models ranging from land suitability models to mineral exploration in the geosciences [Burrough and McDonnell, 1998] [Bonham-Carter, 1994].

The syntax for map algebra follows a mathematical style with statements expressed as equations. These equations use operators to manipulate spatial data types for point and neighbourhoods. Expressions that manipulate a raster surface may use a global operation or alternatively iterate over the cells in a raster. For instance the GRID map algebra [Gao et al., 1993] defines an iteration construct, called *docell*, to apply equations on a cell-by-cell basis. This is trivially performed on columns and rows in a clockwork manner. However, for environmental phenomena there are situations where the order of computations has a special significance. For instance, processes that involve spreading or transport acting along environmental gradients within the landscape. Therefore special control needs to be exercised on the order of execution. Burrough [1998] describes two extra control mechanisms for diffusion and directed topology. Figure 2 shows the three principle types of processing orders, and they are:

- row scan order governed by the clockwork lattice structure,
- spread order governed by the spreading or scattering of a material from a more concentrated region,
- flow order governed by advection which is the transport of a material due to velocity.



a) Row scan order b) Diffuse (spread) order

c) Advective (flow) order

Figure 2. Spatial processing orders for raster

Our implementation of map algebra, called MapScript [Pullar, 2001], includes a special iteration construct that supports these processing orders. MapScript is a lightweight language for processing raster-based GIS data using map algebra. The language parser and engine are built as a software component to interoperate with the IDRISI GIS [Eastman, 1997]. MapScript is built in C++ with a class hierarchy based upon a value type. Variants for value types include numerical, boolean, template, cells, or a grid. MapScript supports combinations of these data types within equations with basic arithmetic and relational comparison operators. Algebra operations on templates typically result in an aggregate value assigned to a cell [Pullar, 2001]; this is similar to the convolution integral in image algebras [Ritter et al., 1990]. The language supports iteration to execute a block of statements in three ways: a) *docell* construct to process raster in a row scan order, b) *dospread* construct to process raster in a spread order, c) *doflow* to process raster by flow order. Examples are given in subsequent sections. Process models will also involve a timing loop which may be handled as a general *while(<condition>)..end* construct in MapScript where the condition expression includes a system *time* variable. This time variable is used in a specific fashion along with a system time *step* by certain operators, namely *diffuse()* and *fluxflow()* described in the next section, to model diffusion and advection as a time evolving front. The evolving front represents quantities such as vegetation growth or surface runoff.

3.2 Ecological Example

This section presents an ecological example based upon plant dispersal in a landscape. The population of a species follows a controlled growth rate and at the same time spreads across landscapes. The theory of the rate of spread of an organism is given in Tilman and Kareiva [1997]. The area occupied by a species grows log-linear with time. This may be modelled by coupling a spatial diffusion term with an exponential population growth term; the combination produces the familiar reaction-diffusion model.

A simple growth population model is used where the reaction term considers one population controlled by births and mortalities is:

$$\frac{dN}{dt} = r \cdot (1 - \frac{N}{K}) \qquad (1)$$

where N is the size of the population, r is the rate of change of population given in terms of the difference between birth and mortality rates, and K is the carrying capacity. Further discussion of population models can be found in Jørgensen and Bendoricchio [2001]. The diffusive term spreads a quantity through space at a specified rate:

$$\frac{du}{dt} = dx \cdot D \frac{du}{dx} \qquad (2)$$

where u is the quantity which in our case is population size, and D is the diffusive coefficient.

The model is operated as a coupled computation. Over a discretized space, or raster, the diffusive term is estimated using a numerical scheme [Press et al., 1992]. The distance over which diffusion takes place in time step dt is minimally constrained by the raster resolution. For a stable computational process the following condition must be satisfied:

$$2D \frac{dt}{dx^2} \leq 1 \qquad (3)$$

This basically states that to account for the diffusive process, the term $2D \cdot dx$ is less than the velocity of the advancing front. This would not be difficult to compute if D is constant, but is problematic if D is variable with respect to landscape conditions.

This problem may be overcome by progressing along a diffusive front over the discrete raster based upon distance rather than being constrained by the cell resolution. The processing and diffusive operator is implemented in a map algebra programming language. The code fragment in Figure 3 shows a map algebra script for a single time step for the coupled reactive-diffusion model for population growth.

```
while (time < 100)
    dospread
        pop = pop + (diffuse(kernel*pop))
        pop = pop + (pop * (1 - (pop/K))
    enddo
end
```

where the diffusive constant is stored in the kernel:

kernel =

·	D	·
D	D	D
·	D	·

Figure 3. Map algebra script and convolution kernel for population dispersion. The variable *pop* is a raster, *K* and *D* are constants, and the kernel is a 3x3 template. It is assumed a time step is defined and the script is run in a simulation. The first line contained in the nested cell processing construct (i.e. *dospread*) is the diffusive term and the second line is the population growth term.

The operator of interest in the script shown in Figure 3 is the *diffuse* operator. It is assumed that the script is run with a given time step. The operator uses a system time *step* which is computed to balance the effect of process errors with efficient computation. With knowledge of the time step the iterative construct applies an appropriate distance propagation such that the condition in Equation 3 is not violated. The level set algorithm [Sethian, 1999] is used to do this in a stable and accurate way. As a diffusive front propagates through the raster, a cost distance kernel assigns the proper time to each raster cell. The time assigned to the cell corresponds to the minimal cost it takes to reach that cell. Hence cell processing is controlled by propagating the kernel outward at a speed adaptive to the local context rather than meeting an arbitrary global constraint.

3.3 Hydrological Example

This section presents a hydrological example based upon surface dispersal of excess rainfall across the terrain. The movement of water is described by the continuity equation:

$$\frac{\partial h}{\partial t} = e_t - \nabla \cdot q_t \qquad (4)$$

where h is the water depth (m), e_t is the rainfall excess (m/s), q is the discharge (m/hr) at time t. Discharge is assumed to have steady uniform flow conditions, and is determined by Manning's equation:

$$q_t = v_t h_t = \frac{1}{n} h_t^{5/3} s^{1/2} \qquad (5)$$

where q_t is the flow velocity (m/s), h_t is water depth, and s is the surface slope (m/m). An explicit method of calculation is used to compute velocity and depth over raster cells, and equations are solved at each time step. A conservative form of a finite difference method solves for q_t in Equation 5. To simplify discussions we describe quasi-one-dimensional equations for the flow problem. The actual numerical computations are normally performed on an Eulerian grid [Julien et al., 1995].

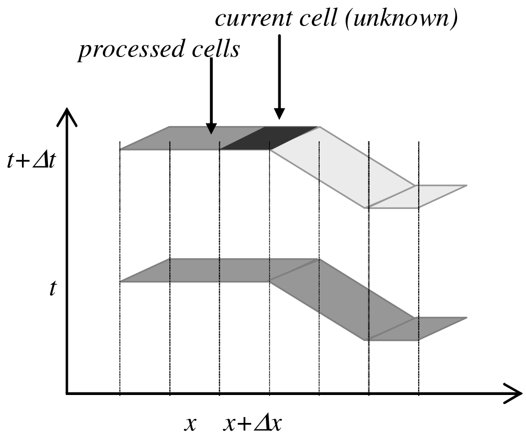


Figure 4. Computation of current cell ($x+\Delta x, t, t+\Delta$).

Finite-element approximations are made to solve the above partial differential equations for the one-dimensional case of flow along a strip of unit width. This leads to a coupled model with one term to maintain the continuity of flow and another term to compute the flow. In addition, all calculations must progress from an uphill cell to the down slope cell. This is implemented in map algebra by a iteration construct, called *doflow*, which processes a raster by flow order. Flow distance is measured in cell size Δx per unit length. One strip is processed during a time interval Δt (Figure 4). The conservative solution for the continuity term using a first order approximation for Equation 5 is derived as:

$$h_{x+\Delta x, t+\Delta t} = h_{x+\Delta x, t} - \frac{q_{x+\Delta x, t} - q_{x,t}}{\Delta x} \Delta t \qquad (6)$$

where the inflow $q_{x,t}$ and outflow $q_{x+\Delta x,t}$ are calculated in the second term using Equation 6 as:

$$q_{x,t} = v_{x,t} \cdot h_t \qquad (7)$$

The calculations approximate discharge from previous time steps. Discharge is dynamically determined within the continuity equation by water depth. The rate of change in state variables for Equation 6 needs to satisfy a stability condition where $v \cdot \Delta t / \Delta x \leq 1$ to maintain numerical stability. The physical interpretation of this is that a finite volume of water would flow across and out of a cell within the time step Δt. Typically the cell resolution is fixed for the raster, and adjusting the time step requires restarting the simulation cycle. Flow velocities change dramatically over the course of a storm event, and it is problematic to set an appropriate time step which is efficient and yields a stable result.

The hydrological model has been implemented in a map algebra programming language Pullar [2003]. To overcome the problem mentioned above we have added high level operators to compute the flow as an advancing front over a landscape. The time *step* advances this front adaptively across the landscape based upon the flow velocity. The level set algorithm [Sethian, 1999] is used to do this in a stable and accurate way. The map algebra script is given in Figure 5. The important operator is the *fluxflow* operator. It computes the advancing front for water flow across a DEM by hydrological principles, and computes the local drainage flux rate for each cell. The flux rate is used to compute the net change in a cell in terms of flow depth over an adaptive time step.

```
while (time < 120)
    doflow(dem)
        fvel = 1/n * pow(depth,m) * sqrt(grade)
        depth = depth + (depth * fluxflow(fvel))
    enddo
end
```

Figure 5. Map algebra script for excess rainfall flow computed over a 120 minute event. The variables *depth* and *grade* are rasters, *fvel* is the flow velocity, *n* and *m* are constants in Manning's equation. It is assumed a time step is defined and the script is run in a simulation. The first line in the nested cell processing (i.e. *doflow*) computes the flow velocity and the second line computes the change in depth from the previous value plus any net change (inflow – outflow) due to velocity flux across the cell.

4. CONCLUSION

The level set method provides the following benefits:

- it more directly models motion of spatial phenomena and may handle both expanding and contracting interfaces,
- is based upon differential equations related to the spatial dynamics of physical processes.

Despite the potential for using level set methods in GIS and land-surface process modeling, there are no commercial or research systems that use this approach. Commercial systems such as GRID [Gao et al., 1993], and research systems such as PCRaster [Wesseling et al., 1996] offer flexible and powerful map algebra programming languages. But operations that involve reaction-diffusive processing are specific to one context, such as groundwater flow. We believe the level set method offers a more generic approach that allows a user to program flow and diffusive landscape processes for a variety of application contexts. We have shown that it provides an appropriate theoretical underpinning and may be efficiently implemented in a GIS. We have demonstrated its application for two landscape processes – albeit relatively simple examples – but these may be extended to deal with more complex and dynamic circumstances. The validation for improved environmental modeling tools ultimately rests in their uptake and usage by scientists and engineers. The tool may be accessed from the web site www.gpa.uq.edu.au/MapScript/ (version with enhancements available April 2004) for use with IDRSIS GIS [Eastman, 1997] and in the future with ArcGIS. It is hoped that a larger community of users will make use of the methodology and implementation for a variety of environmental modeling applications.

5. REFERENCES

Bonham-Carter, G.F., Geographic Information Systems for Geoscientists: Modelling with GIS Pergamon Elsevier Science, New York, 1994.

Burrough, P.A., Dynamic Modelling and Geocomputation. In: P.A. Longley et al., Geocomputation: A Primer, Wiley, England, 165-191, 1998.

Burrough, P.A., and R. McDonnell, Principles of Geographic Information Systems, Oxford University Press, New York, 1998.

Eastman, J.R., IDRISI for Windows Version 2.0, Clark University, Worcester, 1997.

Gao, P., C. Zhan, and S. Menon, An Overview of Cell-Based Modeling with GIS. In M.F. Goodchild et al. (eds), Environmental Modeling with GIS, Oxford University Press, 325-331, 1993.

Goodchild, M., A Geographer Looks at Spatial Information Theory. in: Goos G., Hertmanis J. and van Leeuwen J. (eds), COSIT – Spatial Information Theory, LNCS 2205, 1-13, 2001.

Jørgensen, S. and G. Bendoricchio, Fundamentals of Ecological Modelling. Elsevier, New York, 2001.

Julien, P.Y., B. Saghafian, and F. Ogden, Raster-Based Hydrologic Modelling of Spatially-Varied Surface Runoff. *Water Resources Bulletin*, 31(3), 523-536, 1995.

Moore, I.D., A. Turner, J. Wilson, S. Jenson and L. Band, GIS and Land-Surface-Subsurface Process Modeling. in: M.F.Goodchild et al. (eds), Environmental Modeling with GIS, Oxford University Press, New York, 1993.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling, Numerical Recipes in C: The Art of Scientific Computing, 2nd Ed. Cambridge University Press, Cambridge, 1992.

Pullar, D., MapScript: A Map Algebra Programming Language Incorporating Neighborhood Analysis. *GeoInformatica* 5(2), 145-163, 2001.

Pullar, D., Simulation Modelling Applied To Runoff Modelling Using MapScript. *Transactions in GIS,* 7(2):267-283, 2003.

Ritter, G., J. Wilson, and J. Davidson. Image Algebra: An Overview. *Computer Vision, Graphics, and Image Processing*, 4, 297-331, 1990.

Sethian, J.A., Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge, 1999.

Sklar, F.H. and R. Costanza, The Development of Dynamic Spatial Models for Landscape Ecology: A Review and Progress. in: M.G.

Turner and R. Gardner (eds), Quantitative Methods in Ecology, Springer-Verlag, New York, 239-288, 1991.

Tilman, D., and P. Kareiva, Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions. Princeton University Press, Princeton, New Jersey, USA, 1997.

Wesseling C.G., D. Karssenberg P.A. Burrough and W.P. van Deursen, Integrating Dynamic Environmental Models in GIS: The Development of a Dynamic Modelling Language. *Transactions in GIS*, 1(1), 40-48, 1996.

Predicting the Hydraulic and Morphological Consequences of River Rehabilitation

S. Schweizer, **M. E. Borsuk, and P. Reichert**
Swiss Federal Institute for Environmental Science and Technology (EAWAG)
P.O. Box 611, 8600 Dübendorf, Switzerland. Email: steffen.schweizer@eawag.ch

Abstract: Decisions about flood protection and river rehabilitation require prediction of the consequences of each possible management alternative. To provide such predictions, an integrative model is required that represents the cause-effect relations between revitalisation measures and morphologic, hydraulic and ecological consequences. This paper describes the hydraulics submodel of such an integrative model. This submodel is subdivided into four modules predicting channel morphology, gravel transport, velocity distribution, and riverbed clogging. The channel morphology module predicts natural channel form based on a simple regression model of its dependence on easily available influence factors (valley slope, annual discharge, and median gravel size). Morphology predictions are then corrected by considering width constraints. The gravel transport module tests whether sufficient gravel is available for the development of gravel bar features or if a straight, incising river will result. A parameter describing the spatial velocity distribution is again estimated with a simple regression on relevant influence factors. Finally, estimates of the extent and severity of bed clogging are based on a model estimating the retention of fine particles carried by water infiltrating the gravel bed. This estimate is currently very uncertain as it depends on a number of uncertain input parameters. A preliminary application of the hydraulics submodel to a reach of the Thur River in Switzerland demonstrates its utility for predicting important consequences of river channel widening. The full integrative model will be used, together with quantitative assessments of stakeholder preferences, for decision support concerning revitalisation alternatives.

Keywords: hydraulics; channel morphology; probability network; integrative modelling; velocity distribution, clogging; bed load

1. INTRODUCTION

In the last 200 years, many river systems throughout the world have been regulated and channelised. These alterations have been conducted mainly to enhance agricultural and urban areas, to enable or facilitate river navigation, and to reduce flooding risk. This has resulted in a dramatic reduction of river floodplain area and loss of hydraulic and morphological variability. This uniformity decreases the habitat quality for organisms living in or near a regulated river (including algae and macrophytes, meio- and makro-zoobenthos, fish, and terrestrial flora and fauna). Thus, the biodiversity, abundance, and biomass of organisms are reduced and the functionality of the river ecosystem is impaired.

In Switzerland, only about 10% of all rivers remain in a natural or near natural state [BUWAL 1997]. Therefore, there is a high need for ecological rehabilitation, although most funding for river construction has been granted for the purposes of additional flood control. However, a recent federal requirement to include ecological rehabilitation measures in flood control projects has provided new opportunities. To understand the ecological and socio-economical consequences of river construction projects and provide advice for future efforts, the interdisciplinary "Rhone/Thur River Rehabilitation Project" was recently initiated [Peter et al. 2004]. One subproject of this research program is the development of an integrative model (IM) predicting the hydraulic-morphological situation after river rehabilitation and the resulting changes in the ecosystem. The IM is in the form of a probability network [Pearl 1988] and represents the relevant cause-effect-relations within and among the relevant biotic and abiotic factors, leading to attributes of concern to the river system's stakeholders (Figure 1). Together with a model of the preference structure for different levels of these attributes, the IM provides a comprehensive basis for decision support [Reichert et al. 2004].

In this paper, we describe the hydraulic and morphological submodel that provides the foundation for predicting all the biotic and abiotic attributes of interest. We begin by outlining the model approach, then describe the model equations and implementation, and finally demonstrate application to a section of the Thur River, Switzerland.

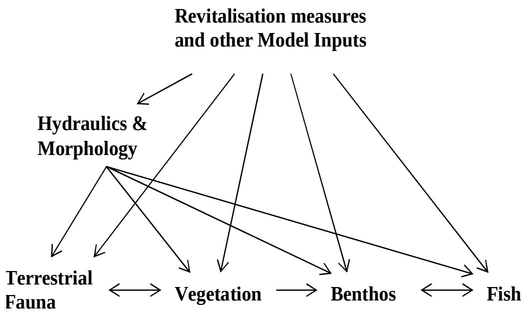


Figure 1. Submodels and structure of the integrative river rehabilitation model

2. MODEL DESCRIPTION

The development of an IM requires scientific knowledge in a variety of forms including literature findings, experimental and field data, other models, and, in the absence of other information, expert assessment. The principal motivations for implementing the IM as a probability network are the simplicity of combining different sources of information to represent cause-effect relations, the ability to simultaneously consider different spatial and temporal scales, and the ability to explicitly include uncertainties in model inputs, structure and outcomes [Borsuk et al. 2004].

Because all biotic endpoints of interest (including terrestrial fauna, riparian vegetation, aquatic benthos, and fish) are influenced by hydraulics and river morphology (see Fig. 1), model construction began with this abiotic submodel. The focus was on predicting variables that would be required as inputs for the biotic submodels including channel morphology, gravel transport, velocity distribution, and gravel bed clogging. These modules are described in the following subsections.

2.1 Channel Morphology

River channel planform is an important model endpoint on its own and is also a fundamental determinant of hydraulic habitat characteristics. Whether a river will be single- or multi-threaded depends on the balance between stream power, bed composition, and artificial width constraints [van den Berg 1995]. While a number of researchers have developed diagrams separating channel patterns based on flow-related parameters [e.g. da Silva 1991], these have generally been descriptive, in that they require advance knowledge of the channel geometry, which is pattern-dependent. Van den Berg [1995] developed a truly predictive method for distinguishing between multi- and single-thread rivers that requires only the pattern-independent properties of annual discharge, gravel size, and valley slope. Bledsoe and Watson [2001] made this approach probabilistic by fitting a logistic regression to data from 127 unconstrained, gravel-bed rivers. Of the several fitted relationships, we chose one which gives the probability, p_m, of a multi-thread pattern as,

$$p_m = \frac{\exp\left[10.35 + 5.71 \cdot \log_{10}\left(J_V \sqrt{\frac{Q_a}{d_{50}}}\right)\right]}{1 + \exp\left[10.35 + 5.71 \cdot \log_{10}\left(J_V \sqrt{\frac{Q_a}{d_{50}}}\right)\right]} \qquad (1)$$

where J_V is valley slope, Q_a is annual discharge (m^3s^{-1}), and d_{50} is median gravel diameter (m) [Bledsoe and Watson 2001].

We used equation (1) to predict the natural tendency of a river in the absence of width constraints, such as along-channel dikes. To determine the effect of constraints, the hypothetical unconstrained width was first estimated for each channel pattern from a regression on discharge, valley slope, and gravel diameter using the data of Bledsoe and Watson [2001]. The resulting model (n=153) yielded, for multi-thread rivers,

$$w_{bf} = 2.61 \cdot Q_a^{0.49} \cdot d_{50}^{-0.76} \cdot \varepsilon_w$$

and, for single-thread rivers, (2)

$$w_{bf} = 1.85 \cdot Q_a^{0.49} \cdot \varepsilon_w$$

where w_{bf} is bankfull width (m), assumed to occur at annual discharge, and ε_w is a lognormally-distributed error term with median zero and geometric standard deviation of 1.75. Valley slope was not a significant predictor for either river pattern, and gravel size was only significant for multi-thread rivers. The exponent on discharge was not significantly different for the two river patterns (all at the 0.01 significance level).

Single-thread rivers may be either straight, meandering, or sinuous with alternating side bars. In most locations, the space required to restore a meandering pattern is impractical given present land use. Therefore, we expect that rivers predicted to be single-threaded according to equation (1) will be sinuous with alternating side bars unless the constrained width is narrower than the width predicted by equation (2), in which case the river will be straight.

Rivers predicted to be multi-threaded according to equation (1) might yet be single-threaded if width constraints are too severe. This can be checked using the pattern diagram of da Silva [1991] for a known gravel size, channel geometry and mean depth at annual discharge. Width at annual discharge is estimated from equation (2), accounting for width constraints, and mean depth is estimated using the equation of Strickler [1923],

$$J = \frac{1}{k_{st}^2}\left(\frac{P}{A}\right)^{4/3}\left(\frac{Q}{A}\right)^2 \quad (3)$$

where J is channel slope (assumed here to equal valley slope, J_v), k_{st} is Strickler's coefficient ($m^{1/3}s^{-1}$), P is wetted perimeter (m), A is cross-sectional area (m^2), and Q is discharge (m^3s^{-1}). In applying equation (3) we assumed that each channel of a multi-thread river carries an equal amount of the total flow and has a triangular cross-section that is filled at annual discharge. For a single-thread river, we assumed a trapezoidal cross-section with a known angle of repose. Strickler's coefficient, k_{st}, is calculated as,

$$k_{st} = \frac{A_{st}}{\sqrt[6]{d_{90}}} \quad (4)$$

where A_{st} is a constant with values reported in the literature between 21 and 26 $m^{1/2}s^{-1}$.

The number of channels expected in an unconstrained multi-thread river was predicted using the relation identified by Robertson-Rintoul and Richards [1993] in an analysis of 21 braided rivers,

$$n_b = round[1 + 5.52 \cdot (Q_a J_v)^{0.40} d_{84}^{-0.14}] \quad (5)$$

where n_b is the number of braids and d_{84} (m) is the 84th percentile of gravel size. To account for width constraints, we multiply the number of channels by the ratio of the constrained width to the natural width predicted by equation (2).

Regardless of predicted morphology, mid- or side-channel bars will not develop if gravel transport out of the reach exceeds upstream gravel supply. In such cases, we assume that an incising, single-thread channel will eventually result. Incision may, however, be prevented by the installation of weirs and other bed stabilization measures or the reduction of upstream gravel retainment.

2.2 Gravel Transport

As mentioned above, the formation of gravel structures in a widened river reach depends on net deposition of gravel. The upstream input is treated as known, while the transport capacity within the reach, Q_b (m^3s^{-1}), is calculated as the product of a specific transport capacity, q_b (m^2s^{-1}), and the width, w (m),

$$Q_b = w \cdot q_b \quad (6).$$

The specific transport capacity is estimated [Meyer-Peter and Müller 1948] as,

$$q_b = \Phi\sqrt{(s-1)gd_{50}^3} \quad (7)$$

where Φ is a dimensionless transport capacity, s is the ratio of sediment to water density, and g is the gravitational constant 9.81 $m \cdot s^{-2}$.

The dimensionless transport capacity Φ can be estimated from the bed load formula of Meyer-Peter and Müller [1948],

$$\Phi = 8 \cdot (\theta - 0.047)^{1.5} \quad (8)$$

where θ is the dimensionless bottom shear stress,

$$\theta = \frac{h \cdot J}{(s-1) \cdot d_{50}} \quad (9)$$

and h is water depth (m), which can be estimated for a given discharge from equation (3) with an assumed channel geometry. To derive the annual gravel input, we cumulate the daily inputs estimated using daily discharge.

2.3 Velocity Distribution

The quality of habitat for aquatic biota is strongly influenced by velocity characteristics. Both average and spatially distributed velocities are of relevance. Spatial mean velocity, v_m is calculated from discharge and cross-sectional area as $v_m=Q/A$. The spatial distribution of velocity can then be estimated for a given mean velocity using the method of Lamouroux *et al*. [1995]. In a statistical analysis of data collected from a diversity of streams, they found that the spatial frequency distribution of measured relative velocity (v/v_m) could be modelled as a mixture of a centred (Gaussian) and a decentred (mixture of exponential and Gaussian) distribution with fixed distributional parameters. A parameter describing the mixture between the centred and decentred distributions could then be expressed as a linear function of the relative roughness (d_{50}/h) and the logarithm of the Froude number ($v_m/(gh)^{1/2}$). An increasing relative roughness leads to a more decentred distribution, while an increasing Froude number leads to a more centred distribution. The accuracy and robustness of the analysis of Lamouroux *et al.* [1995] give us confidence in directly applying their results to our model.

2.4 River Bed Clogging

Fish and benthic species depend on the interstitial gravel zones. Therefore clogging and clearance of the bed matrix are crucial ecological processes. Additionally, the content of fine particles in the riverbed influences water exchange between surface and ground water, thus affecting groundwater regeneration.

Conceptually, we model gravel bed clogging as a process that occurs over time at a rate which depends on hydraulic and bed characteristics. The clogging process is disrupted by the occurrence of high floods which are accompanied by high bottom shear stress. This disturbs the gravel bed matrix and clears it of fines. This flushing occurs with a calculable frequency for a particular river, and the frequency together with the rate of clogging will determine temporal extent and severity of clogging.

The temporal progression of the build up of fines between floods can be estimated from a calculation of the volume of water filtered through the gravel bed according to a simplified version of the formula given by Schälchli [1993],

$$V_A = \sqrt{\frac{\Delta h_w \cdot g \cdot \left(\frac{d_{10}}{d_{50}}\right)^3 \cdot \mathrm{Re}^{1.5} \cdot i \cdot t}{2.5 \cdot 10^{12} m^2 kg^{-1} \cdot \nu \cdot \theta^{0.5} \cdot C}} \quad (10)$$

where V_A is the volume of filtered water per unit area (m3·m^{-2}), Δh_w is the pressure head between channel and groundwater level (m), *Re* is the Reynold's Number, *i* is the hydraulic gradient, *t* is the time since the last flushing event (s), ν is kinematic viscosity (m2·s^{-1}), and *C* is the concentration of suspended particles (kg·m^{-3}).

The mass of fine particles retained in the bed matrix, m_{fines}, is calculated as the product of the volume of filtered water and concentration of suspended particles. The fraction of fines, f_{fines}, in the riverbed is then calculated as,

$$f_{fines} = \frac{m_{fine}}{m_{fines} + m_{coarse}} = \frac{V_A C}{V_A C + (1-\phi) H \rho_{sed}} \quad (11)$$

where m_{coarse} is the mass of coarse bed material (kg), ϕ is porosity, *H* is the depth of the bed layer (m) (usually 0.1 to 0.3m, see Schälchli 1993), and ρ_{sed} is the gravel density (kg·m^{-3}). The percentage of fines can be used as a measure of the degree of gravel bed clogging.

A bottom shear stress of sufficient magnitude to initiate bed disturbance and gravel flushing, θ_D, is calculated according to Günther [1971] as,

$$\theta_D = \theta_{Cr} \cdot \left(\frac{d_{50D}}{d_{50}}\right)^{2/3} \quad (12)$$

where d_{50D} is the median diameter of the upper gravel bed layer (m), d_{50} is the median diameter of particles lying on the river bed (m), and θ_{Cr} is the critical shear stress, assumed to equal 0.05 [Meyer-Peter and Müller 1948].

The water depth associated with a bottom shear stress value of θ_D can be calculated from equation (9) and then related to a critical discharge using Strickler's formula (equation 3).

2.5 Model Implementation

The model described above was implemented using a software program for evaluating probability network models, e.g. Analytica [Lumina, 1997]. A sample of one thousand realizations was drawn for each probability distribution representing uncertainty using Latin hypercube sampling. The major inputs to the model (Figure 2) can be derived from historical data for the river system of interest, and the decision variables can be set to values corresponding to current conditions, decision alternatives, or scenarios used for sensitivity analysis.

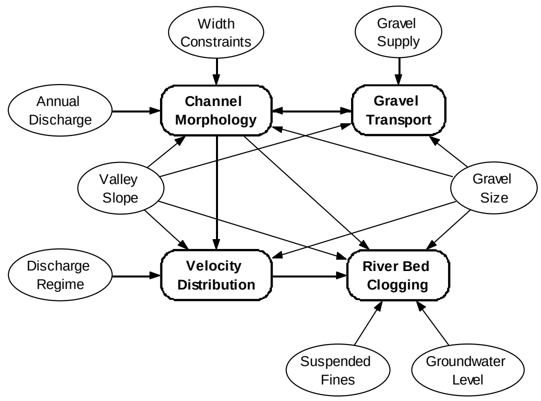


Figure 2. The hydraulic and morphologic submodel as implemented in Analytica. Round nodes indicate important input variables and bold nodes indicate submodel components.

3. FIRST RESULTS OF A CASE STUDY

A case study at the Thur River between the towns of Weinfelden and Bürglen, Switzerland, demonstrates an application of the hydraulic submodel. Two scenarios are considered: (i) the present conditions (leave the straight river width at 50m) and (ii) possible river widening up to 200m. Table 1 summarises the principal site characteristics for these scenarios (model inputs) and their uncertainties.

Table 1: Mean, standard deviation and distribution of model inputs at Weinfelden-Bürglen

Model input	Mean	Std. Dev.	Distribution
Slope J [‰]	2.0	0.2	Lognormal
1-year flood Q_a [m^3/s]	410	40	Normal
d_{50} [cm]	2.9	0.5	Lognormal
d_{90} [cm]	6.8	1	Lognormal
Porosity ϕ [%]	25	2	Lognormal
Bankfull width w_{bf} [m]	50 / 200	-	-
Angle of repose γ (single-thread) [°]	45	-	-

As the logistic regression approach for predicting river form (equation 1) assumes no width constraints, predictions of the natural river form tendency is the same for all possible alternatives: a 24% probability of a multi-thread river and corresponding 76% probability of a single-thread. While, in general, the final channel morphology additionally may depend on channel width constraints, in this case the probabilities above are maintained after considering a 200m constraint (Figure 3).

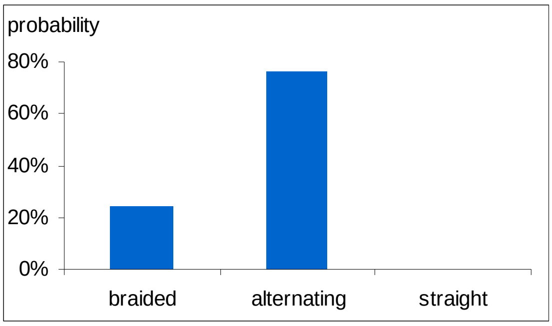


Figure 3: Probability distribution of possible river forms for a 200m river-widening.

Compared to the present state, the alternating river reach would respond to an annual flood regarding mean width and depth and dimensionless bottom shear stress nearly identically (Table 2). However, if a braided river reach were to develop, major differences with respect to these hydraulic quantities can be expected.

Table 2: Predicted hydraulic properties for the present state and the two possible outcomes of the widening alternative at a one-year flood (Q_a=410m^3/s), assuming sufficient gravel supply.

	Present State (50m)	Widening Alternative (200m)	
	Straight	Alternating	Braided
Probability	-	76%	24%
Mean Width	36m	38m	200m
Mean Depth	4.6m	4.5m	1.2m
Mean θ	0.2	0.19	0.05
# of Braids	-	-	2 - 3

Mean water depth, mean velocity and its spatial distribution can also be calculated for discharges below the annual flood, such as the mean discharge (Q_m=40m^3/s) or dry weather discharge (Q_{347}=8m^3/s). For flows equal to or lower than mean discharge no differences in mean velocity and depth between the present state and a widening with the consequence of alternating gravel bars will occur (Table 3). This applies also for the spatial distribution of velocity (Figure 4a,b). If a braided river reach develops, significant distinctions in comparison to the present state can be expected for flows similar to mean discharge, whereas for low stages the differences become indistinct.

Finally, it should be stated that some important differences between the two morphologies straight and alternating exist: alternating gravel bars provide some refuges for benthic organisms (this can be important during floods) and serve as important pioneer areas for terrestrial flora and fauna. This information will be propagated to the biological models.

Table 3: Predicted hydraulic properties for the present state and the two possible outcomes of the widening alternative for mean and dry weather discharge, Q_m= 40m^3/s and Q_{347}=8m^3/s, respectively.

	Present State (50m)	Widening Alternative (200m)	
	Straight	Alternating	Straight
mean depth (Q=40m^3/s)	1.0m	1.0m	0.5m
mean velocity (Q=40m^3/s)	1.5m/s	1.5m/s	1.0m/s
mean depth (Q=8m^3/s)	0.4m	0.4m	0.3m
mean velocity (Q=8m^3/s)	0.8m/s	0.8m/s	0.7m/s

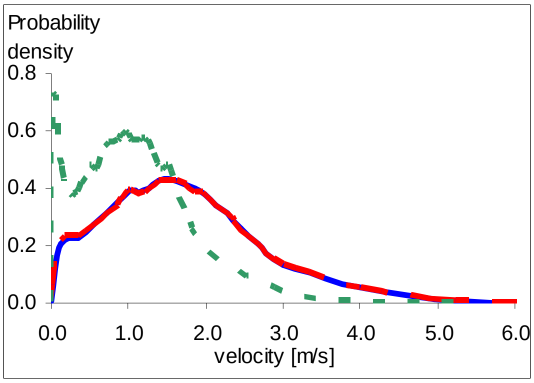


Figure 4a: Spatial velocity distribution for Q=40 m^3/s (solid line=present state; dashed line=200m widening single-thread (almost identical to solid line); dotted line=200m widening, multi-thread).

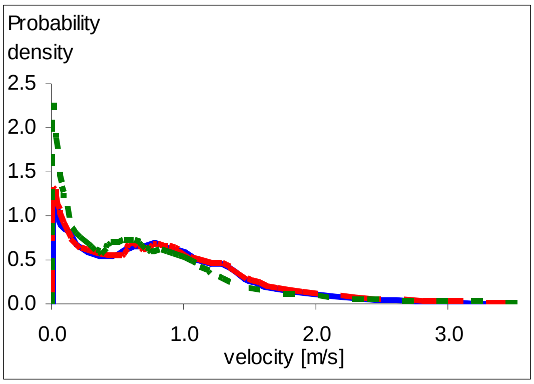


Figure 4b: Spatial velocity distribution for Q=8m^3/s (solid line=present state; dashed line=200m-widening,single-thread (almost identical to solid line); dotted line=200m-widening, multi-thread).

4. DISCUSSION AND CONCLUSIONS

The submodel presented in this paper predicts the morphological and hydraulic changes resulting from rehabilitation alternatives. In addition these predictions will be used in a next step as inputs to the other submodels of an integrated river rehabilitation model of benthic and fish populations, terrestrial vegetation, and shoreline fauna. Reichert et al. (2004) describe how the results of that integrated model, together with value assessments of possible outcomes of rehabilitation measures, will support decision making. The validation of this hydraulics submodel with real data will be started in summer 2004 by another group of the "Rhone/Thur River Rehabilitation Project" at both the studied reach and at an adjacent reach of the river Thur. For detailed planning of construction work, a more detailed hydraulic study is required.

5. ACKNOWLEDGEMENTS

This project was supported by the Rhone-Thur project, which was initiated and is funded by the Swiss Federal Office for Water and Geology (BWG), the Swiss Federal Institute for Environmental Science and Technology (EAWAG) and the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL). For stimulating discussions, comments, advice and for sharing valuable datasets, we thank the many Rhone-Thur project partners, the Laboratory of Hydraulic, Hydrology, and Glaciology (VAW), N. Lamouroux, P. Julien, and B. Bledsoe.

6. REFERENCES

BUWAL (Swiss Agency for the Environment, Forests and Landscape), *Umwelt in der Schweiz*, Bern, Switzerland, 1997.

Bledsoe, B. P. and C. C. Watson, Logistic analysis of channel pattern thresholds: meandering, braiding, and incising, *Geomorphology 38*:281-300, 2001.

Borsuk, M. E., C. A. Stow, and K. H. Reckhow, A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*: In press, 2004.

da Silva A.M.A.F., Alternate bars and related alluvial processes, Master of Science Thesis, Queen's University, Kingston, Ontario, Canada, 1991.

Günther A., Die kritische mittlere Sohlenschubspannung bei Geschiebemischung unter Berücksichtigung der Deckschichtbildung und der turbulenzbedingten Sohlenschubspannungsschwankung, Mitteilung Nr.3 der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie, ETH Zürich, 1971.

Lamouroux, N., Y. Souchon, and E. Herouin, Predicting velocity frequency distributions in stream reaches, *Water Resources Research 31*, 2367-2375, 1995.

Lumina, *Analytica for Windows, User's Guide*, *Version 1.1*, Lumina Decision Systems, Denver, US, 1997

Meyer-Peter, E. and R. Müller, Formulas for bedload transport. Proc. IAHR 3rd Congress, Stockholm, 1948.

Pearl, J., *Probabilistic Reasoning in Intelligent System*, Morgan Kaufmann, San Mateo, CA, 1988.

Peter, A., F. Kienast, and S. Nutter, The Rhone-Thur River project: a comprehensive river rehabilitation project in Switzerland, submitted, 2004.

Reichert, P., M. Borsuk, M. Hostmann, S. Schweizer, C. Spörri, K. Tockner, and B. Truffer, Concepts of decision support for river rehabilitation. Submitted to the International Environmental Modelling and Software Society conference (iEMSs 2004), Osnabrück, Germany, 2004.

Robertson-Rintoul M.S.E. and K.S. Richards, Braided channel pattern and palaeohydrology using an index of total sinuosity. *In* Best J.L. & Bristow C.S. (eds): *Braided Rivers.* Geological Society, special publication No. 75: 113-118, 1993.

Schälchli U., Die Kolmation von Fliessgewässersohlen: Prozesse und Berechnungsgrundlagen, Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie der ETH Zürich Nr. 124, 273 pp, 1993.

Strickler A., Beiträge zur Frage der Geschwindigkeitsformel und der Rauhigkeitszahlen für Ströme, Kanäle und geschlossene Leitungen, Mitteilung Nr. 16 des Amtes für Wasserwirtschaft; Eidg. Departement des Inneren, Bern 1923.

Van den Berg, J.H., Prediction of alluvial channel of perennial rivers. *Geomorphology 12*: 259-279, 1995.

Lake and Climate Models Linkage: A 3D Hydrodynamic Contribution

Luis F. Leon[1a,b], David Lam[2], William Schertzer[2] and David Swayne[1b]

[1a] University of Waterloo, Waterloo, Ontario, N2G3L1 (lfleonvi@uwaterloo.ca)
[1b]University of Guelph, Computing Research Laboratory for the Environment, Guelph, Ontario, Canada N1G 2W1
[2]Environment Canada, National Water Research Institute, Burlington, Ontario, Canada L7R 4A6

Abstract: Under a Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) project, targeted to study the feasibility to link regional climate models with lake models, one of the tasks was to consider such a coupling in large lakes. The objective is to provide detailed information on temperature and circulation distributions of the lake to take into account the spatial variability for temperature and the heat exchange through the water's surface. The major contribution of this work is focused on realistic representation of the heat fluxes and temperature distributions to and from lakes especially during the thermally stratified ice-free periods. This paper presents the detailed 3D model applied in Lake Erie in order to produce and verify, at the surface layer of the lake, the spatial distribution of temperature and heat exchanges that eventually can be coupled with a regional climate model. Preliminary results will be presented on how this lake model may improve the regional climate models which currently do not consider such large lake circulation effects.

Keywords: Climate Models; Lake circulation; Model integration; 3D Hydrodynamics

1. INTRODUCTION

Large lakes are recognized as an important agent which influences the circulation of the atmosphere; the atmospheric forcing also affects the lake thermal structure. This interaction is complex and continues to be a critical issue considering the millions of lakes in Canada, many of which are large, and are unaccounted for in current climatic models. The inclusion of a fully interactive coupling of a lake model with an atmospheric model for regional climate modelling is one option and an important objective of current Canadian Regional Climate Model (CRCM) development.

Only recently, has the CRCM implemented a lake model [Goyette et al., 2000] as an initial attempt to simulate the evolution of the water temperature and ice cover on the Laurentian Great Lakes. Another example is the implementation of a physically-based column model which includes the effect of ice [Hostetler et al., 1993]. These efforts so far pertain to simple one column (1-dimensional) and two layer (2-dimensional) models. In these attempts for the CRCM, the lake is assumed to have uniform surface lake temperature. For small lakes, such assumption may be valid.

For large lakes such as the Laurentian Great Lakes, the lake surface temperature is not uniform and therefore requires a hydrodynamic model with finer spatial resolution architecture [Lam and Schertzer, 1999]. Swayne et al. [2003] proposed to use a hierarchy of 1-D, 2-D and 3-D lake thermodynamic models depending on the size of the lake and available information to advance regional climate modelling. It aimed at conducting an independent evaluation of predictive capabilities of lake models for heat flux and thermal characteristics.

Such evaluations are a critical and essential element in the progression to development of fully coupled lake/CRCM model across lakes of different spatial scales. The purpose of our study is to evaluate the predictive capability of the 3D models with the focus on large lakes in the context of linking them with the CRCM. The intention is to conduct an evaluation using high quality databases for selected large lakes in Canada, e.g. Lake Erie, Lake Ontario and Great Slave Lake, that have different morphometry over contrasting climatic regions. We will report on the preliminary results for Lake Erie as an example of how the 3D model can be used to integrated with the CRCM.

2. TECHNICAL CHALLENGES

The idea of modeling lake-atmosphere interactions will require new research in: aggregating and distributing spatial quantities (e.g. how to fit a lake model into the CRCM grid resolution); in synchronizing temporal scales (e.g. time steps of 15 minutes in CRCM vs. diurnal cycles produced by lake temperature models); and generalizing processes for application over lakes of different spatial scales. A major contribution of this work is focused on methodological procedures to evaluate heat fluxes and temperature distributions to and from lakes. Databases are required which are valid for model verification and uncertainty analyses. This study provides independent feed-back from lake modelers to CRCM modelers.

Such interaction is required to effect improvements in CRCM for application in the priority issues of climatic impacts research on aquatic ecosystems (e.g. on water quality, nutrient and dissolved oxygen concentrations). It also requires the lake modelers to adapt the model results required by the regional climate model, specifically the heat, mass and momentum exchanges at the air-water interface, surface air and water temperature, etc.

The main challenge is how to make these quantities consistent between the lake and the climate model. Figure 1 represents the schematic of the proposed coupling. It shows different surface layers as simulated time progresses and on top, a grid representing the CRCM mesh. The 3D results in a finer grid (2x2 km) must be averaged at every time step for the larger grid resolution (25x25 km) of the climate model.

A technical issue to resolve will be the coupling of both models by exporting output data from the 3D hydrodynamic model to be used as input for the next time step in the climate model. Currently the bulk of the work is focused on attempting to couple these two models by emphasizing the consistency at the air-water interface, both in spatial and temporal resolution. Specifically, for heat transfer, the incoming and outgoing heat fluxes at the air-water interface should be the same at both models while the different time steps are set for a common value.

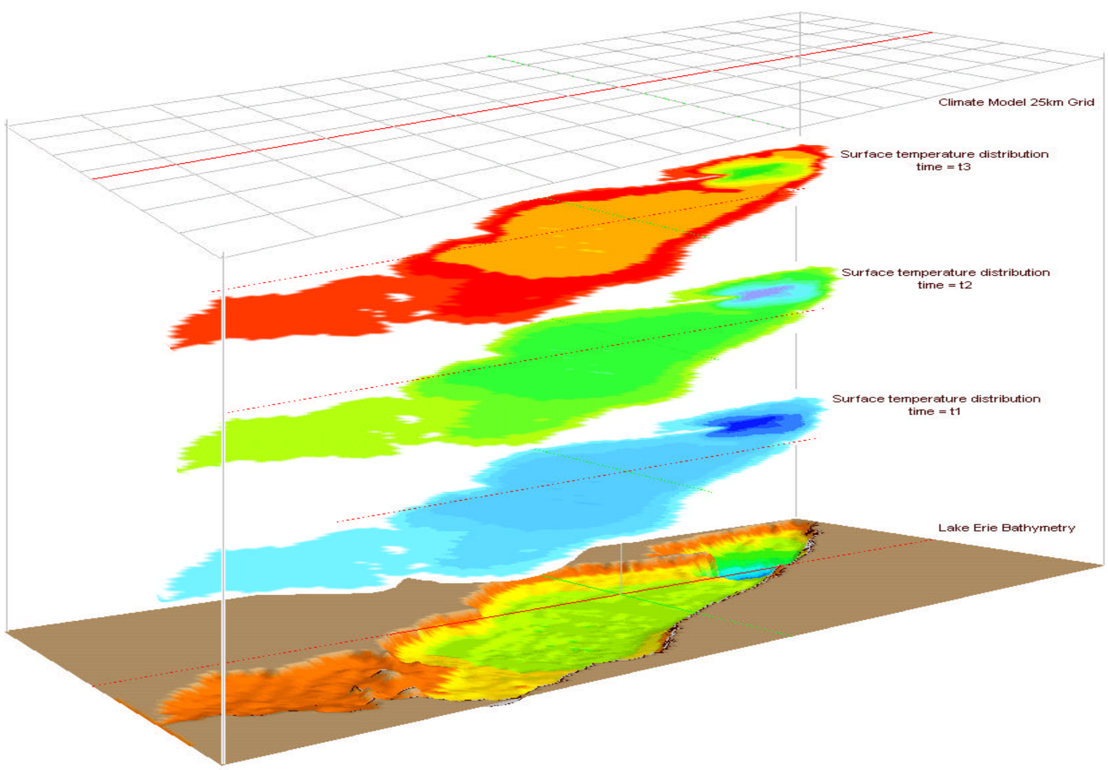


Figure 1. Schematic representation of the proposed model coupling

3. SPATIAL DISTRIBUTION IN LAKES

Canada contains several of the largest lakes in the world within the Laurentian Great Lakes (Lakes Superior, Michigan-Huron, Erie, Ontario), the Mackenzie Great Lakes. (Great Bear Lake, Great Slave Lake and Lake Athabasca) as well as others such as Lake Winnipeg. These lakes have large spatial extents, can have complicated bathymetric characteristics and can also have significant cross-lake variability in meteorological and limnological components [Schertzer and Croley 1999]. For example, Lake Erie is nearly 5-deg of latitude and 2-deg of longitude in size and has three distinct basins with maximum depths of 10m (west basin), 25m (central basin) and 64m (east basin).

The heat transfer at the lake-atmosphere interface is influenced by such meteorological variables as air temperature, humidity, wind and solar radiation. The lake responds through both radiative and turbulent heat transfers and heating/cooling. Due to the high heat capacity of such large lakes, they can have a pronounced seasonal and regional influence as a result of seasonal lags in the heat transfers compared to the surround in land.

Figure 2a shows a pronounced difference between the mean monthly air temperatures recorded from land stations in the west basin of the lake compared to the other basins especially in the spring and again in the fall. The differences in air temperature and other meteorological variables during the spring warming phase combined with the shallow basin characteristics results in significant differences in the surface temperatures between basins (Fig. 2b).

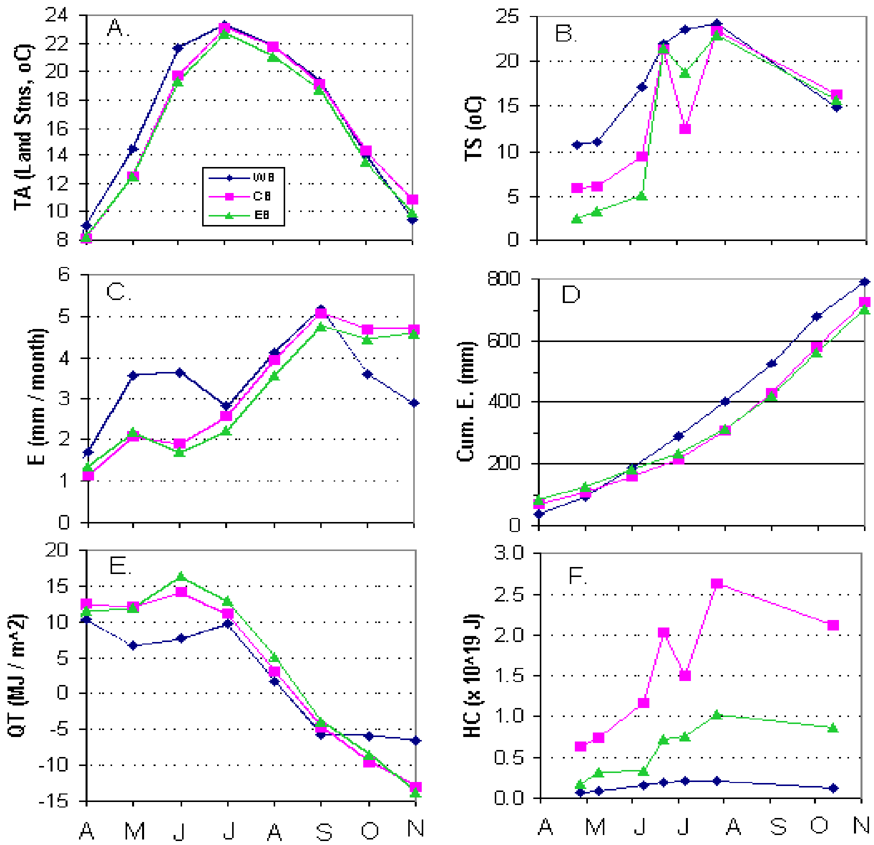


Figure 2. Comparison of selected meteorological and limnological fields for the west, central and east basin of Lake Erie (a) land-based air temperature, (b) surface water temperature, (c) evaporation, (d) cumulative evaporation, (e) total heat flux, and (f) basin heat content.

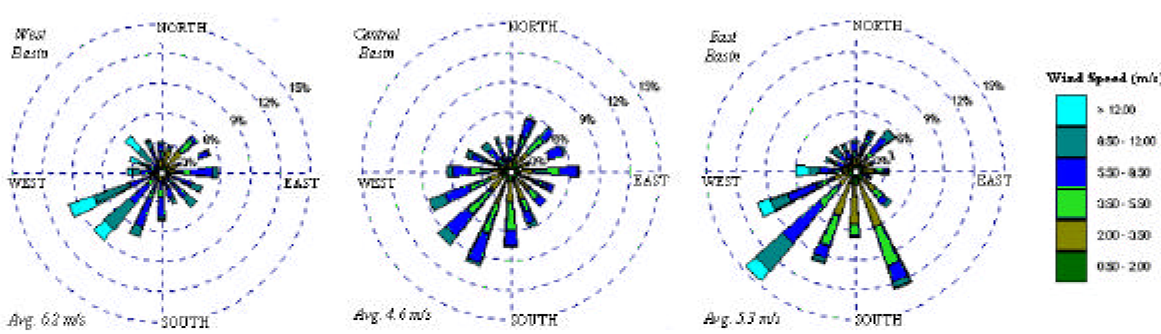


Figure 3. Wind roses for the different basins in Lake Erie for 1994

In Lake Erie, the warmer conditions in the west basin during spring (Fig. 2c) results in significantly higher evaporation compared to the other basins. In fact, on large deep lakes such as Lake Erie and Great Slave Lake [Schertzer 2003] condensation can occur in the cooler mid-lake while evaporation can be high in the warmer and shallow nearshore.

Over the ice-free period, the large central basin in Lake Erie has the highest cumulative evaporation compared to either the west or east basins (Fig. 2d). The cross-lake differences in meteorological and lake response characteristics can be significant especially in the transition seasons. Because of the large spatial extents of such lakes, a 3-dimensional hydrodynamic model linked with a 3-dimensional atmospheric model is critical for realistic simulation of the influence of these lakes within the regional climate model.

Figure 3 shows characteristics of the wind field across the lake. As can be seen from the wind roses, the dominant wind direction is from the SW in all three basins, and the average wind speed is in the order of 5 m/s. The combined differences in air temperature, surface water temperature and wind influences the total radiative and turbulent heat exchange (Fig. 2e), the vertical temperature structure and lake heat content (Fig. 2f) [Schertzer et al. 1987]. On a lake-wide basis, the latent heat flux is often a dominant component compared to other turbulent exchanges, generally small in the spring but high in the fall the lake heat content is released to the atmosphere.

4. THE 3D MODEL

During the last decades a series of three-dimensional hydrodynamic models have been developed at different research institutes. An overview can be found in Lynch and Davies [1995] which includes the well known Princeton Ocean Model [Blumberg and Mellor, 1987], Schwab and Bedford [1994]; Simons [1975]. The model used in this application is ELCOM (Estuary and Lake COmputer Model). ELCOM is a three-dimensional hydrodynamics model for lakes and reservoirs, and is used to predict the variation of water temperature and salinity in space and time [Hodges et al., 2000].

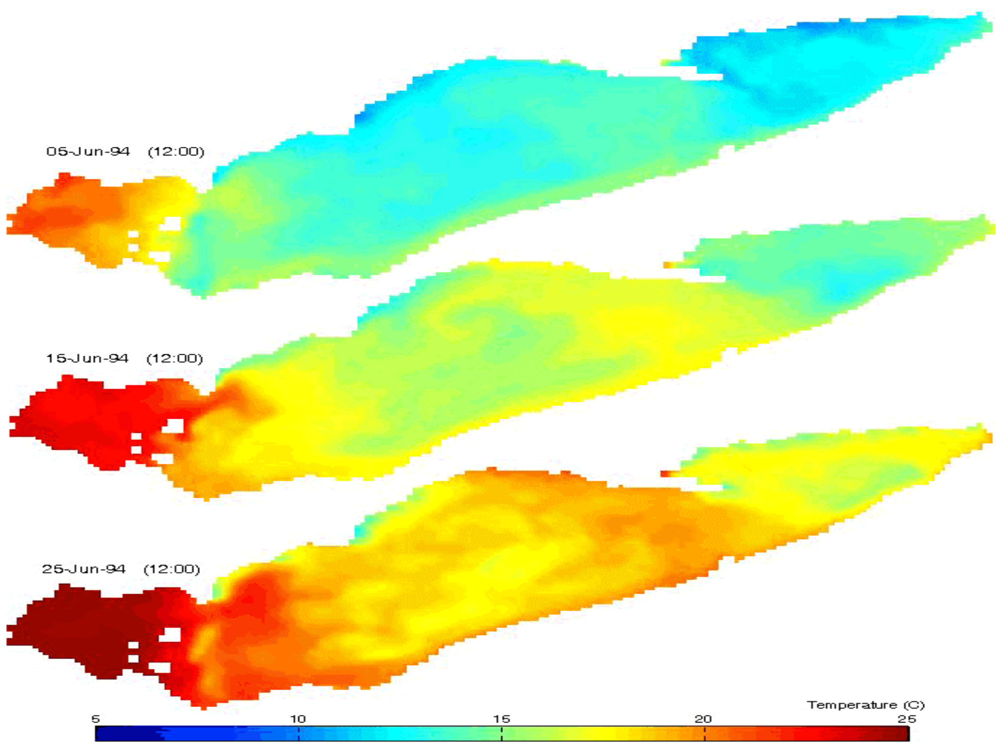


Figure 4. Surface temperature in Lake Erie (3 time slices of the 1994 simulation)

The heat exchange through the water's surface is governed by standard bulk transfer models found in the literature [Schertzer et al. 1987]. Energy transfer across the free surface is separated into non-penetrative components of long-wave radiation, sensible heat transfer, and evaporative heat loss, complemented by penetrative short-wave radiation. Non-penetrative effects are introduced as sources of temperature in the surface-mixed layer, whereas penetrative effects are introduced as source terms in one or more grid layers on the basis of an exponential decay and an extinction coefficient.

ELCOM computes a model time step in a staged approach consisting of introduction of surface heating/ cooling in the surface layer. The solution grid uses rectangular Cartesian cells with fixed Dx and Dy (horizontal) grid spacing, whereas the vertical Dz spacing may vary as a function of z but is horizontally uniform. The solution is based in the Arakawa C-grid stencil where velocities are defined on cell faces with the free-surface height and scalar values on cell centers. The free-surface height in each column of grid cells moves vertically through grid layers as required by the free-surface evolution equation. To be confident enough with the results of the 3D model, several field data for 1994 were used to validate the hydrodynamics on the lake (Leon et al., 2004). To demonstrate the importance of the spatial distribution on Lake Erie, Figure 4 presents the model output for surface temperature during the warming period of the 1994 simulation.

Figure 5 is an extract of such results showing the surface temperature gradient from west to east and selected profile comparisons of measured and calculated values at three lake sites. As can be seen from Fig. 5, the spatial distribution of temperature in such a great lake as Lake Erie, is quite important (temperature differences between west and east up to 8°C). The CRCM has a grid resolution of 25km, then for day 215, the average temperature to pass to the climate model are in the order 25°C, 20°C and 18°C for the west, central and east basins. For the coupling, distributions of temperature, evaporation and heat fluxes can be computed directly from the 3D model results at any time step. The results so far indicated that we achieved internal consistency in the lake model in that the west-east gradient of surface temperature gradient is consistent with the heat content (Fig. 3) and with the internal thermal structure as depicted in Fig. 5.

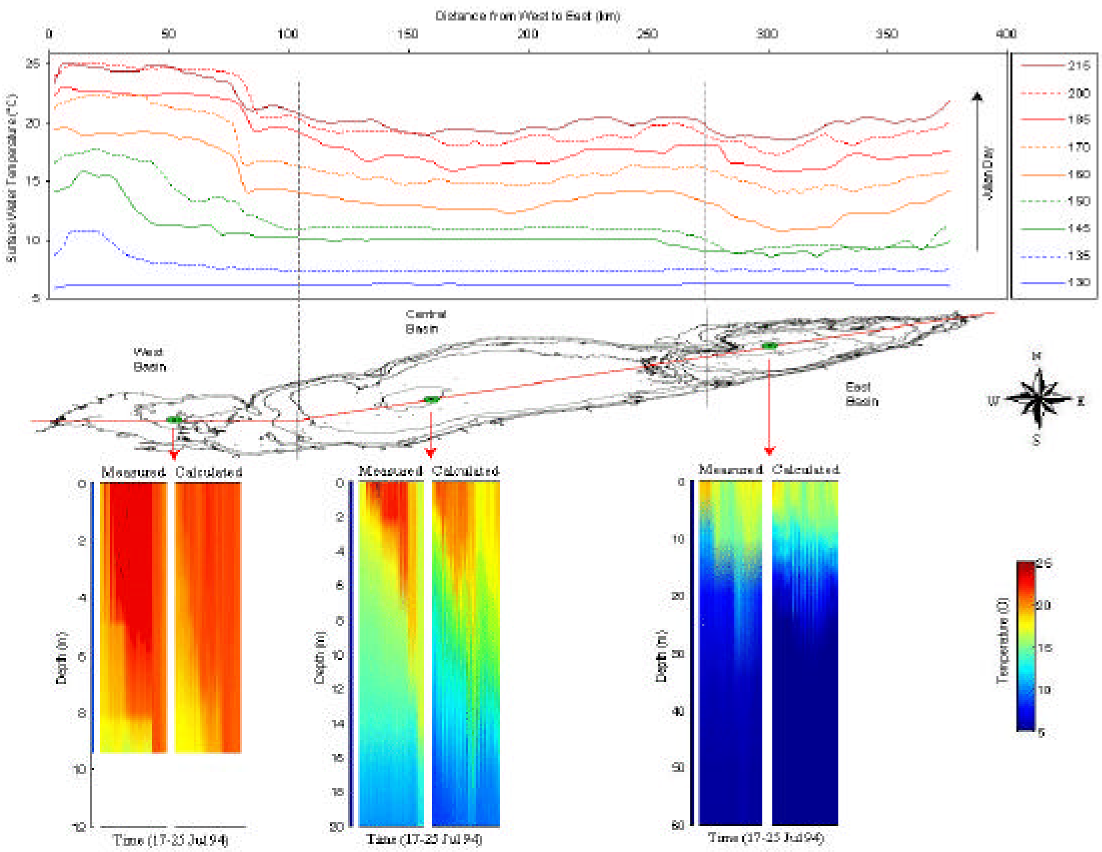


Figure 5. Temperature results of the 3D model. Top: surface temperature gradient on the lake Bottom: profiles for a selected week in 1994 (extracted from Leon et al., 2004)

These results, however, were obtained with measured air temperature, wind and heat fluxes. The next challenge will be to resolve the issue of coupling both models by exporting output data from one model (i.e. 3D hydrodynamics) to be used as input for the next time step in the other model (CRCM), and vice versa, without the assistance of measurements. Currently we are attempting to couple these two models by emphasizing the consistency at the air-water interface. Specifically, for heat transfer, the incoming and outgoing heat fluxes at the air-water interface should be the same at both models. In this case, the fluxes from the CRCM will be used as input to the lake model and new updated fluxes are fed back from the lake model to the CCRM. Some iterations may be required if necessary. Similarly, the mass transfer is made consistent via the evaporation mechanism for water loss from the lake and via precipitation for water gain at the air-water interface. For momentum transfer, the wind stress should be made consistent in both models.

CONCLUSIONS

We have successfully applied a 3D hydrodynamic model to Lake Erie. The next step will be to directly connect the 3D results with the CRCM model with its selected grid resolution and time step. This will have the advantage of seamlessly rendering the air and water models in tandem in both spatial and temporal steps and thereby achieving comparable numerical stability and accuracy, as well as generating detailed information on temperature and circulation distributions of the lake. However, it is also recognized that the 3D approach currently has a disadvantage of high computational cost and requires detailed input data, not to mention the difficulty of having sufficient data for model calibration and verification. Thus, we propose that the current generation of the 3D hydrodynamic model will be limited to only a few of the larger lakes lakes such as Laurentian Great Lakes and others such as Great Slave Lake in Canada for this study.

ACKNOWLEDGMENTS

The authors wish to thank the support from Canadian Foundation for Climate and Atmospheric Sciences, the Natural Sciences and Engineering Research Council of Canada, and Environment Canada. We thank the support and discussion with M. McKay at the Meteorological Service of Canada, W. Rouse from McMaster University and Rene Laprise from the Université du Québec à Montréal. Also acknowledged is the support from R. Smith at the University of Waterloo and the support of J. Imberger at the University of Western Australia for the use of the ELCOM model.

REFERENCES

Blumberg A., and Mellor, G. (1987). A description of a three-dimensional coastal ocean circulation model. In *Three-Dimensional Coastal Ocean Models,* 1-16, AGU, Washington, DC.

Goyette, S., McFarlane, N., Flato, G. (2000). Application of the CRCM to the Laurentian Great Lakes Region: Implementation of a Lake Model, *Atmos.-Ocean*, 38, 481-503.

Hodges, B. R., Imberger, J., Saggio, A., and Winters, K. B. (2000). Modeling basin-scale internal waves in a stratified lake. *Limnol. Oceanogr*., 45(7), 1603-1620.

Hostetler, S.W., Bates, G.T. and Giorgi, F. (1993). Interactive coupling of a lake thermal model with a regional climate model. *J. Geophysical Research*, 98:5045-5057.

Lam, D.C.L. and Schertzer, W.M. (1999). Potential Climate Change Effects on Great Lakes Hydrodynamics and Water Quality, American Society of Civil Engineers (ASCE) Press, Reston, Virginia, 232 p.

León, L.F., Imberger, J., Smith, R.E.H., Hecky, R.E., Lam, D.C.L. and Schertzer, W.M., (2004) Modeling as a Tool for Nutrient Management in Lake Erie: A Hydrodynamics Study, *J. Great Lakes Research*, Lake Erie Trophic Status Special Issue (Submitted)

Lynch, D.R. and Davies, A.M., (1995) Assessment for Coastal Ocean Models, *Coastal and Estuarine Studies*, Volume 47, American Geophysical Union.

Schertzer, W.M., and Croley II, T.E. 1999. Climate and Lake Responses. Chapter 2, p.1-74. In. (eds.) D.C.L. Lam, Potential Climate Change Effects on Great Lakes Hydrodynamics and Water Quality, ASCE Press, Reston, Virginia, USA, 232 p.

Schertzer, W.M., Saylor, J.H., Boyce, F.M., Robertson, D.G., and Rosa, F. (1987). Seasonal thermal cycle of Lake Erie. *J. Great Lakes Res*. 13(4):468-486.

Schertzer, W.M., Rouse, W.R., Blanken, P.D., and Walker, A.E. (2003), Over-lake meteorology and estimated bulk heat exchange of Great Slave Lake in 1998-99. *J. Hydrometeor.* 4: 649-659.

Schwab, D.J., and Bedford, K.W. (1994). Initial implementation of the Great Lakes Forecasting System: a real-time system for predicting lake circulation and thermal structure. *Water Poll. Res. J.* Canada, 29: 203-220.

Simons, T.J. (1975). Verification of numerical models of Lake Ontario: Part II: stratified circulation and temperature changes. *J. Phys. Oceanogr*. 5: 98-110.

Swayne D.A., Lam D.C.L., Mackay M., Rouse W, Schertzer, W.M. (2003). Preliminary Assessment of the Interaction Between the Canadian Regional Climate Model and Lake Thermal-Hydrodynamic Models. Proc. *5th International Symposium on Environmental Software Systems* (ISESS '03), May 27-30, 2003, Semmering, Austria. IFIP Conference Series. IFIP ISBN: 3-901882-16-2, pp. 161-177.

A Generalised Conceptual Framework for Integrated Assessment Modelling of Water Resource Management Issues

Letcher, R.A.[1], Jakeman, A.J.[1,2], Croke, B.F.[1,2]

[1]Integrated Catchment Assessment and Management Centre, and [2]Centre for Resource and Environmental Studies, The Australian National University, Canberra ACT 0200, Australia.
rebecca.letcher@anu.edu.au

Abstract: Nodal network approaches are a common framework for considering water allocation problems. In this type of model framework, a river basin is represented as a series of nodes, where nodes generally represent key points of extraction along the stream. Agricultural production and other water use decisions generally interact with the stream system in two ways: they can affect the generation of runoff and thus the volume of water reaching the stream; or, they may involve direct extraction or use of streamflow once it has reached the stream. This paper provides a generalised conceptual framework for considering these types of interactions and their representation in integrated water allocation models.
Keywords: water allocation, conceptual framework, integrated assessment, integrated water resources management

1 INTRODUCTION

Nodal network approaches are a common framework for considering water allocation problems (see for example McKinney et al. [1999]; Rosegrant et al. [2000]; Merritt et al. [2004]; Letcher et al. [in press]; Letcher and Jakeman [2003]; Jakeman and Letcher [2003]). In this type of model framework, a river basin is represented as a series of nodes. Nodes represent points where extraction and other activities impacting on the stream are aggregated for a region and modelled. Regions refer to land or users attached to a node. These may be defined by physical boundaries (eg. subcatchment areas) or by social, economic, technical or political boundaries, depending on the problem being addressed by the model. An example of this type of boundary may be the property areas of irrigators extracting along a reach of the stream between two nodes. Flows are generally routed from upstream nodes to downstream nodes and thus impacts of upstream land and water use activities on downstream users are modelled.

Three recent projects conducted at the Australian National University have developed nodal network models for considering very different land activities, scales and management issues (see Jakeman and Letcher [2001]; Letcher et al. [in press]; Letcher and Jakeman [2003]; Gilmour et al. [under review]). Experiences gained in these projects have led to the development of a general framework for integrated assessment modelling of water allocation issues. This paper develops this framework and outlines several examples of the way in which it can be used to consider various activities and water related management options. Limitations of the current framework and avenues for future development are also discussed.

2 INTEGRATED ASSESSMENT

Integrated assessment is a holistic approach for assessing the impacts and trade-offs related to various land and water related management options. The need for integrated assessment of such issues has been well documented (see for example Letcher and Jakeman [2003]; Jakeman and Letcher [2003]).

In terms of water allocation, integrated assessment models must be able to consider a wide range of land use and management activities that impact on catchment yields. Aspects of the catchment system that must be represented include agricultural and other types of decision making that affect water use or rainfall-runoff generation (socio-economic decision making), the impacts of changed vegetation cover including forest area, farm dam capture and extractive use on the stream, issues of water availability and its impact on crop and livestock production, and the impacts of changed water and land management policy on households, farms and regional communities. The detail with which these system components are considered will depend on the scale at which the management questions are to be answered, the types of land and water use activities present in the catchment and the types of management options to be considered. Several common component models however can be considered.

2.1 Socio-economic decision and impact components

For the socio-economic sub system, two main components must be considered by the model. These are the decision-making component and the socio-economic impact component.

2.1.1 Decision Models

The decision-making component must represent the key land and water use and management decisions being made in the catchment. These may include agricultural production decisions, industrial and urban water use decisions as well as decisions to plant areas of the catchment to forestry or to capture runoff for productive purposes before it reaches the stream. The specific decisions to be simulated and the types of models used to represent these decisions will depend on the spatial and temporal scales at which these decisions are to be modelled as well as on the types of activities present in the catchment. For example, even where extractive uses such as irrigation direct from the stream are considered, this decision may be modelled differently depending on whether the decision is posed as a short-run decision, considering capital to be constrained, or a long-run decision where capital investment decisions are included in the model. Additionally for some issues a representative farm model, simulating decisions by an individual farm, may be used, whereas for larger scale studies, or studies where trade-offs between different industry users are to be considered, aggregated regional production models may be used. In either case, it is the relevant land and water use decisions that are being represented. Frequently used methods for simulating decisions include optimisation-based approaches, based on the assumption that individuals and firms act to maximise profits or utility, and decision tree approaches, where decisions are simulated using empirically derived ‘rules of thumb’.

In general two types of decisions may be made: those based on perfect knowledge of water and land availability; and, those made on the basis of uncertain expectations.

2.1.2 Impact Models

The second component of the socio-economic system that must be represented is the impact component. This component consists of the relevant social and economic impacts of changes in other system components. This may include impacts on farm profits and financial viability, impacts on the regional economy, and on individuals and communities. In some cases local impacts are aggregated and passed to a separate regional scale model (eg. an input-output model) to estimate second order impacts. Again, the scale and range of impacts to be considered dictates the type of modelling approach used.

2.2 Biophysical modelling components

Other aspects of the catchment system that must be represented are relevant biophysical system components. The biophysical components which must be considered will depend on the scale of modelling undertaken, the land use and management activities represented by the model and the types of policy scenarios to be considered. In all cases the hydrological component of the system must be represented in some way so that water allocation can be appropriately considered. The representation of this component must be made so that it has the appropriate sensitivity to various land use activities and policy options being considered. For example, if farm forestry is a land use option for the catchment, then the sensitivity of water yields to forest area must be represented. Croke and Jakeman [2001] provide a good overview of the limits to prediction in hydrology. Other components may include crop and livestock growth models, sensitive to the availability of both land and water and farm dam capture components. For the purposes of the general framework developed here land use activities are considered to be extractive (eg. irrigation from stream) or non-extractive (eg. farm forestry, farm dam capture) in their interaction with the stream.

3 GENERALISED FRAMEWORK

This generalised framework for integrated assessment modelling of water allocation options provides a generic conceptual model for a nodal network approach to considering water allocation. The form of this conceptual model differs on the basis of the types of decisions being made: those based on perfect knowledge of water and land availability; and, those made on the basis of uncertain expectations. Two types of productive uses are also considered. Extractive uses are those that impact on the stream system after rainfall-runoff has reached the stream. Non-extractive uses are considered to be those that influence rainfall-runoff generation before it reaches the stream (revegetation, capture of rainfall-runoff in on-farm storages for livestock or irrigation uses).

The Biophysical Modelling Component (BMC) consists of crop and hydrological modelling components. Where land use activities are extractive, a 'policy filter' is generally required to translate streamflow into available extraction limits, and to translate lumped extractions to extraction on the same time scale as used by the hydrological component. When land use activities are non-extractive, the hydrological component must have the appropriate sensitivities to these non-extractive uses.

3.1 Perfect knowledge based decisions

Where decisions are simulated under conditions of perfect knowledge, the decision and impact components are represented by a single model component. This is because decisions are made with full knowledge of the impacts they will incur on the social and economic subsystems, and are usually taken into account in the decision simulation. That is, an agricultural production decision is generally made to trade-off the economic and social impacts on the production unit (household or farm).
For perfect knowledge based decisions, regardless of whether the land use is extractive or non-extractive the general conceptual framework is shown in Figure 1.

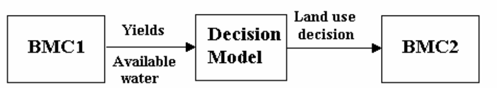


Figure 1. Conceptual framework for decisions based on perfect knowledge

In this case, BMC1 and BMC2 may not contain the same models. BMC2 must contain a hydrological model that is sensitive to changes in the land use activity considered by the decision model. BMC1 may contain only very simple models of crop/activity yield and climate impact of available water. For extractive uses this will be a hydrological model simulating pre-extraction flows, whereas for non-extractive uses this may be a filter on rainfall or evaporation to determine dam capture, or forest yields. In any case BMC1 generally represents the biophysical systems knowledge assumed to be known by the decision maker before making the decision. Several examples of the use of this framework for considering extractive and non-extractive activities are given in the next sections.

3.1.1 Non-extractive land use activities

One possible non-extractive land use activity that has impacts on the availability of water in-stream is forestry. Where forestry decisions are made under the assumption of perfect knowledge the conceptual framework may be applied to produce a model integration structure as shown in Figure 2.

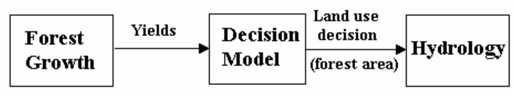


Figure 2. Application of framework for perfect knowledge, forestry activities

In this case BMC1 simulates forest growth (and thus yields). The level of complexity of this component is variable. It may range in complexity from a full forest growth simulation model to a simple empirical model to a look-up table approach. This forest growth component (ie. BMC1) passes forest yields to the decision model. The decision model then links to BMC2, a hydrological model, using the land use decision, or chosen forest area. This hydrological component must be sensitive to changes in forest area.
Another example of a non-extractive land use that has impacts on streamflow yield is the use of farm dams to capture runoff for activities such as viticulture. The integrative framework that results for this activity is illustrated in Figure 3.

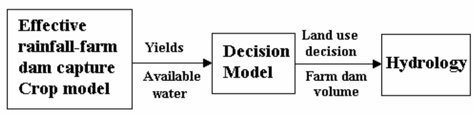


Figure 3. Application of framework for perfect knowledge, runoff capture activities

In this case BMC1 consists of a component to estimate the capture of runoff in farm dams for a given climatic series and a crop growth component. The estimate of farm dam capture may be on a per ha or ML of storage basis where capacity is a decision, or for total capacity where capacity is a constraint to decision making. Links between BMC1 and the decision model comes through simulated crop yields and water available in farm dams for irrigation purposes. BMC2 consists of a hydrological modelling component, which must be sensitive to changes in farm dam capacity in the catchment. BMC2 is integrated using the total volume of farm dam capacity and the area based land use decision, which are outputs of the decision component.

3.1.2 Extractive land use activities

Application of the conceptual framework for considering extractive activities, such as the production of irrigated crops using water extracted from streams results in the integrative framework shown in Figure 4.

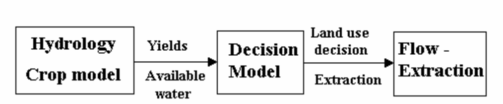


Figure 4. Application of framework for perfect knowledge, extractive activities

In this case, BMC1 consists of both hydrological and crop model components. A policy filter, which models extraction policy rules, must be used to determine water available for extraction on an appropriate time step. BMC1 is linked to the decision model by yields (or productivity) and available water. BMC2 is then a second hydrological component that removes extracted water from 'natural' flow. The decision model integrates with BMC2 through the extraction implied by the area based land use decision.

3.2 Expectations based decision making

Where decisions are assumed to be based on uncertain expectations, a separate decision model and impact model need to be used because the actual impact can only be determined or modelled after the land use decision has been made. Regardless of whether or not the land use is extractive or non-extractive, the general framework for integration is shown in Figure 5.

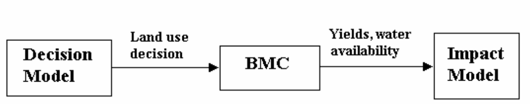


Figure 5. Conceptual framework for decisions based on uncertain expectations

Several examples of the application of this conceptual framework to various activities are given in the next sections.

3.2.1 Non-extractive land use activities

Application of the conceptual framework to forestry plantation would result in model integration of the form shown in Figure 6.

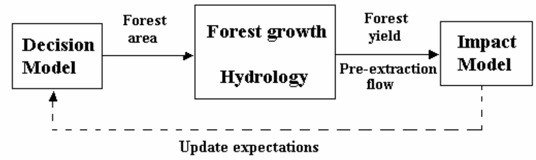


Figure 6. Application of expectations based framework, forestry activities

In this case the BMC consists of a hydrological modelling component, which must be sensitive to changes in forest area, and a forest growth simulation component, which may be as simple as a look up table of forest yields. In this case the link between the decision model and the BMC is the forest area, which is an output of the socio-economic decision. The BMC then passes the yield of forest products and the pre-extraction (often referred to as 'natural') streamflow to the impact model.

Application of the conceptual framework to expectations based decisions on activities involving runoff capture in farm dams results in model integration as shown in Figure 7.

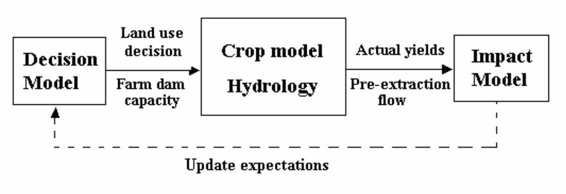


Figure 7. Application of expectations based framework, runoff capture

In this case the hydrology must be sensitive to changes in the capacity of farm dams for capturing runoff. Crop yields must also be sensitive to the availability of water in farm dams for irrigation. The link between the decision model and the BMC is again the area based land use decision and the corresponding farm dams capacity. Where the farm dam capacity is treated as a constraint rather than a decision variable then this capacity is a fixed input to the BMC. It is not produced as an output of the decision model in this case.

3.2.2 Extractive land use activities

If the land use activity considered is extractive, then the BMC will contain a hydrological modelling component and may contain a crop model (this may be simple/empirical crop model, or a simple look-up table of relevant crop yields and water use). The integrated model structure using the conceptual framework is shown in Figure 8.

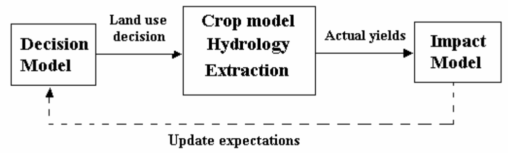


Figure 8. Application of expectations based framework, extractive activities

As before the link between the decision model and the BMC is the area based land use decision. The BMC then links to the impact model through simulated yields from the crop modelling component. In this case extraction and water allocation between alternative crops for the node must be handled using a water allocation module within the BMC, based on a set of predefined prioritisation rules. This is different from the extractive framework under perfect knowledge assumptions (see Section 3.1.2), where water allocation decisions internal to the node were handled by the decision model.

4 MULTIPLE ACTIVITIES AND OTHER INTEGRATION CONSIDERATIONS

In many cases the integrated assessment will be expected to consider a range of extractive and non-extractive land uses at a single node. In this case the concept of a land modelling unit (LMU) may be used to disaggregate decision making at the node.

4.1 Land Modelling Units

A Land Modelling Unit (LMU) is a 'homogenous' area used to disaggregate a catchment for the purposes of modelling. The concept of 'homogenous' is applied in terms of various ecological, physical, social or economic characteristics, usually defined by the model question being considered. Common characteristics underlying the definition of LMU in the model are topography, climate, soils, geology, ecological community, farm production or industry type and policy scales. LMUs are generally considered to be intersections of these key characteristics so that each region or modelling unit considered by the model is 'relatively homogeneous' in terms of these characteristics. LMU are generally associated with a set of activities that interact with the hydrological cycle in a defined way. More than one LMU can be linked to each node.

4.2 Use of the conceptual framework for considering multiple LMU at a node

Consider a situation where two separate sets of policy interventions are likely to impact on the stream system: areas of the upper catchment currently under a grazing system are being considered for reforestation to manage salinity problems in the catchment; and, access to irrigation water extracted direct from the stream is to be changed to manage environmental flow outcomes. Figure 9 demonstrates an example of the break down of this type of catchment into two separate LMU types based on current land use and its interaction with these two policies. LMU 1 correponds to areas currently under pasture or forest and is directly affected by the first policy. LMU 2 corresponds to current and potentially irrigated areas of the catchment, which are areas affected by the second policy. The node is placed at the subcatchment outlet.

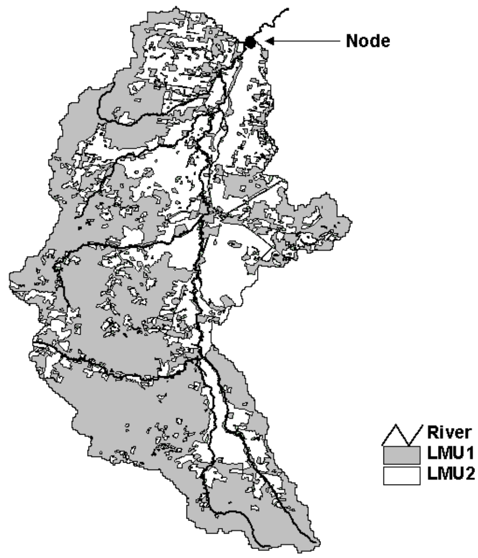


Figure 9. Example LMU break down of catchment

A conceptual framework for modelling this set of issues and activities, where it is assumed that farmers have perfect knowledge, would then be as shown in Figure 10.

The generalised conceptual framework is implemented first for LMU 1 since this affects non-extractive activities. The BMC components for this LMU include a forest growth model and a rainfall-runoff component where effective rainfall is dependent on forest area. The implementation for LMU 2 then links through this rainfall runoff component and a crop modelling component to the decision model for LMU 2. The impact of extractions on streamflow is then considered in a third BMC.

Figure 11 demonstrates the conceptual modelling framework for this problem where both sets of decisions are assumed to be expectations based.

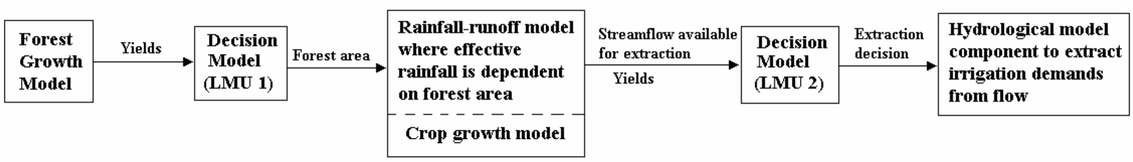


Figure 10. Conceptual framework for multiple LMU under perfect knowledge assumption

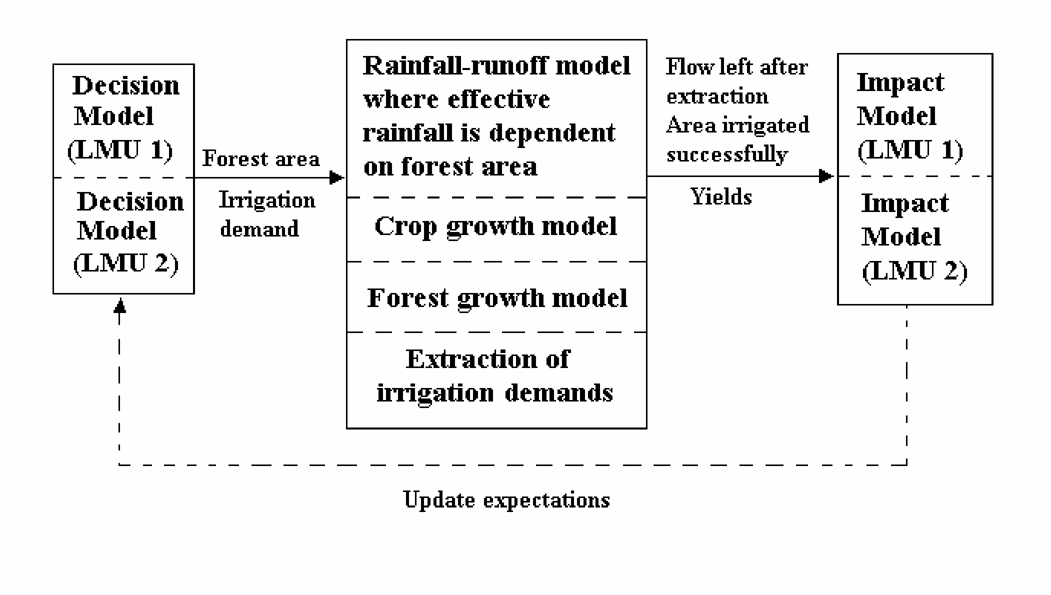


Figure 11. Conceptual framework for multiple LMU under uncertain expectations

This figure demonstrates that both decision models (ie. for LMU 1 and 2) are able to be run concurrently since expectations are all that is required for the decision. These models then pass forest area and irrigation demands to the BMC. This component includes a rainfall-runoff where effective rainfall is dependent on forest area, crop and forest growth models and extraction of actual irrigation demands from streamflow. The area successfully irrigated, crop and forest growth and flow left after extraction are then passed to the impact model to consider the impact of actual values of these on the decision-maker. This information is then used to update expectations for the next set of decisions.

5 DISCUSSION AND CONCLUSIONS

The conceptual integrative framework described in this paper has been successfully applied to consider water quantity based issues such as water allocation, runoff capture and the impacts of forestry plantation on flows (see Jakeman and Letcher [2003] for examples of its application). It should be able to be adapted for a range of land and water based management issues, including water quality and ecological management issues relatively easily. It is expected that these modifications would be made primarily during definition of the LMU for land based activities and impacts, and during definition of nodes for in-stream impacts.

6 REFERENCES

Croke, B. F., and Jakeman, A. J. (2001). "Predictions in catchment hydrology: an Australian perspective." *Marine Freshwater Research*, 52, 65-79.

Gilmour, J. K., Letcher, R. A., and Jakeman, A. J. (under review). "Analysis Of An Integrated Model For Assessing Land And Water Policy Options." *Mathematics and Computers in Simulation*.

Jakeman, A. J., and Letcher, R. A. (2003). "Integrated Assessment and Modelling: Features, Principles and Examples for Catchment Management." *Environmental Modelling and Software*, 18, 491-501.

Letcher, R. A., Croke, B. F., and Jakeman, A. J. (in press). "Model Development for Integrated Assessment of Water Allocation Options." *Water Resources Research*.

Letcher, R. A., and Jakeman, A. J. (2003). "Application of an Adaptive Method for Integrated Assessment of Water Allocation Issues in the Namoi River Catchment, Australia." *Integrated Assessment*, 4(2), 73-89.

Merritt, W. S., Croke, B. F., Jakeman, A. J., Letcher, R. A., and Perez, P. (2004). "A biophysical toolbox for assessment and management of land and water resources in rural catchments in Northern Thailand." Ecological Modelling, 171, pp. 279-300.

McKinney, D. C., Cai, X., Rosegrant, M. W., Ringler, C., and Scott, C. A. (1999). "Modeling water resources management at the basin level: Review and future directions." International Water Management Institute, Colombo, Sri Lanka.

Rosegrant, M. W., Ringler, C., McKinney, D. C., Cai, X., Keller, A., and Donoso, G. (2000). "Integrated economic-hydrologic water modeling at the basin scale: the Maipo river basin." *Agricultural Economics*, 24, 33-46.

Metadata and Modeling Frameworks: The Object Modeling System Example

O. David[a] , **I. W. Schneider**[b], **and G. H. Leavesley**[c]

[a] *Colorado State University, Fort Collins, CO, U.S.A.*

[b] *USDA Agricultural Research Service, Great Plains System Research, Fort Collins, CO, U.S.A*

[c] *US Geological Survey, Denver CO, U.S.A*

Abstract: The main motivation for the usage of modeling frameworks for environmental simulation software is to manage and simplify the interoperability of (loosely) coupled simulation components. Conventional approaches in collaboration are using an Application Programming Interface (API). Recent developments in simulation frameworks focus on introspecting architectures for simulation components, where components become passively explored and integrated in to the framework. Such solutions seem to be more flexible to support the framework evolution because components are less tight to a specific framework API. The Object Modeling System (OMS) is an introspecting simulation framework, which uses metadata in annotated components such as (i) spatial and temporal constraint specification, (ii) data annotation for variables and parameters to specify simulation related data like runtime constraints for range validation, unit conversion, or automated testing. The OMS utilizes metadata annotation (i) at model construction time to support proper spatial and temporal model assembly (ii) and at model runtime to support proper data linkage. This paper will emphasize on metadata access to support model linkage to simplify the development of simulation components for environmental scientists and will give application examples based on the Object Modeling System

Keywords: Modeling; Framework; Metadata; Components; Java

1. INTRODUCTION

Meta data in the context of modelling and simulation traditionally represents knowledge about the application of simulation models. For example, the proper setup of parameter sets to run a model is only possible with the correct knowledge about parameter range constraints, physical units, etc. Such information usually resides in documentation or is a part of the user interface.

For modelling frameworks meta data has further meaning and importance. Meta data can be used to support the process of model (i) construction, (ii) testing and (iii) application.

This paper introduces the Object Modeling System with respect to its underlying component concept.

2. THE OBJECT MODELING SYSTEM

The Object Modeling System (OMS) is a Java-based modeling framework that facilitates simulation model development, evaluation and deployment. OMS models are treated as hierarchical assembled components representing building blocks. Components are independent and reusable software units implementing processing objects for simulation models. In general, OMS consists of a library of science, control, and database modules and a means to assemble the selected modules into an application-specific modeling package. The framework is supported by data dictionary, data retrieval, GIS, graphical visualization, and statistical analysis utility modules. Current challenges in natural resource management have created demand for integrated, flexible, and easily parameterized hydrologic models. Most of these monolithic models are not modular, thus modifications (e.g., changes in process representation) require considerable time, effort, and expense.

OMS addresses these needs by using an object-oriented, component based approach for model development to:

- Reduce duplication of modeling efforts
- Improve the quality and currency of model code
- Make natural resource models much easier to build, access, understand and use;
- Facilitate long-term maintainability of existing and new natural resource models;
- Lead to greater consistency of modeling for particular problems and scales;
- Enhance response and delivery times in scientific modeling projects;
- Ensure creditability and security of model implementations; and
- Function on any major computing platform.

2.1 OMS Components

Components are the basic building blocks of a model. They represent usually a unique concept in a model like a physical process, a management practice, a remote data input, etc. Components can be fully understood just by exploring the component metadata. Therefore OMS defines a comprehensive set of metadata which should be bundled with the component.

Meta data attachment to components is a technique which was introduced to the Java Programming language by means of formal documentation annotation support. Classes, fields, and methods can have documentation header, which are comments according to the Java Language specification but contain useful additional info for the java documentation tool 'javadoc' [???]. Javadoc annotations are primarily used to generate API documentation (HTML and other formats) out of java sources. Such an approach helps managing a consistent source code documentation by keeping code and code documentation in the same file.

The javadoc style annotation of java language elements was also designed to be extensible. It is legal to define custom meta data annotations for java classes. Parsing tools for java such as javadoc, 'qdox', or the Netbeans Java Source API provide comprehensive support to lookup and manage documentation annotations.

The javadoc approach was used to support the meta data annotation of components. The meta data can be categorized in two conceptual groups

- *Meta data to document the component in an informal way*. A component contains for example metadata about its creation time, the author, the organization, or references to publications related to the code.

1. *Formal meta data to annotate a component for model integration or testing.* Such meta data is required by OMS to support the proper component integration into a model at design time. Such meta data enables OMS to deal with formal processing requirements.

There are two main levels of meta data annotation for an OMS component.

1. Component meta data annotates the entire component. It contains information about its overall purpose, authorship, version control, literature references, temporal or spatial scale etc.
2. Attribute meta data annotates data requirements of the component per public attribute. Such metadata captures additional information about the attribute, such as physical units, data constraints, data flow, default values, etc.

Figure 1 shows the component meta data for a forage component. OMS related meta info is tagged with a '@oms.' at the beginning to avoid 'namespace' conflicts with potential other tag definitions.

```
/**
 * Forage.
 *
 * This module estimates daily
 * plant growth for five functional
 * groups in rangelands.Daily growth is
 * driven by average daily temperature
 * relative to the optimum and base
 * temperatures for cool season grasses
 * (C3), warm season grasses (C4),
 * legumes, shrubs, and weeds. The module
 * is adapted from GPFARM.
 *
 * @oms.author  Allan Andales
 * @oms.version $Id: Forage.java 294 2004-
 *              05-24 21:11:15Z david $
 * @oms.created April 1, 2004, 2:39 PM
 * @oms.name    Forage Component
 * @oms.category   Plants
 * @jni.files RunForage.f90 PlantMod.f90
 *            ForSite.f90 Forage.f90
 */
public class Forage extends …
```

Figure 1: 'Forage' Component Definition

Component meta data in OMS affects several phases in component integration and model development which are mentioned briefly.

2.1.1 Component Documentation.

Meta data annotations of components are used in OMS to generate documentation. Unlike the default javadoc tool, OMS generates XML Docbook (Welsh, 1999) to document components. Docbook was chosen as the main document format because of its flexibility to transform component documentation into other formats more easily. Formal and informal meta data is used for documentation generation. Resulting component XML specification can be processed with any tool supporting docbook or they can be transformed in other XML representations. The inclusion of component specifications into an XML component library data base is one of many examples.

2.1.2 Component Testing

Formal meta data is being used to support automated testing. OMS facilitates black box testing of components. Each component attribute can be annotated with test related information for its input and output. OMS consumes these annotations in a testing phase to test the component by

1. Generating input data according to given test annotations. Test data will be generated using several available random or sequence generators.
2. Checking output data according to given test annotations. Output data is usually expected to be in a certain range.

2.1.3 Component Integration/Model Design Time

Model design refers to the process of model construction based on model building blocks, known as components. A component needs to tell the modelling framework if it fits into model, which the model developer is going to built. The component has to offer information about application scales, data dependencies and other information, which are important for a consistent integration of the component into the model. If for example data requirements cannot be resolved, the modelling framework will guide the user to select alternative or additional components for the model.

2.1.4. Component Execution/Model Runtime

At model run time meta data availability about the models data is useful as well. If for example the components data has additional meta data about physical units of the components input data, unit conversion can be performed dynamically. Run time checks of model state variables can be specified with metadata.

Metadata affects the communication between the modeling framework and the component. Components want to see data handled by the modelling system under a certain name, in a certain unit, etc. Components interact with the framework to get the data in the right structure and format.

2.2 API vs. Introspection

There are two major communication principles of components. In an *API* based communication the component interacts with the system by using a well known API. This is the common and traditional method for interoperability. Component developers are using framework API calls to get data in a requested format. Such an approach can be used in any kind of programming language. A component using a framework API is then tightly coupled, technically and conceptually. Such a component can hardly be reused in another context.

The communication between the modeling framework and the component can also be based on a more flexible schema for interaction, called introspection. Here, the modeling framework is capable to explore the structure and content of a component with respect to data and metadata. This approach works only on architectures, which do have building support for such introspection and exploration. Java and C# are the examples for such architectures.

2.2.1 OMS Attributes

An attribute is a basic data type in OMS. It extends the concept of a typed data element with its ability for metadata annotation and management.

Attributes are representing model state variables or model parameter. They are component input or output. A component interacts with attributes only. The usage of attributes in a model determines, if such an object is considered to be a parameter or a variable. There is no such terminology like a parameter or state variable at the component level in OMS.

OMS uses introspection to deal with meta data for design and run time purposes. All data and meta data belonging to the attribute called '*nitstress'* is shown in Figure 2. The object '*nitstress'* represents a stress factor for nitrogen in a OMS component called 'Forage'. There is the data declaration given as object of type 'Attribute.Double'. The does have such classes for all reqired basic data types (floating point, fixed point, boolean). These are actually Java interfaces and no classes.

The data declaration is preceded by a section containing all the metadata about the attribute. This section is enclosed in '/** .. */'. It is actually a comment according to the Java Language specification as mentioned before.

```
/** Nitrogen stress factor.
 * @oms.name_de        Sickstoff     Stress
                       Faktor
 * @oms.unit           kg/kg
 * @oms.access         read
 * @oms.constraint     0.0..1.0
 * @oms.default        1.0
 */
private Attribute.Double nitstress;
```

Figure 2: Attribute Definition

The metadata annotation section starts with the name of the attribute in English. It is followed by OMS annotation tags all starting with "@oms.". There are several tags recognized by OMS:

- `@oms.name_<loc>`
  This optional tag contains the localized name. If the design or run time environment is localized for a certain lanuage, this name gets used.
- `@oms.unit`
  This optional tag refers to the unit of the attribute. It uses the UCAR units notation specification (Emmerson et al. 2001) .
- `@oms.access`
  This required tag specifies the data flow. Either the component reads or writes this attribute.
- `@oms.constraint`
  This optional tag controls additional value constraints of an attribute. In the example above nitstress is a factor wich has a valis range of 0 to 1. Any other value is considered to be incorrect according to the definition of nitstress. Constrains become handy for run time verification of a model. If a value violates the constraint specification the system will notify the user and might indicate a problem with input data. Other valid values for the constraint tag are: '<=0' or '>-273.15'
- `@oms.default`
  This optional tag contains the default value for this attribute.

There are other metadata annotation tags not used in this example.

- `@oms.dim`
  If the attribute is an Array (eg. Attribute.DoubleArray) this tag points to a numerical constant or to another attribute representing the dimension of that array.
- `@oms.test`
  This specific tag supports the automated testing of components. The content of this tag is similar to the constraint tag. OMS perform a component test, which generates data according to the requirements specified in this tag. In the given example a useful test related to 'nitstress' would be '@oms.test 0.0..1.0'. A randomizer would generate input data in this range. The component could then be verified automatically.

Attribute types such as 'Attribute.Double' are realized as Java interfaces as supposed to classes. There are several internal classes in OMS which are implementing these interfaces. The system choses the right implementation for the given set of metadata and will pass the system generated attribute to the component. It uses reflectively the property getter and setter methods to access the attribute. Setter and Getter are implemented according to the Java Beans Standard.

5. CONCLUSIONS

Meta data is far more than a semantic add-on for modelling frameworks. Properly designed, a strong support for meta data helps modelling frameworks to support model developer and user in the process of model construction and application.

The Object Modeling System OMS uses annotations of components to capture information about the models data in addition to data names and types. Such annotations provide the benefit of keeping data and metadata close together and allow for the application of different tools for executing testing, and documenting.

The metadata annotations in OMS were used for the development of model components out of the Root Zone Water Quality Model (RZWQM), the Precipitation Runoff Modeling System (PRMS).

The future Java annotation support specified in the JSR 175 and implemented in JDK 1.5 offers the a more straightforward implementation of metadata management in the Java virtual machine, which is expected to be used for OMS.

6. ACKNOWLEDGEMENTS

The authors wish to thank the Agricultural Research Service ARS, the National Resource Conservation Service NRCS and the U.S. Geological Service USGS for the support of the OMS development.

7. REFERENCES

David O., S.L. Markstrom, K.W. Rojas, L.R. Ahuja, and I.W. Schneider (2002). The Object Modeling System, In: Agricultural System Models in Field Research and Technology Transfer, L. Ahuja, L. Ma, T.A. Howell, Eds., Lewis Publishers, CRC Press LLC, 2002: 317—331.

Leavesley, G. H., R. W. Lichty, et al. (1983). Precipitation-runoff modeling system--User's manual, U.S. Geological Survey: 207.

Leavesley, G. H., G. E. Grant, et al. (1996). A modular modeling approach to watershed analysis and ecosystem management. Watershed '96 A National Conference on Watershed Management, U.S. Government Printing Office.

Bloch J. : A Program Annotation Facility for the Java Programming Language. JSR-175 Public Draft Specification. [http://www.jcp.org/en/jsr/detail?id=175]

Emmerson S. el al. (2001): Units Specification. JSR-108 Public Draft Specification. [http://www.jcp.org/en/jsr/detail?id=108]

Walsh, N., Muellner L. (1999). DocBook: The Definitive Guide. O'Reilly & Associates, Oct 1999.

JavaDoc, Core JavaDoc Tool [http://java.sun.com/j2se/javadoc/]

Arnold,K.; J. Gosling, D. Holmes (1998): The Java Programming Language, Addison Wesley, 1998

Integrating Water Quality Models in the High Level Architecture (HLA) Environment

K.-E. Lindenschmidt, F.B. Hesser and M. Rode

UFZ Centre for Environmental Research Leipzig - Halle, Department Hydrological Modelling, 39114 Magdeburg, Germany; karl-erich.lindenschmidt@ufz.de

Abstract: HLA (High Level Architecture) is a computer architecture for constructing distributed simulations. It facilitates interoperability among different simulations and simulation types and promotes reuse of simulation software modules. The core of the HLA is the Run-Time Infrastructure (RTI) that provides services to start and stop a simulation execution, to transfer data between interoperating simulations, to control the amount and routing of data that is passed, and to co-ordinate the passage of simulated time among the simulations. The authors are not aware of any HLA applications in the field of water resources management. The development of such a system is underway at the UFZ – Centre for Environmental Research, Germany, in which the simulations of a hydrodynamic model (DYNHYD), eutrophication model (EUTRO5) and sediment and micro-pollutant transport model (TOXI5) are interlinked and co-ordinated by the HLA RTI environment. This configuration enables extensions such as (i) “cross-model” uncertainty analysis with Monte Carlo Analysis: time synchronisation allows EUTRO5 and TOXI5 simulations to be made after each successive simulation time step in DYNHYD, (ii) information transfer from EUTRO to TOXI to compute organic carbon fractions of particulate matter in TOXI, (iii) information transfer from TOXI to EUTRO to compute extinction coefficients in EUTRO and (iv) feedback from water quality simulations to the hydrodynamic modeling.

Keywords: High Level Architecture (HLA); Water Quality Simulation Program (WASP5); Model Coupling

1. INTRODUCTION

This paper conceptualizes the integration of computer models used to simulate river water quality in the High Level Architecture (HLA) platform. There has been an impetus to configure models into modeling systems due to the integrative and more holistic approach science has taken to tackle environmental management problems. This is particularly the case for the management of large river basins (e.g. European Water Framework Directive [2000/60/EC]).

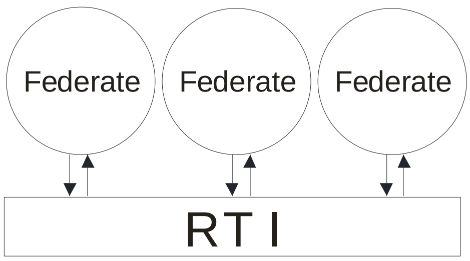


Figure 1. The HLA environment (adapted from Petty [2002].

HLA is a program specifically developed for the coupling of models into one system (see Figure 1). The models are represented as Federates to and from which information is communicated by the Run Time Infrastructure (RTI), which is the core of the HLA. It provides services to start and stop a simulation execution, transfer data between interoperating simulations, control the amount and routing of data that is passed and co-ordinate the passage of simulated time among the simulations [Kuhl et al., 1999]. One particular strength of the RTI is its ability of sequencing command intercommunication amongst the Federates in a very efficient manner. An additional advantage is its ease of communication of Federates on different computer systems via a network.

HLA has been used for a number of applications such as:

- communication and synchronisation media for modelling of hybrid systems (e.g. water level control in a two tank system) [Borshchev et al., 2002],

- modelling and simulation of processes associated with software system acquisition activities which concerns the funding, management, engineering, system integration, deployment and long-term support of large software systems [Choi and Scacchi, 2001],
- developing multi-agent systems for applications in mobile robotics [Das and Reyes, 2002],
- providing online/real-time location information of streetcars of the public transportation company in Magdeburg, Germany [Klein, 2000],
- traffic reduction schemes based on space-based quantization [Lee and Zeigler, 2002],
- designing simulation environments for human training (esp. military personnel) [Maamar, 2003].

The authors are unaware of any applications in water quality simulation. A project has been initiated in Germany to develop a flood forecasting system integrating a hydrological model, GIS (Geographical Information System) and up-to-date rainfall data [Schulze et al., 2002] but the project is still in its infancy.

In this paper the Water Quality Simulation Program version 5 (WASP5) is used to illustrate the capabilities gained in the simulation process by integrating its submodels into a HLA federation. The submodels include the hydrodynamics (DYNHYD), eutrophication (EUTRO) and sediment and micro-pollutant transport (TOXI) of surface waters. The HLA implementation induces better interactive transfer of information between the submodels to extend predictive ability and uncertainty analyses. The paper concludes with a comparison of the HLA method with conventional model coupling and object-oriented approaches.

2. MODELLING METHODS

The WASP5 modeling system [Ambrose et al., 1993] is used to simulate the hydrodynamics, water quality and transport of sediment and micro-pollutants in a river. It is a conglomeration of three models:

i) DYNHYD – uses the St. Venant equation (full dynamic wave) to calculate the flow of water through a water body. With the volume continuity equation and Manning's equation of flow through a channel, mean interfacial flows $\overline{Q}$, water volumes V, mean current velocity $\overline{v}$ and mean water depth $\overline{d}$ of each discretized unit can be simulated dynamically. These variables are transferred as input to the two models, EUTRO and TOXI, described next:

ii) EUTRO – simulates the transport and transformation of variables important in describing the eutrophication processes (high productivity) in surface waters. Oxygen plays a central role in the system and is balanced between sources (production through photosynthesis by phytoplankton growth and re-aeration from the atmosphere via the water surface) and sinks (oxygen demand by organic matter and bottom sediments; respiration; nitrification). The dynamics of nutrients (phosphorus and nitrogen) complement the phytoplankton growth.

iii) TOXI – simulates the transport of sediments and micro-pollutants. Transport processes induce advection, dispersion and sedimentation and resuspension of particulate matter to and from the bottom sediments and diffusion of dissolved matter between the water body and bottom sediments. These substances may also undergo transformations such as equilibrium sorption, volatilization, ionization, photolysis, hydrolysis and biodegradation.

The sequence of simulation flow is illustrated in Figure 2. The hydrodynamic simulation must be completed first to generate interfacial flows, water volumes, mean velocities and mean depths of each discretized unit. These variables are stored in a hydrodynamic file (*.hyd) for subsequent input to EUTRO and TOXI.

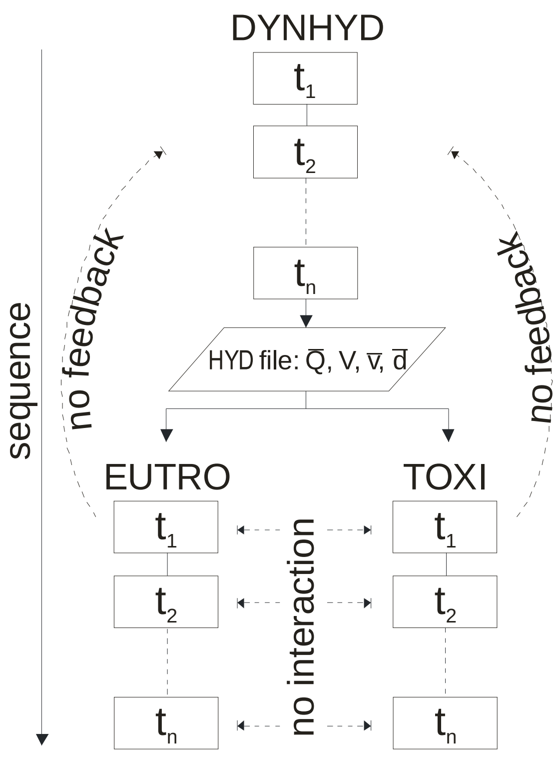


Figure 2. Sequence of submodel simulations in the original WASP5.

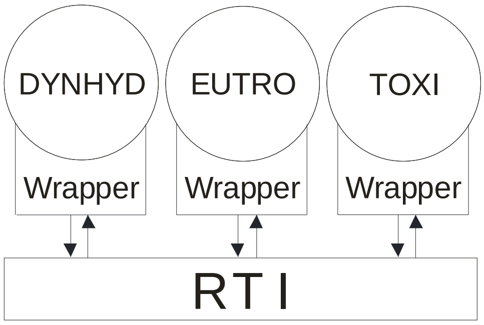


Figure 3. Implementation of the WASP5 submodels DYNHYD, EUTRO and TOXI in HLA.

In its original version, no interaction occurs between EUTRO and TOXI and feedback from these two models to DYNHYD is also not possible. More flexibility is obtained when these models are coupled together in the HLA platform as shown in Figure 3. Since the RTI is programmed in the C++ language and the three WASP5 models in Fortran, a wrapper needs to be integrated into each model to allow calls to be made to and from the models by the RTI.

3. HLA IMPLEMENTATION

An initial configuration of the simulation sequencing in the HLA environment is shown in Figure 4. Storage of the hydrodynamic variables is avoided and can be passed after each time step directly to EUTRO and TOXI. All three models can now be run simultaneously and the simulation time steps are synchronized by the RTI. Hence, for example, the simulation of EUTRO and TOXI at time step 2 (t_2) occurs as soon as the t_2 simulation of DYNHYD is completed. DYNHYD than immediately begins simulation of its time step 3 (t_3) while EUTRO and TOXI are computing their t_2. There is a saving in computation time, especially when the three models are networked together on three different processors or on a multi processor computer. Also, Intels® Hyperthreading expedites the simulations.

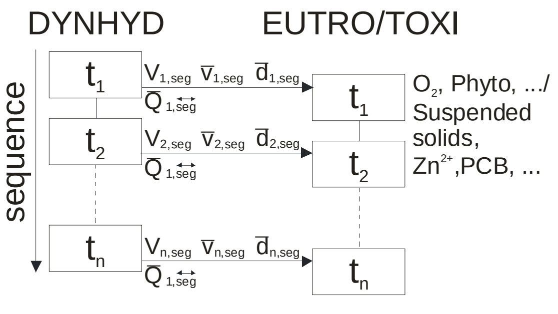


Figure 4. WASP5 submodel simulation sequence in HLA.

3.1 Monte-Carlo uncertainty analysis

An important advantage of such a configuration is the extensive capabilities it provides for Monte-Carlo uncertainty analysis. The effect of certain variations of change in the parameter values (e.g. *a* and *b* in Figure 5) in DYNHYD on the variables in EUTRO and TOXI can now be explored. Many 100s and 1000s of runs of the same simulation are carried out, in which a different parameter setting is chosen for each run. The DYNHYD parameters are randomly selected from a probability distribution of its values. This evokes a distribution of the EUTRO and TOXI variables (e.g. oxygen O_2 and suspended sediments *SS*), from which the contribution of uncertainty of each parameter on the water quality description can be made. In the conventional simulation sequence (Figure 2) only an uncertainty analysis of the parameters on the variables of the same model could be made [see for example Lindenschmidt et al., 2003; Lindenschmidt et al., 2004]. With the HLA implementation, "cross-model" uncertainty analysis is made possible.

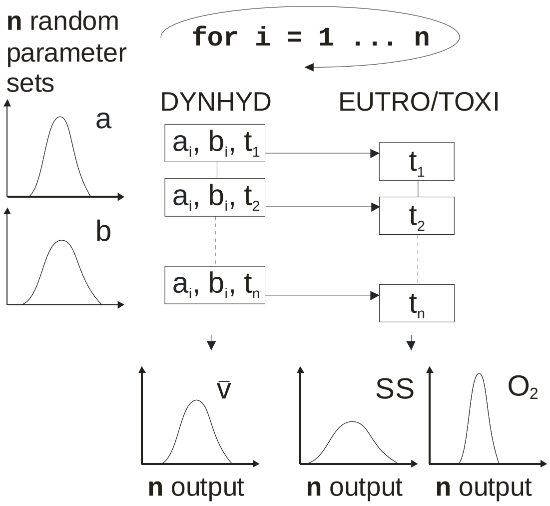


Figure 5. Monte-Carlo analysis of the WASP5 federation.

3.2 EUTRO to TOXI interactions

Coupling EUTRO and TOXI together in the HLA environment allows ease of communication between the two models. An example of data transfer from EUTRO to TOXI is shown in Figure 6.

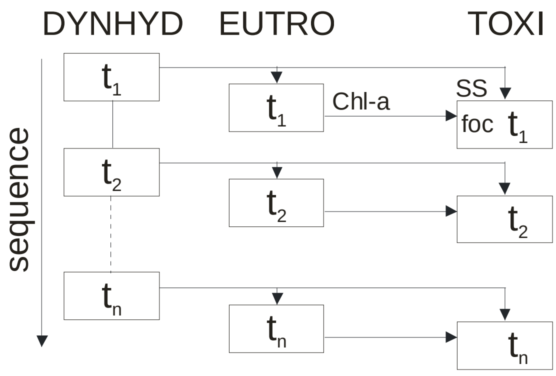


Figure 6. Flow of information from EUTRO to TOXI to calculate *foc*.

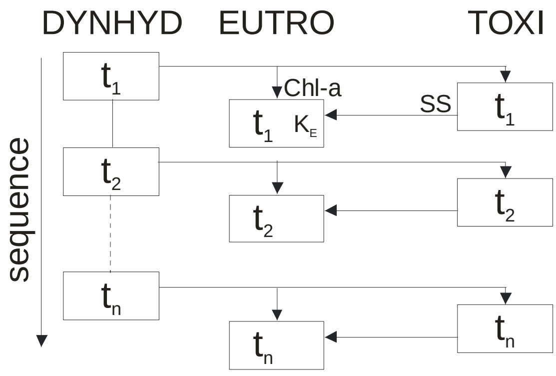


Figure 7. Flow of information from TOXI to EUTRO to calculate K_E.

The concentration of chlorophyll-a (*Chl-a*) is a variable in EUTRO which, in many cases, correlates well with the particulate organic carbon (*POC*) content in the water. By dividing *POC* with the concentration of suspended sediment (*SS*) simulated in TOXI, the weight fraction of organic carbon in suspended matter (*foc*) is obtained [Ambrose et al., 1993, Chapra, 1997]:

$$foc = \frac{POC}{SS}$$

The partitioning of heavy metals M in the water can now occur between the dissolved phase M_{DIS} and the organic carbon of the particulate phase M_{OC} using the partition coefficient K_{OC}

$$K_{OC} = \frac{K_D}{foc}$$

This is an extension of the simplified partitioning of heavy metals between the dissolved and total particulate fractions of heavy metals, M_{DIS} and M_{PART}, respectively, using the partition coefficient K_D:

$$M_{DIS} = \frac{1}{1 + K_D \cdot SS} \qquad M_{PART} = \frac{K_D \cdot SS}{1 + K_D \cdot SS}$$

where *SS* is the concentration of suspended solids.

3.3 EUTRO to TOXI interactions

The transfer of information from TOXI to EUTRO is shown conceptually in Figure 7. In the original WASP5 version, the extinction coefficient K_E of light passing through the water column is a constant parameter implemented for each discretized unit of the modelled river. With the communication between TOXI and EUTRO, K_E can now vary depending on the chlorophyll-a concentrations *Chl-a* (µg/l) and phytoplankton biomass *Phyto* (mg/l) computed in EUTRO and the suspended solids *SS* (mg/l) calculated in TOXI [equation modified from Schöl et al., 2002]:

$$K_E = 0.052 \cdot (SS - Phyto) + 0.013 \cdot Chla + 1.06$$

3.4 EUTRO/TOXI feedback to DYNHYD

An extension of the WASP5 model by Shanahan [2001] includes the variable macrophytes. This allows interacting feedback constructs from EUTRO and TOXI to DYNHYD, shown in Figure 8. Water bodies laden with many submerged plants may alter the hydrodynamics of the water course. High concentrations of particular micro-pollutants may inhibit the growth of phytoplankton and macrophytes. Alternatively, areas of greater or less sediment deposition may also influence the water flows and velocities, especially for shallow water bodies.

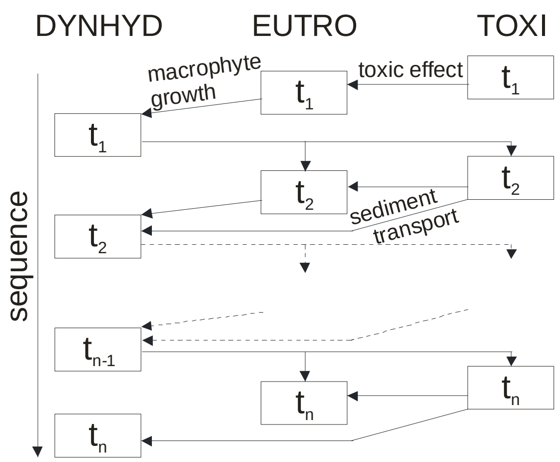


Figure 8. Feedback from EUTRO and TOXI to DYNHYD.

4. DISCUSSION: MODEL COUPLING

Figure 9 shows a comparative exercise between three different model coupling variants:
1) conventional – models are loosely coupled, in that information is transferred from one model to another by file storage and retrieval. Additional programming in the model source codes is not necessary. Sequence control is managed by batch files or an external program.
2) object oriented – in the open source project Object Modeling System (OMS) [David, 1997; http://oms.ars.usda.gov/; Hesser and Kralisch, 2003] models are refracted to single processes and only called when required for simulating a particular modeling exercise. The processes act on single represented objects called entities. A central kernel controls iteration in time and space and the data exchange is realised by global variables without the aid of buffer storage files. The time required for integrating new components is low but is extensive for the development of the entire system. Flexibility in the configuration of the modeling exercise is high.
3) coupling platform (HLA) – entire models are easily and rapidly integrated into the simulation environment, but efficiency may be compromised since the execution of service and support routines that are similar in several coupled models need to be repeated. Programming in the source code is required to include RTI functionality, which eliminates the need for buffer storage of data for model interaction.

In order to obtain all the advantages of both OMS and HLA environments, a HLA component may be integrated into OMS to allow communication between the RTI and OMS (see Figure 10). This HLA component would provide the necessary docking station for quick and easy testing of model's performance and capabilities.

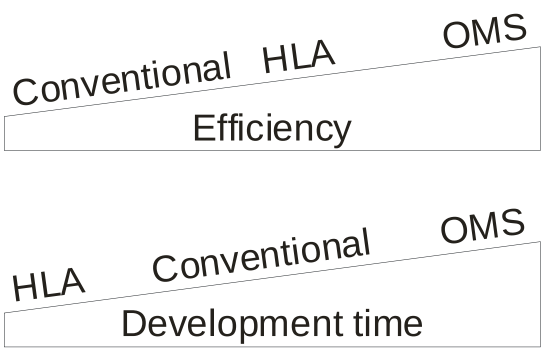


Figure 9. Pros and cons of three model coupling approaches: conventional, HLA and OMS.

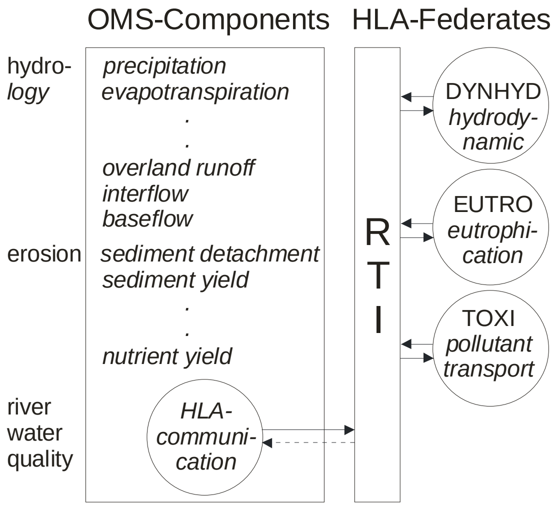


Figure 10. HLA docking port in OMS.

5. CONCLUSIONS

- The HLA platform provides a simple and fast way of coupling models together in a simple modeling system.
- Additional capabilities are extracted from modeling systems with several submodels, such as "cross-model" uncertainty analysis and submodel interaction and feedbacks.
- The RTI of a modeling system may serve as a docking mechanism to other modeling systems (such as OMS).

6. REFERENCES

Ambrose, R.B., T.A. Wool and J.L. Martin, The Water Quality Simulation Program, WASP5: model theory, user's manual, and programmer's guide, U.S. Environmental Protection Agency, Athens, GA, 1993.

Borshchev, A., Y. Karpov and V. Kharitonov, Distributed simulation of hybrid systems with AnyLogic and HLA, *Future Generation Computer Systems 18*(6), 829–839, 2002.

Chapra, S. C., Surface water-quality modeling, McGraw-Hill Companies Inc., 1997.

Choi, S.J. and W. Scacchi, Modeling and simulating software acquisition process architectures, *Journal of Systems and Software 59*(3), 343–354, 2001.

Das, S.K. and A.A. Reyes, An approach to integrating HLA federations and genetic algorithms to support automatic design evaluation for multi-agent systems, *Simulation Practice and Theory 9*(3–5), 167–192, 2002.

David, O., A kernel approach for interactive-oriented model construction in Java, *Concurrency: Practice and Experience 9*(11), 1319–1326, 1997.

European Union, the European Parliament, the Council: Directive of the European Parliament and of the Council concerning establishing a framework for community action in the field of water policy (2000/60/EC), 2000.

Hesser, F. and S. Kralisch, Das "Object Modeling System" zur Kopplung von Modellen mit unterschiedlicher Zeit- und Raumdiskretisierung, in: Hohmann, R. (ed.): *Simulationstechnik*, ASIM 2003, 439–444, SCS-European Publishing House, ISBN 3-936150-27-3, 2003.

Klein, U., Simulation-based distributed systems: serving multiple purposes through composition of components, *Safety Science 35*(1–3), 29–39, 2000.

Kuhl, F., R. Weatherly and J. Dahmann, *Creating computer simulation systems*, Prentice-Hall, Englewood Cliffs, NJ, 1999.

Lee, J.S. and B.P. Zeigler, Space-based communication data management in scalable distributed simulation, *Journal of Parallel and Distributed Computing 62*(3), 336–365, 2002.

Lindenschmidt, K.-E., D. Schlehf, C. Hesse, M. Baborowski and H. Guhr, Gewässergütemodellierung der Saale, in: R. Hohmann (ed.), *Simulationstechnik* - ASIM 2003, 297–302, SCS-European Publishing House, ISBN 3-936150-27-3, 2003.

Lindenschmidt, K.-E., K. Poser, M. von Saleski, and J.J. Warwick, Uncertainty analysis of a hydrodynamic model of a regulated river, *River research and applications*, (submitted), 2004.

Maamar, Z., Design of a simulation environment based on software agents and the high level architecture, *Information and Software Technology 45*(3), 137–148, 2003.

Petty, D., Comparing high level architecture data distribution management specifications 1.3 and 1516, *Simulation Practice and Theory 9*(3-5), 95–119, 2002.

Schöl, A, V. Kirchesch, T. Bergfeld, F. Schöll, J. Borcherding and D. Müller, Modelling the chlorophyll a content of the river Rhine – interrelation between riverine algal production and population biomass of grazers, rotifers and zebra musel, *Dreissena polymorpha*, *Internationaler Revue der gesamten Hydrobiologie 87*(2–3), 295–317, 2002.

Schulze, T., A. Wytzisk, I. Simonis and U. Raape, Distributed spatio-temporal modeling and simulation, in: E. Yücesan, C.-H. Chen, J.L. Snowdon and J.M. Charnes (eds.), *Proceedings of the 2002 Winter Simulation Conference*, 2002.

Shanahan, P. and M.M. Alam, The Water Quality Simulation Program, WASP5: model theory, user's manual, and programmer's guide - Version 5.2 – MDEP (update), Hydraulic and Water Resources Engineers, Inc., Waltham, MA, 2001.

Agents, Bayes, and Climatic Risks – A Modular Modelling Approach

Armin Haas, Carlo Jaeger[a]

[a]Potsdam Institute for Climate Impact Research
P.O. Box 60 12 03, D-14412 Potsdam, Germany
armin.haas@pik-potsdam.de, carlo.jaeger@pik-potsdam.de

Abstract: When insurance firms, energy companies, governments, NGOs, and other agents strive to manage climatic risks, it is by no way clear what the aggregate outcome should and will be. As a framework for investigating this subject, we present the LAGOM model family. It is based on modules depicting learning social agents. For managing climate risks, our agents use second order probabilities and update them by means of a Bayesian mechanism while differing in priors and risk aversion. The interactions between these modules and the aggregate outcomes of their actions are implemented using further modules. The software system is implemented as a series of parallel processes using the CIAM^n approach. It is possible to couple modules irrespective of the language they are written in, the operating system under which they are run, and the physical location of the machine.

Keywords: Learning agents; heterogeneous expectations; climatic risks; modular modelling.

1 Introduction

The experience of the Kyoto protocol shows how difficult it is to establish a plausible climate policy in the short-term (Michel [2004]). At the same time even if the Kyoto protocol had been implemented quickly and successfully, it would be far from solving the long-term challenge of climate change (Hasselmann et al. [2003]). For the future evolution of climate policy, it will be vital to develop complementary approaches to trigger a sustained process of social learning in the face of this global challenge.

So far, scientific thinking about climate policy has been mostly framed in terms of a single goal function that is assumed to be somehow appropriate for humankind as a whole. We will call such a function a Leviathan function, following Hobbes' classical view of a super-agent taking care of co-ordination problems shared by society as a whole. In research on climate policy, two kinds of Leviathan functions are currently used. On the one hand, natural scientists often assume that some quantitative measure of a stable climate can be given as a guideline for policy-making. In this spirit, increases in global mean temperature of 1/10 degree Celsius per decade have been proposed as a limit for the stress ecosystems can cope with (WMO [1988]). A total increase of one to two degrees Celsius has been proposed as an additional limit (Rijsberman and Swart [1990]). The resulting goal function is pretty simple, yielding only one of two values – 1 and 0, or 'acceptable' and 'inacceptable' – for possible trajectories of the climate system. On the other hand, economists often assume that a slightly more complex goal function is appropriate. In this case, a discounted time-integral of global GDP is proposed as Leviathan function. Climate change then is to be avoided to the extent to which it lowers GDP. Non-market effects are taken into account by monetary correction terms, sometimes including highly controversial estimates of different economic values of human lives in different countries.

In the future, it will be increasingly important to model the climate problem in a multi-agent setting where different goal functions are associated with different agents. As Tesfatsion [2003] argues, agent based computational economics has reached a degree of maturity where this seems a feasible goal. However, two kinds of challenges have to be met if useful insights are to be obtained. The first kind arises from the need to identify and represent social agents in a suitable way (section 2). The second is the need to find a reasonable representation of learning processes in the face of far-reaching uncertainties (section 3). On this basis, it is possible to gain substantial new insights on pitfalls and opportunities in the management of climate risks (section 4).

2 AGENTS

The definition of agents for the present purpose requires a balance between standard economic concepts of utility, probability, and constraints on one hand, and more AI oriented concepts of rule following on the other hand (section 2.1). The design of suitable modelling tools in turn requires a series of choices about how to implement software agents in a modular and integrated way (section 2.2).

2.1 Human agents

There are two major approaches for modelling human agents, be they individuals or organisations: maximising expected utility and rule following.

In economics, the standard concept states that agents do have preferences about the possible consequences of alternative choices. In practice (although not necessarily in theory) it is assumed that these preferences can be represented by an ordinal utility function. When confronted with an uncertain outcome from choices, agents are depicted as combining their preferences with their expectations for the occurrence of different outcomes. This is represented by a cardinal utility function combined with a probability distribution. Human agents are supposed to be able to identify and choose a single action which maximises their expected utility under constraints given by the situation they find themselves in.

In this approach, the agent's attitude towards risk can be handled in two ways. First, the shape of the utility function can be used to represent risk averse, neutral, or risk seeking behavior. Second, when maximising utility, additional constraints can be added to depict an agent's behavior towards risk. We use, for example, a non-ruin constraint stating a maximally accepted probability for going into ruin.

In information sciences, agents are often conceived as rule following entities. When confronted with a decision situation, they choose an action using a set of rules. These rules typically are if-then statements which trigger an action once a condition is met. In a decision situation, the core challenge for this approach is to come up with a single action as typically several conditions are met and the resulting set of stated actions has more than just one element. A standard mechanism to tackle this problem is to attach weights to the rules and apply the one with the highest weight. The weights themselves are adjusted in a learning process like the bucket brigade mechanism suggested by Holland [1992].

These two ways of modelling the decision processes of agents are most often perceived as alternatives. We interpret them slightly differently. Each constraint of a maximisation problem can be interpreted as a rule the agent has to obey. In the case of the non-ruin constraint applied in our work, the rule is: whatever you do, the probability of going into ruin shall not exceed a specified threshold. The expected utility then can be used to come up with a single action amongst those satisfying the rules relevant for the situation at hand.

It is important to realise that even if each agent comes up with a single action in a given situation, this by no means implies the existence of a single optimal equilibrium for the whole range of agents. This hints at the potential richness of such a perspective as it opens up the way to hybrid models achieving greater descriptive and normative content by linking standard approaches of economics and of information sciences.

2.2 Software Agents

Human agents can be modelled by suitable software agents. We have done so with the model family LAGOM (LAGOM is a Swedish word denoting a sense of balance and harmony, perhaps akin to the chinese "Tao"). One way of designing the required software agents is to look at them as functions linking specific data types. It is useful to think about the building blocks of the program in terms of a hierarchy of such data types. At the most general level, we can distinguish between perception types and action types. A human agent, then, is implemented as a function whose domain consists of one or several perception types while the range consist of one or several action types.

In order to represent the uncertainties the agent is faced with, a third data type is needed, namely expectations. Human agents form expectations by comparing a given state of affairs with their previous expectations. Expectations are a state variable of the agent in question. This is reminiscent of Turing machines, as these map a perception into an action on the basis of an inner state which is redefined at each step of their operation. Accordingly, we use functions mapping perceptions and expectations to actions and expectations:

Human Agent:
[[Perceptions], [Expectations]]
→ [[Actions], [Expectations]]

The perceptions of standard economic agents are characterised by two data types: the prices found on the market and the resources owned by the agent. Their actions relate to two more types: supply quantities and demand quantitites. Economic agents can then be further specified into firms and households of various kinds. In a typical version of LAGOM, the economic agents are producers, households, and insurance companies. LAGOM relies on a general purpose optimisation algorithm (produced by Dan Ontanu, Bukarest, and Cezar Ionescu, Potsdam) that can handle an extremely large class of optimisation problems. In the future, additional modules may implement "fast and frugal" searching algorithms needed to model choices of non-optimising agents (Gigerenzer and Selten [2001]).

Human actions have consequences, and in social interactions these consequences depend on the actions of several agents. This pattern can be modelled by a second kind of software agent representing various kinds of social and biophysical contexts. A context can be implemented as a function mapping states of affairs and actions into new states of affairs:

Context:
[[State of Affairs], [Actions]]
→ [State of Affairs]

The standard economic context is a market mapping supply and demand under given prices and a given resource allocation into new prices, traded quantities and the resulting resource allocation.

Finally, a third kind of software agent is needed to get a fully specified model. These are the program components enabling the software agents representing human agents and contexts to run in an appropriate way. They are bundled in a driver performing the following steps. First, the driver launches the other modules as parallel processes. Second, it gives to each one of them the addresses it is supposed to talk or listen to. All modules have variables for input and output ports. The driver assigns values to these (modifying them at runtime if needed), thereby making it easy to exchange modules of the same kind in a plug and play mode. Third, the driver listens to selected modules in order to relaunch a next iteration until the goal of the simulation has been reached. And finally, it shuts down the whole process.

Figure 1 shows a typical LAGOM implementation (producers have been left out to avoid overcrowding the diagram, the expectation manager will be discussed in the next section). There are two layers of communication between the different modules. On one layer, the driver performs his business as described above. On the other layer, the different modules interact in a structured manner.

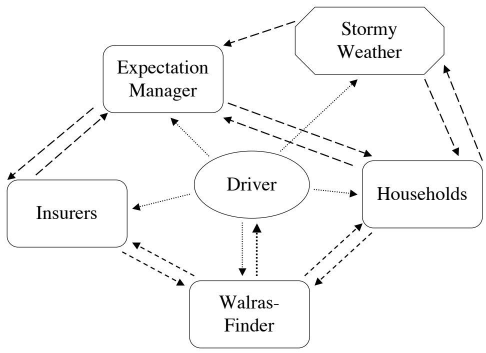


Figure 1: Scheme of typical LAGOM implementation.

In the present example, there is a climate module (dubbed "stormy weather" in the figure) that damages households and is observed by the expectation manager and is influenced by actions of households. The expectation manager helps insurers and households to up-date their respective – usually different – expectations on the basis of new observations. The Walras finder, representing the "invisible hand" of the market, finds prices matching demand and supply for weather insurance by running through a fast (hence the short dashes) sequence of interactions with households and insurers. LAGOM operates at multiple time scales. Market interchanges can be and usually are much faster than climate change. Industrial production introduces a third time scale, usually lying between those two. By running a set of such agents recursively – i.e., feeding their output of step n as input into step n+1 – one gets a dynamic system. Computer experiments with LAGOM have shown that this format can be used to represent economic agents according to standard economic theory. An economic equilibrium then is a fixed point of the recursive mapping defined above. In systems with several fixed points, non-equilibrium dynamics can obtain when the initial conditions are not identical to a fixed point. They are particularly interesting when combined with some stochastic process (Haas [2001]).

The design pattern of LAGOM is based on the CIAM^n software platform (CIAM^n is an acronym for ”Community Integrated Assessment Modules to the power of n” and refers to a software platform that we developed together with other researchers at PIK and in the broader context of the European Climate Forum – cf. the link ”CIAM^n” at <www.european-climate-forum.net>). The platform combines an understanding of software management as a social process (Jaeger et al. [2002]) with a logical analysis of the plurality of domains of discourse required for integrated assessment modelling (Jaeger [2003]). It provides algorithms for coupling modules programmed in a dynamical simulation mode with modules programmed in intertemporal optimisation mode – a key problem when combining natural science and economic models (Leimbach and Jaeger [2004]).

Using this platform requires the user to do three things: assign each piece of code to exactly one module, provide a driver module suitable for the problem at hand, and program input and output to each module according to a shared protocol for typed data transfer – TDT.

The TDT protocol (see the CIAM^n link indicated above as well as www.pik-potsdam/~linstead) provides a simple way of handling input and output to each module by means of two functions, one for writing and one for reading. These functions take addresses as parameters and work across different programming languages, different operating systems, and different machines in different physical locations. Computer experiments with LAGOM and other models have shown that drivers using this protocol can effectively co-ordinate highly heterogeneous multi-agent systems.

3 BAYES

As mentioned above, human agents often act on the basis of expectations that they try to improve in the course of action. To model such processes of expectation formation, some kind of Bayesian reasoning is needed.

A reasonable management of climatic risks is impossible without some way of making the relevant expectations of various agents – including scholars engaged in climate research – explicit. With LAGOM, we tackle this problem in two steps. First, we represent heterogeneous expectations of a variety of agents, accepting that they may have perfectly reasonable ways to entertain different expectations. Second, given these expectations, we model how learning agents can update their expectations on the basis of further experience. Whether this updating will lead to converging expectations, and if so, at what speed, then becomes a question amenable to scientific inquiry.

For this purpose, it is necessary to consider not only frequentist probabilities, but also guesses about them. Every child learns to deal with relative frequencies: it learns to distinguish fast rythms from slow ones, it learns that certain teams win more often than other ones, it learns to roll a dice, etc. Relative frequencies can be observed, they can be forecast, and they matter for many actions. The mathematics of probability measures can be used to deal with relative frequencies where a more in-depth analysis is required. Probabilities then appear as limits of relative frequencies defined over sequences of events.

Children also learn to make guesses about things they do not know, and to take action on the basis of such guesses – one may guess that a certain branch of a tree will not crack and climb on it, etc. Often, guesses have nothing to do with relative frequencies, as when one guesses what is the capital of some foreign country. But sometimes, we need to take action in the face of a situation where unknown relative frequencies matter. In these cases, guesses about frequentist probabilities occur.

The mathematics of probability measures turns out to be helpful for analysing guesses about frequentist probabilities, too. This is due to a result known as de Finetti’s theorem (see Bernardo and Smith [1994] for an exposition). The theorem shows how an unknown probability distribution can be approximated if additional samples become available step by step. For practical purposes, an important point is that the approximation can start from very different initial guesses as long as these guesses do not exclude the limiting distribution.

To apply this framework, one must distinguish between first order and second order probabilities and consider updating schemes for the latter ones. If two possibilities are given with unknown probabilities, then the relevant first order probabilities are given by the open interval (0,1). Complete ignorance then implies a distribution of second order probabilities corresponding to the probability density function y=1 over this interval.

An agent can improve upon this initial distribution under two conditions. First, she must be able to gather additional evidence. Second, she must have some opinion about the structure of the underlying process, which may be quite complex. If she is willing to consider key parameters of the process as characterized by some unknown frequentist probability, then the following updating scheme can be shown to be efficient.

$$p_{2,t+1}(p_i(s))$$

$$= \quad p_{2,t}(p_i(s)) \cdot \frac{p_i(s)}{\int_{i \in I}(p_{2,t}(p_i(s)) \cdot p_i(s))dp_i(s)}.$$

p_2 stands for second order probabilities, t for time, p for first order probabilities, P for the set of all first order probabilities suitably indexed by an index set I, p_i for a specific member of the set of first order probabilities, and s for a specific situation obtained at time t out of a set of all possible events S.

In LAGOM, this updating mechanism has been implemented in a separate module, the expectation manager. It may be used as a skeleton for representing scientific communities along with mass-media amplifying their claims, as both decision-makers and the general public up-date their expectations concerning climate change by interacting with these institutions.

The updating process is similar to the formula known as Bayes' rule. But there is an important difference between Bayes' rule as applied to a static situation and this dynamic updating process. In a static situation, Bayes' rule follows immediately from the definition of conditional probabilities. The updating mechanism introduced above can be shown to be efficient for dynamic processes with a limiting frequentist probability, but this is far from being trivial.

Therefore, it is important to notice that the model design used for the updating of expectations in LAGOM can also be applied to non-Bayesian updating schemes, as when one agent imitates the behavior of another one.

4 CLIMATIC RISKS

The various IPCC publications (for an overview see Watson [2001]) are the authoritative source on climatic risks. They show beyond reasonable doubt that humankind is altering the global climate system in ways that can lead to serious damages to humans and to things they value – by sea-level rise, droughts, floods, storms, and the like. These publications derive their authority to a considerable extent from the fact that their production is controlled by a carefully crafted consensus-building mechanism. This, however, makes it impossible to reach an in-depth assessment of these risks. There are two reasons for this state of affairs: the difficulty to reach an agreement on the value to be attached to non-marketable goods like human lives or the beauty of coral reefs, and the difficulty to reach an agreement on the likelihood to be attached to events without statistical track record.

A good example for these difficulties as well as for their importance is the Pentagon study on abrupt climate change (Schwartz and Randall [2003]). The study considers the possibility that anthropogenic climatic change will lead to a shut-down of the thermohaline ocean circulation in the North Atlantic. It does so by taking as an analog a similar shift in ocean currents that occurred about 8.000 years ago, looking at its effects then, and asking what might be the worst conceivable implications of these events in today's world. This leads to scenarios of social turmoil and international tensions, including violent conflict and dramatic losses of welfares in Europe and North America. Similar scenarios, geared to events like shifting monsoon patterns, melting permafrost, intensified El Nino events, and the like, can be produced for other parts of the world. While the available evidence is sufficient to reach agreement on the fact that these scenarios describe possible courses of events, there is no way to attach any frequentist probability measures to them; nor is there a meaningful Leviathan function providing the one and only yardstick for the evaluation of such risks.

With the model of learning agents presented in this paper, another approach is feasible: One can study the options of agents acting on the basis of heterogeneous expectations and evaluations in the face of climatic risks. In particular, one can study to what extent their expectations may converge on the basis of additional evidence and what sort of compromise they may find in order to reduce the risks about which they care most. While it is extremely unlikely that problems of this kind have a single optimal solution, it is almost certain that stepwise improvements can be achieved through a combination of negotiations and learning by doing.

Along these lines, the model can be used to study the management of climatic risks without the need to assume an unwarranted Leviathan function. This is not to say that no agent can play the role of Hobbes' Leviathan. Quite the opposite, the model can be used to study how a specific agent may acquire that role, which is then not assumed, but explained. The model can also be used to study multilateral constellations where climate risks are managed without any agent assuming the Leviathan role. Simulations show that such constellations are hardly ever characterised by a single fixed point of the relevant dynamics. Therefore, a multilateral solution of the climate problem will need to rely on some selection mechanism, most likely of the kind described by Schelling as a focal point (Schelling [1960]).

This state of affairs implies that processes of supply and demand – even if amended by instruments like carbon taxes and emissions permits – will be insufficient to bring about a solution. For this purpose, economic mechanisms and policy instruments must be complemented by suitable processes of expectation formation. While Bayesian learning is by no means the only mechanism of this kind, it provides a useful starting point for multi-agent modelling of climate risks.

5 CONCLUSION

Modular multi-agent models representing learning agents faced with climatic risks are both desirable and feasible. They are desirable because climate risks are not amenable to a purely frequentist analysis and because the use of a Leviathan function dodges the question of how to reach agreement over the valuation of non-marketable goods. They are feasible because software tools are available to couple highly heterogeneous software agents, because mathematical concepts like second-order probabilities can be used to represent learning agents, and because the scientific evidence on climatic risks is sufficient to identify salient risks. In addition, modular modelling has the advantage of facilitating cooperation between researchers with different fields of expertise, thereby improving the reliability of the resulting computations.

REFERENCES

1 Bernardo, J. M. and A. F. M. Smith. *Bayesian Theory*. Wiley, New York, 1994.

2 Gigerenzer, G. and R. Selten. *Bounded Rationality: The Adaptive Toolbox*. MIT Press, Cambridge, Massachusetts, 2001.

3 Haas, A. *Kartell und Marktprozeß*. Duncker & Humblot, Berlin, 2001.

4 Hasselmann, K., M. Latif, G. Hooss, C. Azar, O. Edenhofer, C. C. Jaeger, O. M. Johannessen, C. Kemfert, M. Welp, and A. Wokaun. The challenge of long-term climate change. *Science*, 302: 1923–1925, 12 December 2003.

5 Holland, J. H. *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, Massachusetts, 2nd edition, 1992.

6 Jaeger, C. C. A note on domains of discourse. Logical know-how for integrated environmental modelling, www.pik-potsdam.de/~cjaeger/working_papers/pr86.pdf, 2003.

7 Jaeger, C. C., M. Leimbach, C. Carraro, K. Hasselmann, J.-C. Hourcade, A. Keeler, and R. Klein. Integrated assessment modeling: Modules for cooperation. FEEM Nota di lavoro 53-2002, www.european-climate-forum.net/ciamn.html, 2002.

8 Leimbach, M. and C. C. Jaeger. A modular approach to integrated assessment modeling. *Environmental Modelling and Assessment*, in press, 2004.

9 Michel, D., editor. *Climate Policy for the 21st Century. Meeting the long-term Challenge of Climate Warming*. Johns Hopkins University, Washington, D.C., 2004.

10 Rijsberman, F. and R. Swart. *Targets and indicators of climate change*. Report of Working Group II of the Advisory Group on Greenhouse Gases, WMO/UNEP. Stockholm Environment Institute, Stockholm, 1990.

11 Schelling, T. *The Strategy of Conflict*. Harvard University Press, Cambridge, Massachusetts, 1960.

12 Schwartz, P. and D. Randall. An abrupt climate change scenario and its implications for united states national security, www.icomm.ca/survival/pentagon_climate_change.pdf, 2003.

13 Tesfatsion, L. Agent-based computational economics. ISU Economics Working Paper No. 1, Iowa State University, Revised August 24, 2003.

14 Watson, R. T., editor. *Climate change 2001: synthesis report*. Cambridge University Press, Cambridge, 2001.

15 WMO. Developing policies for responding to climatic change. WMO/TD no. 225, World Meteorological Organization, Geneva, 1988.

OpenMI:
The Essential Concepts and their Implications for Legacy Software

J. B. Gregersen[a] **and M. Blind**[b]

[a] *DHI Water & Environment, Denmark*

[b] *RIZA Institute for Inland Water Management and Waste Water Treatment, The Netherlands*

Abstract: Information & Communication Technology (ICT) tools such as computational models are very helpful in designing river basin management plans (rbmp-s). However, in the scientific world there is consensus that a single integrated modelling system to support the Water Framework Directive cannot be developed and that integrated systems need to be very much tailored for the local situation. As a consequence there is an urgent need to increase the flexibility of modelling systems, such that dedicated model systems can be developed from available building blocks. The HarmonIT project aims at precisely that. Its objective is to develop and implement a standard interface for modelling components and other relevant tools: The Open Modelling Interface (OpenMI). The architecture for OpenMI is complete and has been documented. It relies entirely on the "pull" principle, where data are pulled by one model from the previous model in the chain. There are, of course, complications in satisfying the needs for iteration, buffering and feedback loops. This paper and presentation gives an overview of the architecture, explains the foremost concepts and the rational behind them. Most importantly it will provide workable insight in the consequences for existing software. The paper contains information for water model developers about how they would put an OpenMI "wrapper" around their model for incorporation in integrated suites of models that use the OpenMI standard. This should give modellers opportunities to make their models available to a wider range of users world-wide.

Keywords: Model linking, decision support systems, interface standard, legacy model

1. INTRODUCTION

The concept of integrated catchment management has arisen because managing environmental processes independently does not always produce sensible decisions when the wider view is taken. Therefore, it becomes important to be able to model not only the individual catchment processes such as ground water, river flow, irrigation, etc, but also their interactions. However, most existing models tend to address only single issues. The objectives of the HarmonIT project address these problems through the development of an open modelling interface (OpenMI) that will facilitate easy linking of existing and new models.

The HarmonIT project will, apart from the development of the OpenMI standard, migrate about 30 existing legacy models to the OpenMI platform. We believe that the number of OpenMI compliant models will keep growing also beyond the timeframe of the HarmonIT project. Most existing hydrological decision support systems use combined hydrological models as the one of the main building blocks. Creation of such systems has so far been prerogative to the model suppliers. Now with OpenMI compliant models available also third parties can create such systems for their specific needs.

2. EXISTING MODEL SYSTEMS

Before going into detail about OpenMI some definitions of the existing model systems are given.

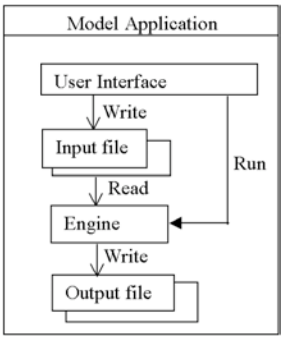


Figure 1: Model application pattern

A *model application* is the entire model software system that you install on your computer. Normally a model application consists of a user interface and an engine. The engine is where the calculations take place. The user supplies information through the user interface upon which the user interface generates input files for the engine. The user can run the model simulation e.g. by pressing a button in the user interface, which will deploy the engine (see figure 1). The engine will read the input files and perform calculations and finally the results are written to output files. When an engine has read its input files it becomes *a model*. In other words a model is an engine populated with data. A model can simulate the behaviour of a specific physical entity e.g. the River Rhine. If an engine can be instantiated separately and has a well-defined interface it becomes *an engine component.* An engine component populated with data is a *model component*. There are many variations of the model application pattern described above, but most important from the OpenMI perspective is the distinction between model application, engine, model, engine component, and model component.

3. OpenMI

Basically, a model can be regarded as an entity that can provide data and/or accept data. Most models receive data by reading input files and provide data by writing output files. However, the approach for OpenMI is to access the model directly at run time and not to use files for data exchange. In order to make this possible, the engine needs to be turned into an engine component and the engine component needs to implement an interface through which the data inside the component is accessible. OpenMI defines a standard interface for engine components (ILinkableComponent, see figure 2) that OpenMI compliant engine components must implement. When an engine component implements the ILinkableComponent interface it becomes a LinkableComponent.

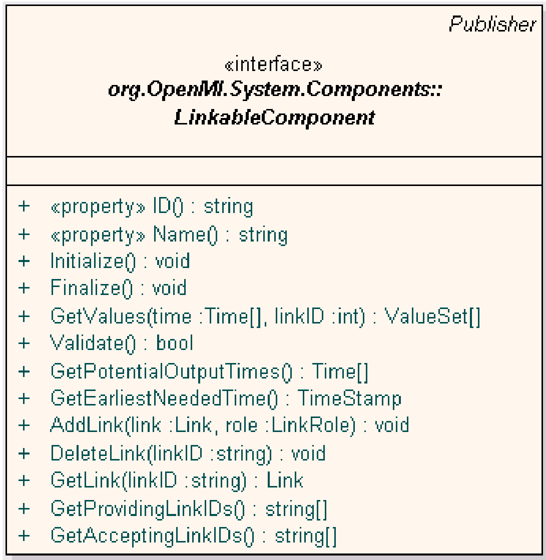


Figure 2: Linkable component interface

One LinkableComponent can retrieve data from another LinkableComponent by invocation of the GetValues method. However, this is only possible if the two components have information about each others existence and have a clear definition of the kind of data that is requested. This information is contained in the OpenMI Link class. Before invocation of the GetValues method a Link object must be created, populated and added to the two components by use of the AddLink method.

The Link object holds a reference (handle) to the two linked components. The Link object also contains information about *what* is requested, *where* the requested values apply, and *how* the requested data should be calculated. This information is included in the OpenMI Quantity class, the OpenMI ElementSet class, and the OpenMI DataOperation class, respectively. Figure 3 shows how all this information is organized into the Link class.

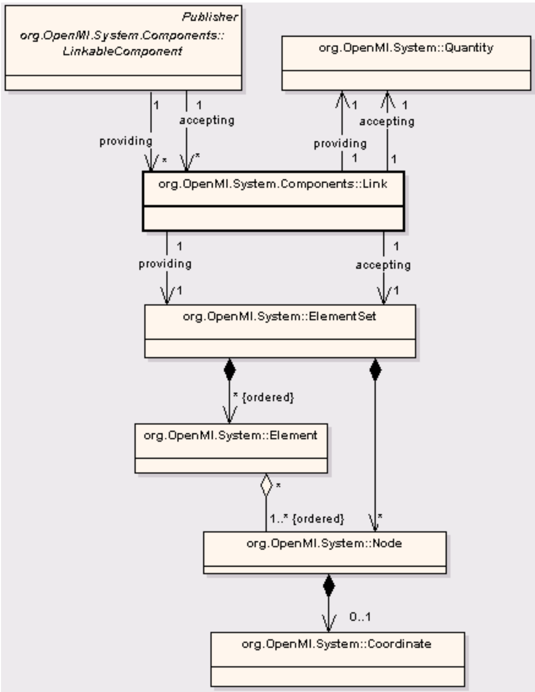


Figure 3: The Link class and associated classes

The Quantity object defines what should be retrieved. This could be e.g. water level or flow. The Quantity class represents this information simply as a text string (quantity description). OpenMI does not provide any naming convention for quantities. The Link class has an accepting Quantity object and a providing Quantity object. The quantity description in the providing Quantity object must be recognizable by the providing LinkableComponent and the quantity description in the accepting Quantity object must be recognizable by the accepting LinkableComponent. It is the responsibility of the person that configures the linked system to ensure that the combination of the two particular quantities makes sense physically.

The ElementSet object defines where the retrieved values must apply. A ground water model may be requested for either the ground water level at a particular point or the ground water level as an average value over a polygon. A river model may be requested for the flow at a particular calculation node. These locations are defined in the ElementSet. The ElementSet is a collection of Elements, where each element can be an ID-based entity like a particular node or a geometrical entity. A geometrical entity is either a point, a polyline, a polygon, or a 3D shape. The GetValues method returns a ValueSet, which is an array of values. Each value in the returned ValueSet applies to one Element in the accepting ElementSet.

The DataOperation object defines how the requested values should be calculated. Examples of data operations could be time accumulated, spatially averaged, maximum values etc. There are no OpenMI conventions for data operations. As for the quantities, the data operations are simply defined by a text string, which is recognizable by the providing LinkableComponent.

The Link class defines a specific connection between two LinkableComponents. For two specific LinkableComponents many possible links may exist.

When model links are created and populated, information about which quantities, locations and operations each LinkableComponent will support is needed. Therefore, it is required that a LinkableComponent is associated with an XML document that holds this information. These XML documents must comply with an OpenMI exchange model schema definition and they must contain information about potential output and input quantities, potential output and input ElementSets, potential output DataOperations, and the time frame for the model. This information is organized hierarchically. This means that the XML document can be queried for potential output quantities, then queried for potential output ElementSets for a particular quantity, and then queried for potential data operations for a particular quantity and ElementSet. The same type of queries can be made for input quantities and input ElementSets.

Time in OpenMI is defined either by a TimeStamp class or a TimeSpan class, both classes inherited from the abstract OpenMI Time class. A time stamp is a single point in time whereas the time span is a period from a begin time to end time. Each of these times is represented by the Modified Julian Date. A modified Julian date is the Julian date minus 2400000.5. A modified Julian date represents the number of days since midnight November 17, 1858 Universal Time on the Julian Calendar.

Now let us move back to the essence of OpenMI, the GetValues method. When one LinkableComponent invokes the GetValues method of another LinkableComponent, the providing LinkableComponent must return the values for the specified quantity, the specified time stamp or time span, and at the specified location. If the LinkableComponent is of the time stepping kind of numerical model it will do no calculation until it receives a GetValues call. Once the

GetValues method is invoked, it will calculate as long as it is necessary to obtain the needed data. Usually it will be necessary for the providing component to interpolate or extrapolate its internal data in time and space before these can be returned.

The OpenMI architecture puts a lot of responsibilities on the LinkableComponents. One of the reasons for this is that we feel that any data conversion like interpolations can be done in the optimal way by the providing component. If the providing component is e.g. a ground water model any interpolations of the ground water levels are most safely done by the ground water model itself rather than some external tool.

The OpenMI framework is very simple or you may say that there is no framework. All there is, is LinkableComponents. Once the system of linked model components is created, the invocation GetValues methods from one model component to another is driving the calculations. Since the ILinkableComponent interface (figure 2) does not have any methods that can be used to start the chain of calculations, a trigger component is needed. The trigger component is a Linkable component that has an additional method for starting calculations (see example below).

4. EXAMPLE

Let us look at a simple example: A conceptual lumped rainfall runoff (RR) model provides inflow to a river model. The populated link class is shown in figure 4.

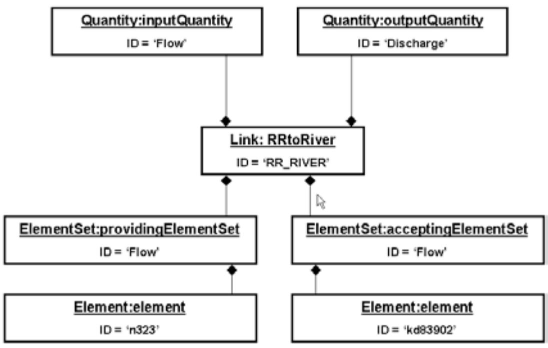


Figure 4: The polulated Link class

The sequence diagram in figure 5 shows how the calling sequence for a configuration with a river model linked to a rainfall-runoff model is set-up. The diagram demonstrates how things would look if a "hardcoded" configuration were used. For normal usage of OpenMI a configuration editor would create a configuration for you.
The sequence diagram has the following steps:

1. Each LinkableComponent must have an associated LinkableComponentFactory. The Factory class will construct the LinkableComponent. As a part of the construction the LinkableComponent object will be initialized, which typically involves reading input files. The LinkableComponentFactory object can then return a handle to a LinkableComponent object that is initialized and ready to go.

2. The Link object between the RR model and the river model is created and populated (see figure 4).

3. A trigger link to the river model is created and populated. The purpose of this link is to enable the first GetValues call to the RiverModel. This call is what triggers the calculation chain.

4. Add the trigger link to the river model component.

5. Add links to the model components.

6. Invoke the GetValues method in the River model component. The time argument defines the time for which results are expected from the river model component.

7. The river model will evaluate whether its internal time (RMtime) is before or after the requested time (t1). While RMtime is less than t1 the river model will perform time steps. The river model will, before each time step, retrieve the runoff from the RR model by invoking the GetValues method in the RR model.

8. The RR model will perform as many time steps as necessary in order to calculate the requested value. Note that the two models do not need to have matching time steps. It is the responsibility of the delivering model to do any needed interpolation in order to return a value that represents the time argument in the GetValues call.

9. When the calculations are completed the Finalize method in each component is called. Typically the components will free the memory, close the file and the like.

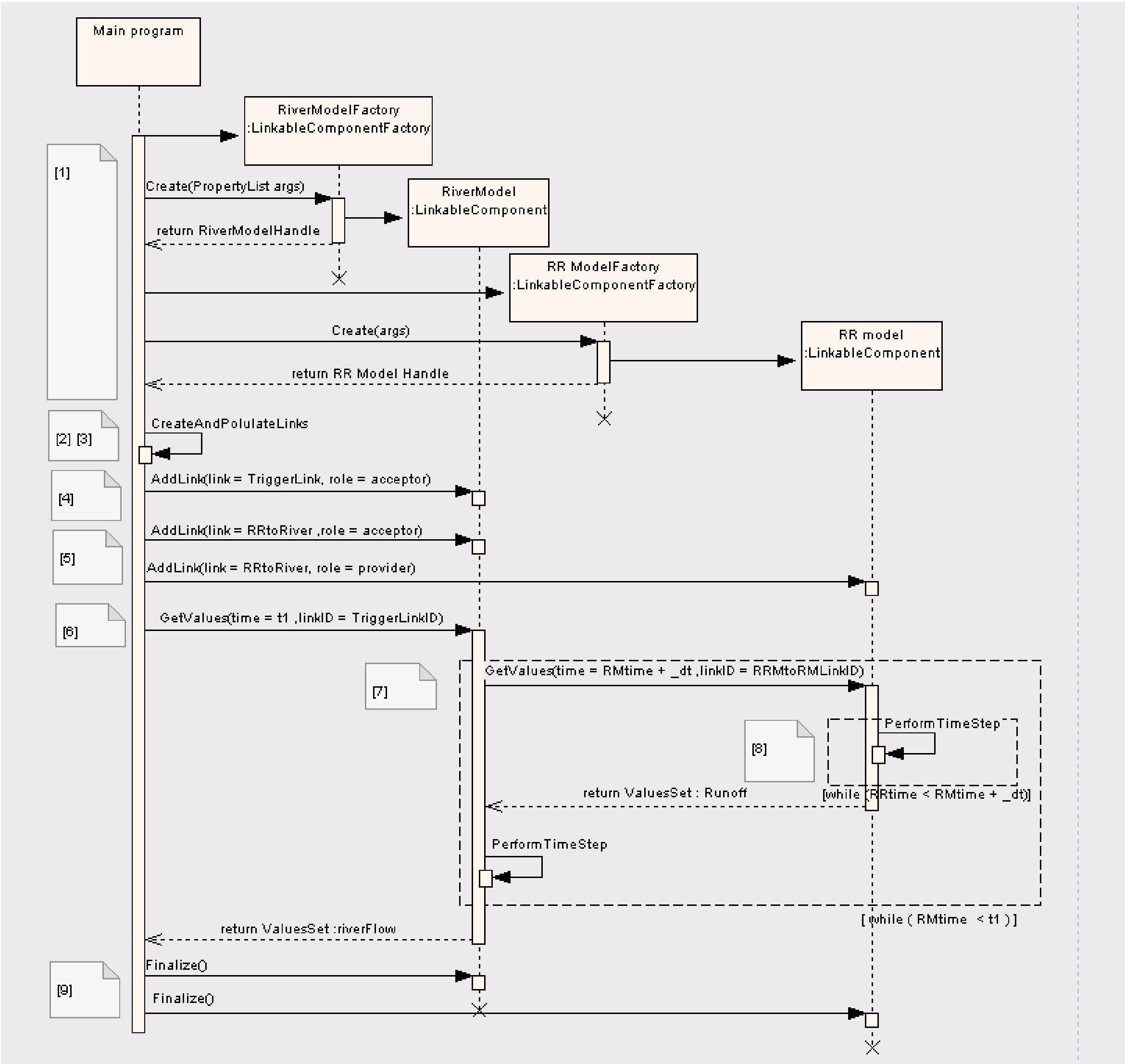


Figure 5: Sequence diagram for the example

5. OpenMI COMPLIANCE

An OpenMI compliant model must provide:

1. A component that implements the ILinkableComponent interface

2. A LinkableComponentFactory class that can construct and initialize the LinkableComponent

3. A model description XML – file that follows the OpenMI model description schema

4. An exchange model XML file that follows the OpenMI exchange model schema.

6. OpenMI UTILITIES AND TOOLS

At first hand it may seem like a huge challenge to turn a model engine into an OpenMI compliant LinkableComponent. However it is not so difficult. OpenMI provides guidelines for migration of models and a great number of software tools and utilities that will make migration easier. These tools and utilities can be used by anyone who is migrating a model but are not required in order to comply with the OpenMI standard.

For existing model engines wrapping is the recommended technology for migration. If a model engine is e.g. a numerical model programmed in Fortran this engine can be compiled into a dynamic link library (dll). If this dll is organized with a specific Win32API (see figure 6), an OpenMI utility wrapper class (SmartWrapper) can be used.

This class will take care of all the bookkeeping associated with handling links, and all the interpolation in time and space. We estimate that migration of models will take between a few weeks to a few months of work, depending on how well organized the program code is.

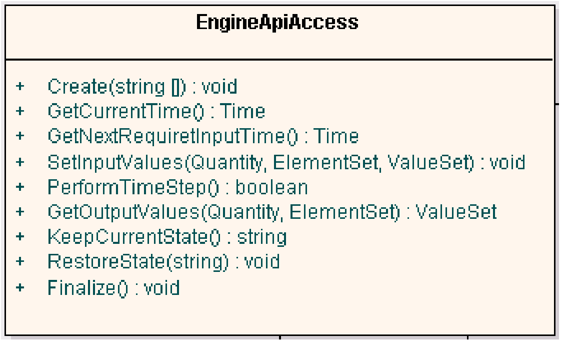


Figure 6: Required WIN32 dll API for model engines, when the SmartWrapper is used.

The SmartWrapper can be used for most numerical time stepping engines. Implementation of the KeepCurrentState and the RestoreState methods are optional since they are only used for systems where iterations are needed. So in order to use the SmartWrapper you need to make sure that your engine complies with the following:

1. Initialization must be a separate function or subroutine.

2. The time stepping loop must be broken up so that a function or subroutine performs only one single time step.

3. A function or a subroutine that can get values from the engine core must be created.

4. A function or a subroutine that can set values in the engine core must be created.

OpenMI also provides software class libraries for working with exchange models (the XML files and associated data). The configuration package will also support creation of configured systems and the deployment of these. For end-users that only need to link and deploy existing OpenMI compliant models, a configuration editor is provided.

7. CONCLUSIONS

We feel that the OpenMI architecture and associated tools and utilities will be an attractive way for model providers and model users to create systems for integrated catchment management. More information about OpenMI and HarmonIT is provides on our web site [1]

8. REFERENCES

[1] The OpenMI web site: http:\\www.OpenMI.org

A Multi-Modelling Approach to Help Agricultural Stakeholders Design Animal Wastes Management Strategies in the Reunion Island

Jean-Michel Médoc[1], François Guerrin[1], Rémy Courdier[2] and Jean-Marie Paillat[3]
[1]Cirad-Tera, Gdor Team, Saint-Denis, La Réunion, France (medoc@cirad.fr)
[2] University of La Réunion, Iremia, Saint-Denis, La Réunion, France
[3]Cirad-Tera/Inra-EA, Umr Sas, Rennes, France

Abstract: Recent concentration of indoor livestock farming in the Highlands of the Reunion Island generates environmental risks. Livestock effluent management is often difficult due to the lack of suitable spreading areas close to the farms. However, crop farms in the Lowlands are demanding organic matter to maintain soil fertility. This is the case of the locality of *Grand-Ilet* where pig and poultry effluents are intensively produced that should be exported towards the coastal zone where large areas of sugar-cane are available. This paper presents how we are supporting Grand-Ilet stakeholders to devise territorial management strategies of their wastes. To this end, a stepwise approach is being tested. First, the supply of animal wastes and the management by farmers are characterized through farm surveys and agronomic expertise. Then, the crop demand is determined by using a GIS considering a large zone surrounding Grand-Ilet onto which several constraints are applied to eliminate unsuitable spreading areas. Comparing supply and demand enables one to draw general conclusions about the strategic choices to be considered. In the case of Grand-Ilet, distances between livestock farms and crop locations, and amounts of liquid manure to be processed, led one to consider implementing a treatment plant collectively managed by farmers. To devise sustainable management strategies for such a trade-off, we are developing a participatory approach based on several simulation models (spreadsheet, hybrid dynamical system, agent-based). These models, developed in our team, aim at tackling several decision-making issues: (i) Which treatment process should be chosen? (ii) How to supply it with raw wastes? (iii) How to best organise organic material fluxes amongst farms at the district level? Together with the description of those tools, each step of the approach is exemplified and the way of integrating the use of simulation models within the stakeholders' decision process is outlined.

***Keywords**:* Animal waste management; Simulation models; Decision support; Reunion Island.

1 INTRODUCTION

The development of livestock farms in the 1980's (pig, poultry, cattle) in the Highlands of the Reunion Island was aimed at increasing the food self-sufficiency and maintaining people in difficult rural zones. Despite of its success, this policy has led to a significant concentration of effluent in these areas, particularly pig slurry. Because of the lack of suitable spreading areas, the risks of pollution are high. In 2001, the annual pig slurry production in Reunion was estimated at 162 060 m^3, i.e., 678 tons N. We focus here on the case of *Grand-Ilet*, a small locality in the cirque of *Salazie* in the north-eastern part of the island, which produces more than 15% of local pork meat. No suitable spreading areas and difficult access make this locality the most critical in environmental terms as it generates large surpluses of animal wastes. Currently, the State policy wants the stockbreeders to put their enterprises at the standards and to transform their effluents, particularly pig slurry, to comply with regulatory constraints.

Our aim is to support Grand-Ilet stakeholders to design territorial management strategies of their wastes. First, we assessed the livestock effluent production, and the nutrient demand by crops was determined by using a GIS, considering a large zone surrounding Grand-Ilet onto which several constraints were applied to eliminate unsuitable spreading areas. In order to help Grand-Ilet stakeholders to define reliable management options, we planned using several models devised for different tasks [Guerrin and Paillat, 2003a]: (i) the *Macsizut* model for evaluating and choosing slurry treatment processes, (ii) the *Approzut* model for testing various supply policies of slurry treatment plants and (iii) the *Biomas* model for simulating effluent transfers between distinct farms and evaluating the interest of collective treatment units. After presenting how we assessed supply and demand of effluents and describing the models, the paper ends with a discussion about how we are intervening in Grand-Ilet stakeholders' decision process.

2 SUPPLY OF LIVESTOCK EFFLUENTS

There are about 70 pig farmers in Grand-Ilet holding small enterprises (< 50 sows) among which two-thirds have also poultry or cattle (< 5 LU[1]). In 1998, these farms added up to 848 breeding sows and 11 000 m² of livestock buildings annually producing 17 000 m^3 of pig slurry[2], 3 300 tons of poultry manure and 2 700 m^3 of hen slurry [Renault and Paillat, 1999]. To date the pig and poultry herds are evaluated to 2 050 LU and other livestocks (cattle, sheep, rabbit…) to 170 LU. In front of this animal production, 186 ha of agricultural land only are available (75 ha as cultivated crops and 112 ha as fallow land). Hence, the global N application rate is about 871 $kgha^{-1}year^{-1}$ or, considering only the crops that can be fertilized by raw wastes (fodder, fruit crops), 2 160 $kgN\ ha^{-1}year^{-1}$. This represents 6 times the maximum amount authorized for fodder and 11 times for fruit crops. Moreover, many crop plots are unsuitable for slurry spreading because of incompatibility with market gardening, strong slopes, proximity of dwellings and rivers. The sugar-cane irrigation project in Western Reunion finds part of its resource in the *Fleurs Jaunes* stream down by Grand-Ilet. In order to preserve this resource from pollution by nitrates, the agricultural authorities summoned the farmers to standardize their enterprises according to livestock buildings and environmental specifications. From now, maintaining livestock farming in Grand-Ilet is strongly dependent upon the implementation of an efficient management strategy of animal wastes. Since 2002, Grand-Ilet farmers are trying to devise a strategy considering tradeoffs like exporting the total amount of slurry, either raw or transformed, towards the coastal zone. This requires to better know potential consumption areas and to assess their demand in terms of quality and quantity of organic material (OM).

3 DEMAND OF OM BY CROPS

The OM demand by crops was estimated for a large zone of 46 290 ha of which agricultural lands cover 12 400 ha. Within this zone we considered the 21 districts, including Grand-Ilet, used for agricultural censuses and statistics purposes. First, we identified the spatial distribution of the main crops: sugar-cane, fruit and banana crops. Then, after a review of data on NPK supply by the soil, we drew up the crop requirements in nitrogen fertilization (both organic and mineral) according to the nature of each plot's crop and its average output. We also evaluated the amount of livestock effluent specifically produced in each district. All these data were stored into the MapInfo GIS. Applying the French spreading regulations, we highlighted the suitable areas for spreading liquid and solid effluents. Finally, by merging the crop requirements and their spatial distribution, the GIS delivers for each plot the quantity of organic and mineral fertilizers that could be applied. Applying regulatory constraints thus exhibits three spreading categories for land:

- 3 197 ha can receive both liquid and solid manure (25% of agricultural land);
- 7 076 ha can receive only solid manure (58% of agricultural land);
- 36 017 ha are not suitable for spreading (of which 2 119 ha of agricultural land, i.e. 17%).

Transferring livestock effluents from Grand-Ilet can be envisaged if the authorized spreading area of the coastal zone is not saturated by other sources. For each type of effluent (liquid or solid) we calculated the balance between supply and demand of N at the district scale. A negative balance denotes a deficit of OM to fertilize the crops, a positive balance surpluses that need be exported. For liquid effluents, the balance is highly positive for all the districts. This highlights the fact that the suitable area for liquid manure is too cramped to absorb all the slurry produced in the zone and, thus, does not allow a transfer of raw slurry from Grand-Ilet. Conversely, the solid manure produced in the zone does not saturate the area (Figure 1).

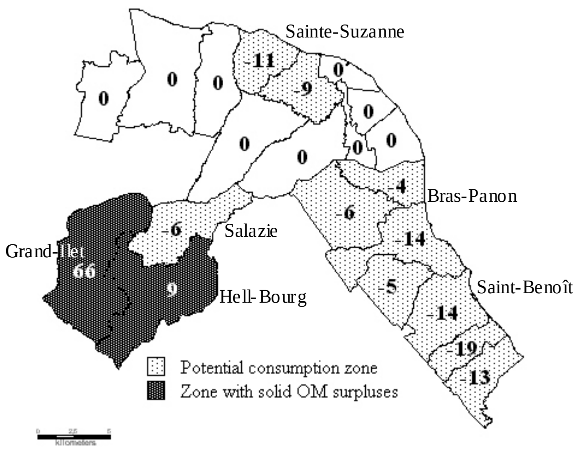


Figure 1. Nitrogen balance for solid effluent in tons and by district.

[1] LU: livestock units.
[2] However, a census made by the agricultural authorities in 2003 consider a pig slurry production of about 22 000 m^3year^{-1}.

The current surpluses for the zone, located in the districts of Grand-Ilet and Hell-Bourg, amount up to 75 tons N whereas the potential demand is 101 tons N. This shows that an additional amount of 26 tons N of solid effluent could be spread over the zone. Thus, a pig slurry treatment plant possibly implemented in Grand-Ilet could reasonably export at most this quantity as solid organic matter.

4 CHOOSING SLURRY TREATMENT

Macsizut is a spreadsheet model devised to calculate matter balances and assess investment and running costs of pig slurry treatment plants [Guerrin and Paillat, 2003b]. Given the situation in Grand-Ilet and taking into account farmers' wishes, the goal of treatment should be to produce easily transferable solid products (composts, fertilizers) with sufficient nutrient content to give them a commercial value and a liquid phase treated enough to be directly disposed in the environment. Figure 2 shows the characteristics of the liquid phase issued from the treatment of the pig slurry produced in Grand-Ilet according to various treatment processes (TP; see appendix).

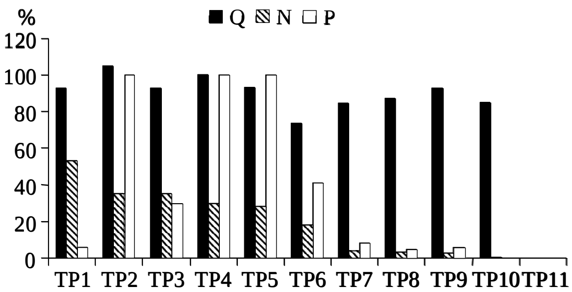


Figure 2. Quantity (Q) and quality (N, P) of liquid effluents according to various treatment processes (TP) as % of initial slurry.

Putting aside TP11 (i.e. composting, which generates no liquids), the liquid phase to manage after treatment is between 75-100% of the input slurry with N and P contents respectively ranging within 0-53% and 0-41%. Whereas TPs 2, 4 and 5 have no efficiency for P removal, TPs 7-10 are rather efficient in terms of N and P abatement. Of these four processes, TP10 only allows the liquid phase to be directly released in the environment because its residual NP content is null. The other three processes (TPs 1, 3 and 6), due to relatively high NP contents in the liquid phase, would necessitate crop irrigation to be disposed off.
From an economic viewpoint, Figure 3 compares the treatment costs of Grand-Ilet pig slurry according to the TPs. They vary between 11 and 32 € per m^3 of processed slurry. In the case of TP10, energy cost (8 €m^{-3} *vs.* 1 to 4 €m^{-3} for the other processes) increases the total cost of treatment (32 €m^{-3}).

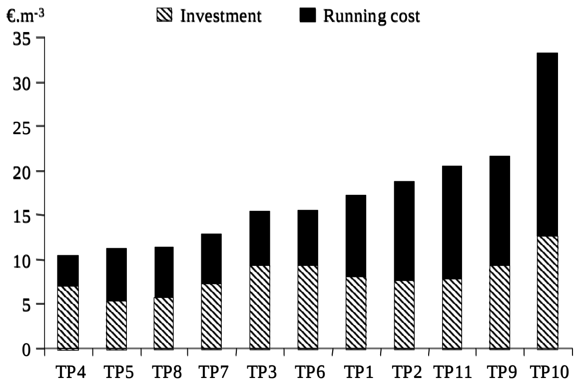


Figure 3. Investment and running costs for pig slurry treatment in Grand-Ilet.

According to these results, the stakeholders hesitated between TP10 (best treatment efficiency despite of a high cost) and TP8 (reasonable cost/efficiency compromise). They finally adopted TP8 to undertake feasibility studies.

5 TESTING SUPPLY POLICIES OF A TP

Our aim is to coordinate the deliveries made by various farms producing pig slurry at different rates so that the TP be adequately supplied and no stock overflows occur. For this, we need to implement policies capable of regulating the stocks. These policies must address the basic questions in stock management: Who should supply? How much? When? To answer these questions, we may use a simulation approach that we intend to test on the case of Grand-Ilet. We illustrate here the approach by defining three tentative policies:

- A *Rigid* policy, so that a delivery plan is imposed on the farmers that must deliver fixed quotas all on the same date.
- A *Flexible* policy, so that multiple deliveries are possible for a single quota and delivery dates are not the same for all the farmers.
- A *Free* policy, so that, at any instant, farmers may independently initiate multiple deliveries of their quota.

The *Approzut* model [Guerrin and Ranaivosolo, 2001] enables one to simulate such policies and optimise the control parameters to achieve our aims. The system is modelled as a set of farm stocks connected to a single TP stock. Each farm stock is supplied by its own animal waste production and emptied (i) by the export to the TP and (ii) by possible overflows. Integrating the difference between inflow and outflows over time gives the evolution of the stock level. Similarly, the TP is modelled as a stock with inflow (sum of exports from the farms delayed by transportation time) and outflows, the rate at which the treatment process consumes the stored slurry and stock overflows.

Taking as main input the delivery period (i.e., time lapse during which at least one delivery is made) determined according to the policy to be simulated, the model calculates for each farmer a delivery quota equal to the quantity of slurry produced during this period. A "clock" function is then used to determine the exact delivery dates within each period. To take into account possible risks, random functions are used to introduce noise on both the amount of slurry delivered to the TP and its consumption rate. In the Free scenario, a random function is also used, in addition, to simulate arbitrary choices made by each farmer on their own delivery periods. In order to detect possible long-term effects, simulations are performed over 10 years.

First, simulating the system without introducing noise shows that no stock overflow occurs for neither simulated scenario. For the TP, it appears that the Flexible and Free scenarios better regulate the stock than the Rigid one (Figure 4A). Note that, as all the deliveries are concentrated in time, the Rigid scenario exhibits wider and sharper fluctuations than others where deliveries can be made at different dates.

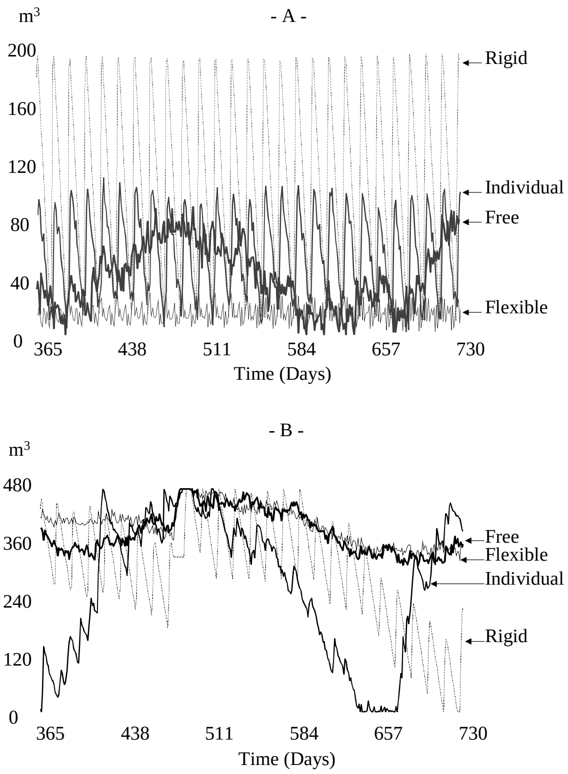


Figure 4. Evolutions of the TP stock without (A) or with (B) random noise for the three basic scenarios plus an "Individual" new scenario.

With farmers' perspective, the result is more varied: for one farm the Free scenario should be preferred, for two the Rigid scenario gave best results while the Flexible was best for the other two. To reconcile both individual and collective viewpoints, a simulation based on the best-fitted policy for each farmer gave a very acceptable result for the TP ("Individual" scenario; Figure 4A).

Whenever noise is introduced, the results are quite different (Figure 4B). The stocks are not as well regulated and overflows occur both at the TP (for an average of 150-200 m^3 per year), but also to a lesser extent at some farmers'. However, a similar conclusion is drawn: the Flexible scenario is best fitted with the TP dynamics whereas, from an individual point of view, it would be necessary to apply individual scenarios for each farmer. But in that case, this solution has a catastrophic effect on the TP: there is a very large overflow (275 $m^3 year^{-1}$ in average) and the stock levels fluctuate widely between excess and shortage (Figure 4B).

Starting from the Flexible scenario (with noise for greater realism) with an optimisation procedure minimising stock overflows, we can devise a new solution. Compared with the 'regular' Flexible scenario this scenario has the following features: farmers may triple the amount to be delivered if they can, the TP consumption rate is doubled, and feed-back control is introduced to prevent all deliveries from farmers when the TP stock is nearly full (e.g. 90% of maximum capacity). This optimised solution (close to 'just-in-time' policies) provides us with the following benefits (Figure 5):

- No stock overflows;
- All stock levels are minimised except for short periods, due to random interruptions in the consumption of the TP (see the plateau on figure 5) which impose to stop farmers' deliveries;
- Storage capacities can be reduced by 67% for the farmers and 90% for the TP (see figure 5).

This solution is likely to provide economical gains and fits our management goals.

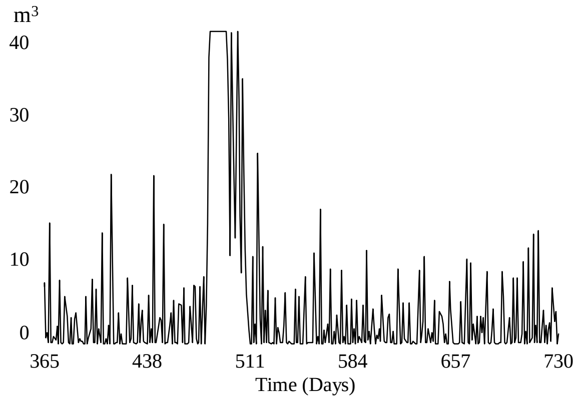


Figure 5. Evolution of the TP stock in the optimised Flexible scenario with random noise.

This example illustrates how such a simulation approach may enable new solutions to be derived from experimenting with preconceived policies, in a way that may be crucial to support the decision process of stakeholders. This is performed, here, by highlighting the importance of accounting for risk in the system's operation and introducing some feedback mechanism to improve the system's robustness.

6 SIMULATING EFFLUENT TRANSFERS AT A TERRITORIAL SCALE

Here, our aim is to simulate possible transfers of organic material between farms and test organizational alternatives open to agricultural stakeholders at a territorial scale. For this, we devised a multi-agents system, called *Biomas*, developed as an application layer of a simulation platform called Geamas [Courdier et al., 2002]. In Biomas, a territory is viewed in regard to waste management. It is composed of several livestock farms, crops, treatment plants, transportation and storage facilities. Livestock farms create a supply of OM and crop farms create a demand that may be met subject to quality, quantity and availability conditions of OM. The process enabling the matching of supply and demand is called a negotiation. Any negotiation can be initiated either by supplier or client-user. It may result into a transfer of organic material from the place of storage (livestock buildings) to the place of use (crop plots) if and only if a suitable shipper can be found.

Three abstraction levels are implemented in Biomas: micro-level describing atomic agents; medium-level describing intermediate structures of organisation; macro-level describing the whole system.

At the micro-level two types of agents are differentiated. The first is composed of the Farmer agents who possess the capacity of negotiation. Their function is defined by the composition of roles of OM producer, consumer, and shipper. Acquaintanceship networks determining potential ways of transfer may link them. The second type is made of physical entities like Crop, Livestock, Type of shipping, Storage facilities, Treatment plants. These agents are subordinated to the Farmer agents to transmit an alarm, respond to inquiries, and execute actions. Satisfaction degrees are defined for Crop and Livestock agents and come into play as criteria in the conclusion of a negotiation between Farmer agents: the satisfaction of a Livestock agent resides in its capacity to export its OM stock when a given threshold is reached; the satisfaction of a crop is to cover its needs in OM specified by three attributes: nature of acceptable OM, quantity to be applied and period of application. The TPs are represented as mass balance models describing, for each type of treatment process, the outflows of the different products in relation to the inflows of untreated OM. These physical agents are represented in space by 'situated objects' characterized by their position relative to some map. This enables OM surpluses or deficits to be spatially located and distances separating Crops from Livestock to be accounted for in farmers' negotiations.

At the medium level, are specified the Group agents, representing collective management structures like farmer associations aimed at treating animal wastes. Each farmer can belong to one or several groups. A group can impose constraints upon its members (e.g., delivery of specified quotas of OM to the TP it manages).

We are intending to use Biomas to evaluate the flux of OM produced in Grand-Ilet in anticipation of the creation of the collective TP and test some management parameters (treatment process, supply, product destination). Such simulated scenarios may involve several hundred interacting "cognitive" agents (i.e. with reasoning abilities) representing the entities interacting within the system.

7 SUPPORTING STAKEHOLDERS

Our goal is to support Grand-Ilet stakeholders to devise sustainable management strategies of their wastes based on the use of simulation models. The main issue is: How?

First, an organisation framework with three levels (action, coordination, decision) was set up to gather the various stakeholders concerned by the issue of pig slurry treatment.

At the first level, four action groups (AG) were organized to discuss and elaborate tentative solutions on specific topics: implementation of the treatment plant, process choice and management, legal status of the organisation and administration, biogas unit to be coupled to waste treatment. These action groups gather farmers, agricultural advisers and representatives of farm cooperatives, agricultural services, city council and, when needed, any kind of experts. The main results issued from the groups are presented to other participants via the coordination group. The decision committee, gathering farmers, cooperatives and private investors representatives, endorses all the final decisions.

It is within the AG2 that simulation models are being used and discussed. To date, our intervention was about the choice of a treatment process for pig slurry using the Macsizut model. A first simulation was carried out to launch the decision process, comparing all the TPs on criteria of efficiency and costs, whose outputs (diagrams in Figures 2 and 3) were discussed among the stakeholders. This resulted in a first shortlist (TPs 7-10). Then, a study trip in French Brittany (where the environmental impact of pig farming is very critical) was organized with some pig farmers and their technical advisers. This trip allowed the participants to meet institution representatives, engineers, manufacturers and farmers experienced in pig slurry treatment and to get elements of method to apprehend globally the multiple aspects of their project: choice of a TP, supply and management modes, destination of output products, organisation rules, etc. Finally, the choice of TP8 was made after retrospective discussions about the trip among all the farmers of Grand-Ilet, based on slide presentations of the visits and detailed cost analysis of each process.

Concerning the use of the Approzut and Biomas models (that have not yet been used with the stakeholders) we intend to follow an iterative loop:

1. Test the models on generic farm types [Guerrin and Paillat, 2003b] representing the full range of OM management situations found in Grand-Ilet and devise tentative management strategies to be discussed with the stakeholders in order to build a common representation of the main issues to be dealt with.
2. Run the models on real situations and discuss the simulation outputs with the stakeholders to make operational decisions.
3. If necessary, adapt the models to fit them with unexpected situations or specific stakeholders' demands.
4. Return to steps 1 or 2.

This iterative approach based on frequent exchanges between the model development and its use is a central point of our methodology [Courdier et al., 1998].

8 CONCLUSIONS

The originality of this (still ongoing) project lies in the use of multiple simulation models devised to deal with complementary aspects of a complex problem. These models enable one to test animal waste management strategies at a territorial scale with agricultural stakeholders of different types (farmers, technical advisers, policy makers and researchers). If decision-making is obviously one of the main goals, models by themselves are not conceived to prescribe decisions. They rather are intended to be used as tools to explore, test, and iteratively devise management strategies. In keeping with other approaches (e.g. the Farmscape project by CSIRO in Australia), model usefulness is expected in fostering discussion and learning among stakeholders [McCown, 2002]. Model validation is thus conceived as checking stakeholders' satisfaction in using models, directly or indirectly, rather than mimicking the 'real' behaviour of some sets of variables. The open issue is to assess the place models should take in stakeholders' decision processes. This is our main research goal to be achieved with Grand-Ilet stakeholders.

9 REFERENCES

Courdier, R., F. Guerrin, F. H. Andriamasinoro, J.-M. Paillat, Agent-based simulation of complex systems: application to collective management of animal wastes, *Journal of Artificial Societies and Social Simulation, 5*(3), 2002.

Courdier R., P. Marcenac, S. Giroux, Un processus de développement en spirale pour la simulation multi-agents, Revue L'Objet, 4(1), 73-86, 1998.

Guerrin, F., P.W. Ranaivosolo, Managing the supply of a slurry treatment plant by means of a hybrid dynamical system, paper presented at the 2nd IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications, Bali, Indonesia, August 22-24, 2001.

Guerrin, F., J.-M. Paillat, Modelling biomass fluxes and fertility transfers: animal wastes management in Reunion Island, paper presented at Modsim 2003, International Congress on Modelling and Simulation, Townsville, Australia, July 14-17, 2003a.

Guerrin, F., J.-M. Paillat (Eds.), Modélisation des flux de biomasse et des transferts de fertilité - Cas de la gestion des effluents d'élevage à l'île de la Réunion. Séminaire Cirad, Montpellier, France, CD-rom, 2003b.

Mc Cown, R.L., Changing systems for supporting farmers' decisions: problems, paradigms and prospects, *Agricultural Systems, 74* (1), 179-220, 2002.

Renault, D., J.-M. Paillat, Analyse de la production et de l'utilisation des effluents porcins à Grand-Ilet, localité de l'île de La Réunion (cirque de Salazie). Cirad-Tera Tech. Report #16/99, Réunion, 1999.

10 APPENDIX: PIG SLURRY TREATMENT PROCESSES

TP1: coagulation, flocculation, pressing to concentrate N and P in a solid filtrate.

TP2: centrifugation, N concentration by stripping and acid washing.

TP3: centrifugation, N catalytic combustion.

TP4: NDN (= nitrification-denitrification).

TP5: coagulation, filtration, NDN.

TP6: filtration on straw, NDN.

TP7: filtration, NDN on biofilters.

TP8: centrifugation, NDN.

TP9: coagulation, flocculation, filtration, NDN.

TP10: dehydratation on scraped discs, N concentration by stripping and acid washing.

TP11: slurry composting on straw.

General requirements for LCA software tools

N. Unger, **P. Beigl and G. Wassermann**
Institute of Waste Management, BOKU – University of Natural Resources and Applied Life Sciences, Vienna, Austria

Abstract: In recent years Life Cycle Assessment (LCA) software tools have become increasingly important. Today a large number of LCA programs are available. The large diversity of LCA software tools on offer makes it necessary to first pinpoint some general requirements that determine the quality of an LCA software tool and secondly to describe these requirements qualitatively. The authors last year assessed four different software tools by using each of these programs to model waste management scenarios . This modelling process revealed many differences in the quality of the LCA software tools. These differences are unrelated to the kind of modelled scenario and therefore are relevant for all kinds of LCAs (e.g. product LCA). Using these experiences and by conducting additional research in the literature, we have deduced some general requirements considered essential for good quality LCA software tools. These will be presented in this paper.

Key words: LCA, software, features, quality

1 INTRODUCTION

Life Cycle Assessment (LCA) is today an often used method for assessing the potential environmental impact of products, services or proceedings. Software tools were developed to make the processing and calculation of LCAs easier. The first steps were taken about two decades ago, with the main focus often on the assessment of production processes. Over time LCA software was also applied to other fields such as waste management.

There is a large variety of LCA software tools on the market. The foremost – and for the potential user also often prohibitive – property of a software tool is the price. The price of an LCA software tool can vary between several thousand euros and free of charge. Some tools offer a wider range of features than others. Some are focussed on a specific field of LCA, e.g. LCA in waste management, while others try to cover different application fields of LCA. Also the data and data quality can have an effect on the price of a software tool.

Depending on the purpose for which the user has selected the software, different LCA tools are more suitable for particular applications. However, a number of properties and features are essential for any good quality LCA software tool regardless of the kind of user and the kind of LCA it is being used for. This paper will discuss which features are important and which requirements are desirable for a good LCA application.

The content of this paper is mostly gleaned from experiences with four software tools for LCAs in waste management, as well as from test versions of other LCA software tools.

2 WHO USES LCA SOFTWARE TOOLS?

Different groups of LCA software users can be distinguished. The first group includes scientists and researchers. Users in this group are often experienced with LCA and have a good knowledge and understanding of the context and the features of the LCA method. Thus they make high demands on LCA software tools: They need a flexible software tool that enables them to model "common" often-modelled scenarios as well as scenarios that diverge from the standard. Also the tool should support modelling of complex process chains. The provided data need to be of good quality (see 4.3) and adequate, particularly because, in contrast to business users, scientists usually do not have their own data. It should be possible to create new data sets. In addition, scientists need the freedom to make their own improvements and modifications to existing data, specifications and parameters.

Industry, on the other hand, uses LCA software to improve its environmental performance, for process optimisation and product development. The users want “ready-to-use” software, where many of the specifications are already pre-set with only a few parameters needing to be determined.

Also decision makers use LCA to compare different solution options and hence also LCA software tools. Decision makers generally want an easy-to-understand presentation of the results in terms of which option is the best.

The developers of LCA tools aim to serve both groups of users: scientists and practical users from industry. It is very expensive to develop a software tool and thus it can only pay off when it is sold to the widest possible audience [Rizzoli and Young, 1997].

Not all of the mentioned requirements need to be fulfilled by a software tool in order to be acceptable to a specific user group.

3 WHY ARE LCA SOFTWARE TOOLS USED?

Environmental processes are often very complex and convoluted. This makes it difficult to model an LCA. Additionally LCA is often data intensive. Computers and adequate software tools are thus used to support the user in managing and editing these amounts of data. LCA software further helps to structure the modelled scenario, displaying the process chains and presenting and analysing the results. LCA software tools can be used whenever the method of LCA is applied.

The main reason for using LCA is to calculate the environmental aspects and potential impact associated with a product (ISO 14040). Also environmental hot spots (processes that have a large impact on the environment) can be identified. A more environmentally-friendly production process can thus be developed where they are most effective. LCA can also be used for a cleaner approach to production. It can help to improve and optimise resource management, which leads to a more efficient use of materials and energy.

LCA therefore is used mainly for comparing different options and for deciding which option is best for the environment. LCA and LCA software are thus used as a support tool in decision taking.

4 EXAMPLES OF TECHNICAL AND METHODOLOGICAL REQUIREMENTS

People who wish to use an LCA software tool often face the dilemma of which tool is best for their purposes. There are some software comparisons available that can help (cf. Jönbrink et al., 2000; Frühbrodt, 2002; Unger, 2003). An overview of some properties and features of commercial LCA software tools is provided here. Additional desirable features are pointed out. These can be seen as general requirements that need to be fulfilled by a good LCA software tool.

4.1 Structure and display of processes

A software tool generally consists of a database and a modelling module. The data are handled and modelled on an interface.

The modelling consists mainly of connecting successive processes with material flows. They build the process chain. Each process represents a stage in production and is defined by its input and output (see 4.4). The output from a preceding process builds the input for the next process. Simple process chains can be modelled in one layer. To handle more complex process chains, a hierarchical structure, as displayed in Figure 1, is needed. The main process stages, e.g. extraction, production and disposal, are modelled in the top layer. Each of these stages can be specified more exactly in their own sub-layer. Thus very long and complex processes can also be modelled and displayed in a clear way.

In assessing the life cycles of products the main focus is often placed on the output. The main question is: How can a certain amount of output (product) be produced with a minimum environmental impact (output-orientated calculation)? However, to assess other proceedings, other approaches are more appropriate. For example, in waste management the question “How can a certain amount of waste be treated with a minimum of environmental impact?” is of importance which is an input orientated approach. Good software tools offer the possibility of orienting the calculation towards any process within the process chain.

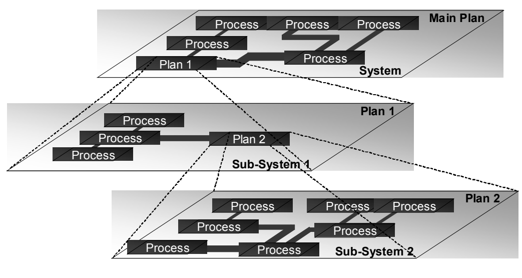


Figure 1. Schematic figure of a hierarchical structure [IKP, PE, 2002].

Some output-oriented software tools allow only one output of a process for the follow-up. Other outputs (by-products) are then addressed as negative inputs, which cannot be followed in the same process chain.

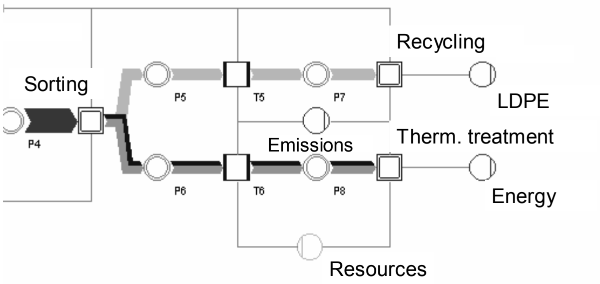


Figure 2. An example of a process chain. (Umberto)

But the user will sometimes encounter process chains with more than one output. Good software offers the possibility of following up different outputs. A simplified example from waste management is given in Figure 2. A sorting process has two outputs: LDPE films, which should be recycled (upper stream), and waste and other plastics that are designated for incineration (lower stream; the two colours indicate two different materials). The software enables the user to continue the process chain using both outputs.

4.2 Transparency, flexibility and user-friendliness

The structure of a software tool is partly responsible for its transparency. The calculation modus is important for a transparent compilation of an LCA. The user should be able to trace back each result in order to find mistakes. The proceedings from the modelling should be chronological and logical. Of vital importance in this context is the user interface, which should be clearly structured and self-explanatory. Modelling the process chain on a graphical interface is a very transparent way of modelling. There processes are arranged (e.g. as plan, network, assembly) and connected with material flows (arrows). The drag-and-drop feature is very helpful in this context.

To improve the user-friendliness many user interfaces have been designed similar to MS Office applications. The user will feel familiar with a number of features from the start. This improves the working quality of the software.

The user will often have to present his results to different groups of people (such as purchasers or the scientific community). Thus the software tools should offer a good presentation toolbox. A sankey diagram (see Figure 2) is a good option for presenting a process chain. The hierarchical structure can help to present the results clearly. Also it should be possible to create diagrams. A uniform layout for printed reports can enhance the software quality. Further the compatibility with other applications such as MS Office is important.

The implementation of a documentation feature is recommended to comply with the ISO 14040 standard, which defines the LCA.

The user should feel comfortable using the software tool. Little features, such as the possibility to change the entry unit, a zoom function for modelling the process chain or the possibility to show the input/output inventory in different gradations of detail, can make modelling more convenient for the user.

It is not easy for software developers to comply with often contradicting features such as user-friendliness and flexibility. Eriksson et al. [2002], for example, state that their software should be seen as a service rather than a computer program. Further they point out that they are continuously working to enhance user friendliness without losing flexibility. This is true for many software tools.

An important aspect in the context of user-friendliness is the time needed to learn how to use the software. The amount of time that has to be invested should be appropriate relative to the level of detail of the LCA.

4.3 Database

Data should be stored separately from the modelling module. It should be created and managed in some kind of database or library. This storage base has to be structured clearly. Very convenient for most users is a database structure similar to the one in MS Windows Explorer, where data can be edited without working in a modelled scenario. Also the import from and export to other applications is easier.

Apart from processes and flows, a database also contains modelled process chains. It should also be possible to file sub-layers in a process chain. They can be reused to model other scenarios. Further it should be possible to file separate data for a specific project so that the user does not need to search the entire database when looking for a specific process.

The data in the database need to be of good quality. They should be up-to-date and from a reliable source. More than one source is desirable in order to limit the danger of making mistakes. The user needs to clearly define the conditions under which the data are valid as well as the region for which they can be applied (e.g. energy for different countries). It can be helpful to include a data quality index to indicate the level of data quality.

An automatic update should be provided as soon as new data or data of better quality are available.

Good quality data should contain following information in the documentation:

- original data source
- age of the data
- composition of the data (number of companies or different literature, where the data are generated).

The user, particularly the scientist, will often use data from the database as well as from his own generated data sets. Processes and especially material flows have to be named carefully. Problems occur if different entries are created for one flow. For example, a process from the database produces the output “CO2.” Then the user creates a new process with the output “carbon dioxide.” The result is that two different names stand for the same flow. A feature that defines these flows as equivalent is necessary. This is especially important when the user creates his own data and the valuation. A very user-friendly way of communicating that two names stand for the same flow is to define synonyms.

Sometimes the user may want to connect processes where the output and the successive input are different. This should also be possible. An example of this is when the output of a process is “miscellaneous plastics” and the input in the next process is “waste” (plastics).

4.4 Calculation methods, uncertainty and variability analyses

Software tools offer different options for defining the proportion of inputs and outputs of a process. The simplest is to define a mass balance, e.g. the inputs are 1 kg of A and 2 kg of B and the output is 3 kg of C. However, mass balances are usually insufficient. Linear equation systems are an adequate way of modelling processes most of the time. Some tools also offer scripts, enabling the user to calculate non-linear systems like iterations.

Up to now LCA software tools have not usually considered the factors of uncertainty and variability. This refers mainly to parameter uncertainty (e.g. inaccuracy of emission measurements or of normalisation data) as well as the variability between sources (e.g. different emissions of comparable processes) and objects [Huijbregts, 2001].

The spectrum of tools to deal with these potential distortions ranges from simple parameter variations and sensitivity analyses to sophisticated methods, such as fuzzy logic computations, Bayesian statistics or probabilistic simulations.

In particular, simulations based on statistical modelling methods seem to be a promising technique for making uncertainty operational. Two approaches – the Monte Carlo and Latin Hypercube simulations – are currently implemented in LCA software tools [Weidema and Mortensen, 1997].

To perform the Monte Carlo simulation, the uncertainty distribution (normal or rectangular are usually available) of each parameter has to be specified. All the parameters vary randomly within the limits of the given distribution. The randomly selected values are inserted in the output equation. After repeated calculations, the output is represented by a predicted distribution of each output parameter. The Latin Hypercube simulation works in similar way. The main difference is, that the uncertainty distribution of a parameter is segmented in a number of non-overlapping intervals with equal probability. This fact leads to generally more precise random samples than the Monte Carlo simulation [Huijbregts, 2001].

In LCA practice the application of these methods is useful for assessing the influence of the parameter uncertainty on the uncertainty of the model output. The most important consequence of such analyses is the identification of parameters that cause a large spread in the model output. This can help to increase the accuracy of the overall model.

4.5 Methodological Properties

For waste management questions LCA normally leads to the comparison of different treatment options for waste streams with a reference scenario (e.g. landfilling) that provides a functional equivalence. This equivalence can be achieved either by given credits outside of the system or by expanding every system to achieve the same benefits. To use the LCA software tool comfortably it is necessary to provide both methods (the “credit method” and the “basket-of-benefits” method). Especially complex scenarios cannot really be addressed with the basket-of-benefits method. If only “credits” can be provided by inverting existing primary production processes, the assessment will not be comfortable, because outputs are shown in the input table and the other way round. In a good software credits are automatically subtracted from the outputs.

At the international level two impact assessment methods have been established and are most commonly used in Life Cycle Assessment: an operational guide to the ISO Standards (CML 2001 method [Guinée et al., 2001] and Eco-

Indicator 99 [Goedkoop et al, 2000]). Less often used methods, particularly in the German language area, include the Swiss Eco-factors 1997 [BUWAL, 1998] and the German Federal Environmental Agency (UBA) method [UBA, 1999]. The software should at least provide both internationally used methods because they follow different general approaches: problem-oriented methods (CML) and damage-oriented methods (Eco-indicator).

Especially when the CML method is used for the impact assessment, the software needs to provide another aid for interpretation of the results. Weighting the results according to their relative importance often is necessary for the results interpreter. One possibility for results aggregation is normalisation, where calculating the magnitude of indicator results relative to reference information is possible. The software should provide different normalisation parameters.

In general a different quality of results should be given, e.g. a thorough inventory, different valuation results, aggregated values of different impact categories, or a summarisation to just one parameter to afford a ranking of options.

4.6 Service and Support

Service and support are very important aspects of LCA software and should not be underestimated. Software needs continuous maintenance.

The database especially needs a great deal of attention to keep it up to date. The software should also be continuously improved to eliminate malfunctions and improve user-friendliness and software ergonomics.

A telephone or e-mail hotline should be provided to ensure that the user receives qualified help for technical as well as methodological problems. A detailed manual is essential. Many LCA software providers offer special training sessions to introduce the software to the new user. Demonstration versions and tutorials to demonstrate the functionality and features are very helpful in providing a quick overview of the properties of a software tool. Such demo versions should be available for free to demonstrate the advantages of a software tool to potential new users.

Another essential aspect of service is getting relevant information about the software. This aspect should take into account that there are at least two different kinds of users. On one hand there is the LCA newcomer: He needs some general information about LCA and about the advantages of the particular software. This information can normally be found on the software homepage. On the other hand there is the professional LCA user: He needs more detailed information about the different features the software provides and the assumptions included in the database or the methodological solutions, such as which assessment methods are provided and where the database is from. At resent there is a lack of information in this area. Normally one sees this information only after purchasing the software. More detailed information is needed on the Internet for LCA software.

4.7 Other features

As mentioned before, many LCA software tools offer additional features. One group of them focuses on analysing data. One example is a sensibility analysis, which should be implemented in each good software product. The feature of comparing different scenarios can also be called a standard feature.

The cost consideration is also important. Although there are major methodological differences between Life Cycle Cost analysis (LCC) and an LCA, they can be tightly, logically and practically integrated with one another [Norris, 2001]. Some software tools also consider time aspects and social parameters such as working time.

5 CONCLUSION

Many LCA software tools can be considered of good quality. They were often developed for a specific application of LCA but were then improved for a wider scope. Sometimes although the software is generally designed for a wide scope, it is not possible to use this wide scope due to e.g. inadequate calculation methods or an unsuitable structure. Thus it is not enough for single features to be implemented in an LCA tool, but the whole package of features needs to fit together in a good quality software tool. Basic requirements need to be fulfilled by the software to be suitable for a wide audience.

Generally a software tool should operate smoothly and quickly, without errors due to mistakes in the software programming. The hardware requirements should also be adequate. A hierarchical structure is essential for good quality software, in order to be able to work on more complex problems as well. A clear structure ensures transparency and modelling comfort. The starting point of the calculation should be of free choice. Also the modelling of different outputs should be possible. The results should be transparent. A graphical modelling of the process chain is very convenient for the user.

Compatibility of the software with other application should be provided and the user interface should be designed in such a way that the user finds his way around easily and feels comfortable working with the program (e.g. if designed similar to MS Office applications). A good toolbox to present the results is desirable.

The database should be managed and edited separately (creating, deleting, modifying of data). The data should be of good and transparent quality. There should be a possibility of separately saving and organising data used for single projects. The names of processes and materials need to be clear and logical and the problem of synonyms should be taken into consideration.

It should be possible to choose between different methodological approaches for the impact assessment and the aggregation of results as well as for the comparison of scenarios with different outputs.

An Internet homepage with detailed information should be provided for an LCA software tool. It should contain information for newcomers as well as experts. Different versions and a free demo version of the software should be available.

Additional features that help the user to analyse results and allow further calculations are important requirements for some users.

To define the proportion of input and output, linear equation systems will most often be sufficient, although scripts can be essential for some processes.

Good software tools featuring uncertainty and variability analyses such as the Monte Carlo simulation enable the user to identify parameters, which cause a large spread in the model outcome. Thus the accuracy of the model can be increased through support of a more selective procedure.

It is important that an LCA software tool be continually improved and updated with new developments in the field of LCA. Maybe they can even give an incentive to new developments since most life cycle assessments are calculated with an LCA software tool.

6 REFERENCES

BUWAL, Bewertung in Ökobilanzen mit der Methode der ökologischen Knappheit – Ökofaktoren 1997. Schriftenreihe Umwelt Nr. 297, 1998.

Eriksson O., B. Frostell, A. Björklund, G. Assefa, J.-O. Sundqvist, J. Granath, A. Baky, and L. Thyselius, ORWARE – A simulation tool for waste management, *Resources, Conservation and Recycling,* 36, pp. 287-307, 2002.

Frühbrodt E., LCA Software Review – An up-to-date overview of the European market, available on http://europe.eu.int/comm/enterprise/electr_equipment/eee/workshop9-10-02/present/lcasoftwarere.pdf .

Goedkoop M., S. Effting and M. Collignon, The Eco-indicator 99 – A damage oriented method for Life Cycle Impact Assessment. PRé product ecology consultants, second edition, 2000.

Guineé J., M .Gorrèe, R. Heijungs, G. Huppes, R. Kleijn, A. de Koning, L. van Oers, A.W. Sleeswijk, S. Suh, H.A.U. de Haes, H. de Bruijn, R. van Duin, M.A.J. Huijbregts, E. Lindeijer, A.A.H. Roorda, B.L. van der Ven and B.P. Weidema, Life cycle assessment – An operational guide to the ISO standards. Ministery of Housing, Spatial Planning and the Environment and Centre of Environmental Science - Leiden University, 2001.

Huijbregts, M.A.J., Uncertainty and variability in environmental life-cycle assessment, thesis, University of Amsterdam, 2001.

IKP, PE, GaBi 4 Software-System and Database for Life Cycle Engineering (User Guide) Stuttgart, Echterdingen 1992-2002.

Jönbrink A.K., C. Wolf-Wats, M. Erixon, P. Olsson and E. Wallén, LCA Software Survey, IVL rapport No. B 1390, IVL Svenska Miljöinstitute AB, Stockholm, 2000.

Norris G.A., Integrating Life Cycle Cost Analysis and LCA, *International Journal of Life Cycle Assessment*, 6(2) pp. 118-121, 2001.

Rizzoli A.E. and W.J. Young, Delivering environmental decision support systems: software tools and technique, *Environmental Modelling & Software*, Vol. 12(2-3), pp. 237-249, 1997.

UBA (Umweltbundesamt), Bewertung in Ökobilanzen, Version 1999, UBA Texte 92/99, 1999.

Unger N, Eignung unterschiedlicher Softwaretools für Ökobilanzen bei abfallwirtschaftlichen Fragestellungen, diploma thesis, Boku-University, Vienna, 2003.

Weidema B.P., and B. Mortensen, Results of a test of LCA-software with statistical functionality, Note for the Danish Environmental Protection Agency, Institute for Product Development, Lyngby, 1997.

Grid-based Environmental Risk Analysis System

Thilo Ernst, Andreas Hoheisel, and Steffen Unger

Institute for Computer Architecture and Software Technology (Fraunhofer FIRST), Berlin, Germany (steffen.unger@first.fraunhofer.de)

Abstract: A computer-supported system for analyzing and managing pollutant-related environmental risks is being developed in the ERAMAS (Environmental Risk Analysis and Management System) project. The objective is to use sophisticated simulation programs to forecast and evaluate the dispersion of carcinogenic and chemically toxic substances in the atmosphere, the soil and the groundwater and to calculate the risk they pose to human beings. The system is being implemented using the technology of the Fraunhofer Resource Grid (FhRG). This technology facilitates coupling of the various project partners' local hardware and software components via the Internet, thus making the system extremely flexible. Depending on the problem in hand, it can select the most suitable models and system information and thus give optimal consideration to the respective constraints, whether in terms of the concrete area affected or the type of substance released. The design of ERAMAS as a distributed system also guarantees that it is dynamically adapted to the current state of the art because the different simulation models are monitored by their respective developers. ERAMAS provides lightweight web-based user interfaces (also suitable for low-bandwidth clients such as mobile devices) implemented using the VirtualLab platform developed in collaboration with DLR (German Aerospace Center).

Keywords: Simulation Models; Environmental Risk analysis; Grid Computing; e-science Platform

1. INTRODUCTION

ERAMAS is an acronym for the **Environmental Risk Analysis and Management System** being developed by Fraunhofer FIRST in collaboration with *Ingenieurbüro Beger für Umweltanalyse und Forschung* and the *Dresdner Grundwasser Consulting GmbH* in Germany [ERAMAS, 2004]. ERAMAS is a simulation-based analysis framework for risks caused by chemically toxic or carcinogenic substances released during accidents in industrial installations, the transport of dangerous goods or by terrorist attacks. It is designed to be applicable both for real-time emergency management and for risk mitigation activities such as simulation-aided studies concerning the design of approval procedures or emergency plans.

Figure 1 shows an overview of the simulation models involved in ERAMAS regarding the different transportation paths in the atmosphere and the soil. In the environmental simulation domain, this is a rather typical scenario – nevertheless, with the software technology widely used in this domain today, integrating and coupling even a small number of heterogeneous simulation models and making them available to technically unskilled users amounts to a very complex and time-consuming effort.

In ERAMAS, this task is made much easier by a Grid-based component model which allows the integration of legacy applications without making changes to their internals, and by a Petri-net-based workflow model coordinating the collaboration of the software components involved in the simulation (Figure 3).

2. IMPLEMENTED SIMULATION MODELS

The analysis and forecast components employed in ERAMAS are based on approved simulation packages. The following components are available:

- A *diagnostic windfield model (dwm)* that calculates three-dimensional realistic wind fields considering topography from sparse input data, e.g., measurements from a weather station such as wind speed and direction, temperature, stability class, etc.
- Several *source modules* for the simulation of the pollutant emission through a chimney, jet release, line source, or evaporation from a puddle. To consider uncertainties in the input data, we use a Monte Carlo Simulation for these source modules.

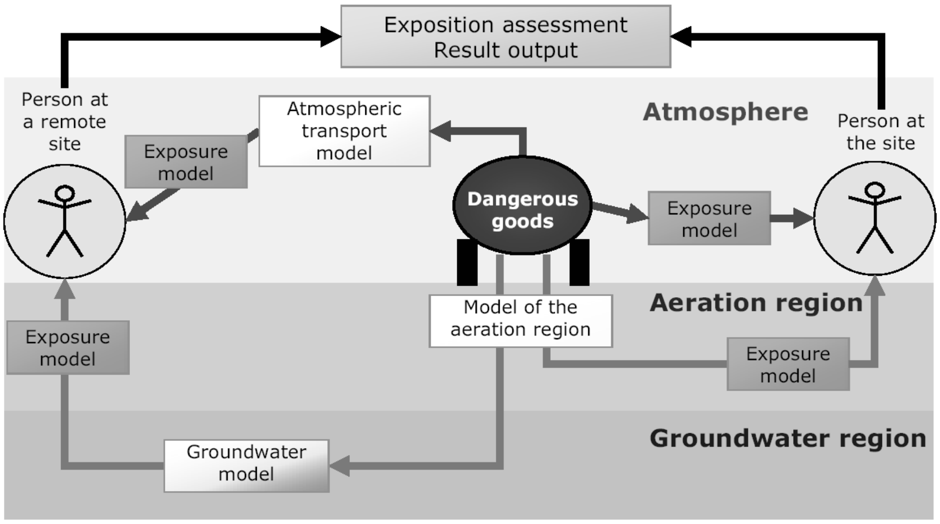


Figure 1. ERAMAS integrates a variety of physical simulation models in order to calculate the transport of carcinogenic and chemically toxic substances in the atmosphere, the ground, and the exposure to humans.

- A *Lagrange model* for atmospheric pollutant transport simulation. From the emissions and the three-dimensional wind field it calculates the spatial and temporal distribution of the pollutants. As our Grid framework is not optimized for tightly coupled applications, this model uses MPI for parallel execution on a Linux cluster.
- A one-dimensional simulation model for the calculation of the pollutant transport in the aeration region (*HYDRUS*).
- A three-dimensional simulation model for the calculation of the pollutant transport in the water saturated ground (*MODFLOW*).
- Several models for the simulation of the pollutant exposure to the human being for different kinds of ingestion and inhalation paths.
- Modules for the risk analysis as a function of the land usage and the resulting distribution of pollutants on the ground.
- An open-source geographical information system (*GRASS*)
- A MySQL database server

More background on ERAMAS is available in German in [ERAMAS, 2004, and Unger et al., 2003].

3. ERAMAS ARCHITECTURE

ERAMAS is being developed using the technology of the *Fraunhofer Resource Grid (FhRG)* that simplifies the coupling of the heterogeneously distributed software, hardware, and data resources (see Figure 2). The Fraunhofer Resource Grid is a Grid initiative of five Fraunhofer institutes funded by the German federal ministry of education and research with the main objective to develop and to implement a stable and robust Grid infrastructure within the Fraunhofer-Gesellschaft, to integrate available resources, and to provide internal and external users with an easy-to-use interface for controlling distributed applications and services on the Grid [FhRG, 2004]. The component environment supports loosely coupled software components where each software component represents an executable file (command-line application) that reads input files and writes output files. (This smallest unit of execution is also referred to as an *atomic job*). Communication between components mainly takes place via file transfer (GridFTP). Arbitrary legacy code can be integrated easily using shell scripts to encapsulate programs or fixed combinations of them.

Up to now, the FhRG architecture does not directly support tight coupling schemes like CORBA, MPI, or HLA. However, tightly coupled applications can be included as a whole, forming one atomic job each. The FhRG Petri-net-based workflow model and the Grid Job Handler implementing it, is described more detailed in Section 5.2 and [Hoheisel, 2004].

ERAMAS uses the VirtualLab platform [Ernst et al., 2003] for providing a convenient portal-based

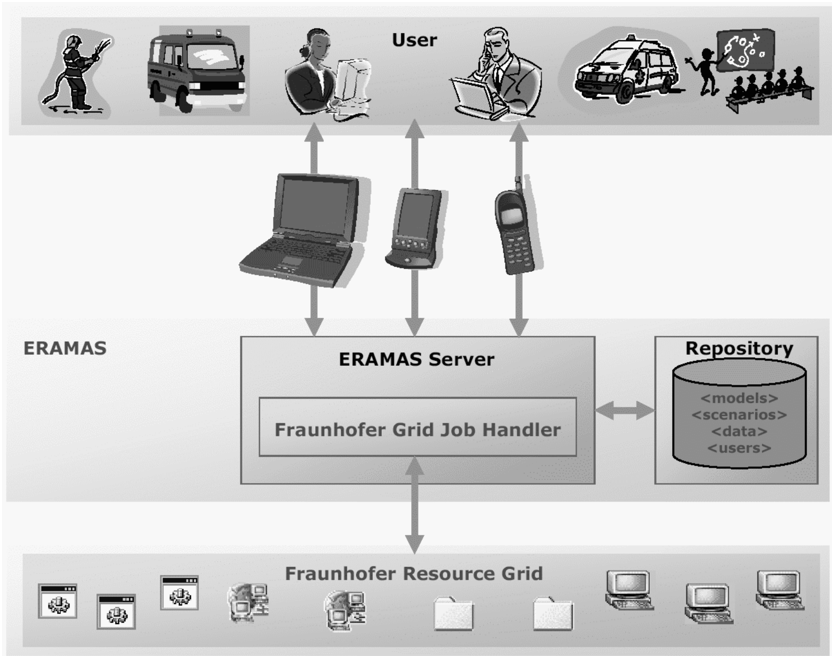


Figure 2. The ERAMAS architecture. The user connects to the ERAMAS server by HTTP. The ERAMAS Server uses the Fraunhofer Grid Job Handler in order to enact the ERAMAS workflow.

web user interface. The genericity of this user interface mechanism extends the ease of integration provided by the underlying FhRG Grid component model up to the UI level.

Currently, ERAMAS has the status of a demonstration prototype. We plan to use it as a production system in the near future.

4. WHY DO IT ON THE GRID?

ERAMAS is a system with considerable resource demands which arise not only from the inner complexity of its components, but also from complex workflows and usage scenarios in which a substantial number of component instances need to be executed – for instance in parameter studies. Such a system cannot be expected to run on a single workstation; the need to exploit parallel and distributed computing techniques is obvious. However, the primary advantage of using Grid technology in ERAMAS is not mainly the performance gain from being able to use additional resources, but rather the organizational advantages in building and maintaining a distributed, highly heterogeneous simulation system. We use the Grid to organize the workflow of coupled simulations and to provide uniform access to a wide variety of hardware and software (including data) resources.

The component abstractions offered by the Fraunhofer Resource Grid make coupling of a wide range of models (and data sources) very easy – detailed knowledge of the internals of the components is no longer needed. In the long term, we are convinced that the Grid will provide the perfect exchange and transfer medium for simulation models and scientific software (and data) in general – once the Grid technology is sufficiently mature, standardized and openly available, it can be expected to become widely deployed, at least throughout the research and technology sector. For first-time or casual users, invoking a program directly on the Grid (most often through a web portal) will be the quickest and most convenient way to use it. Provided that it will get easy to publish software on the Grid as well, this can enable a much smoother and unobstructed flow of knowledge manifested in scientific software into its potential application domains in industry, engineering and government than we see it today. The consequences of this transition must be considered not only on the technical level. Lowering the barrier-to-entry both on the "producer" and on the "consumer" side, makes lots of application scenarios realistic that today are simply not economically viable: For instance, a small, specialized company will be able to work with a large selection of programs that would have been prohibitively expensive to purchase "rented" as application services on the Grid; a research institute can realistically offer a specialized software package as a Grid service that would otherwise never get used outside the department it was developed in. Of course this era of broad e-Science [Ernst and Wauer, 2003]

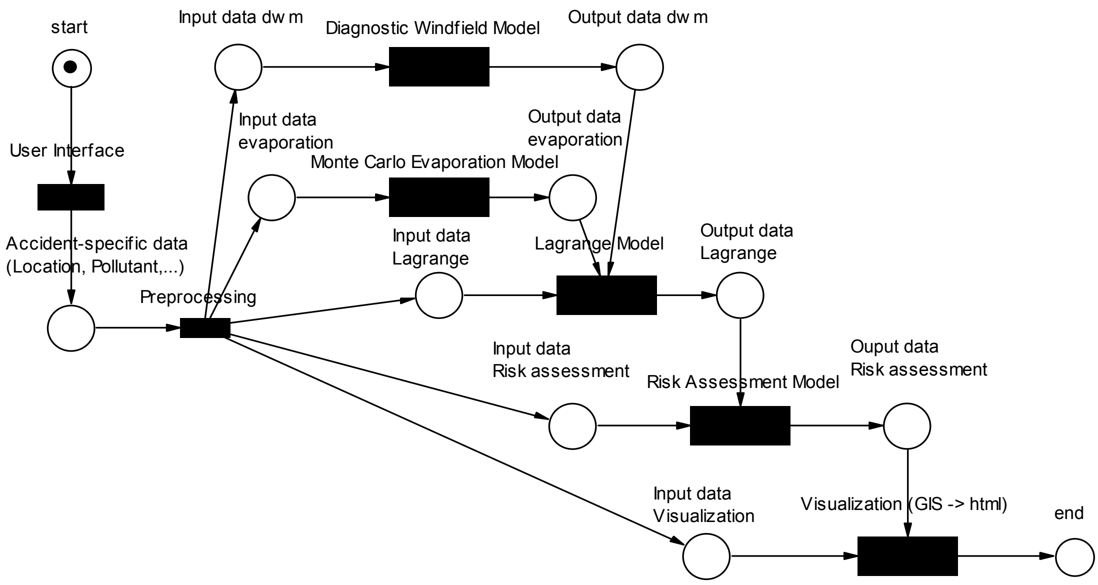


Figure 3. Petri net describing the Grid job for simulating atmospheric pollutant transport in the atmosphere.

will not arrive all by itself – ERAMAS represents a step in this direction. In other words, ERAMAS can be viewed as introducing Grid- and web-based e-science methods into the environmental simulation and risk management community and to develop and deploy a dedicated platform for the purpose.

With respect to the risk management-related project goals it is also mandatory to take care of quality-of-service aspects. However, there are many open questions in this area today, so the ERAMAS prototype does not provide such a guarantee right now. On the positive side, the Grid technologies it relies on are in our view an ideal base for addressing these issues.

5. REALIZATION WITHIN THE FRAUNHOFER RESOURCE GRID

Software Component Framework. Most of the software components that are integrated within ERAMAS are provided as legacy code. In order to make these components available to the Grid architecture they have to comply with the software component model [Hoheisel, 2002]. This component model is rather simple – it just specifies mandatory command line parameters and defines the input and output files that are processed by the component. Furthermore, the standard output, standard error and standard input of the component can be used by the Grid architecture. Arbitrary executables can easily be wrapped using shell scripts in order to enforce compliance with the software component model.

Following this, the software component is packed together with its static data and the encapsulating shell script into a tar.gz archive that is stored somewhere on the Grid, so the software component can be deployed automatically on a suitable hardware resource on demand (a planned extension will also allow replicated component storage in order to improve reliability and availability). Each software component is described separately by an XML-based resource description (GResourceDL) in order to support task-mapping mechanisms for problem-solving environments and resource discovery with regard to the dependencies between Grid resources. The FhRG component framework supports loosely coupled communication between the software components. This communication mainly takes place by transferring the corresponding input and output files via GridFTP. The main advantage of this approach is that it does not require any modification or recompilation of the software components, so existing legacy code can be included easily in the framework ("black-box integration"). The coordination and synchronization of the execution of multiple application tasks is done by the so-called *Grid Job Handler* that acts as a higher-level service on top of existing Grid middleware.

Workflow Orchestration and Enactment. Our approach is centered on a Petri-net-based workflow model that allows the graphical definition of arbitrary workflows with only few basic graph

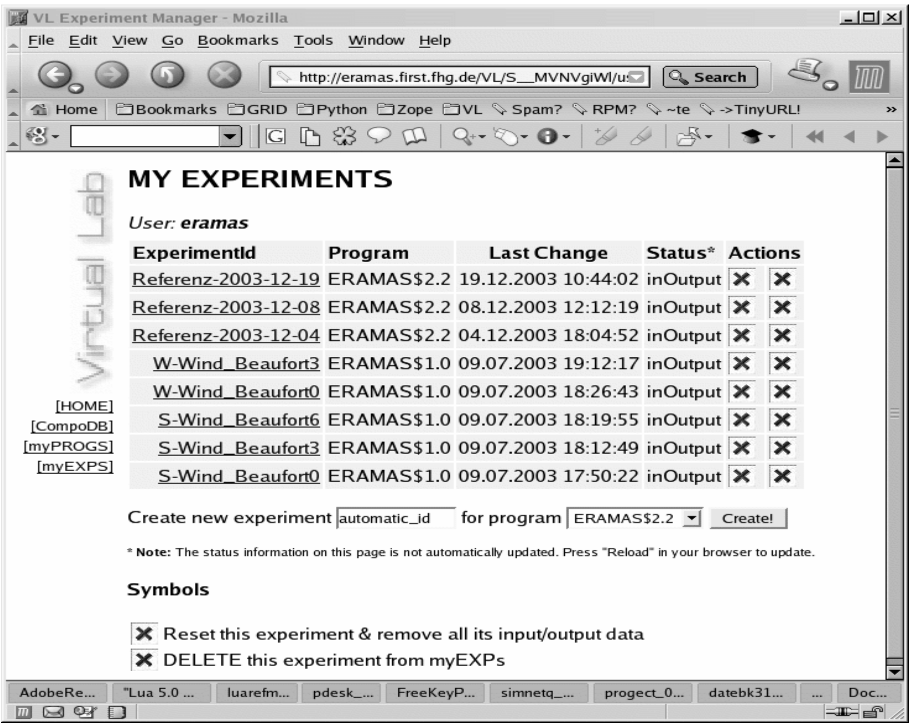


Figure 4. Within VirtualLab, each simulation of a specific scenario is managed as a virtual experiment. The user can access his data from past simulation runs and create new experiments using a standard web browser.

elements – just by connecting data and software components. The output of the workflow orchestration process is an XML-based *Grid Job Definition Language (GJobDL)* document, which defines the Grid job. The GJobDL description of a Grid job contains the resource descriptions of the basic resources that are required to define the Grid job and the model of the Grid job workflow and the dataflow using the concept of Petri nets [Petri, 1962]. The GJobDL document can be stored as a file or be transmitted directly to the Grid Job Handler Web Service in order to enact the workflow. Within ERAMAS we use predefined Grid job templates, so the user does not have to take care of these coupling mechanisms. An example Petri net that represents the workflow of a sub part of the ERAMAS simulation is displayed in Figure 3. For further details about the Petri net approach of the Fraunhofer Resource Grid refer to [Hoheisel and Der, 2003].

The Grid Job Handler is responsible for the execution of each Grid job on a set of suitable hardware resources. Therefore, the Grid Job Handler parses the Grid job description, resolves the dependencies between the Grid resources, and searches for sets of hardware resources that fulfill the requirements of each software component. A meta scheduler is used to select the best-suited hardware resource of each set of matching hardware resources. The Grid Job Handler maps the resulting atomic jobs to the Globus Resource Specification Language (RSL) [Globus, 2000] and submits them via GRAM to the corresponding Grid nodes. The Grid Job Handler itself is deployed as a Web Service with possibilities to create, run and monitor Grid jobs remotely.

6. WEB ACCESS TO ERAMAS VIA VIRTUALLAB

The simulation components integrated in ERAMAS are pure command-line applications, i.e., they have no graphical user interface. As specialized simulators usually originate from the research sector, this can be considered normal, but it severely conflicts with the goals of and application scenarios envisioned for ERAMAS, which call for strong support of the less technically skilled user, e.g. users from on-site emergency response teams. We bridged this gap by relying on the VirtualLab web platform architected by one of the authors at the German Aerospace Center (DLR), and now further developed in an ongoing collaboration DLR-Fraunhofer FIRST. VirtualLab [Ernst et al., 2003] can be characterized as a pre-Grid Science Portal which allows making scientific programs available for direct on-line execution as web applications in a set of domain-specific virtual laboratories.

Targeted at facilitating technology transfer and scientific exchange, the system is consistently focused on reducing barriers-to-entry for users (but also component authors) so they no longer need to act as system administrators or developers when the goal is just to use some piece of scientific software or offer it for on-line execution, respectively.

VirtualLab contains a subsystem for dynamically generating flexible and easy-to-use web user interfaces for command-line applications from abstract descriptions of their input datasets. Together with its generic web portal features (protected user areas persistently storing simulation runs, integrated documentation management, web-based administration), VirtualLab is thus able to provide a powerful web access layer for ERAMAS.

Figure 4 shows the VirtualLab-based web user interface. The user provides input data for the input files that are transferred from the VirtualLab web server to a Grid gateway node. Also, a Grid job description (including workflow description) is generated and submitted to the Grid Job Handler Web Service, which takes care of its execution. Progress can be monitored from the user's browser through a Java applet. After completion of the Grid job, the user can access the resulting data using his web browser.

7. CONCLUSIONS AND FURTHER WORK

ERAMAS is a work in progress, but already now demonstrates the vast potential Grid and web-based e-science methods will have when routinely available to the broader scientific computing community. For the near future, it is planned to migrate the FhRG and ERAMAS to Globus Toolkit V3. As our component model does not require the internals of components to be modified, we expect this to proceed rather smoothly. In the mid term, we will further develop ERAMAS and FhRG and will more closely integrate them with central parts of VirtualLab, with the goal of building a Grid and web-based platform for broad e-science. The application domain of environmental simulation and risk management will remain a main focus of our prototyping and case study work.

8. ACKNOWLEDGEMENTS

The project is funded by the *Arbeitsgemeinschaft industrieller Forschungseinrichtungen "Otto von Guericke" (AiF)* in the program "Innovationskompetenz mittelständischer Unternehmen" (PRO INNO).

9. REFERENCES

ERAMAS homepage, http://www.eramas.de/, 2004.

Ernst, T., Rother, T., Schreier, F., Wauer, J., and Balzer, W., DLR's Virtual Lab – Scientific Software Just A Mouse Click Away. *IEEE CiSE,* 5 (1), 2003.

Ernst, T., Wauer, J., Grid Content Evolution and Management – Enabling e-Science for the Rest of Us. In: *Proceedings of the Cracow Grid Workshop '03*, Cracow, Poland, 2003.

Fraunhofer Resource Grid homepage, http://www.fhrg.fhg.de/, 2004.

Hoheisel, A., Ein Komponentenmodell für Softwarekomponenten des Fraunhofer Resource Grid. Internal report, Fraunhofer FIRST, http://www.andreas-hoheisel.de/docs/FhRGSoftwareComponent.pdf, 2002.

Hoheisel, A., Der, U., An XML-based Framework for Loosely Coupled Applications on Grid Environments. In: Sloot, P.M.A. et al. (eds.), ICCS 2003. *Lecture Notes in Computer Science*, 2657, 245–254, 2003.

Hoheisel, A., User Tools and Languages for Graph-based Grid Workflows. Proc. GGF10-Grid Workflow Workshop (to appear), 2004.

Petri, C. A., Kommunikation mit Automaten. Ph.D. dissertation, Bonn, 1962.

The Globus Project, The Globus Resource Specification Language RSL v.1.0, http://www-fp.globus.org/gram/rsl_spec1.html, 2000.

Unger, S., Hoheisel, A., Beger, E., Beims, U., ERAMAS – Analyse- und Managementsystem von schadstoffbedingten Umweltrisiken, *Technische Überwachung,* 44 (4), Springer-VDI-Verlag Düsseldorf, 46–49, 2003.

Software Environment of a Grid-based Virtual Organization for Flood Prediction

L. Hluchy, O. Habala, J. Astalos, V. D. Tran, B. Simo, M. Dobrucky, E. Gatial and M. Maliska[a]

[a] *Institute of Informatics of the Slovak Academy of Sciences,Dúbravká cesta 9, Bratislava, Slovakia*

hluchy.ui@savba.sk

Abstract: In this paper we present architecture and current state of implementation of a collaborative computation environment based on Grid infrastructure, used as a mean of support for large scientific virtual organizations. The environment consists primarily of a collaboration-supporting user interface, workflow system capable of submission of jobs to the Grid and a Grid-based data management suite.
The whole system is controlled via a web-based user portal, which enables to design and control simulations, to preview results and to communicate with other users. The workflow management module is responsible for executing a set of consecutive computation jobs and linking their input/output data. This computation engine is backed by a data management module, which performs data storage/retrieval tasks and metadata queries.
A prototype of such an environment is deployed and tested for a flood forecasting system. The system consists of workflow system for executing simulation cascade of meteorological, hydrological and hydraulic models, data management system and a set of web portals.

Keywords: Flood prediction; Grid computing; Problem Solving Environment; Modelling

1. INTRODUCTION

The problems of flood prediction are very actual every summer in Europe, with vast catastrophes in several countries each year. Better than acting after the flood occurs is to predict it, prevent it, or at least minimize the damage it may cause.

We are developing a software suite for modelling and prediction of flood, using state-of-the-art technologies. The whole system is a Problem Solving Environment (PSE) [Gallopoulos et al., 1994] consisting of a Grid-based simulation core, monitored and controlled by a comfortable user interface, accessible through a stand-alone Java application or a web portal.

This paper describes the architecture of this system, its usage scenario and possible interactions with its environment. Then the most important components are described in more detail, together with their history and directions of future development.

2. ARCHITECTURE OF THE SOFTWARE ENVIRONMENT FOR FLOOD PREDICTION

The software of our PSE consists of several hierarchically organized modules (Figure 1). We may say that this architecture is principally similar for every PSE supporting a scientific virtual organization (VO).
The user accesses the system through a comfortable interface. He may choose between a Java application and a web-based portal, depending on his preferences for security and mobility. Underlying this interface are three tools for the management of the Grid systems and applications of the core computational facility – the workflow management system, data management and collaborative tools.
The workflow system is responsible for the management of the set of applications (simulation models) used for flood prediction (these will be described later). Because of the large amounts of data consumed and produced by the simulation

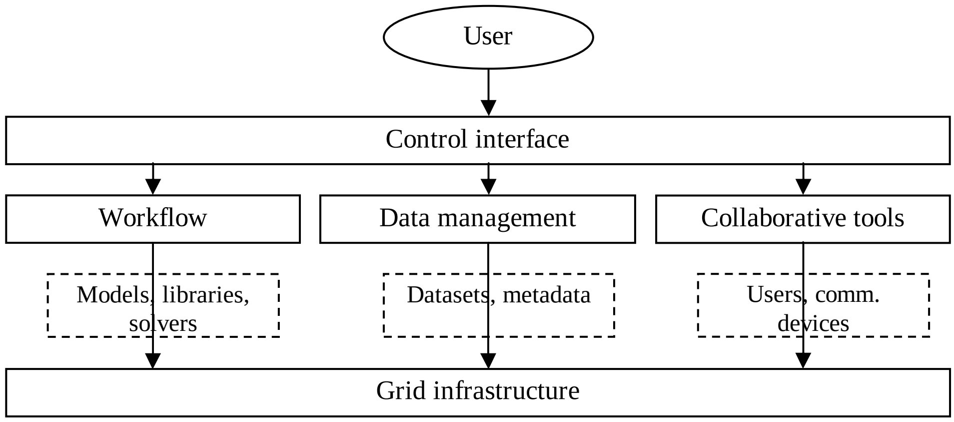


Figure 1. Architecture of a Grid infrastructure for scientific VO

processes involved, an automated data management suite is also necessary. To actually enable two or more scientists to work together towards achieving a common goal, a collaboration and communication suite is integrated in the infrastructure.

The rest of this paper describes the parts mentioned above, as well as the set of simulations used for flood prediction.

3. WORKFLOW MANAGEMENT

3.1 Introduction

As the grid infrastructure matures it is being used by scientists for more and more complex computations. Each such computation can include executions of several applications and transfers of required data. The complexity of the process is becoming too high to be handled manually. Therefore the employment of workflow concept seems quite natural. The business community is widely using the workflow concept and in this context it has been defined as follows:

"Workflow is the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules."

Similar to the definition of a workflow in a business process management, a grid workflow is an automation of a grid process, during which documents, information or data are passed from one grid service to another for action, according to set of procedural rules.

3.2 Existing systems

Most of grid workflow systems being developed focus on the web services, al-though there are older systems that do not use the web services paradigm.

One example of system not using web services can be the Condor DAGMan – a meta-scheduler for the Condor workload management system. It uses a directed acyclic graph (DAG) to represent a set of programs where the input, output, or execution of one or more programs is dependent on one or more other programs. Pegasus is a system for transforming abstract workflow descriptions into concrete workflows, which are then executed using the Condor's DAGMan.

Web services oriented workflow systems are mostly in early stage of development as can be seen in the Scientific Workflow Survey web page [Survey]. There are two main specifications of workflow languages: Web Services Flow Language (WSFL) targeting web services and the Grid Services Flow Language (GSFL), which builds upon Open Grid Services Architecture (OGSA) [Foster et al. 2003]. OGSA is based on grid services, which are web services with additional grid-oriented features, and allows distributed resources to be shared over a network. Currently, we are not aware of any existing workflow system using either of these languages.

3.3 Workflow management for flood prediction

We need an interactive portal-based workflow system that enables the user to construct a workflow or to choose from predefined ones.

As for the interactivity, it means the possibility to view the results of each task (activity) instantly after it has finished without waiting for the whole workflow to finish and ability to clone existing (possibly running) workflow and submit it with modified parameters. The modification may cover one or more tasks.

Important feature is the ability to replace selected step or steps in the workflow with user selected or defined “output” in order to let the user perform various parameter studies. Such replacement must be possible both during workflow definition and during workflow execution.

A workflow system that we designed for our flood prediction system enables the user to define whole cascade execution in advance as a workflow and run it with the possibility to inspect every step.

The whole flood simulation uses three main steps – meteorology, hydrology and hydraulics - to produce the final result – the prediction of the parts of the target area that are going to be flooded. When the expert wants to use already computed results or does not need to compute the last step of the cascade, just parts of the cascade are required. The run of a single simulation model represents the simplest case.

So we have several possible workflow templates that may be executed. We have decided to constrain the workflow selection to several predefined workflows in the first version. Workflow is defined for each target area based on the computation dependencies for that particular area. The changing part of the workflow is mainly hydrology because the run-off in the target catchment is computed from several sub-catchments.

An expert who wants to perform a simulation chooses a target area and time for which to make the prediction, then the workflow template from the list of templates available for the area of interest and a model to be used in each step. The possibility to select more models for the same step or even to enter user defined values instead of running a particular simulation step makes it possible to have several parallel in-stances of a workflow giving several results for the same time and area.

4. COLLABORATIVE TOOLS

The need of cooperation between scientists and users from many organizations in Grid projects requires sophisticated tools for collaborations in portals. The scientists need to access and share data, analyze them, and discuss with other scientists via the collaborative tools. Therefore, collaborative tools are one of the key elements of virtual organizations.

Collaborative tools may be mailing lists, instant messaging, file-sharing tools, discussion groups, etc. However, one single tool cannot provide all features necessary for the collaborations. Therefore, there are several projects that aim to provide an integrated and extensible collaborative environment via portals. One of such projects is CHEF [CHEF].

CHEF (CompreHensive collaborativE Framework) is a collaborative environment based on Jespeed portal engine. The collaborative tools (teamlets) are written as portlets in Jetspeed [Jetspeed] that are extended to special features for multi-user group work nature of collaborative tools:

- Resource access security: The users can only view and modify what they have permissions to.
- Automatic updating of displays: as a user makes changes that affects the display other users are viewing, their display is automatically updated.
- Multi-user safe: if several users are using the same tool at the same time, they work together to avoid conflicts.
- Presence: every user can see who else is using the same tool in the same area at the same time.
- Notification: every user can request to be notified of changes made through the tool by other users.

For accessing to the collaborative tools in CHEF, users need a standard web browser and access to the portal.

5. DATA MANAGEMENT IN SCIENTIFIC VIRTUAL ORGANIZATIONS

5.1 Introduction

The increasing needs for volume and accessibility of data in scientific computations in the last decade leads also to increased demands on better data management tools. The main responsibilities of such software are:

- To track available datasets in the virtual organization.
- To store and maintain these datasets in a coherent fashion.
- To publish their properties and enable their discovery.
- To enable their download and usage.

The data stored in a virtual organization’s data storage facilities has two main parts - the actual datasets and their metadata (meaning their description by another layer of data). Especially the metadata management is evolving rapidly in recent years, as the already quite mature global network infrastructure enables to create larger virtual organizations and data collections with more complicated search and discovery of relevant datasets. Thus, also the data management efforts are divided into two main streams – replica management and metadata storage/lookup.

5.2 Replica management

The actual storage and maintenance of a coherent dataset collection is performed by replica management software. It keeps track of the datasets, potentially stored at multiple places redundantly (replicated). The creation of replicas of a single dataset may be well used for better security and protection against an unwanted loss of the dataset because of a sudden storage device failure, as well as for better access to the file by making it more local to the place that requires it. Although the term replica management may be pertinent to several areas distributed computing research, we deal mainly with the Grid and Grid computing paradigm. For the Grid, a replica management suite has been developed in the European DataGrid Project [Hoschek et al.]. The software developed in work package 2 of the DataGrid [Kunszt et al. 2003] covers the registration, lookup, transfer and replication tasks of a mature replica management suite, with sufficiently distributed control. Its last implementation is based on the modern paradigm of web services and OGSA [Foster et al. 2003] architecture. Anyway, it is lacking a modern and scalable metadata repository.

5.3 Data sources for flood prediction

The general schema of possible data sources for the operation of the Virtual Organization for Flood Forecasting - FloodVO was described in previous work [Hluchý et al. 2003] and also included in the virtual organization architecture. From these sources, only some were realized in the prototype stage of FloodVO.

The most important data in FloodVO storage are the boundary condition for the operation of our meteorological prediction model ALADIN. The second type of data implemented in the prototype stage of FloodVO are radar images of current weather conditions in the pilot operation area, and the third type of currently available data are the ground-based water level, precipitation and temperature measurements provided by SHMI's (Slovak Hydrometeorological Institute) network of measurement stations. The last type of data currently under development are satellite images of the pilot operation site.

5.4 Metadata management

The problem of managing and searching the descriptive information of the dataset collection of a virtual organization (especially for large international scientific Grid-based virtual organizations) is in its nature very similar to the problems of recent peer-to-peer computing efforts. Potentially, the space of storage nodes in such an organization is very large and the especially the distributed lookup is a non-trivial problem. Various solutions have been proposed and evaluated [Joseph and Hoshiai 2003], but the more efficient of them pose severe restrictions on the stored metadata. But several peer-to-peer infrastructure problems, connected with the high instability of the whole network may be disregarded in Grid computing, and in such a controlled environment a decentralized and efficient metadata registry may be deployed. Also, considering the better and more reliable network infrastructure available in grids, a certain level of centralization may be tolerable, without the fear of creating a single point of failure of a bottleneck in the metadata lookup middleware.

The situation in metadata management becomes also more complicated with the re-cent advances in semantic web and Grid. The introduction of ontologies into resource description (like the RDF standard) creates a new area of problems, connected with replication, lookup and especially security of ontology graphs. These problems have yet to be solved, but certainly graph representation of metadata will in the near future begin to replace current, relational representation.

5.5 Prototype implementation

Data management in the prototype of FloodVO was implemented mainly using these software tools, provided by the European DataGrid (EDG) IST project (EDG Replica Manager and underlying services). The metadata database was implemented using the MySQL RDBMS and the EDG Spitfire Grid interface to this RDBMS. A service and client application have been implemented. The client enables to add, modify, locate and delete metadata for given file in the FloodVO SE (identified by its GUID). The metadata service is also accessible through the portal interface, where a user can locate datasets and their details based on their properties.

6. USER INTERFACES FOR COLLABORATION

There are three different user interfaces in various stages of development that provide access to the grid for the flood application.
We have developed GridPort [Thomas et al. 2001] based application portal, we are developing flood application specific portlets for the Jetspeed portal framework based application portal and we are being integrated with Java based client called Migrating Desktop.

6.1 Application portal with the Jetspeed framework

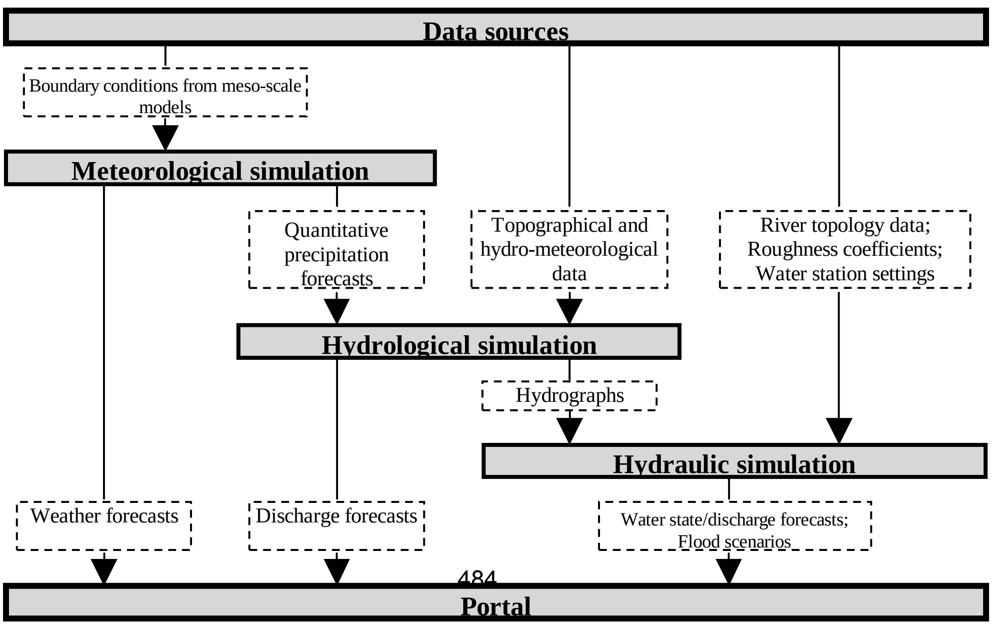
The Jetspeed portal framework has been chosen in the CROSSGRID project as a modern powerful platform for creating grid application portal for the applications in the project. This framework is also being used by other grid projects such as Alliance portal and the new version of the GridPort toolkit – GridPort 3.0.
Jetspeed is implemented as a server-side Java based engine (application server). Client services are plugged in using software components called portlets. Each portlet has a dedicated space on the screen, which it uses for communication with users. Portlets are independent from each other and user can arrange their position, size and visibility.
Jetspeed provides framework for building information portals (pluggable portlets mechanism, user interface management, security model based on permissions, groups and roles, persistence of information etc.) but does not provide any support for grid services and applications. Common Grid portlets that can be used in Jetspeed are being developed in CROSSGRID and other projects.
Portlet for submission of specific simulation models of flood application has been developed and now we are focusing on automation of a computation of the flood simulation cascade by employing workflows. We are also investigating the possibility of using groupware portlets from the CHEF project.

6.2 Migrating desktop

Migrating Desktop is a Java client being developed in the CROSSGRID project. The idea was to create user interface with greater interactivity than could be possible to achieve by using web technology.
Current version provides access to basic Grid services such as authentication, job management, and file management. Support for specific application features is addressed by application and tool plugin interfaces that enable to plug in code handling application specific parameter definition and visualization. We have implemented both plugins for the flood application.

7. SIMULATION CASCADE

Flood forecasting requires quantitative precipitation forecasts based on the meteorological simulations of different resolution from meso-scale to storm-scale. Especially for flash floods, high-resolution (1 km) regional atmospheric models have to be used along with remote sensing data (satellite, radar). From the quantitative precipitation forecast hydrological models are used to determine the discharge from the affected area. Based on this information hydraulic models simulate flow through various river structures to predict the impact of the flood. The results of simulations can be interactively reviewed by experts using the PSE, some of them accepted and forwarded to the next step, some rejected or re-run with modified inputs.
So the cascade of simulations, which we have just described, starts by running a meteorological



model. The plan is to use several of them, but the primary one will be the ALADIN model, currently used in SHMI. This model has been modified so that it can be used in a cluster environment, with MPI as a messaging interface. The model will be run twice a day, creating forecasts for the next 48 hours. The forecasts will consist of precipitation, wind speed and direction in several heights.

The output from the ALADIN model will be used to run hydrological simulations. Here we will again use more models, some simple and some more elaborated (and more demanding). These models will be run in a high-throughput environment, using a pool of workstations and the Condor management toolkit. The output from these simulations will be a hydrograph, representing estimated water level in the target area. This hydrograph will be used in the last stage of the simulation cascade, in the hydraulic simulation (Figure 2).

8. FUTURE WORK

The core of our workflow system has been implemented and is running there are some important features that are yet to be implemented. For example, the visualization portlet is not directly connected to visualization jobs, so the user has to browse to the output directory manually.

We also plan to create a new portlet for creation and modification of the workflow templates, and implement features, which would enable a user to see standard error and standard output interactively during the job run.

The metadata service will be more tightly integrated into the user interface and workflow system, so the user may choose data for simulations based on a group of constraints, rather than its physical location.

However, the most important goal is to integrate our portlets into a collaboration tool in order to enable the exchange of ideas and results among the experts working with the portal. We are currently investigating the possibility to integrate with the CHEF collaboration portal that is based on the Jetspeed framework and therefore also uses the portlet paradigm.

9. CONCLUSIONS

The software system we present here is a specialized simulation tool for flood prediction, which uses a cascade of meteorological, hydrological and hydraulic models to predict water flow in a river basin. The system is controlled by a user interface with integrated workflow and data management and tools for user collaboration. The current prototype implementation will be further refined and enriched by new models and features.

10. ACKNOWLEDGEMENTS

This work is supported by EU 5FP CROSSGRID IST-2001-32243 RTD project and the Slovak Scientific Grant Agency within Research Project No. 2/7186/20.

11. REFERENCES

Gallopoulos, S., Houstis, E., Rice, J., Computer as Thinker/Doer: Problem Solving Environments for Computational Science. *IEEE Comp. Science and Engineering Magazine*, 1994, Vol. 2, pp. 11-23.

Hoschek, W., et. al., Data Management in the European DataGrid Project. *The 2001 Globus Retreat*, San Francisco, August 9-10 2001.

Foster, I., Kesselman, C., Nick, J. M., Tuecke, S., The Physiology of the Grid; An Open Grid Services Architecture for Distributed Systems Integration. *The Globus Project*, 2003.

Hluchý L., Habala O., Simo B., Astalos J., Tran V.D., Dobrucký M., Problem Solving Environment for Flood Forecasting. *Proc. of The 7th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2003)*, July 2003, Orlando, USA pp. 350-355.

Kunszt, P., Laure, E., Stockinger, H. and Stockinger, K., Advanced Replica Management with Reptor. *In 5th International Conference on Parallel Processing and Applied Mathematics*, Czestochowa, Poland, September 7-10, 2003. Springer Verlag.

Joseph, S., Hoshiai, T., Decentralized Meta-data Strategies: Effective Peer-to-Peer Search. *IEICE Trans. Commun.*, Vol. E86-B, No. 6. June 2003.

Thomas, M., Mock, S., Boisseau, J., Dahan, M., Mueller, K., Sutton, D., The GridPort Toolkit Architecture for Building Grid Portals. *Proceedings of the 10th IEEE Intl. Symp. on High Perf. Dist. Comp.* August 2001.

CHEF Information site. *http://www.chefproject.org* (visited Dec. 2003)

The Jetspeed portal framework.
http://jakarta.apache.org/jetspeed/site/index.html (visited October, 2003)
Scientific Workflows Survey.
http://www.extreme.indiana.edu/swf-survey/ (visited December, 2003)

Optimization of mode of operations of hydropower stations with reservoir

I.Sh. Normatov

Institute of water problems, hydropower and ecology Academy of Sciences Republic of Tajikistan, 33, Rudaki Avenue, Dushanbe, 734025, Tajikistan

Abstract: Optimization of work of electric station with reservoirs is even more difficult task. The main criterion of optimization of energy mode of reservoirs is to maximize energy production in that period of year while there is a shortage of water.
For instance, Kayrakkum hydroelectric station, which is the only hydroelectric station of the most developed northern region of the republic this period, continues 6-7 months.

Keywords: Optimization, Hydropower, Kayrakkum reservoir

1. INTRODUCTION

What can be the role of reservoir here? It seems that this task is easy. According to accepted point of view the more works reservoir the more water will go through turbine and as a result hydroelectric station will produce more energy. Then, the most optimal mode should be such at which reservoir will be filled up to the greatest possible mode to the beginning of considered period and is completely worked at the end.

2. MODELS FOR OPTIMIZATION

Let' s consider two variation of mode of hydroelectric station work without any function of reservoir (on transit flow). Appropriate to this variation the work of energy on hydroelectric station will be equal:

$$Э_1 = 9{,}81 \cdot \eta \cdot Q_0 \cdot H_1 \cdot T \cdot 24 \qquad (1)$$

Q_0-the transit charge through reservoir and hydroelectric station provided the tributary of the river, $H_1= \Delta^{max}$ upp.b. $-\Delta_{low..b.}$ pressure on hydroelectric station, T-Duration considered period in day.
In second variation reservoir in regular intervals works for the considered period up to a mark Δ^{t} upp.b. Having accepted that the area of water s mirrors in reservoir at all its level is equal S (m^2) we consider that in this variation the additional volume of water passed through turbines of hydroelectric station is equal:

$$V_{add.} = S(\nabla^{max}_{upp.b} - \nabla^{t}_{upp.b}) \qquad (2)$$

and additional average consumption for entire period:

$$Q_{add.} = \frac{V_{add.}}{t} = \frac{S(\nabla^{max}_{upp.b.} - \nabla^{t}_{upp.b.})}{T \cdot 86400} \qquad (3)$$

Having accepted with sufficient for estimated accounts mark of tail-water in all cases constant, $\Delta_{low..b.}$ the equation (3) we write:

$$Q_{add.} = \frac{S(H_1 - H_2)}{86400 \cdot T} \qquad (4)$$

Where: $H_2=\Delta^{t}_{low.b.} - \Delta_{low..b.}$ - pressure on hydroelectric station, at the end of the work of reservoir.
In evenly work of reservoir it may be accepted that in second variation the hydroelectric station all the time worked with average pressure:

$$H_{aver} = \frac{H_1 + H_2}{2} \qquad (5)$$

And its energy production will be equal to:

$$Э_2 = 9{,}81 \cdot \eta \cdot (Q_0 + Q_{add.})H_{aver} \cdot T \cdot 24 = \\ = 9{,}81 \cdot \eta(Q_0 + \frac{S(H_1 - H_2)}{86400T})\frac{H_1 + H_2}{2} \cdot T \cdot 24. \qquad (6)$$

With consideration Э=9,81ηQHt
Where Э-energy (kW/h), η-efficiency, Q-water consumption (m^3/sek), H-pressure (m), t-time (h) and (6) condition at which the work of reservoir brings to additional effect can be written as:

$$\frac{Э_2}{Э_1} = \frac{9{,}81\eta(Q_0 + \frac{S(H_1 - H_2)}{86400T}) \cdot \frac{H_1 + H_2}{2} \cdot T \cdot 24}{9{,}81\eta_0 H_1 T24} \geq 1 \quad (7)$$

As a result of elementary transformation from equation (7) we can get two following criterion:

$$T \leq \frac{S(H_1 + H_2)}{86400\ Q_0} \quad \text{and}$$

$$Q_0 \leq \frac{S(H_1 + H_2)}{86400T}. \quad (8)$$

Kayrakkum hydroelectric station with reservoir has following main parameters and can be served as an example.

$S = 200\ km^2$; $H_1 = H_{max} = \nabla^{max}_{upp.b.} - \nabla_{low.b.} = 347{,}5-327{,}5 = 20$ m; $9\ m \leq H_2 \leq 20$ m ; $0 \leq Q_0 \leq 900\ m^3/s$; $Q_{aver.\ long.stand.} = 600\ m^3/s$.

The appropriate calculation at formulas (7) and (8) are given in the tables 1,2,3. They show that for really having a place for Kayrakkum hydro electric station of range of change ($400 \leq Q_0 \leq 900$) of the average consumption and pressures of work ($9 \leq H_2 \leq 20$) the efficiency of use of reservoir takes place only at it fully draining during no more than 70-120 that is much less shortage of the electric power of the autumn-winter period, which is seven months (October- April).

Thus for representing practical interests of the periods of time (T>180) the efficiency of work of reservoir for real pressures is provided only at average charges 300-400m^3 that is essentially smaller valid.

Concrete sizes of losses of production of the electric power at work of reservoir shows Fig. 1. At usual for today's practice $H_2 \cong 10$ m and T =180 days they comprise 10% or in absolute meanings 70-80mln.kWt/hour a year.

The given analysis even in view of the made simplifications show that the mode of operations of Kayrakkum reservoir accepted today in practice, is not optimum from the point of view of power. The loss of energy is obvious during such mode. Moreover, it is not effective for irrigation because a big work of reservoir would result in additional expenses of energy at pumping station.

3. OPTIMIZATION OF KAYRAKKUM HYDROPOWER

As an example we consider the concrete task of optimization of Kayrakkum hydropower station at Syrdarya River in Tajikistan. Its capacity is 126 MWt; the volume of reservoir is 4,6km^3 and useful-2,6km^3. The reservoir of Kayrakkum hydro electric station is the biggest for all northern zone of republic hydro electric station, which is isolated from its basic energy system and carries out seasonal regulation of flow in the most intense river pool of Aral Sea in the interests of following republics: Kazakhstan, Turkmenistan and Uzbekistan .The additional need in water for these republics in vegetation period comprises 2,2km^3

For the defining of national and regional interests we consider two regimes: for national-power and for regional-irrigation. Thus all accounts in models we shall carry out for time units equal to one month with use of monthly average of parameters.

Basic settlement formula for production of energy on hydroelectric station after transformations shall present as:

$$Э = 9{,}81\,\eta \frac{H}{3600}(Q3600\,t) = \frac{9{,}81H\eta}{3600}W \quad (9)$$

Table 1.The maximal meaning of the period of time T/ day ensuring efficiency of work of Kayrakum reservoir at various meanings Q m^3/sec and H,m.

Q\H	9	11	13	15	17	19
400	167.8	179.4	191.0	202.5	214.1	225.7
500	134.3	143.5	152.8	162.0	171.3	180.6
600	111.9	119.6	127.3	135.0	142.7	150.5
700	95.9	102.5	109.1	115.7	122.4	129.0
800	83.9	89.7	95.5	101.3	107.1	112.8
900	74.6	79.7	84.9	90.0	95.2	100.3

Table 2.The maximal meaning of water charge Q m^3/sec and ensuring efficiency of work of Kayrakkum reservoir at various period of time T and H,m.

T\H	9	11	13	15	17	19
10	6713.0	7175.9	7638.9	8101.9	8564.8	9027.8
30	2237.7	2392.0	2546.3	2700.6	2854.9	3009.3
60	1118.8	1196.0	1273.1	1350.3	1427.5	1504.6
90	745.9	797.3	848.8	900.2	951.6	1003.1
120	559.4	598.0	636.6	675.2	713.7	752.3
150	447.5	478.4	509.3	540.1	571.0	601.9
180	372.9	398.7	424.4	450.1	475.8	501.5
210	319.7	341.7	363.8	385.8	407.8	429.9

Table 3.The attitude of production of the electric power at work to pressure H_2 to a maximal pressure H=20m at different time of work/day and Q=const=600m^3/sec.

$T\backslash H_2$	9	11	13	15	17	19
10	3.80	3.47	3.05	2.56	2.00	1.35
30	1.75	1.67	1.57	1.44	1.28	1.10
60	1.24	1.22	1.20	1.16	1.10	1.04
90	1.07	1.07	1.07	1.06	1.04	1.02
120	0.98	1.00	1.01	1.02	1.01	1.01
150	0.93	0.95	0.97	0.99	1.00	1.00
180	0.90	0.92	0.95	0.97	0.98	1.00
210	0.87	0.90	0.93	0.96	0.98	0.99

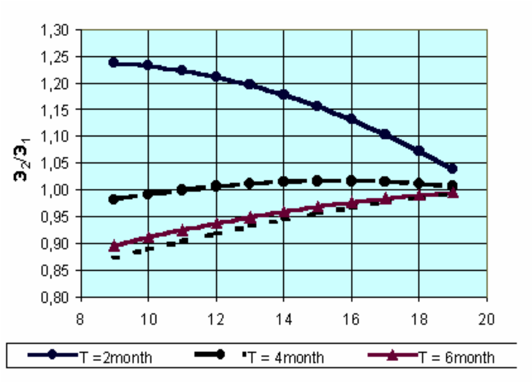


Figure1. Effectiveness of working out of Kayrakkum reservoir for production of power energy in dependence of final pressure and time

From the last formula we get:

$$\frac{W}{Э} = \frac{3600}{9{,}81\eta H} = q \qquad (10)$$

Where: q-water consumption on energy production on hydroelectric station, m^3/kwt/h.

The volume of water through hydroelectric station Wi is calculated by formula:

$$Wi = Q^i_{ch}\ 3600 \cdot N_i \qquad (11)$$

Where: Q^i_{ch}-average monthly consumption of water through turbine of hydroelectric station, Ni-quantity of days in 1-month, Wi-volume of water passed through turbine of hydropower station in 1-month.

The main entry condition for our model is an initial volume of reservoir. For this we should define the settlement period. We take it equal to 1 year (12months) not consider it as usual calendar year, but we take it from the start of vegetation period from 1st October till 30 September when the vegetation period is finished. Thus, it makes easy to consider irrigational issues and also the issue of energy so as shortage of energy and vegetation period are coincide.

Both in hydropower and in irrigational models the natural restrictions having clear physical sense are used.

1.The entry conditions on volume of reservoir (on 1October) should be reproduced by the end of the period, considered in models, (by September 30).

2.The consumption of water through turbines of hydroelectric station should be more or equal to 0.

3.The volume of reservoir in any considered period of time should not be less than minimally possible and more than maximally possible

$$W_{min} \leq W^i_{reserv.} \leq W_{max}$$

4.The consumption of water through turbines of hydroelectric station should not be more as possible allowed on conditions for small pressures.

$$Q_{HPS} \leq Q^{мах}_{HPS}$$

As a changing parameters in both models the consumption of water through turbines of hydroelectric station are accepted.

The criterion of optimization of operational regime of Kayrakkum hydroelectric station from a position of national interests of water-power engineering is defined practically unequivocally i.e. maximization of production of energy in winter: from October till April (May) without any other conditions for other period of year.

A criterion of irrigational regional operational mode of Kayrakkum reservoir is defined by need of water in vegetation period required by countries Uzbekistan and Kazakhstan.

Both models were developed with the use of personal computers in framework of programs Microsoft Office1998 or Excel and menu Service.

The algorithms of a simplex method and method ’’branch-and-bound” for the decision of linear and nonlinear tasks with restrictions are developed by Yohn Watson and Dan Fulstra, Frontline Systems, Ync.

For conducting of concrete calculation on mathematic models the fact sheet, which were in use for the last years (1998-2001) have been used.

The results of calculation are shown in Fig. 2.

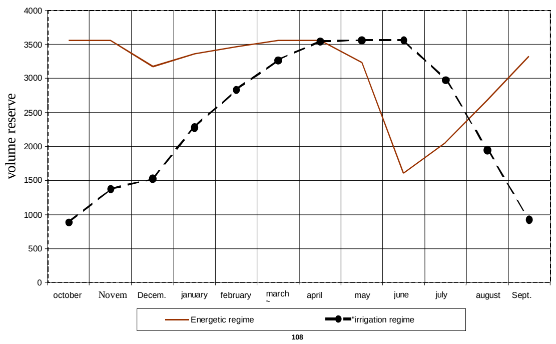
Comparison of results of these two models allow to define those direct losses of produced energy, which bears Tajikistan rendering service on irrigational regulation of flow for Uzbekistan and Kazakhstan. They are equal to:

719,183-472,334=246,849 mln.kwt.hour

As characteristics of this mode it is possible to note that for maintenance of the maximal production of energy in autumn-winter period there is no necessity of he complete function of reservoir up to its minimal volume 881,1mln.m^3.The minimal volume of reservoir during its energy mode is1600, 7mln.m^3.

Received results allow that by using of market approach to optimize the use of water energy resources of river on international level.

a



b

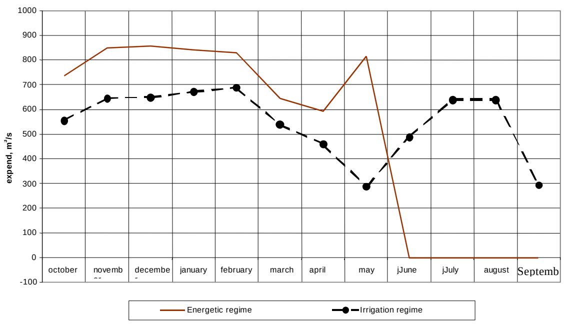


Figure2. Energetic and irrigation scheme of operation and accumulation of Kayrakkum reservoir (a),Monthly average drawn downs from Kayrakkum reservoir in energetic and irrigation regime(b)

4. CONCLUSION

To certain extend it can provide considering by us the maximization of regional benefit but with taking into consideration the social aspects.

It is that: In conditions of priority for life-support of Central Asian Countries and irrigational farming in practice will be implemented irrigational regulation of flow while the criterion will be the maximum provision of water for irrigated agriculture in vegetation period. Countries of lower flow should compensate the loss of energy, which bears the countries of upper flow, getting by result of such scheme the additional water for irrigation.

Such a scheme was implemented in Central Asia for the most intense river pool –Syrdarya of Aral Sea. In 1998 among Kirgizstan, Tajikistan Turkmenistan and Uzbekistan the ”Agreement of using of water-energy recourses of the river pool Syrdarya’’ was signed. This agreement successfully worked during 5 years and in 2003 was prolonged by Central Asian Republics for five years.

Next Generation GRIDs for Environmental Science

Keith G. Jeffery[1] **(k.g.Jeffery@rl.ac.uk)**

[1] *Director, IT; CCLRC Rutherford Appleton Laboratory: Chilton; Didcot; Oxfordshire OX11 0QX UK*

Abstract: GRIDs technology has developed from first generation, supplier-specific and configuration-specific systems through second generation systems providing metacomputing facilities over various underlying platforms. The 2.5 generation added facilities for fast data transfer (GRIDFTP) and directories (LDAP). The third generation embraced from the W3C (World Wide Web Consortium) community the idea of services, leading to OGSA (Open Grid Services Architecture). Under GGF (Global Grid Forum) auspices, a team from UK has developed OGSA/DAI (Data access interface) which starts to make 3G GRIDs usable. However, others consider the existing architecture unsuitable for widespread acceptance beyond the metacomputing / supercomputing community. A small group of ERCIM experts was invited to give keynote presentations at an EC-sponsored workshop in January 2003 which had well over 200 participants. This led to a call from the community for documentation of NGG (Next Generation GRIDs). The three experts with the EC pulled together a team of 14 (including three more from ERCIM) which produced the 'NGG Report' in June 2003; the report is provided as an adjunct document to FP6 Calls for proposals. The key point is that the proposed NGG architecture is more suitable for environmental applications than the conventional 3G GRID. The characteristics and implications are discussed.

***Keywords**:* GRID; data; information; knowledge; systems engineering; middleware

1. A BRIEF INTRODUCTION TO GRIDS

The idea behind GRIDs is to provide an IT environment that interacts with the user to determine the user requirement for service and then satisfies that requirement across a heterogeneous environment of data stores, processing power, special facilities for display and data collection systems thus making the IT environment appear homogeneous to the end-user. As an environmental example, a user – possibly an engineer in the control centre for water distribution - might request to know the effect of rainfall (current and expected in the next day or so) on lower valley flooding in Southern England. This requires retrieval of data on rainfall from all relevant recording weather stations and from all relevant river gauges, both historical (to assist in prediction from long-term historical records) and current (to give the current state).

It will also require retrieval of information on past flood events by region and date to determine the liability to flooding of an area of land, given certain river heights and rainfall conditions. It will also need to retrieve any information on civil engineering works undertaken (with dates by location) to alleviate flooding such as embankments or levées, relief channels or sluice gates and pumping facilities. Finally it will need to retrieve both recent and long-term (historical) meteorological data to assist in prediction of – for example – rainfall from a given intensity of cyclone at a given time of year. Clearly, such a user request will require access to data and information sources, resources to integrate the heterogeneous information both spatially and temporally, computation resources for the modelling and visualisation facilities to display the correlated results. It will involve data collection from detectors to obtain the latest data on all relevant variables. The idea of the GRIDs IT environment is to make all that transparent to the end-user.

2. WHERE ARE WE NOW

2.1 Introduction

Unfortunately, existing GRIDs technology is not yet close to providing the kind of architecture and facilities described above. However, many of the components have been developed independently and using different architectural principles. The challenge is to integrate such developments within architecture appropriate for the envisaged requirements and applications across science, business and culture for the next decade or more.

2.2 A brief History of GRIDS

The initial GRID (not GRIDs) concepts arose from North American ambitions to provide massive compute power – essentially a metacomputing surface provided by linking several supercomputers with scheduling and load-sharing. The early systems (First Generation) were developed independently, were proprietary and had few or no facilities for managing data, information, knowledge, visualisation, user interface, security, resilience – all the aspects required for a generally useful architecture.

By the late nineties, marked by the publication of the 'GRID Bible' [1], a framework (Second Generation) existed based dominantly on the GLOBUS middleware for scheduling and resource management across supercomputers. With the addition of GRID FTP for fast data transfer, SRB (Storage Request Broker) from SDSC (San Diego Supercomputer Centre) for persistent data storage with metadata descriptions and the addition of LDAP for directory services, a platform of use to computational scientists, usually known as 2.5 Generation, was made available in 2001.

In a parallel timescale in UK the Director General of the Research Councils was not content with the proposed IT components in the strategic plans of the Research Councils, which fund research in UK universities. In response to his request, the author in 1999 prepared a paper describing a strategy for IT for R&D, meeting the scientific requirements (e-science) but with a path involving the IT industry to support also e-business, e-culture, e-learning, e-health etc. In the environmental science field example applications whose requirements were discussed included climate change and global warming; meteorology; ocean circulation; local pollution (city areas); coastal pollution and natural disaster (e.g. earthquake, fire, flood) management.

The CCLRC (Central Laboratory of the Research Councils) BITD (Business and Information Technology Department) runs the UK & Ireland W3C (World Wide Web Consortium) Office and has been involved heavily in WWW development. The team joined the pressure on GGF (Global GRID Forum) to adopt Web Services which was eventually agreed late in 2002 with the publication of Open Grid Services Architecture based on W3C Web Services. This marked the emergence of the third generation GRID architecture.

Meantime, returning to the original UK proposal, the GRIDs architecture proposed consists of three layers; the lowest computation and data layer corresponding approximately to the US GRID and having above an information layer providing homogeneous access over heterogeneous data and information sources and at the top a knowledge layer which contains the results of data mining in the information layer (scientific hypotheses automatically developed) and user-proposed hypotheses – with links to appropriate data and software to provide the evidence for the hypothesis. This layer is cross-linked with e-publications so making the connection between peer-reviewed scientific knowledge and grey literature with the data and software involved in its generation.

The initial idea also recognised the coming emergence of ambient computing – the connection of mobile devices which may have associated sensors / detectors to a pervasive wireless or wired network.

Thus the overall vision provides for the environmental scientist a consistent IT surface from data collection fieldwork, advised by expert systems and a control centre through to the control centre receiving information from laboratories, simulations and past data. The current third generation GRID architecture does not meet these requirements. However, the NGG (Next Generation GRIDs) specification does. Unsurprisingly, this specification is close to the original UK paper.

2.3 The Architecture Components

The major components external to the GRIDs environment are:

- users: each being a human or another system;
- sources: data, information or software
- resources: such as computers, sensors, detectors, visualisation or VR (virtual reality) facilities.

Each of these three major components is represented continuously and actively within the GRIDs environment by:

- metadata: which describes the external component and which is changed with changes in circumstances through events
- an agent: which acts on behalf of the external resource representing it within the GRIDs environment.

As a simple example, the agent could be regarded as the answering service of a person's mobile phone and the metadata as the instructions given to the service such as 'divert to service when busy' and/or 'divert to service if unanswered'. Finally there is a component which acts as a 'go between' between the agents. These are brokers which, as software components, act much in the same way as human brokers by arranging agreements and deals between agents, by acting themselves (or using other agents) to locate sources and resources, to manage data integration, to ensure authentication of external components and authorisation of rights to use by an authenticated component and to monitor the overall system. From this it is clear that they key components are the metadata, the agents and the brokers.

Metadata

Metadata is data about data [2]. An example might be a product tag attached to a product (e.g. a tag attached to a piece of clothing) that is available for sale. The metadata on the product tag tells the end-user (human considering purchasing the article of clothing) data about the article itself – such as the fibres from which it is made, the way it should be cleaned, its size (possibly in different classification schemes such as European, British, American) and maybe style, designer and other useful data. The metadata tag may be attached directly to the garment, or it may appear in a catalogue of clothing articles offered for sale (or, more usually, both). The metadata may be used to make a selection of potentially interesting articles of clothing before the actual articles are inspected, thus improving convenience. Today this concept is widely-used. Much e-commerce is based on B2C (Business to Customer) transactions based on an online catalogue (metadata) of goods offered. One well-known example is www.amazon.com .

What is metadata to one application may be data to another. For example, an electronic library catalogue card is metadata to a person searching for a book on a particular topic, but data to the catalogue system of the library which will be grouping books in various ways: by author, classification code, shelf position, title – depending on the purpose required. It is increasingly accepted that there are several kinds of metadata. The classification proposed is gaining wide acceptance and is detailed below.

Schema metadata constrains the associated data. It defines the intension whereas instances of data are the extension. From the intension a theoretical universal extension can be created, constrained only by the intension. Conversely, any observed instance should be a subset of the theoretical extension and should obey the constraints defined in the intension (schema). One problem with existing schema metadata (e.g. schemas for relational DBMS) is that they lack certain intensional information that is required [8]. Systems for information retrieval based on, e.g. the SGML (Standard Generalised Markup Language) DTD (Document Type Definition) experience similar problems. It is noticeable that many ad hoc systems for data exchange between systems send with the data instances a schema that is richer than that in conventional DBMS – to assist the software (and people) handling the exchange to utilise the exchanged data to best advantage.

Navigational metadata provides the pathway or routing to the data described by the schema metadata or associative metadata. In the RDF model it is a URL (universal resource locator), or more accurately, a URI (Universal Resource Identifier). With increasing use of databases to store resources, the most common navigational metadata now is a URL with associated query parameters embedded in the string to be used by CGI (Common Gateway Interface) software or proprietary software for a particular DBMS product or DBMS-Webserver software pairing.

The navigational metadata describes only the physical access path. Naturally, associated with a particular URI are other properties such as:

- security and privacy (e.g. a password required to access the target of the URI);
- access rights and charges (e.g. does one have to pay to access the resource at the URI target);
- constraints over traversing the hyperlink mapped by the URI (e.g. the target of the URI is only available if previously a field on a form has been input with a value between 10 and 20). Another example would be the hypermedia equivalent of referential integrity in a relational database
- semantics describing the hyperlink such as 'the target resource describes the son of the person described in the origin resource'.

However, these properties are best described by associative metadata which then allows more convenient co-processing in context of metadata describing both resources and hyperlinks between them and – if appropriate - events. In the data and information domain associative metadata can describe:

- a set of data (e.g. a database, a relation (table) or a collection of documents or a retrieved subset). An example would be a description of a dataset collected as part of a scientific mission
- an individual instance (record, tuple, document). An example would be a library catalogue record describing a book

- an attribute (column in a table, field in a set of records, named element in a set of documents). An example would be the accuracy / precision of instances of the attribute in a particular scientific experiment ;
- domain information (e.g. value range) of an attribute. An example would be the range of acceptable values in a numeric field such as the capacity of a car engine or the list of valid values in an enumerated list such as the list of names of car manufacturers;
- a record / field intersection unique value (i.e. value of one attribute in one instance) This would be used to explain an apparently anomalous value.

In the relationship domain, associative metadata can describe relationships between sets of data e.g. hyperlinks. Associative metadata can – with more flexibility and expressivity than available in e.g. relational database technology or hypermedia document system technology – describe the semantics of a relationship, the constraints, the roles of the entities (objects) involved and additional constraints. In the process domain, associative metadata can describe (among other things) the functionality of the process, its external interface characteristics, restrictions on utilisation of the process and its performance requirements / characteristics.

In the event domain, associative metadata can describe the event, the temporal constraints associated with it, the other constraints associated with it and actions arising from the event occurring. Associative metadata can also be personalised: given clear relationships between them that can be resolved automatically and unambiguously, different metadata describing the same base data may be used by different users. Taking an orthogonal view over these different kinds of information system objects to be described, associative metadata may be classified as follows:

- descriptive: provides additional information about the object to assist in understanding and using it;
- restrictive: provides additional information about the object to restrict access to authorised users and is related to security, privacy, access rights, copyright and IPR (Intellectual Property Rights);
- supportive: a separate and general information resource that can be cross-linked to an individual object to provide additional information e.g. translation to a different language, super- or sub-terms to improve a query – the kind of support provided by a thesaurus or domain ontology.

Most examples of metadata in use today include some components of most of these kinds but neither structured formally nor specified formally so that the metadata tends to be of limited use for automated operations – particularly interoperation – thus requiring additional human interpretation.

Agents

Agents operate continuously and autonomously and act on behalf of the external component they represent. They interact with other agents via brokers, whose task it is to locate suitable agents for the requested purpose. An agent's actions are controlled to a large extent by the associated metadata which should include either instructions, or constraints, such that the agent can act directly or deduce what action is to be taken. Each agent is waiting to be 'woken up' by some kind of event; on receipt of a message the agent interprets the message and – using the metadata as parametric control – executes the appropriate action, either communicating with the external component (user, source or resource) or with brokers as a conduit to other agents representing other external components.

An agent representing an end-user accepts a request from the end-user and interacts with the end-user to refine the request (clarification and precision), first based on the user metadata and then based on the results of a first attempt to locate (via brokers and other agents) appropriate sources and resources to satisfy the request. The proposed activity within GRIDs for that request is presented to the end-user as a 'deal' with any costs, restrictions on rights of use etc. Assuming the user accepts the offered deal, the GRIDs environment then satisfies it using appropriate resources and sources and finally sends the result back to the user agent where – again using metadata – end-user presentation is determined and executed.

An agent representing a source will – with the associated metadata – respond to requests (via brokers) from other agents concerning the data or information stored, or the properties of the software stored. Assuming the deal with the end-user is accepted, the agent performs the retrieval of data requested, or supply of software requested. An agent representing a resource – with the associated metadata – responds to requests for utilisation of the resource with details of any costs, restrictions and relevant capabilities. Assuming the deal with the end-user is accepted the resource agent then schedules its contribution to providing the result to the end-user.

Brokers

Brokers act as 'go betweens' between agents. Their task is to accept messages from an agent which request some external component (source, resource or user), identify an external component that can satisfy the request by its agent working with its associated metadata and either put the two agents in direct contact or continue to act as an intermediary, possibly invoking other brokers (and possibly agents) to handle, for example, measurement unit conversion or textual word translation. Other brokers perform system monitoring functions including overseeing performance (and if necessary requesting more resources to contribute to the overall system e.g. more networking bandwidth or more compute power). They may also monitor usage of external components both for statistical purposes and possibly for any charging scheme.

The Components working together

Now let us consider how the components interact. An agent representing a user may request a broker to find an agent representing another external component such as a source or a resource. The broker will usually consult a directory service (itself controlled by an agent) to locate potential agents representing suitable sources or resources. The information will be returned to the requesting (user) agent, probably with recommendations as to order of preference based on criteria concerning the offered services. The user agent matches these against preferences expressed in the metadata associated with the user and makes a choice. The user agent then makes the appropriate recommendation to the end-user who in turn decides to 'accept the deal' or not.

Ambient Computing

The concept of ambient computing implies that the computing environment is always present and available in an even manner. The concept of pervasive computing implies that the computing environment is available everywhere and is 'into everything'. The concept of mobile computing implies that the end-user device may be connected even when on the move. In general usage of the term, ambient computing implies both pervasive and mobile computing. The idea, then, is that an end-user may find herself connected (or connectable – she may choose to be disconnected) to the computing environment all the time.

The computing environment may involve information provision (access to database and web facilities), office functions (calendar, email, directory), desktop functions (word processing, spreadsheet, presentation editor), perhaps project management software and systems specialised for her application needs – accessed from her end-user device connected back to 'home base' so that her view of the world is as if at her desk. In addition entertainment subsystems (video, audio, games) should be available.

A typical configuration might comprise:

- a headset with earphone(s) and microphone for audio communication, connected by bluetooth wireless local connection to
- a PDA (personal digital assistant) with small screen, numeric/text keyboard (like a telephone), GSM/GPRS (mobile phone) connections for voice and data, wireless LAN connectivity and ports for connecting sensor devices (to measure anything close to the end-user) in turn connected by bluetooth to
- an optional notebook computer carried in a backpack (but taken out for use in a suitable environment) with conventional screen, keyboard, large hard disk and connectivity through GSM/GPRS, wireless LAN, cable LAN and dial-up telephone.

The end-user would perhaps use only (a) and (b) (or maybe (b) alone using the built in speaker and microphone) in a social or professional context as mobile phone and 'filofax', and as entertainment centre, with or without connectivity to 'home base' servers and IT environment. For more traditional working requiring keyboard and screen the notebook computer would be used, probably without the PDA. The two might be used together with data collection validation / calibration software on the notebook computer and sensors attached to the PDA. The balance between that (data, software) which is on servers accessed over the network and that which is on (one of) the end-user device(s) depends on the mode of work, speed of required response and likelihood of interrupted connections. Clearly the GRIDs environment is ideal for such a user to be connected. Such a configuration is clearly useful for a 'road warrior' (travelling salesman), for emergency services such as firefighters or paramedics, for businessmen, for production industry managers, for the distribution / logistics industry (warehousing, transport, delivery), for scientists in the field…. and also for leisure activities such as mountain walking, visiting an art gallery, locating a restaurant or visiting an archaeological site.

3. THE KEY IT CHALLENGES

3.1 Metadata

Since metadata is critically important for interoperation and semantic understanding, there is a requirement for precise and formal representation of metadata to allow automated processing. Research is required into the metadata representation language expressivity in order to represent the entities user, source, resource. For example, the existing Dublin Core Metadata standard [2] is machine-readable but not machine-understandable, and furthermore mixes navigational, associative descriptive and associative restrictive metadata. A formal version has been proposed [4] and updated [5].

3.2 Agents

There is an interesting research area concerning the generality or specificity of agents. Agents could be specialised for a particular task or generalised and configured dynamically for the task by metadata. Furthermore, agents may well need to be reactive and dynamically reconfigured by events / messages. This would cause a designer to lean towards general agents with dynamic configuration, but there are performance, reliability and security issues. In addition there are research issues concerning the syntax and semantics of messages passed between agents and brokers to ensure optimal representation with appropriate performance and security.

3.3 Brokers

A similar research question is posed for brokers – are they generalised and dynamic or specific? However, brokers have not just representational functions, they have also to negotiate. The degree of autonomy becomes the key research issue: can the broker decide by itself or does it solicit input from the external entity (user, source, resource) via its agent and metadata? The broker will need general strategic knowledge (negotiation techniques) but the way a broker uses the additional information supplied by the agents representing the entities could be a differentiating factor and therefore a potential business benefit. In addition there are research issues concerning the syntax and semantics of messages passed between brokers to ensure optimal representation with appropriate performance and security.

3.4 Security

Security is an issue in any system, and particularly in a distributed system. It becomes even more important if the system is a common marketplace with great heterogeneity of purpose and intent. The security takes the forms:

- prevention of unauthorised access: this requires authentication of the user, authorisation of the user to access or use a source or resource and provision or denial of that access. The current heterogeneity of authentication and authorisation mechanisms provides many opportunities for deliberate or unwitting security exposure;
- ensuring availability of the source or resource: this requires techniques such as replication, mirroring and hot or warm failover. There are deep research issues in transactions and rollback/recovery and optimisation;
- ensuring continuity of service: this relates to the former point but includes additional fallback procedures and facilities and there are research issues concerning the optimal (cost-effective) assurance of continuity.

In the case of interrupted communication there is a requirement for synchronisation of the end-user's view of the system between that which is required on the PDA and / or laptop and the servers. There are particular problems with wireless communications because of interception. Encryption of sensitive transmissions is available but there remain research issues concerning security assurance.

3.5 Privacy

The privacy issues concern essentially the trade-off of personal information provision for intelligent system reaction. There are research issues on the optimal balance for particular end-user requirements. Furthermore, data protection legislation in countries varies and there are research issues concerning the requirement to provide data or to conceal data.

3.6 Trust

When any end-user purchases online (e.g. a book from www.amazon.com) there is a trust that the supplier will deliver the goods and that the purchaser's credit card information is valid. This concept requires much extension in the case of contracts for supply of engineered components for assembly into e.g. a car. The provision of an e-marketplace brings with it the need for e-tendering, e-contracts, e-payments, e-guarantees as well as opportunities to re-engineer the business process for effectiveness and efficiency. This is currently a very hot research topic since it requires the representation in an IT system of artefacts (documents) associated with business transactions.

3.7 Interoperability

There is a clear need to provide the end-user with homogeneous access to heterogeneous information sources [3]. His involves schema reconciliation / mapping and associated transformations. Associated with this topic are requirements for languages that are more representative (of the entities / objects in the real world) and more expressive (in expressing the transformations or operations). Recent R&D [6],[7] has indicated that graphs provide a neutral basis for the syntax with added value in graph properties such that structural properties may be used.

3.8 Data Quality

The purpose of data, especially when structured in context as information, is to represent the world of interest. There are real research issues in ensuring this is true – especially when the data is incomplete or uncertain, when the data is subject to certain precision, accuracy and associated calibration constraints or when only by knowing its provenance can a user utilise it confidently.

3.9 Performance

The architecture opens the possibility of knowing the characteristics of data / information, software and processing power on each node, and thus generating optimal execution plans. Refinements involve data movement (expensive if the volumes are large) or program code movement (security implications) to appropriate nodes.

4. NGG: NEXT GENERATION GRID

4.1 Introduction

Based partly on work by the author with the EC DG INFSO Unit for Environmental Science, the EC set up a GRIDs Unit under Wolfgang Boch late in 2002. The unit organised a workshop in January 2003 where the author gave a keynote talk. A group of experts, including the author, was set up and produced the NGG report in June 2003 which was published as an adjunct to the Framework 6 calls for proposals. In March 2004 the group was reconvened; their first conclusion was that the NGG report required no changes but perhaps some annexes to cover what had happened since: the accepted proposals for projects in the area and some events worldwide.

4.2 The NGG Principles

The NGG Report includes the following desirable properties of a NGG: Transparent and reliable; Open to wide user and provider communities; Pervasive and ubiquitous; Secure and provide trust across multiple administrative domains; Easy to use and to program; Persistent; Based on standards for software and protocols; Person-centric; Scalable; Easy to configure and manage.

These properties are much different from those of the US GRID and are much more aligned with the requirements of science, business and culture. In particular the architectural principles, end-user principles and software development principles derived from these properties are well-suited to the needs of environmental science. It is the opinion of the author that the requirements are such that exising third generation GRID technology will not satisfy the requirement, and even great extensions to it will not satisfy the requirement. The way forward is to design an architecture based on the properties of NGG and implement it, providing backward-compatible paths to allow integration of the architecturally inferior GRID technology.

4.3 NGG for environmental science

The current EC-funded projects embracing NGG are developing the appropriate middleware using architectural principles based on those in the original UK paper and described above. Thus, environmental science – like all the other applications – will benefit from the provision of a uniform environment for the end-user over data, information, knowledge, computation and detector / sensor access. This will open up a new world for environmental science: in particular the integration of information and computational models from different environmental disciplines will allow for much more accurate simulation and modelling. Furthermore, the provision of advanced systems with detectors on PDAs and an immediate feedback from a control centre will improve data quality. The provision of cooperative working environments based on shared data, information, knowledge and software will improve the output quality and quantity of environmental science. Finally, the integration with digital library technology will improve the scientific process and the relationship of publications to software and data.

5. ACKNOWLEDGEMENT

Some of the material presented here has appeared in previous papers by the author. Although the author remains responsible for the content, many of the ideas have come from fruitful discussions not only with the author's own team at CCLRC-RAL but also with many members of the UK science community and the UK Computer Science / Information systems community. The author has also benefited from discussions in the contexts of ERCIM (www.ercim.org), W3C (www.w3.org) and the EC NGG Group.

6. REFERENCES

(1) Foster and Kesselmann (EDS). The Grid: Blueprint for a New Computing Infrastructure. Morgan-Kauffman 1998

(2) K G Jeffery. 'Metadata': in Brinkkemper,J; Lindencrona,E; Solvberg,A: 'Information Systems Engineering' Springer Verlag, London 2000. ISBN 1-85233-317-0.

(3) K G Jeffery, E K Hutchinson, J R Kalmus, M D Wilson, W Behrendt, C A Macnee, 'A Model for Heterogeneous Distributed Databases' Proceedings BNCOD12 July 1994; LNCS 826 pp 221-234 Springer-Verlag

(4) K G Jeffery: 'An Architecture for Grey Literature in a R&D Context' Proceedings GL'99 (Grey Literature) Conference Washington DC October 1999

(5) A Asserson, K G Jeffery, 'Research Output Publications and CRIS' Proceedings CRIS2004 Conference, Leuven University Press (in press)

(6) Kohoutkova, J; 'Structured Interfaces for Information Presentation' PhD Thesis, Masaryk University, Brno, Czech Republic, 2003

(7) Skoupy,K; Kohoutkova,J; Benesovsky,M; Jeffery,K G: 'Hypermedata Approach: A Way to Systems Integration' Proceedings Third East European Conference, ADBIS'99, Maribor, Slovenia, September 13-16, 1999, Published: Institute of Informatics, Faculty of Electrical Engineering and Computer Science, Smetanova 17, IS-2000 Maribor, Slovenia,1999, ISBN 86-435-0285-5, pp 9-15

Workflow Management for Loosely Coupled Simulations

A. Hoheisel

Fraunhofer Institute for Computer Architecture and Software Technology (Fraunhofer FIRST), Berlin, Germany (andreas.hoheisel@first.fraunhofer.de)

Abstract: In the distributed computing domain, several approaches have been evolved that allow the user not only to execute single programs on single hardware resources but also to support workflow schemes that enable the composition and execution of complex applications on distributed resources. Within the Fraunhofer Resource Grid, we developed a Grid computing architecture for composing and executing loosely coupled simulations by means of an abstract dataflow and control flow description. We use a Petri-net-based workflow model that allows the graphical definition of arbitrary workflows with only few basic graph elements – just by connecting data and software components. The communication between the distributed simulation components is realized via file transfer in the actual implementation, but we plan to include tighter coupling schemes using Web Service technology. We validated our approach with applications from the environmental simulation domain, such as the Environmental Risk analysis and Management System ERAMAS.

Keywords: Grid Computing; Coupled Simulations; Workflow Management

1. INTRODUCTION

Currently, there is a high demand of building complex simulation systems that are composed of several distributed and coupled models. One reason for this trend is the increasing specialization of researchers and research organizations to a specific scientific topic on the one hand and the wish to combine the knowledge included in the models to an overall picture on the other hand. Environmental simulation projects often involve a variety of research groups from different organizations that are distributed all over the world. In the past, a common way to couple several models was to merge the models to one monolithic program and to use direct FORTRAN or C function or subroutine calls to realize the communication between the models. Nowadays, this approach is reaching its limits because of the complexity of the models and the difficulty to bundle the know-how of all the models in one working group. Furthermore, the implementations of the models require more and more computational power as well as data storage capacity, and the simulation models are often written in different programming languages for specific operating systems that run on special hardware platforms, such as Linux clusters or MS-Windows PCs.

One approach that promises to satisfy all these requirements is the new paradigm of distributed computing called *Grid computing*. The main vision of Grid computing is to realize a unified interface for arbitrary computational resources – including hardware, software and data – that everybody can use without having to care about the hardware infrastructure and the implementation details of the software components; just as easy as getting electricity through a standardized plug from the *electric power grid* [Foster and Kesselman, 2003]. While it is unsure if this overall vision will ever be accomplished, it can be said that the Grid computing hype is coming to maturity and that important parts of the Grid computing technology have reached production status. In contrast to *cluster computing*, Grid computing considers heterogeneous and non-reliable computing environments as well.

In the simulation community, the research and development of Grid computing techniques was first driven by the demand for computational power and data storage [NASA, 2004; Segal, 2000]; meanwhile nowadays organizational issues – such as building virtual organizations, enabling complex workflows and collaborative working, as

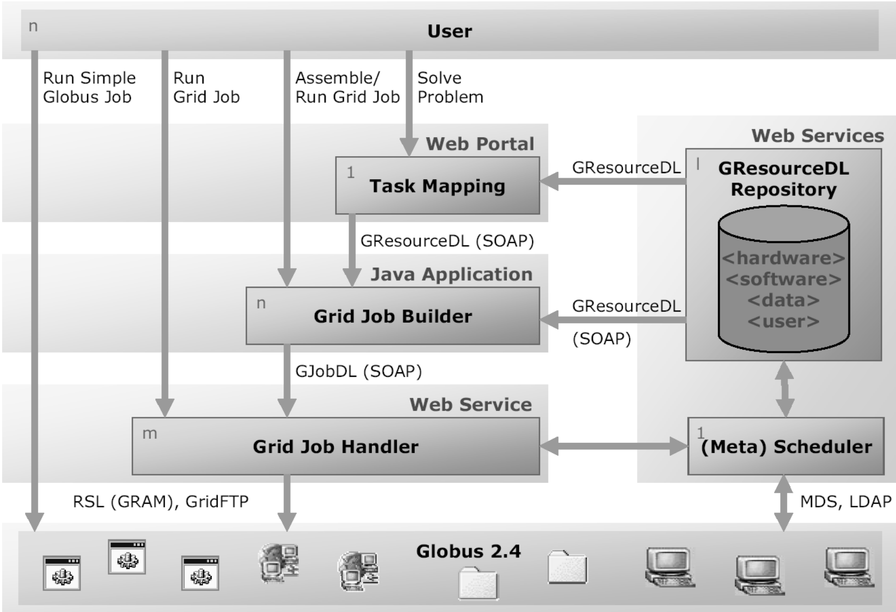


Figure 1. The layered Grid architecture of the Fraunhofer Resource Grid.

well as security, accounting, and billing aspects – are gaining more importance.

Several techniques have been established in the Grid community in order to define the workflow of coupled applications. A very promising approach is the usage of graphs for this purpose as they possess very intuitive ways of visualization that can be handled easily even by non-expert users. This article focuses on the workflow management system for loosely coupled simulation models being developed within the Fraunhofer Resource Grid (FhRG) [Fraunhofer, 2004]. In contrast to other workflow approaches which usually are based on directed acyclic graphs, the FhRG workflow is built on the more expressive formalism of Petri nets [Petri, 1962]. *Dynamic workflow graph refinement* is introduced as a technique to transform abstract workflow graphs into the concrete ones needed for execution and to automatically add fault tolerance to the workflows [Hoheisel and Der, 2003b].

The following section describes the architecture of the Fraunhofer Resource Grid. Section 3 focuses on the techniques we use to define the workflow of coupled simulation models and their data. The enactment of the workflow on a Grid computing environment is presented in section 4, whereas section 5 shortly lists some applications that are being ported to the Fraunhofer Resource Grid as case studies. I complete this paper with conclusions and future work in section 6.

2. FRAUNHOFER RESOURCE GRID

The Fraunhofer Resource Grid (FhRG) is a Grid initiative of five Fraunhofer institutes funded by the German federal ministry of education and research with the main objective to develop and to implement a stable and robust Grid infrastructure within the Fraunhofer-Gesellschaft, to integrate available resources, and to provide internal and external users with an easy-to-use interface for controlling distributed applications and services in the Grid environment [Fraunhofer, 2004].

The component environment supports loosely coupled software components where each software component represents an executable file that reads input files and writes output files. The execution of such a software component we call *atomic job*. The communication is realized via file transfer. Legacy code can be integrated easily using shell scripts to encapsulate the program. We plan to include *Web Service invocations* as atomic jobs in future releases of the FhRG framework in order to make it OGSA compatible [Foster et al., 2002]. Up to now, the workflow architecture does not support tight coupling schemes like CORBA [OMG, 2002], MPI [1997], or HLA, but tightly coupled applications can be included as a whole like an atomic job. Most of the software developed within the Fraunhofer Resource Grid will be made Open Source (GPL) under the label *eXeGrid* [2004].

Figure 1 depicts the Grid architecture of the Fraunhofer Resource Grid that is currently built on top of the Globus 2.4 toolkit [Globus, 2003]. The

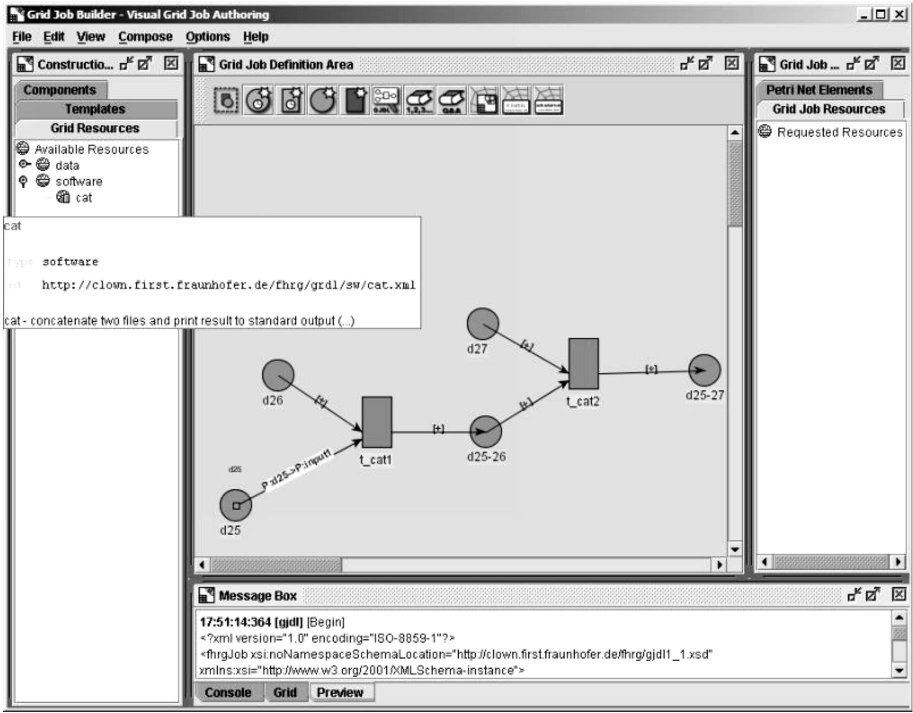


Figure 2. A screenshot of the Grid Job Builder, developed by Fraunhofer IGD, including a Grid resource browser (*left*), a composition panel for Petri-net-based workflows (*middle*), and a job inspector (*right*). The Grid Job Builder supports drag and drop to introduce new components to the Grid job workflow.

user has four alternatives to access Grid resources within this architecture:

1. The user can directly use the standard *Globus services* like GRAM or GridFTP in order to run simple Globus jobs (atomic jobs) on a specified Grid node.
2. If the user wants to run a predefined Grid job, e.g., a coupled simulation, he can use the *Grid Job Handler Web Service*. In this case, the user must provide an XML document that specifies the Grid job. The selection of suitable Grid resources is done during runtime based on current information.
3. The user may use the graphical *Grid Job Builder* to assemble and configure the resources to form a coupled Grid job.
4. If the user does not know which resources to use in order to solve his problem, he may invoke the *Task Mapping* of the web portal. There, the user navigates through a task tree in order to restrict the application area of the problem and to map it onto a suitable set of Grid resources.

3. COMPOSING COUPLED SIMULATIONS

We define the term *Grid job* as a Grid application that is composed of several Grid resources with a specified workflow. A Grid job can, e.g., represent a complex simulation run or sequential data processing steps, and it may induce a variety of single tasks (atomic jobs) that are indivisible components of a Grid job. According to our definition, *Grid resources* are either abstract classes or concrete instances of software, hardware or data.

There are several possibilities to provide a workflow management that coordinates the execution of Grid jobs. The workflow is either defined inherently by the software components themselves, or by software agents that act on behalf of the software components, resulting in a self-organizing or hard-wired Grid job. Another alternative is to define the workflow on a meta level on top of the software components, providing a complete view of the workflow. To describe this kind of workflow it is very important to have suitable semantics.

In our approach, we use a Petri-net-based workflow model that allows the graphical definition of arbitrary workflows with only few basic graph elements – just by connecting data and software components. Figure 2 shows a screenshot of the Grid Job Builder, a Java application providing a graphical user interface for assembling such Grid jobs [Jung et al., 2004]. The output of the Grid Job Builder is a *Grid Job Definition Language (GJobDL)* document, which defines the Grid job.

This GJobDL document can be saved as a file or transmitted directly to the Grid Job Handler Web Service in order to enact the workflow. The GJobDL description of a Grid job contains the resource descriptions of the basic resources that are required to define the Grid job and the model of the Grid job workflow using the concept of Petri nets [Petri, 1962] as shown in Figure 3.

The idea to use Petri nets to control the workflow of complex applications has been borrowed from the Graphical Simulation Builder that is being developed by the Potsdam Institute for Climate Impact Research (C. Ionescu, pers. comm.). Marinescu [2002] describes a similar approach in his book about internet-based workflow management.

Petri nets belong to a special class of directed graphs. The type of Petri nets we introduce here corresponds to the concept of Petri nets with individual tokens (colored Petri net) and constant arc expressions which are composed of places, denoted by circles (○), transitions, denoted by boxes (□), arcs from places to transitions (○→□), arcs from transitions to places (□→○), individual and distinguishable objects that flow through the net as tokens (•), an initial marking that defines the objects which each place contains at the beginning, and an expression for every arc that denotes an individual object. A place *p* is called input place (output place) of transition *t* if an arc from *p* to *t* (from *t* to *p*) exists. A brief introduction to the theoretical aspects of colored Petri nets can be found, e.g., in Jensen [1994]. The standardization of the Petri net concept is currently in progress as an ISO 15909 committee draft [ISO, 1997]. Van der Aalst and Kumar [2000] give an overview of how to describe different workflow patterns using Petri nets. Petri nets are suitable to describe the sequential and parallel execution of tasks with or without synchronization; it is possible to define loops and the conditional execution of tasks.

We use Petri nets not only to *model*, but furthermore to *control* the workflow of Grid jobs. In most cases, the workflow within Grid jobs is equivalent to the dataflow, i.e., the decision when to execute a software component is taken by means of availability of the input data. Therefore, the tokens of the Petri net represent real data that is exchanged between the software components. In this case, we use Petri nets to model the interaction between software resources represented by software transitions, and data resources represented by data places. In some cases, however, the workflow is independent from the dataflow, and in addition to the data places and software transitions we have to introduce control places and control transitions. The corresponding tokens contain the exit status of the process (e.g., done, failed). Control transitions evaluate logical conditions. For further details about this Petri net approach refer to Hoheisel and Der [2003a; 2003b].

We introduced a dedicated XML syntax – similar to the Petri Net Markup Language (PNML) developed by Weber and Kindler [2002] – in order to describe Petri nets within the GJobDL. The job description consists of the declaration of the places, transitions, and arcs that build the Petri net of the Grid job. Transitions and places may be linked to external or internal resource descriptions. Control transitions may possess conditions that are evaluated prior to the firing of activated transitions.

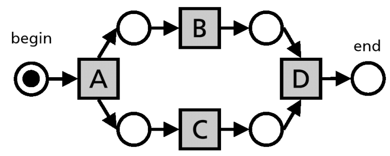


Figure 3. A Petri net representing the parallel execution of atomic jobs *B* and *C* and the sequential execution of *A, B, D* and *A, C, D*.

4. EXECUTING COUPLED SIMULATIONS

The Grid Job Handler is responsible for the enactment of the coupled simulation. Therefore, the Grid Job Handler parses the Grid job description, resolves the dependencies between the Grid resources, and searches for sets of hardware resources that fulfill the requirements of each software component. A meta scheduler (see Figure 1) selects the best-suited hardware resource of each set of matching hardware resources according to a given scheduling policy (fastest, cheapest, etc.). In the current implementation, the Grid Job Handler maps the resulting atomic jobs onto the Globus Resource Specification Language (RSL) [Globus, 2000] and submits them via GRAM to the corresponding Grid nodes. For the communication between the Grid Job Handler and the Globus Grid middleware, we use a patched version of the Java Commodity Grid Kit [von Laszewski et al., 2001]. The Grid Job Handler itself is deployed as a Web Service with possibilities to create, run and monitor Grid jobs remotely. The desktop version of the Grid Job Handler includes a graphical user interface (see Figure 4) and additional command line tools.

The refinement model of the Petri net theory allows substituting parts of a Petri net by new sub Petri nets. The Grid Job Handler takes advantage of this feature and supplements the workflow during runtime by introducing additional tasks that are necessary to complete the Grid job. The user is not

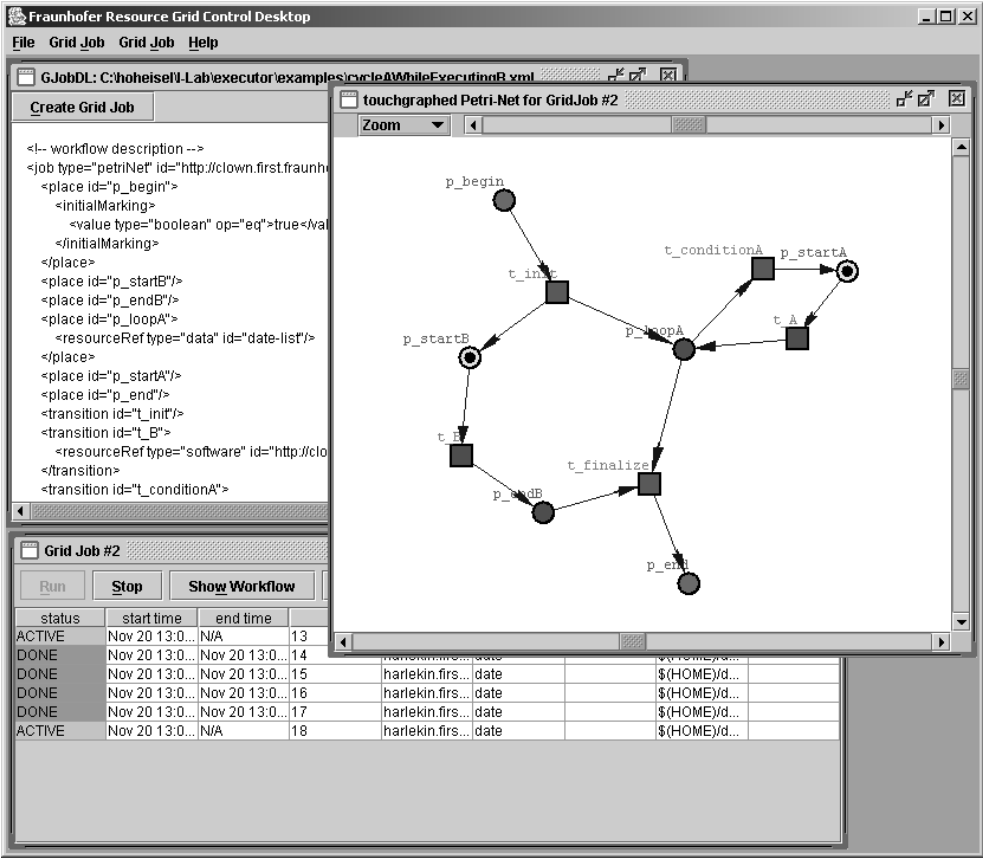


Figure 4. Screenshot of the graphical Grid Job Handler user interface. The upper left panel displays an excerpt of the GJobDL document. The right panel shows a graphical representation of the corresponding Grid job workflow. The lower left panel lists the atomic jobs that are induced by the Grid job with their actual status.

required to model every detail of the workflow – he just has to include the essential transitions and places that are related to the software components and the data he wants to include in his Grid job. Additional tasks that have to be invoked due to specific properties of the Grid infrastructure (e.g., network topology) are detected by the Grid Job Handler and considered by automatically introducing additional transitions and places before or during runtime of the Grid job. In the current version of the Grid Job Handler, data transfer tasks and software deployment tasks are automatically added to the workflow if they are missing in the initial Grid job definition provided by the user. A data transfer task may be introduced to transfer files that are not available on the remote computer via GridFTP. A software deployment task may be introduced to install software components on a remote computer using standard Globus protocols. Further Petri net refinements could concern authorization, accounting, billing and fault management tasks.

5. CASE STUDIES

An example Grid application that takes advantage of the described workflow framework is the *Environmental Risk Analysis and Management System (ERAMAS)*, developed by Fraunhofer FIRST in collaboration with the *Ingenieurbüro Beger für Umweltanalyse und Forschung* and the *Dresdner Grundwasser Consulting GmbH* [ERAMAS, 2004; Unger et al., 2003]. ERAMAS is a simulation-based analysis framework for risks caused by carcinogenic and chemically toxic substances that are released during accidents in industrial installations, the transport of dangerous goods or by terrorist attacks. It is designed to be employed for real-time emergency management as well as for preliminary studies concerning approval procedures and emergency plans.

Other applications that are recently ported to the Fraunhofer Resource Grid are MAGMASOFT [2004], LUMOS [2004], and EIQU.

6. CONCLUSIONS AND FUTURE WORK

Within the Fraunhofer Resource Grid, we have developed a workflow management architecture that is suitable for composing and executing loosely coupled simulation models on a Grid computing environment. The Petri-net-based workflow model seems to be a very promising approach that allows the definition of arbitrary workflows with only three different components: transitions, places and arcs. This enables the easy orchestration of complex workflows, including conditions and loops and regarding the dataflow as well as the control flow of coupled simulations. Within this framework, distributed applications can be defined independently from the hardware infrastructure, just by connecting simulation models and data.

Future work that we planned in the domain of workflow management for coupling simulation models considers the implementation of tight coupling schemes on basis of the Web Service standard. Another important issue that we did not cover so far with the framework presented in this paper is the abstraction from specific data formats. In the present implementation it is up to the user to care about data formats and to ensure that the software components, which interchange data within a Grid job, use compatible data formats. It is planned to make use of special meta data in order to achieve automatic data conversion.

7. REFERENCES

van der Aalst, W., and Kumar, A., XML based schema definition for support of inter-organizational workflow, 2000.

ERAMAS homepage: http://www.eramas.de/, 2004.

eXeGrid homepage: http://www.exegrid.net/, 2004.

Foster, I., and Kesselman, C. (eds.), The Grid, Morgan Kaufmann Publishers, Inc., 2003.

Foster, I., Kesselman, C., Nick, J., and Tuecke, S., The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, Open Grid Service Infrastructure WG, Global Grid Forum, 2002.

Fraunhofer Resource Grid homepage, http://www.fhrg.fhg.de/, 2004.

Globus Project, The Globus Resource Specification Language RSL v.1.0, http://www-fp.globus.org/gram/rsl_spec1.html, 2000.

Globus Project, The Globus Toolkit 2.4, http://www.globus.org/gt2.4/download.html, 2003.

Hoheisel, A., and Der, U., An XML-based Framework for Loosely Coupled Applications on Grid Environments. ICCS 2003. *Lecture Notes in Computer Science*, 2657, 245–254, 2003a.

Hoheisel, A., and Der, U., Dynamic Workflows for Grid Applications, *Proceedings of the Cracow Grid Workshop '03*, Cracow, Poland, 2003b.

ISO 15909, High-level Petri Nets – Concepts, Definitions and Graphical Notation. Committee Draft ISO/IEC 15909, Version 3.4, 1997.

Jensen, K., An Introduction to the Theoretical Aspects of Coloured Petri Nets, *Lecture Notes in Computer Science*, 803, 230–272, 1994.

Jung, C., Einhoff, M., Noll, S., and Schiffer, N., Grid Job Builder – A Workflow Editor for Computing Grids, *IEEE Information Technology*, Coding and Computing, ITCC, 2004.

von Laszewski, G., Foster, I., Gawor, J., and Lane, P., A Java Commodity Grid Kit, *Concurrency and Computation: Practice and Experience*, 13, 643–662, 2001.

LUMOS project homepage, http://www.projekt-lumos.de/, 2004.

MAGMASOFT homepage, http://www.magmasoft.com/, 2004.

Marinescu, D. C., Internet-Based Workflow Management – Toward a Semantic Web, Wiley, 2002.

Message Passing Interface Forum, MPI-2: Extensions to the Message-Passing Interface, http://www.mpi-forum.org/docs/, 1997.

NASA Information Power Grid. http://www.ipg.nasa.gov/, 2004.

Object Management Group, Common Object Request Broker Architecture: Core Specification, 2002.

Petri, C. A., Kommunikation mit Automaten, Ph.D. dissertation, Bonn, 1962.

Segal, B., Grid Computing: The European Data Project, *IEEE Nuclear Science*, Symposium and Medical Imaging Conference. Lyon, France, 2000.

Unger, S., Hoheisel, A., Beger, E., and Beims, U., ERAMAS – Analyse- und Managementsystem von schadstoffbedingten Umweltrisiken *Technische Überwachung*, 44 (4), 46–49, 2003.

Weber, M., Kindler, E., The Petri Net Markup Language. In: Petri Net Technology for Communication Based Systems. *Lecture Notes in Computer Science*, Advances in Petri Nets, 2002.

Testing the Impact of Management Scenarios on Water Quality Using an Ecosystem Model

Gideon Gal, Arkadi Parparov, Udi Wagner, and Tzahi Rozenberg

Israel Oceanographic and Limnological Research, Ltd. *Kinneret Limnological Laboratory, PO Box 447 Migdal, Israel 14950*

Abstract: Lake Kinneret, the only natural freshwater lake in Israel, provides approximately 30% of the national demand for freshwater and is therefore closely monitored and managed by the Israeli water authorities. Domestic water supply and sustaining the lake's ecosystem are prime objectives of the water resource's management. The past decade has been characterized by over-pumping from the lake, aggravated by several years of reduced rainfall, changes to the watershed and alterations to the usual algal succession and seasonal dynamics. The changes have led to concerns regarding the stability of the ecosystem. One approach to studying the environmental effects of different management policies is by the application of a model-based decision support system. The system involves the combination of the hydrodynamical model DYRESM, and the biological-chemical model CAEDYM. In addition, a water quality index (WQI) developed at the Kinneret Limnological Laboratory has been integrated into the model output and the graphical interface. The integration provides managers and scientists, alike, with a readily available and easy to use tool for evaluating the impact of management scenarios on the lake ecosystem. In this study, we used the integrated system to examine the following management measures on lake water quality: 1. variations in P loading into the lake and, 2. variations in water retention time due to changes in lake level. These examples are used to illustrate the advantages and constraints of the proposed approach.

Keywords: Ecosystem modelling; CAEDYM; Water Quality Index; lake level; P loading

1. INTRODUCTION

Water quality models are an excellent tool for the study of large inland water bodies due to the physical, biological and chemical complexity of these ecosystems. Hydrodynamic and water quality models of lake ecosystems have become more abundant and sophisticated since the early models developed in the 1970s as a result of increased demand and improved computing technology [J¬rgensen, 1995]. As a consequence, the use of water-quality models in planning, designing and testing management strategies has become increasingly common since the 1980's [Friedman, et al., 1984]. The use of such models allow the resource manager to study, for example, the causes and consequences of lake eutrophication and potential remedial actions [Tufford and McKellar, 1999] and the potential impact of management decisions on the ecosystem. Hence, models can play both a role in enhancing the ecological understanding of a lake and as a management tool [Hˀkanson and Peters, 1995]. The convenience and value of such management tools depend, to a large degree, on the ability to convey the information they contain in a concise, accurate, understandable and efficient manner. The means for conveying such information between all parties requires a commonly shared language with user-friendly tools.

The task of optimal water resource management requires predefined standards of water quality as integral components of the optimization criteria [Kalceva, et al., 1982; Straskraba and Gnauck, 1985]. Thus, water quality should be expressed in a measurable form to allow analytical expression of relationships between water quality and various driving factors that can affect water quality (e.g., external nutrient loading). The methodological approaches allowing quantitative expression of water quality are well known [Brown, et al., 1970; Horton, 1965; Ott, 1978; D.G. Smith, 1987; D.G. 1990 Smith, 1990]. Practical implementations of these concepts have been concentrated mainly in the US, while examples of quantitative estimates of lake water quality in Europe have been relatively rare. In many cases, particularly in the US, trophic classification was used as a gross indicator of water quality. For example, oligotrophic lakes are considered to be of "good" water quality; and eutrophic lakes of "bad" water quality [Fozzard, et al., 1999]. Indeed, the Clean Water Act of 1972 (section 314) stipulates that all lakes in the US be *classified* according to their trophic state [EPA, 1974]. Such a system of trophic classification allows one to rank lakes according to their productivity, oxygen and nutrient regime and thereby evaluate water quality changes (i.e., trophic state changes) in response to various driving forces.

Water quality assessment, therefore, provides water resource managers the means to ask 1) is the current water quality “good” or “bad”? and, 2) will a management scenario under consideration (e.g., reduction in lake level, changes in nutrient loading,) result in an improvement or deterioration in water quality?

Lake Kinneret is a warm, medium size natural monomictic lake located in the north of Israel. It is a meso-eutrophic lake with a mean annual primary production of 650 g C m^{-2} [Berman, et al., 1995] and Secchi depths varying from 0.7 to 5.5 m. A prominent biological feature of the lake has been the spring bloom of the dinoflagellate *Peridinium gatunense* [Berman, et al., 1992; U. Pollingher and Serruya, 1978], though since 1994, the lake has exhibited uncharacteristic developments in the phytoplankton assemblage [Berman, et al., 1998], including the first-ever bloom of a potentially toxic, N_2-fixing cyanobacteria [U. Pollingher, et al., 1998]. Further details on the basic ecology of Lake Kinneret can be found in Serruya [1978]. Since 1965 (the inception of National Water Carrier of Israel), and with the economic and population growth within Israel, the main uses of Lake Kinneret waters have been consumptive — domestic “drinking water” supply and agricultural irrigation. Owing to its historical significance and climatic characteristics, the lake has large recreational potential. Between the mid-1990s and 2002 there was an increasing deficit between inflows to and withdrawals from the lake leading to the lowest lake level on record in 2002 raising concerns that the ongoing low lake level will lead to a long term decline in water quality and ecosystem stability. Given its role as a major source for drinking water, correct management of Lake Kinneret is vital for sustaining ecosystem stability and water quality.

In order to assist in sustaining the lakes' ecosystem stability and water quality the Israeli Water Commissioner funded a large scale project for the development of a model based management tool. We are currently involved in the implementation of the model-based tool to the lake. The management tool includes a suite of models developed at the Centre for Water Research (CWR) at the University of Western Australia. The models consist of one-dimensional (DYRESM) and three-dimensional (ELCOM) hydrodynamic drivers and a biogeochemical model (CAEDYM).

In addition to the suite of models, the management tool includes the application of a water quality index [WQI, Hambright, et al., 2000; Parparov and Hambright, 1996] developed at the Kinneret Limnological Laboratory (KLL). The WQI defines the permissible ranges, in the epilimnion (0-15 m), for a set of parameters as determined from a long-term database. The permissible ranges were determined under the assumption that between 1969 and 1992 the lake exhibited acceptable water quality and a stable ecosystem. Therefore, acceptable ranges were determined based on the observed ranges, during this period, of a series of variables. Data collected as part of the on-going routine monitoring program are plotted in the framework of the WQI, thereby providing an indication of the status of the lake in terms of the selected parameters.

The main objective of this study was to illustrate an approach in which the benefits of an intergrated support system, that combines ecosystem models along with water quality assessment and a graphical interface, enhances the capabilities of resource managers. We demonstrate this by attempting to determine the permissible ranges of management measures (water pumping regimes and phosphorous loading) based on their impact on chlorophyll a (CHL) and total phosphorous (TP) concentrations.

2. METHODS

2. 1 Models

The one-dimensional model DYRESM was used as the hydrodynamic driver for the biogeochemical model CAEDYM. DYRESM has been used for predicting the vertical distribution of temperature, salinity and density in lakes and reservoirs. First developed over 20 years ago [Imberger and Patterson, 1981], the model recently underwent revisions and was applied successfully to Lake Kinneret [Gal, et al., 2003]. It is assumed that the water bodies comply with the one-dimensional approximation in that the destabilizing forcing variables (wind, surface cooling, and plunging inflows) do not act over prolonged periods of time. DYRESM has been used for simulation periods extending from weeks to decades. Thus, the model provides a means of predicting seasonal, annual and inter-annual variation in lakes and reservoirs, as well as sensitivity testing to long-term changes in environmental factors or watershed properties. The model requires a series of input data that include the meteorological conditions, inflows and withdrawals to and from the lake over the period of simulation in addition to initial temperature and salinity profiles.

Table 1. Initial lake levels, in meters below sea level (mbl), used in the lake level scenarios.

Simulated lake levels (mbl)	
Simulation ID	Lake Level
A	210
B	212
C	214
D	215.5
E	217.5
F	220.9

CAEDYM is an aquatic ecological model that may be run independently or coupled with DYRESM. CAEDYM (Figure 1) consists of a series of mathematical equations representing the major biogeochemical processes influencing water quality [Romero, et al., In press]. The model contains a series of process descriptions for primary production, secondary production, nutrient and metal cycling, oxygen dynamics and the movement of sediment. The model requires the initial vertical distribution of all the simulated state variables and information on the nutrient loading into the lake over the period of the simulation. In the current configuration, the model included five phytoplankton groups representing the major taxa in the lake (*Peridinium gatunense*, *Aulacoseira granulata*, a general nanoplankton group, *microcystis aeruginosa*, and *Aphanizomeno ovalisporum*). In addition, we included 3 size based zooplankton groups; predatory copepods, macrograzers and micrograzers. While the calibration and validation of the model to the lake still requires minor adjustments, the model successfully simulates the seasonal and inter-annual variability observed in the lake including the spring *Peridinium* blooms [Gal, et al., 2002].

2.2 Simulations

Long term DYRESM-CAEDYM simulations of Lake Kinneret were conducted using 20 year template data running from 2010-2029. A one-year template dataset was constructed by calculating daily average values of inflows and withdrawals and hourly meterological conditions based on data collected at the Kinneret Limnological Laboratory between 1997 and 2001. Inflow data included daily volumes, water temperature, nutrient loading (NO3, NH4, PO4), and salinity for the six main surface and subsurface tributaries to the lake. Daily withdrawal volumes included pumped and overflow values at three different sites. The one-year template was then multiplied 20 times (and corrected for leap years) to create a 20-year input dataset.

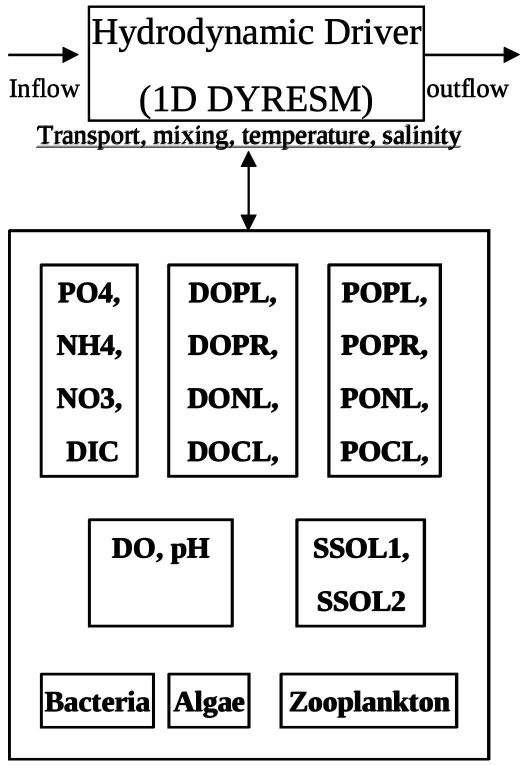


Figure 1. The main state variables simulated by CAEDYM including the dissolved inorganic and the dissolved and particulate organic fractions of C, N, and P, and biological groups. For sake of clarity, the interactions between the various boxes are not shown.

In order to evaluate the impact of lake level on the ecosystem the simulations were conducted at a number of base levels. Base levels were lake levels at which the lake was maintained from year to year throughout the simulations. To do so, the initial lake level was set at the desired level and the inflow volumes were slightly modified in order to maintain a balanced water budget. Six base level simulations were conducted (Table 1). The levels corresponded to the observed and permitted lake levels with the exception of the two lowest lake levels that was used as an example of extreme conditions.

The impact of changes to P loading on the lake was examined by varying P loading to the lake, via the inflows. This was accomplished by applying a multiplication factor to base inflows concentration. The multiplication factor ranged between 0.001- 10. The P loading scenarios were conducted at a series of lake levels in order to assess possible interactions between P loading and lake level. Here we present only the results of changes at one level (level C, Table 1).

the winter-spring period (Jan–Jun) and the summer-autumn period (Jul- Dec). The results presented in this study are based on half-year averages, of the top 15 m of the water column, coinciding with the winter-spring (W-S) and summer- autumn (S-A) period. The results are based on the last four years of simulations, thereby representing a steady-state of the model.

For sake of clarity, we present here results of only two of the WQI variables in order to demonstrate the rating approach used here; CHL in the case of the lake level variations and TP in the case of variations to P loading.

3. RESULTS AND DISCUSSION

Base level simulation results indicated that the lake ecosystem was relatively insensitive to changes in lake level, down to a threshold level. This is clearly indicated by plotting rating values for CHL as a function of the management activity (Figure 3).

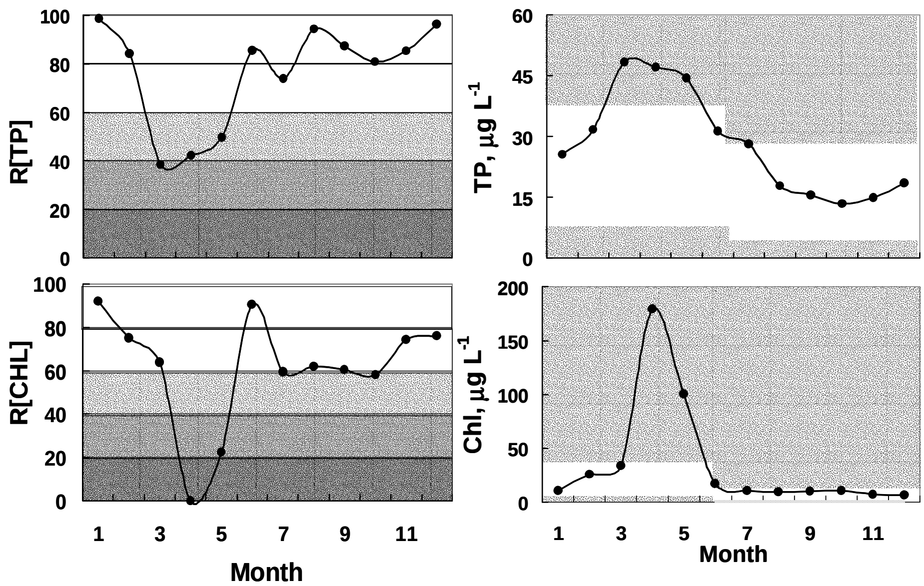


Figure 2. Example of WQI plots: Rating values and concentrations of TP and chlorophyll (CHL) in Lake Kinneret during 2003. Values represent monthly mean values for top 15 m.

2.3 WQI

The WQI provides a graphical display of a series of variables (see Figure 2 for example). The WQI rating values range between 0-100 where values over 60 are considered acceptable. The rating curves for each variable, that convert concentrations to a rating value, differ between

Until the threshold level was reached, biological and chemical variables changed only moderately between simulations. CHL rating values, for example, remained within the permissible levels and varied between 67-92 during the winter-spring and 71-92 during the summer-fall. The drastic change occurred once the base lake level was lowered below approximately 218 mbl resulting in rapid declining rating values.

The existence of a threshold in the reaction of a lake ecosystem to changes has been discussed in the past [Parparov and Hambright, 1996] although, to date, no clear evidence has been put forward. The predicted moderate decline in water quality due to a reduction in lake level, as indicated by the changes in CHL concentrations and rating values are consistent with patterns that have occurred in Lake Kinneret, in the past, as result of the fall in lake level. In order to perform a better comparison however, of the predicted and observed patterns in the lake, real input data and the entire suite of the WQI variables should be used.

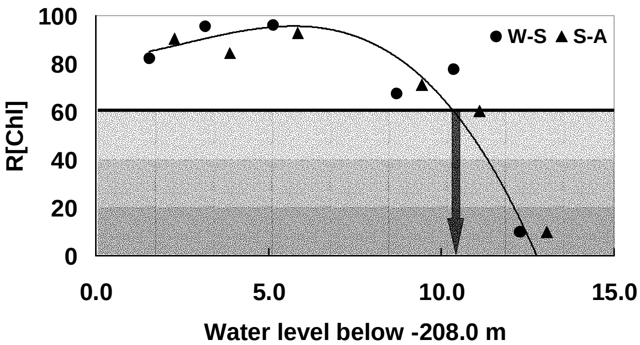


Figure 3. The impact of water level lowering as indicated by the rating values of CHL. Rating values >60 are defined as acceptable. The threshold value is demarked by the vertical arrow. W-S represents the period of Jan-Jun period and S-A represents the second half of the year (Jul-Dec).

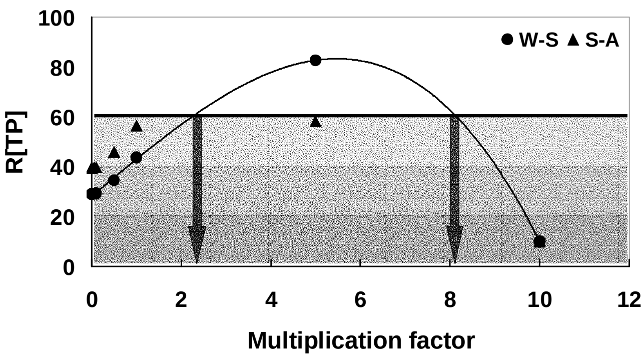


Figure 4. The impact of changes in P loading on the lake water quality as indicated by the rating values of TP. The multiplication factor represents the changes to the base run P loading. Rating values >60 are defined as acceptable and the arrows demark the region of acceptable values. W-S represents the period of Jan-Jun period and S-A represents the second half of the year (Jul-Dec).

Results of simulations, in which P loading into the lake varied, suggested a different pattern from that observed due to changes in lake level (Figure 4). As expected a peak in rating values occurred at an intermediate level indicating that P loading values higher or lower lead to a decline in rating values. Surprisingly, the peak occurred during the W-S seasons as P loading 5 times greater then the loading used in the base run. In the case of he S-A, a peak was not observed due to lack of resolution in the loading values, however, it would have most likely occurred at a loading level approximately 3 times greater than the base run.

As opposed to the impact of the water level scenarios, the P loading simulations do not indicate any clear cut threshold affect but rather a range of acceptable values preceded and followed by values that lead to undesirable rating values. It was expected that at both low and high levels of loading rating values would be low. The reason for this lies with the underlying assumptions governing the WQI rating system. The WQI rating system was constructed assuming that high values of the WQI will be achieved when the concentrations are similar to those observed in the lake prior to 1991 and, as opposed to the traditional trophic classification, .not necessarily when low concentrations exist

4. CONCLUSIONS

We presented examples of management scenarios and the outcome of the scenarios in terms of two WQI variables. The results suggested threshold type behaviour of the predicted chlorophyll concentration due to changes in lake level while variation in P loading resulted in an intermediate range of acceptable TP values. It should be noted, however, that as the input data used do not accurately reflect actual data the precise location of the threshold and peak values should be further checked. Nevertheless, the results demonstrate the effectiveness in the combination of ecosystem modelling and indexing of water quality variables in the form of the WQI. The combination presents an easy to use and effective tool for evaluating the impact of management measures on a lake ecosystem. It further, provides the resource manager and ecosystem modeller with a tool for estimating the optimal operating ranges of the various measures. While our presentation of the model-based decision support system focuses on long-term scenarios, it should be noted that the system is currently implemented on simulations ranging between 2 and 20 years.

5. ACKNOWLEDGEMENTS

Funding for this project was provided by the Israel Water Commissioner. We thank the Mekorot water company for providing inflow and withdrawal data. We thank an anonymous reviewer for comments.

6. REFERENCES

Berman, T., U. Pollingher and T. Zohary. (1998) A short history of stability and change in phytoplankton populations in Lake Kinneret. *Isr. J. Plant Sci.*, **46**, 73-80.

Berman, T., L. Stone, Y. Z. Yacobi, B. Kaplan, M. Schlichter, A. Nishri and U. Pollingher. (1995) Primary production and phytoplankton in Lake Knneret: a long-term record (1972-1993). *Limnology and Oceanography*, **40**, 1064-1076.

Berman, T., Y. Z. Yacobi and U. Pollingher. (1992) Lake Kinneret phytoplankton stability and variability during twenty years 1970-1989. *Aquat Sci*, **54**, 104-127.

Brown, P. M., N. I. McClelland, R. A. Deninger and R. G. Tozer. (1970) A water quality index - do we dare? *Water&Sewage Word*, **Oct. 1970**, 339-343.

Fozzard, I., R. Doughty, R. C. Ferrier, T. Leatherland and R. Owen. (1999) A quality classification for management of Scottish standing waters. *Hydrobilogia*, **395/396**, 443-453.

Friedman, R., C. S. Ansell, S. Diamond and Y. Y. Haimes. (1984) The use of models for water resources management , planning and policy. *Water Resour. Res.*, **20**, 793-802.

Gal, G., J. Imberger, T. Zohary, J. Antenucci, A. Anis and T. Rosenberg. (2003) Simulating the thermal dynamics of Lake Kinneret. *Ecological Modelling*, **162**, 69-86.

Gal, G., T. Zohary and D. Horn. (2002) Lake Kinneret Water Quality Management and Optimization Support System - Phase 2; 2001 Annual Report to the Israel Water Commission, pp. 215. Migdal, Israel: Kinneret Limnological Laboratory.

Hambright, K. D., A. Parparov and T. Berman. (2000) Indices of water quality for sustainable management and conservation of an arid region lake, Lake Kinneret (Sea of Galilee), Israel. *Aquatic Conservation : Marine & Freshwater Ecosystems*, **10**, 393-406.

Horton, R. K. (1965) An index-number system for rating water quality. *Journal of the water pollution control federation*, **37**, 300-306.

Hakanson, L. and R. H. Peters. (1995) *Predictive limnology : methods for predictive modelling*. Amsterdam, The Netherlands: SPB Academic Publishing.

Imberger, J. and J. C. Patterson. (1981) A dynamic reservoir simulation model-DYRESM: 5. In *Transport models for inland and coastal waters* (ed. H. B. Fischer), pp. 310-361. New York: Academic Press

Jorgensen, S. E. (1995) State of the art of ecological modelling in limnology. *Ecological Modelling*, **78**, 101-115.

Kalceva, R., O. J. V., Z. Schindler and M. Straskraba. (1982) An optimization model for the economic control of reservoir eutrophication. *Ecological Modelling*, **17**, 121-128.

Ott, W. (1978) Water quality indices: a survey of indices used in the United States, pp. 128: Environmental Monitoring Series.

Parparov, A. and K. D. Hambright. (1996) A proposed framework for the management of water quality in arid-region lakes. *Internationale Revue der Gesamten Hydrobiologie*, **81**, 435-454.

Pollingher, U., O. Hadas, Y. Z. Yacobi, T. Zohary and T. Berman. (1998) *Aphanizomenon ovalisporum* (Forti) in Lake Kinneret, Israel. *J. Plankton Res.*, **20**, 1321-1339.

Pollingher, U. and C. Serruya. (1978) The phytoplankton of Lake Kinneret. In *Lake Kinneret*, vol. 32 (ed. C. Serruya), pp. 285-295. The Hague: Dr. W. Junk bv Publishers

Romero, J. R., J. P. Antenucci and J. Imberger. (In press) One- and three- dimensional biogeochemical simulations of two differing reservoirs. *Ecol. Model.*

Serruya, C. (1978) *Lake Kinneret*. Monographiae Biologicae. The Hague: Dr. W. Junk.

Smith, D. G. (1987) Water quality indexes for use in New Zealand's rivers and streams: Department of Scientific and Industrial Research, Water Quality Centre.

Smith, D. G. (1990) A better water quality indexing system for rivers and streams. *Water Research*, **24**, 1237-1244.

Straskraba, M. and A. H. Gnauck. (1985) *Freshwater ecosystems. Modelling and simulation.* Development in Environmental Modelling: Elsevier.

Tufford, D. L. and H. N. McKellar. (1999) Spatial and temporal hydrodynamic and water quality modeling analysis of a large reservoir on the South Carolina (USA) coastal plain. *Ecological Modelling*, **114**, 137-173.

OASI: An Integrated Multidomain Information System

M. Andretta[a], **G. Bernasconi**[a], **G. Corti**[b] **and R. Mastropietro**[b]

[a] *Ufficio della Prevenzione dei Rumori, Dipartimento del Territorio, Canton Ticino , Switzerland*

[b] *Department of Innovative Tecnologies – University of Applied Science of Southern Switzerland (SUPSI), 6928 Manno Switzerland*

Abstract: In the frame of a Swiss national project aimed at developing a decision support system for pollution related data, the Canton of Ticino (the Italian speaking canton of Switzerland) has developed an integrated system (OASI) for data collection, management and analysis for noise, air and traffic related data. Purpose of the OASI system is to follow the evolution of air and noise pollution in time and space, making data available and accessible; thus becoming a centre for data verification and analysis and, eventually, a tool for decision making. The OASI system has a scalable architecture as additional domains can be added in future to the same framework. OASI data are regularly and automatically collected from a network of hundreds of sensors distributed on the Ticino territory and undergo an automatic statistical, intersite and interdomain validation process before being available for further human manual validation and use. In this process, existing validated data are used to detect suspicious newly collected measures.

From the ICT view point, the system is developed in a 3-tier architecture: a data layer, a logical layer and a presentation layer. Data of different domains are collected in separated databases but can be transparently integrated for validation and analysis. One of the most challenging aspects of the project is related to the amount of measurements being collected and the size of the databases, which results in interesting storage management and system performance issues. Furthermore, the system allows the definition of users roles and data ownership. An organization may or may not allow its data to be seen or modified by users belonging to other organizations. Data can be accessed by a Java thin client that addresses scientists and researcher needs or by any Internet browser as well as via commonly used wireless devices, for use by the public. Whatever the type of client, they all share the same logic, which is implemented in the application server.

Keywords: pollution, information systems, environmental monitoring

1. INTRODUCTION

One may consider air pollution or noise pollution as two independent domains about which data could be collected and analyzed independently. However, as the environment is a complex and tightly integrated system, this view falls short when it comes to understanding the reasons of some changes in the monitored values.

In fact, it is difficult, if not impossible, to make a correct analysis of air pollution, for example, if data about traffic, locations of industries and other sources of air pollution and meteorological data are not available.

Similarly, to understand the impact of new traffic regulations on the environment, consistent collection of data about environmental parameters, which are influenced by traffic, is required.

As a further example, meteorological conditions, such as rain and snow, influence air pollution parameters (e.g. Ozone levels) or may influence noise levels and traffic.

What might at a first glance appear as independent domains, are actually tightly integrated. Hence, scientists who are supposed to investigate and analyze the environmental conditions, need tools, that provide the capability to collect, validate and analyze data pertaining to different domains in an integrated fashion.

2. THE OASI PROJECT

The idea to develop a multidomain information system arose from the considerations outlined above. The system was named OASI (an acronym for the Italian Osservatorio Ambientale della Svizzera Italiana or Environment Observatory of Southern Switzerland).

The OASI project was actually initiated with the primary goal of becoming a source of realtime, reliable information on environmental conditions in the Canton of Ticino (Switzerland). These conditions are especially linked to the pollution induced by traffic on the A2 motorway, which crosses the entire Cantonal territory north south, and is one of the major European backbones as it links Italy to central European Countries via the Gotthard tunnel.

Furthermore, the OASI system is part of a Swiss national initiative to monitor pollution caused by traffic on the main motorways.

3. MONITORING THE ENVIRONMENT

Environmental monitoring must pursue the following goals: a) detect and track the evolution of environment loads over time and space, b) collect meaningful, precise and high quality data in order to allow a fair and effective information to politicians, citizens and researchers, c) allow the analysis of extraordinary situations and d) the effective control of strategies and decisions in environmental matters.

The methodological choices developed in the frame of the OASI project are aimed at reaching the objectives outlined above.

Among innovations, it is worth to underline the following:

the deployment of coordinated multidomain measurement points. In the first phase, OASI was focused on the air, noise, traffic and meteo data in locations which are along the motorway as defined in the Swiss Federal project. This not only allows to follow the evolution, but also allows to detect correlations among measured parameters, and to analyze the impact of traffic on the environment, in general, and the impact of heavy vehicles traffic on the alpine valleys, especially.

the development of a specific concept for the noise domain, which allow the permanent collection of phonic emissions and immissions. It is actually a relevant technical and scientific novelty at Swiss level. It was decided to install permanent measurement stations along the motorway and to integrate a network of mobile measurement stations, whose task is to collect data at medium-large distance from sources of noise emission, the motorway in this case. The purpose of this mobile measurement stations is to allow to determine the background noise or, from a different view point, to qualify the noise landscape of a given region.

the option to record noise as an acoustic picture, so as to allow later direct comparisons based on the actual perceived noise rather than on physical parameters only.

the joint management of measured data and data about pollution sources (cadastre, streets, industries, railway, antennas, etc.) makes it possible to analyze and correlate activities and their side effects, computed data and measured data.

The goal of the OASI project is not only limited to the building of a tool for observing and recording. It is conceived to get a snapshot of the environmental situation, on the one hand, and to provide support for analysis, validation and exchange of data, on the other.

In fact, data collected in the OASI information system is available to a variety of users depending on their competences. Especially, they are available to authorities of the Cantonal and Federal Administration, that are in charge of working out strategies and making decisions aimed at protecting the environment. In this context OASI becomes both a decision support system and a tool to verify if objectives have been reached.

4. SYSTEM FUNCTIONALITIES

4.1 Data Collection

Data is automatically collected from a variety of sources for the different domains and must be validated before being used for actual reseach and reporting.

4.2 Data Validation

Given the huge amount of data collected, it is not feasible to leave the entire validation work to scientists or operators. Some automatic support for detecting suspicious data must be provided by the system.

For this purpose a unified quality assessment and data validation system has been developed, that can be applied to all continuously sampled data, independently of the domain. On top of this, the system allows each organization to perform human (i.e. manual) validation on their own data, taking advantage of their long term experience in a certain environmental domain. The quality assessment system is built as a succession of automated procedures with increasing complexity, and ends with the manual validation module included in the OASI application.

The automated quality control procedures monitor all the incoming data from automatic stations. If requested, early alerts on possible malfunction of the monitoring network can be sent to the responsible operator and to the network maintenance team by e-mail or sms. Statistical algorithms perform interparameter, intersite and interdomain tests and flag suspect data sets, thus implementing a unified flagging system. At the end of each automated procedure, appropriate flags indicate the status (quality) of the data. Therefore, the status of data is determined by the number of executed procedures and by the results obtained for each test.

The automated procedures are conceived to be at the same time a safety net, for early identification of network problems, and a valuable support for the later manual validation of the data, helping the operator in the difficult task of distinguishing instrumentation problems from unusual, but physically plausible situations.

The fact that data from different domains are simultaneously measured at the same location could lead to a complete change in validation paradigms. As an example, spectral characteristics of traffic noise could be used to confirm certain meteorological events, like rain or wind gusts.

4.3 Integrated Data Analysis

Scientists can access the data repository via the OASI application in order to carry out the key part of their work: performing integrated data analysis.

Users can select locations, parameters, time intervals and diagrams types. Multiple curves can be displayed and compared on a single diagram. Data belonging to different domains can be shown on a single screen, for comparison purposes.

4.4 Users Authorizations

Users of the system may belong to different organizations with different responsibilities and different tasks to perform. Data hosted in the database may also belong to different organizations, as well as pertaining to the different domains, as we have seen already. Therefore, the system provides a mechanism to grant access to data that accounts for this.

Each user has a login name and password and is assigned one or more roles within the system. A role may span multiple domains. Each role has a set of functions associated to it, that the holder of that role can carry out on the data. Read-only or read-write permissions may be granted to each function. The structure is very flexible and allows to configure new roles with grants as needed. User grants will be reflected in the OASI application layout: some menu entries or buttons will be enabled or disabled based on the user profile. This solves the problems of allowing or disallowing certain operations to certain users. The logic for granting access to data is based on data ownership: each data is generated by an instrument. An instrument belongs to an organization. Users belong to organizations. An organization can grant access to its data to users belonging to other organizations.

This architecture makes it possible for an external organization to join the OASI framework and add its data to the system, with the choice to allow or to prevent access to them to all or to a subset of users.

4.5 GIS Integration

Environmental data are obviously geographically referenced, hence an environmental information system must provide some sort of integration with GIS applications.

In the case of the OASI system, the integration is twofold: data can be accessed via a GIS application (ESRI) and via GIS-like functionalities built into the application and conceived to perform simple tasks that would hardly justify the need to install a full blown GIS application.

Georeferenced data as well as GIS maps are stored as Oracle Spatial data types. This allows both the integration with a GIS and to perform georeferenced queries from the application making use of the Oracle Spatial extended functions.

From within the OASI application map, GIS layers can be retrieved and displayed. Additional layers can be built using OASI specific data (e.g. the map of air domain sensors' locations).

The map can be browsed to navigate more detailed data.

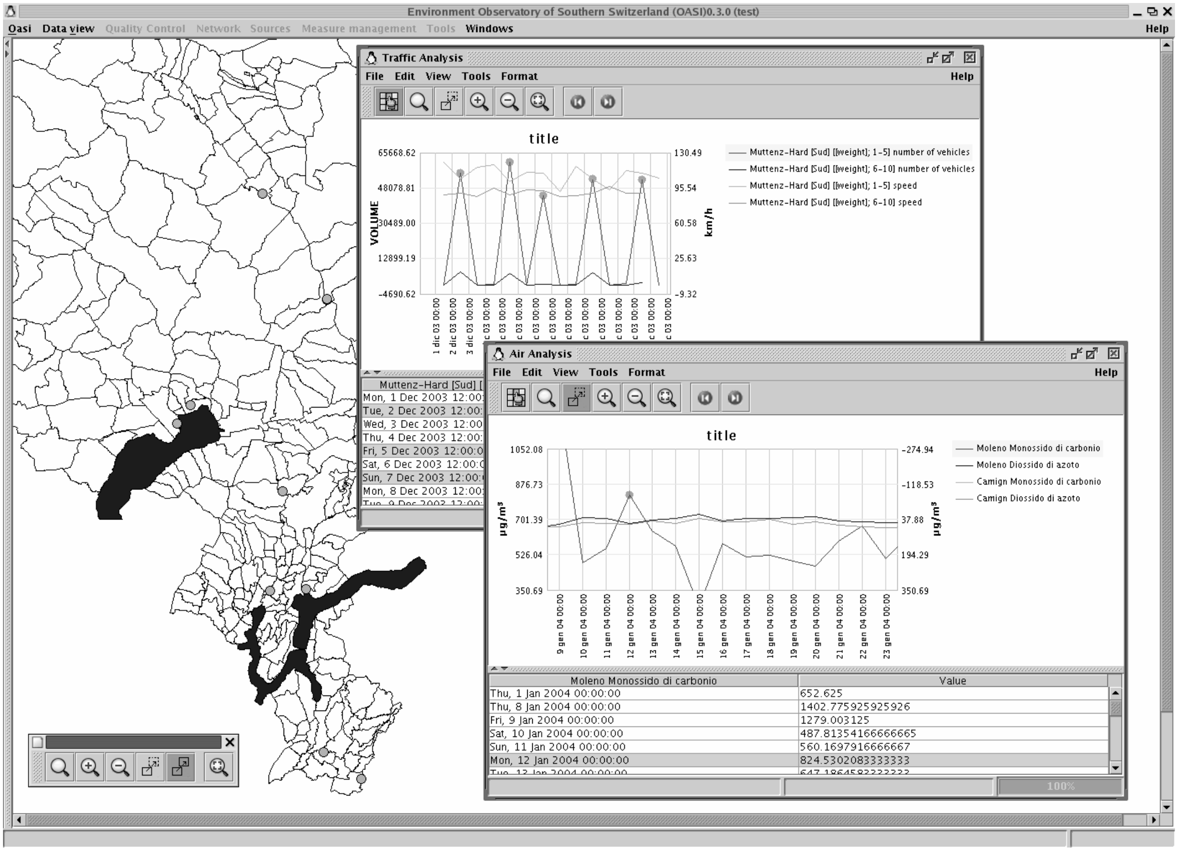


Figure 1. OASI application screenshot

5. SYSTEM ARCHITECTURE

5.1 Database Architecture

Although the OASI information system only deals with pollution related data, these data actually belong to different domains which are dealt with by different Cantonal departments and offices. Each office is consequently responsible for the management of the data of its domain of activity. This organizational aspect would be a driver towards a solution where there is an independent database for each domain. On the other side there are data, administrative data, that are common to all the domains. Moreover, the requirement standing behind the OASI information system is to have an integrated environment where data can be analysed in an integrated fashion and the correlation among events and among different types of pollution can be given evidence. Based on these somewhat conflicting requirements, a mixed architecture was designed.

Data can be grouped according to the following taxonomy:

- data that must be shared by all domains in read/write mode
- data that must be shared in read mode by multiple domains
- data that are private to one domain but have the same data structure in all domains
- data structure and data that are specific to one domain

As a result of this, there is a shared schema which includes data structures that have to be shared among all domains. And there is a schema for each specific domain. Shared structures are directly visible and are referenced by structures in each domain making domains apparently independent from one another.

Today, the available domains are: air, noise, traffic and meteo. Further domains could be added in future and plugged into the existing infrastructure leveraging the design work already done.

Currently the database is composed by over 200 tables for the four domains supported today.

5.2 Application Architecture

The OASI application is an Internet (web) application based on the 3-tier model. The OASI application is written in Java and is self-contained. It is actually composed by a client side and a servlet-based server side. The entire application logic is server side and is run within the application server.

The client side is thin and implements just GUI aspects of the application.
The server side modules are used also to support the accesses via the web site which is available, with limited functionality compared to the application, for simplified public access.
A client, be it a Java stand-alone application or an Internet browser, resorts to a number of Java servlets, which in turn access the database through JDBC (Figure 2).
Servlets are capable of recognizing their caller, thus returning data in the appropriate packaging (i.e., HTML for browsers, serialized Java objects, tunneled into HTTP, for Java stand-alone clients).
The code which is, in terms of function, closer to the database is mostly stored and run within the Data Base Management System in the form of stored procedures. Hence, presentation, application logic and data are actually decoupled in three tiers.

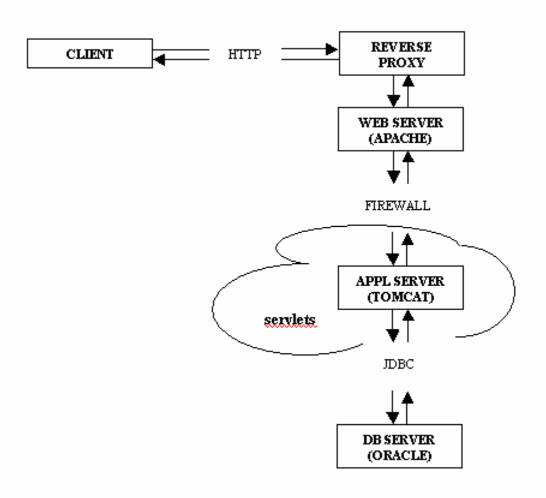


Figure 2. OASI architecture

6. TECHNICAL CHALLANGES

6.1 Database size

The amount of data collected and stored in the databases is one of the challenging aspects of the project. The measurements of the air, meteo, traffic and noise parameters is what takes up the largest portion of the database. The rest is used for calculated values (statistics), metadata (about sensors, persons, users and the like) and geometries (for geographical reference). The number of vehicles passing by a single measurement station on the A2 motorway, for instance, averages to 40'000, daily. Each passing vehicle is recorded along with a number of parameters: speed, length, weight, gap from the previous vehicle, category, etc. This means that the system collects an average of 70 millions measures per measurement station every year, for the traffic domain only. These data end up being turned into an equivalent number of table rows. Similarly, to store the noise spectrum for each noise sample causes several million rows to be added to the database, yearly. The remaining two domains are less demanding. The OASI system therefore falls into the *Very Large Databases* (VLDB) category.

A number of otherwise straightforward issues now need to be considered carefully. Backups, for example, become problematic, both in terms of cost (infrastructure and media), time (backup windows cannot be stretched indefinitely), and data integrity (the DBMS cannot be stopped to run cold backups, as data could be accessed at any time, especially by Internet users). The life cycle of OASI measurements is such that, after their loading, measurements undergo a semi-automatic validation process, after which they are not modified any more. To minimize backup issues, the following physical layout is adopted. Tables which contain measurements are *partitioned by range*. Measurements of a solar year are kept in a dedicated partition. Following this approach there will only be a single partition per table, which will be modified daily and that will therefore require a backup. The other partitions can be kept in read-only tablespaces and will not require regular backup.

Over time, the amount of data will be so large, that it will become impossible to keep everything online. The installation of an automated tape library will then have to be considered along with a Hierarchical Storage Management (HSM) system. With such a configuration, DBMS tablespaces partitions could be managed by the HSM (i.e purged by online storage when not accessed for a certain time and then recalled from offline, or nearline, storage, when access is required). This would allow to limit the need of continously expanding the disk subsystem (or possibly allow operating with an undersized disk subsystem).

As for data themselves, the most space consuming data are treated in a special way to reduce them as much as possible. Vehicle-by-vehicle traffic data are processed daily and aggregated according to some statistics (this is configurable). Average speed per vehicle class, volume per vehicle class, etc are worked out and stored in the database, while their raw data are kept online for one year and then deleted. In case of need older data can be reloaded from datafiles delivered by the measurement stations. The same applies to noise spectrum measurements which can be loaded on demand.

6.2 System Performance

The large amount of data also has an impact on system performance. An interesting aspect of this issues is the performance perceived by users, especially Internet users. While researchers who use the OASI application are aware of the amount of data they are dealing with and are therefore ready to wait for a few seconds for some computations to take place, Internet users behaviour is known to be less tolerant. If the web page is not displayed in a handful of seconds, the user will assume the site is unavailable and will go away.

In order to support fast response to the query submitted through the Web, which are by the way a small subset of all those that may be submitted through the application, it was necessary to store calculated measures (averages, minimums, maximums, etc) into materialized views, that are automatically updated on a daily basis. This allows very fast response as complex computations against several gigabytes of data are run in batch mode and data are readily available for display during interactive activity.

The current system is a Sun E250. It is equipped with 4 GB RAM and an external Fibrechannel attached EMC RAID subsystem and runs the Oracle DBMS. It has so far delivered satisfactory performance. However, should in future the amount of data and users bring the system performance to an unacceptable level, the current configuration allows easy migration to a Storage Area Network based environment, by adding a fibre channel switch and extra servers, which run in a clustered configuration and share the RAID subsystems via the SAN infrastructure. The load could be then distributed among multiple servers allowing for great scalability of the solution. A further alternative solution is to reduce the load on the system by distributing the various databases among multiple database servers. The application server would easily hide this change of topology.

7. CONCLUSIONS

The OASI system currently manages 6 so called *superlocations*, which are permanent and collect data about air, noise, traffic and meteo parameters along the A2 motorway between Chiasso and Basel, and 7 locations where air and meteo parameters only are collected. These locations collect averaged data with a frequency of half an hour. In future the integration of additional locations for the collection of data about the quality of the air is foreseen, as well as the extension of the system to other domains, namely to that of non ionizing radiations.

In terms of functionalities, access to the system via various kinds of mobile devices, from old GSM phones via SMS or from more modern Java-enabled devices via simplified yet powerful interfaces, has already been developed as a prototype.

From the methodological and operational directions described previously, a great built-in potential of the OASI project can be derived, which is not limited to environmental aspects, but can also cover administrative and management aspects.

Due to the its interdisciplinarity, the whole system will have to evolve and adapt to new, upcoming requirements and expectations.

8. REFERENCES

Adam Cannane, Hugh E. Williams: A Compression Scheme for Large Databases. Australasian Database Conference 2000.

Barrow, C. J., Environmental Management: Principles and practice”, Routledge Editor, London, 1999.

Bernhard Thalheim: Component Engineering for Large Database Applications. CAiSE Short Paper Proceedings 2003

Mengjou Lin, Jenwei Hsieh, David Hung-Chang Du, James A. MacDonald: Performance of high-speed network I/O subsystems: case study of a fibre channel network. SC 1994

Michael Lautenschlager, Hannes Thiemann: Connection of a Climate Model Database and Mass Storage Archive(s). IEEE Symposium on Mass Storage Systems 2000

Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, Riccardo Torlone: Database Systems - Concepts, Languages and Architectures. McGraw-Hill Book Company 1999

Stefano Ceri, Piero Fraternali, Stefano Paraboschi: Design Principles for Data-Intensive Web Sites. SIGMOD Record 28(1): 84-89 (1999)

Vassiliadis, P., *Data Warehouse Modeling and Quality Issue*s, Ph.D. Thesis, National Technical University of Athens, Athens, Greece, June 2000.

Wijnand Derks, Willem Jonker: Practical Experiences with Materialized Views in a Very Large Data Store for Telecommunication Service Management. DEXA Workshop 1998

A Data Warehouse for Physical Flows

E. Löfving and A. Grimvall

Department of Mathematics, Linköping University, SE-58183 Linköping, Sweden

Abstract: Data on flows of materials and substances through the economy and the environment are collected by a large number of organizations and are stored in different systems and formats. We outline how a data warehouse can give added value to such data by facilitating the exchange of collected data and enabling new studies of data from different sources. The structure of the proposed data warehouse includes a relational database in which flow data are stored along with structured information about the underlying cell system and relationships between cells in different systems. Data cube representations tied to the database facilitate statistical analysis of collected data.

Keywords: Substance flows; Data warehouse; ER-diagram; Relational database; Data cube; Matrix operations.

1. INTRODUCTION

Information on flows of materials and substances through the economy and the environment can greatly improve our understanding of environmental problems. At present, physical flow data are collected by a large number of organizations and are stored in different systems and formats. To facilitate exchange of such data, it is important to harmonize the systems that are used. In addition, collected data can be given added value by organizing both raw data and relevant background information in an efficient manner.

Examples of the methods currently used to handle physical flow data can be found in a large number of publications in the area of industrial ecology (Voet *et al*., 1995a, 1995b; Voet, 1996; Weisz *et al*., 1998; Burström and Frostell, 2000; Heijungs and Suh, 2002). In general, the data in these reports have been organised in input-output tables or flow matrices that form the basis of both storage and analysis. Some authors have also presented matrix formulae for data on a single substance, material, or commodity (Miller and Blair, 1985; Voet, 1996; Heijungs, 2001; Löfving and Grimvall, 2003). Furthermore, a few attempts have been made to introduce data cubes for multivariate flow data (Rademacher and Höh, 1997; Löfving and Grimvall, 2003), and to develop software that can make it easier to handle flow data (Schröder, 1995; Boelens and Olsthoorn, 1998; Burström and Frostell, 2000). However, thus far, industrial ecologists have paid little attention to the problems that arise when flow data collected by different organisations or teams need to be merged in new studies or meta-analyses.

During the past decade, data warehousing has become a key concept in the handling of multiple data sources. This concept was introduced by Inmon (1992), who defined a data warehouse as a "subject-oriented, integrated, non-volatile, time-variant collection of data in support of management's decisions." In contrast to conventional transactional databases, a data warehouse can support time series analyses of collected data and provide convenient access to a wide range of background information on the subject under consideration. Also, it is worth noticing that large amounts of logistics data, which in several respects resemble physical flow data, are presently being organized in data warehouses.

This article aims to provide a seed to the organization of warehouses for physical flow data and related information. This seed consists of two major parts: a data model illustrated by a so-called ER diagram, and a data cube representation that can provide a link between storage and analysis of collected data. In addition, we discuss possible links to classification systems used in official statistics, as well as some specific problems related to the merging of data from different studies.

2. GENERAL FEATURES OF DATA WAREHOUSES

Practical use of data warehouses entails extracting data from different systems, which in most cases have been created without any intention of sharing data with other systems. Consequently, all data sets made available in a data warehouse must be accompanied by metadata (i.e., data about data). A dictionary of metadata must contain information about data types and coding standards and it must also be able to provide relevant information about measurement techniques, sampling methods, and other features of the data collection process. Data cleaning is another important part of data warehousing, and a minimum requirement is that obvious errors and duplicate records be removed.

Most early information systems were focused on compact storage, and the granularity (level of detail) of collected data was regarded as fixed. In data warehousing on the other hand, granularity is a design issue, and redundancy is permitted if it can facilitate analysis of collected data. Consequently, data are often presented both with the lowest possible granularity and as summaries that are easier to handle and more suitable for communication and decision support.

A data warehouse data model has three major components: a conceptual, a logical, and a physical data model. The conceptual model specifies the entities included in the data warehouse; the logical model defines the relationships between these entities; and the physical model specifies how the data are stored in the physical database.

The relationships between the different entities can be displayed in an entity relationship (ER) diagram (see Figure 5). Entities are shown as rectangular boxes, and attributes are depicted as ovals that are attached to their entities by straight lines. The relationship types are shown as diamond-shaped boxes attached to the participating entities with straight lines. The cardinality of a relationship is also shown in the ER-diagram. For example, a 1:N relationship indicates that one entry in the first entity corresponds to many entries in the other entity.

3. A DATA WAREHOUSE FOR INDUSTRIAL ECOLOGY

3.1. The cell concept

The basic principle of physical flow studies is to divide the technosphere and biosphere into cells and gather information about (i) the flow of substances or commodities between cells and (ii) the stocks in each cell.

A typical cell in the biosphere is a medium, (such as air), a catchment area, or an ecosystem, and it normally has a well-defined physical location. Cells in the technosphere typically represent different industrial branches or activities (such as traffic), or groups of consumer products. The exact location of such cells may be difficult to define and is often considered to be of secondary importance.

From a mathematical point of view, a collection of cells $\mathcal{U} = \{U_i, i = 1, \ldots, m\}$ is a partition of a given set Ω, that is the subsets U_i are non-empty and disjoint, and their union is Ω. Furthermore, a partition $\mathcal{U}_1$ is said to be a refinement of another partition $\mathcal{U}_2$, if each $U_{1j} \in \mathcal{U}_1$ is contained in some $U_{2k} \in \mathcal{U}_2$.

3.2. Hierarchic and non-hierarchic cell systems

Many classification systems used in official statistics are hierarchic, in other words they include a sequence of partitions such that each is a refinement of the previous one. Figure 1 illustrates how the agricultural sector in the Classification of Economic Activities in the European Community (NACE, revision 1.1) is divided into sub-sectors.

SECTION A	AGRICULTURE, HUNTING AND FORESTRY
01	AGRICULTURE, HUNTING AND RELATED SERVICE ACTIVITIES
01.1	**Growing of crops; market gardening; horticulture**
01.11	Growing of cereals and other crops n.e.c.
01.12	Growing of vegetables, horticultural specialties and nursery products
01.13	Growing of fruit, nuts, beverage and spice crops
01.2	**Farming of animals**
01.21	Farming of cattle, dairy farming
01.22	Farming of sheep, goats, horses, asses, mules and hinnies
01.23	Farming of swine
01.24	Farming of poultry
01.25	Other farming of animals

Figure 1. Hierarchic structure of the agricultural sector as it is defined in NACE, revision 1.1.

Other hierarchic cell systems focus on the biosphere. For example, a land area can be divided into major land-use categories, such as arable land and forest, and each of these categories can be further divided into more specific land-use classes. However, merging of partitions of the biosphere and technosphere can result in non-hierarchic systems. For example, the water districts established within the EU Water Framework

Directive have hydrological borders that do not normally coincide with administrative borders.

3.3. Commodity classes and transmission coefficients

Each flow datum refers to a given element, substance or commodity. The classification systems used in official statistics can impose a hierarchic structure on the set of commodities under consideration.

Transmission coefficients must be defined in order to calculate the turnover (metabolism) of a specific element. Each transmission coefficient is related to a certain flow and represents the concentration of a given element in that flow.

3.4. Entities in the data warehouse

The above discussion leads to the following collections of entities.

The **flow entity** forms the core of a data warehouse for physical flows. This entity includes information about the commodity, the magnitude, and the source and destination cells of the flow. Also, information about the time period of the flow is included in the flow entity. A **storage entity** can be defined in a similar manner.

The **Cell entity** includes the description of the different cells, and the **Partition entity** lists the cells forming a partition and how the cells are related to different classification systems. Further information about the relation between different cells, such as if cells are subsets of other cells, is included in the **Cell relation entity.** Information about the hierarchic order of different partitions is stored in the **Partition relation entity**.

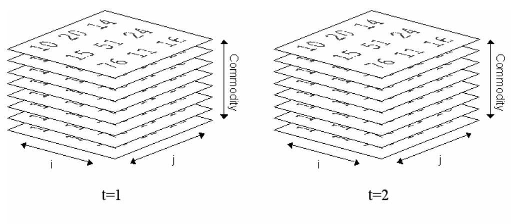
Information about different commodities and their hierarchic structure are stored in the entities **commodity classification** and **commodity**. The concentrations of certain substances in each flow are stored as transmission coefficients in the **transmission entity**.

The relationship between some of the entities in our proposed data warehouse is illustrated in the ER diagram in Figure 5 in the appendix.

4. EXTRACTION OF DATA CUBES FROM THE RELATIONAL DATABASE

A time series of flow data in a given system of cells can be conveniently stored in a three-dimensional data cube. Likewise, it is natural to store flow and transmission data in such data cubes. Four-dimensional flow data cubes arise naturally when we extract data on two or more commodities from two or more time intervals. For example, we can introduce the notation $F = \{f_{ijct}, i = 1, \dots, m, j = 1, \dots, m, c = 1, \dots, p, t = 1, \dots, q\}$, where f_{ijct} designates the flow of a given substance from the source cell U_i to the destination cell U_j that can be attributed to commodity c during the time period t. Figure 2 illustrates how four-dimensional data can be represented as a sequence of three-dimensional data cubes:

Formally, a multidimensional flow data cube F can be defined as an array of non-negative real numbers in which the first two indices reference the source cell and the destination cell, respectively.



$F = \{f_{ijct}, i=1,\dots,a, j=1,\dots,a, c=1,\dots,8, t=1,2\}$

Figure 2. Four-dimensional flow data visualized as a pair of data cubes. Each cube represents a specific time period, and the various layers stand for different materials or commodities.

In a data warehouse, data cubes representing different levels of aggregation in time and commodity and different levels of granularity in the underlying cell system can be formed and coexist with the relational database. We therefore propose an integrated system in which the previously outlined relational database is used for storing data, and data cubes are employed for analysing data and assessing the quality of data before they are included in the data warehouse. The relationship between the different systems is illustrated in Figure 3.

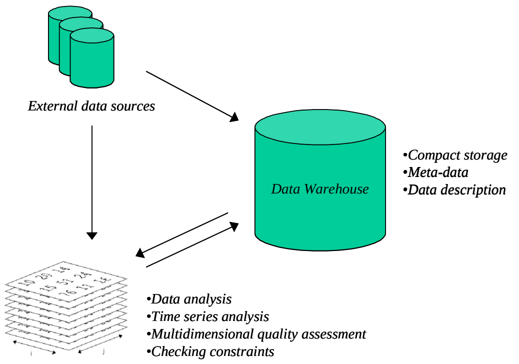


Figure 3. Relationship between data, a data cube, and the data warehouse.

5. OPERATIONS ON DATA

Conventional matrix operations have long been applied to physical flow data. We have recently shown how such operations can be generalized to data cubes of arbitrary dimension (Löfving and Grimvall, 2003). Here, we emphasize that this generalization can provide important tools for the extraction of data from the outlined data warehouse.

The handling of flow data stored in cubes involves simultaneous operations on several layers of data. The key to such operations is to introduce vectors, matrices and data cubes in which the elements are ordinary two-dimensional flow matrices. A three-dimensional flow data cube, for example, can be regarded as a vector of ordinary flow matrices. Likewise, a four-dimensional flow data cube can be regarded as a matrix in which all elements are flow matrices.

In analogy with multiplication of an ordinary matrix with a scalar, we define an elementwise (layerwise) product of a two-dimensional matrix and a data cube of arbitrary dimension. If $\boldsymbol{X}$ depicts the two-dimensional matrix and $\boldsymbol{F}$ is the flow data cube, the layerwise product $\boldsymbol{F}*\boldsymbol{X}$ can be written as

$$\boldsymbol{F} * \boldsymbol{X} = \begin{pmatrix} F_{11}\mathbf{X} & \cdots & F_{1q}\mathbf{X} \\ \vdots & \ddots & \vdots \\ F_{p1}\mathbf{X} & \cdots & F_{pq}\mathbf{X} \end{pmatrix}$$

where the element F_{ij} represents the flow matrix for commodity i and time period j. Similarly, the layerwise transpose of a flow data cube $\boldsymbol{F}$ is defined as

$$\boldsymbol{F}^T = \begin{pmatrix} F_{11}^T & \cdots & F_{1q}^T \\ \vdots & \ddots & \vdots \\ F_{p1}^T & \cdots & F_{pq}^T \end{pmatrix}$$

where the element F_{ij} represents the flow matrix for commodity i and time period j.

The notation introduced above is particularly useful for the transfer of multidimensional flow data to a coarser partition. In fact, it makes the matrix formula for aggregation of multidimensional flow data identical to the formula for a single commodity (Löfving and Grimvall, 2003). To be more precise, let $\boldsymbol{A}$ be an aggregation matrix transforming a partition $\{U_i\}$ of a set $\boldsymbol{\Omega}$ into a coarser partition $\{V_j\}$, and let $\boldsymbol{F}_U$ be a material flow cube associated with $\{U_i\}$. Then the flows on $\{V_j\}$ are given by

$$\boldsymbol{F}_V = \boldsymbol{A}*\boldsymbol{F}_U*\boldsymbol{A}^T$$

where * denotes layerwise multiplication.

6. COMBINING INFORMATION FROM TWO OR MORE STUDIES

Data collected in different studies can easily be merged into a joint study if the underlying cell systems are identical. For example, it is then a trivial task to merge data representing different (non-overlapping) time intervals or commodities. When two different cell systems are involved, joint analyses of collected data must be done with great care. In the following, we discuss two cases: (i) joining two studies to a single study, and (ii) comparison of data collected in different geographic regions.

6.1. Joining two studies to a single study

Let us assume that we have collected flow data about the same commodity at two occasions and that the cell systems used were defined by two different partitions $\mathcal{U}_1$ and $\mathcal{U}_2$. It is then fairly easy to construct an algorithm that creates the finest partition $\mathcal{U}_3$, such that both $\mathcal{U}_1$ and $\mathcal{U}_2$ is a refinement of $\mathcal{U}_3$. Furthermore, we can derive flow data for $\mathcal{U}_3$ from each of the two studies. The information content of the flow data derived in this manner may vary strongly from case to case. If the partitions $\mathcal{U}_1$ and $\mathcal{U}_2$.are very similar, practically all information in each of the original studies can be taken care of. On the other hand, if the two partitions are very different, all information may be lost because $\mathcal{U}_3$ becomes too coarse. In the most unfavourable case, the partition $\mathcal{U}_3$.consists of a single set.

6.2 Comparison of data collected in two different regions

Let us consider two studies that have been carried out in different (non-overlapping) regions, and assume that each study includes the following three cells: biosphere (B), technosphere (T), and surrounding world (S). Even in such a simple case, it is obvious that additional information is needed for a joint evaluation of the two data sets. Figure 5 illustrates how collected data can be coherent with entirely different flows in a system where the two regions have been combined into a single region.

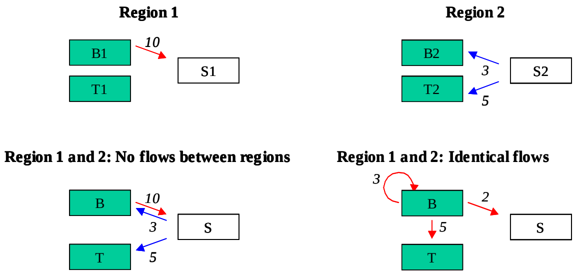


Figure 4. Two possibilities to combine flow data collected in different regions.

8. DISCUSSION

Data warehousing is a process that aims to organize existing data and refine them into information. When properly implemented, this process can provide: (i) easy access to data from disparate sources, (ii) well structured data management, and (iii) a basis for knowledge discovery and effective analysis of collected data.

Thus far, data warehousing is mainly used to support analysis of business and customer activities. This is not surprising, because it is easier to introduce a new system in a single organization than in an area that includes a large number of independent organizations and many disparate activities. However, the general interest in putting data on the web and sharing information with research colleagues shall not be underestimated. Therefore, we believe that a data warehouse in industrial ecology can grow by itself if a proper seed is developed.

This article paid special attention to three aspects of data warehousing in industrial ecology: (i) standardization of the entities stored in the data base; (ii) the use of data cubes for presenting data in a form that supports analysis; and (iii) the establishment of a framework for merging existing data sets in new studies or meta-analyses.

Previous attempts to standardize entities in databases for industrial ecology have focused on the documentation of single studies. Our system of entities was designed to support joint analyses of data from different sources. Data cubes are particularly useful when data have been collected over different time periods, or the flows of several commodities have been examined. Along with the matrix formulae that were presented, the data cube representation can also greatly facilitate the development of software for analysis of physical flow data. The examples of how data from different studies can be merged illustrate the strength and weakness of fully automated procedures. Even though data warehousing aims to provide relevant meta-data, some joint studies will require user input of additional information.

9. ACKNOWLEDGEMENTS

The authors are grateful to the Bank of Sweden Tercentenary Foundation for financial support.

10. REFERENCES

Boelens, J. and X. Olsthoorn. 1998. Substance flow analysis with FLUX. Institute for Environmental Studies, Free University, Amsterdam.

Burström, F. and B. Frostell. 2000. Materialflöden i ett kommunalt miljöinformationssystem: Grundläggande struktur för kommunala materialräkenskaper. Technical report 2000:2, Royal Institute of Technology, Stockholm (in Swedish).

Elmasri, R. and S.B. Navathe. 2000. Fundamentals of database systems, 3rd ed. Reading, Massachusetts: Addison-Wesley.

Heijungs, R. 2001. A theory of the environment and economic systems: A unified framework for ecological economic analysis and decision-support. Cheltenham, UK: Edvard Elgar.

Heijungs, R. and S. Suh. 2002. The computational structure of life cycle assessment. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Inmon, W.H. 1992. Building the data warehouse. New York, USA: John Wiley & sons, Inc.

Löfving E., Grimvall A. and Palm V. (2003): Data cubes and matrix formulae for convenient handling of physical flow data, LiU-MAT-R-2003-03, Linköping University, Sweden.

Miller, P.E. and Blair, P.D. 1985. Input-output analysis: Foundation and extensions. Englewood Cliffs, New Jersey: Prentice Hall.

Radermacher, W. and H. Höh. 1997. Material and energy flow analysis accounting framework and information system. In: S. Bringezu et al.

(Eds.). Proceedings of the ConAccount Workshop, 21-23 January 1997, Wuppertal Institute, Germany.

Schröder, H. 1995. Environmental input-output analysis and the software system steady. Danedi, Helsinge, Denmark.

Voet, E. van der. 1996. Substances from cradle to grave. Development of a methodology for the analysis of substance flows through the economy and the environment of a region, with case studies on nitrogen and cadmium. Ph.D. Thesis, Leiden University.

Voet, E. van der, R. Kleijn, L. van Oers, R. Heijungs, R. Huele, and P. Mulder. 1995a. Studying substance flows through the economy and environment of a region - part I: Systems definition. Environmental Science and Pollution Research 2:89-96.

Voet, E. van der, R. Heijungs, P. Mulder, R. Huele, R. Kleijn, and L. van Oers. 1995b. Studying substance flows through the economy and environment of a region - part II: Modelling. Environmental Science and Pollution Research 2:137-144.

Weisz, H., H. Schandl, and M. Fischer-Kowalski. 1998. OMEN – An operating matrix for material interrelations between the economy and nature. Proceedings of the ConAccount workshop. 21 November 1998, CML Report 148. Leiden University Institute. The Netherlands.

11. APPENDIX

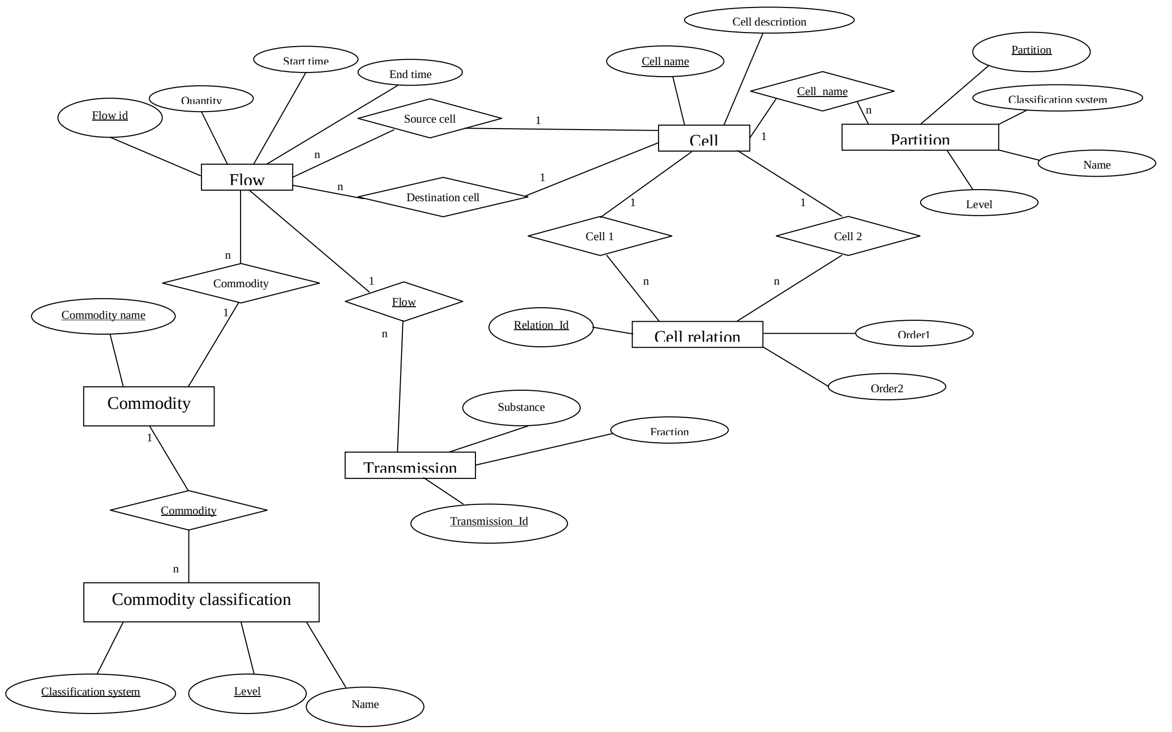


Figure 5. Part of the ER-diagram of the proposed data warehouse

Index of authors

SEA - COMO, Italy

http://www.iemss.org

iEMSs, Switzerland, ISBN 88-900787-1-5