In: iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Artificial-Intelligence-based Modeling for Environmental Applications and Decision Support

Peter Struss

Tech. Univ. Munich, Comp. Sci. Dept., Garching, Germany (struss@in.tum.de)

Keywords: Modeling; Artificial Intelligence; Environmental Decision Support.

1. THE PROBLEM

What we call "environment" or environmental systems is the natural world (or sections thereof) including humans and the artifacts they create. Environmental systems are complex systems, comprising a large number of interacting objects, substances, and processes which are subject to different specialized areas of science, such as biology, chemistry, hydrology, meteorology, and, when it comes to human activities, sociology, economics, even psychology. If we want to intervene in order to achieve certain goals – avoidance of environmental damage, sustainable development of resources, preservation of species – we rely on knowledge about the respective systems and the various relevant phenomena they comprise, both scientific knowledge and traditional knowledge and experience. We need a **model**, a model in a general sense (not simply a mathematical simulation model): a formal representation of concepts and interdependencies that allow explaining what we observe and predicting future developments and the impact of interventions.

Research in the different fields has contributed lots of relevant results, very often as reports, empirical data, and scientific papers, but also delivered executable models of certain systems at different scales, from small water catchments to global weather patterns. Models are accepted as an important means for validating or refuting scientific hypotheses. The corpus of available knowledge and well-supported hypotheses has increased tremendously over the past years. We should be in a much better position to understand environmental problems, develop possible solutions, take well-founded decisions and perform successful interventions.

However, we face a number of fundamental and hard problems that arise from the – inevitable – fact that all relevant scientific results and also other elements of knowledge and insights

- deal with only a limited number of aspects relevant to an environmental system or problem, and
- have been obtained in a particular context.

In order to exploit these results, we have to combine them to "obtain a more complete picture", to form a holistic model of a system. This usually requires crossing the borders of different scientific fields, which is difficult. It also requires assessing whether and which part of certain results can be transferred to a different context (chemical reactions will remain the same in different places, but whether their necessary ingredients are present or whether there is another compensating reaction may have to be established explicitly for each place). This can be stated as the problem of **integrating knowledge** elements and the problem of **context-dependent knowledge** and its re-use and adaptation.

If the results take the form of models, e.g. numerical simulation models, these general problems show up as various difficulties, such as

- Contradictory results and comparison of models. It is often hard to compare two models and identify the origin of discrepancies between the results in a rigorous and formal way (beyond the, usually informally stated or implicit, assumptions of the originators).
- Difficult re-use of models or model elements. It may not be obvious whether a model that has been validated in some area is appropriate for a similar system in another area. Sometimes, it may be known that some changes in the context violate certain assumptions underlying the model, which, hence, has to be modified. But it may be hard to tell which parts of the model are affected and need adaptation.
- Difficult integration of models. Even if we understand that certain insights underlying existing models need to be combined for obtaining a more appropriate model, integrating these models is usually a non-trivial task: each individual model has its boundaries which now have to and can be overcome, but it is not obvious where and how this can be done and how the underlying concepts can be mapped onto each other.

It is important to note that these difficulties are not simply technical issues. They are very deeply rooted in the very nature of the complexity of the systems to be modeled and in the (reductionist) way we can derive knowledge about them. “Divide and conquer” is a good research strategy which delivers insights each under a small set of aspects. However, in order to establish the basis for a broader understanding and well-founded interventions, we need to put the pieces together. We cannot divide the problem world into “improving agriculture”, “maintaining water quality”, “preservation of bio-diversity”, … and conquer them separately.

2. THE ARTIFICIAL INTELLIGENCE ANSWER

2. 1 What Is Needed?

The challenge lies in integrating knowledge from different sources and about different subjects stated in terms of models and adapting models to different contexts. Models can only be integrated if they share some concepts. Otherwise, they will stay unrelated. The problem may resemble the problem of combining data from different data bases, but it is much harder than this, because one has to identify “mappings” between different concepts of describing the behavior of certain environmental, social etc. (sub) systems, which is more complex than identifying corresponding variables in different models.

A fundamental consequence is that such models have to include a **conceptual model**, rather than some mathematical constructions (e.g. ordinary or partial differential equations) relating a number of variables. They need an explicit representation of the knowledge underlying the model, knowledge about the real-world objects whose properties are captured by the variables and their interrelations.

Integration and adaptation of models to a new scenario and/or context involves identifying and including those parts of a model that capture the aspects relevant to the new situation and dropping the irrelevant and inappropriate ones. To enable this, we need to develop **compositional models**, i.e. models that are composed of independent, elementary, combinable model fragments that can be re-used in different contexts, rather than holistic models whose structure does not reflect and preserve the structure of the modeled systems in terms of various interacting phenomena. We need to develop libraries of such elementary models that are building blocks for models of various complex systems and situations that comprise varying sets of objects and processes.

In order to be combinable with others und in different contexts, such model units from a library have to describe basic phenomena independently of others and a specific context.

They have to be **context-independent models**. If they are not, we could not be sure that such a model unit would be appropriate in a new context.

Since building and using compositional, context-independent, conceptual models is a complex task in itself, we need powerful tools supporting the development, validation, and maintenance of models and model libraries, and we need inference systems for reasoning about such models and for solving tasks like interpretation of observations, predicting the evolution of systems, or developing remedial strategies.

2. 1 What Is Offered?

The requirements outlined above are obviously a challenge to knowledge representation and automated problem solving, which belong to the key areas of Artificial Intelligence. Indeed, AI has delivered theories and systems facing this challenge, especially in research on qualitative modeling and model-based systems. The former work has developed formalisms for the kind of reductionist modeling we need, both for artifacts that are established by interconnected components and natural systems comprising different interacting processes. In this approach, **process models** describe configurations of objects or substances and elementary dynamic changes imposed on properties of the participating objects (see Forbus [2008], Heller-Struss [2001]). Examples are a chemical reaction involving certain substances, heat transfer between objects, a prey-predator relation, or the impact of an increase in agricultural productivity on the income of small farms.

Since the conceptual entities (objects and their relations, processes) are explicitly represented and subject to automated reasoning, process models are **conceptual** models in the sense stated above. They are also **compositional**. Each process model explicitly states the preconditions for a certain effect to occur. Different process models can be combined and interact via their impact on quantities of the objects in the system. The effect of one process can trigger another process by establishing its preconditions, and it can destroy process preconditions, even of itself (evaporation may finally reduce the amount of liquid to zero). The modeling formalism cannot guarantee **context-independent** models, because it cannot prevent a modeler from hard-wiring certain assumptions in the model. But it provides a construct, which is quite fundamental and lies beyond other representation of changes, for instance, in ordinary differential equations. An individual process model in a library can only state its own contribution to dynamic changes, but must not assume the non-existence of other, processes that could amplify, compensate for, or overrun this contribution. As a consequence, its impact cannot be described by a differential equation, which deterministically fixes a first-order derivative. The impact of an individual process can only be stated as an **influence** that determines the derivative only in combination with all other processes that have an impact on the same quantity.

The models developed in this field of AI satisfy the requirement for context independence also in a more specific sense, namely in refraining from implementing a particular computational direction and order , which may vary with the task, available data etc.

Models do not have a value unless they are used for some purpose. The field of model-based systems has developed theoretical foundations and systems to perform **problem solving** exploiting models (see Struss [2008]). While most of this work is concerned with artifacts and based on component-oriented modeling, some of the basic ideas and techniques have been transferred to the environmental domain and process-oriented modeling (e.g. Heller-Struss [2002]). Different tasks can be formalized in logic and supported or automated by inference engines that exploit a process library (see Struss [2004]). For instance, **situation assessment** based on a set of observations (and, perhaps, assumptions) can be formalized as the task of composing a model from the library that is (minimal and) consistent with the observations. Since such a model contains the conceptual layer, it also forms the basis for providing **explanations** of the observations, namely in terms of the processes and their preconditions that contribute to certain observed effects. The resulting model can be used for **prediction** of the potential future behavior of systems. **Remedy proposal** in unwanted situations can be formalized as extending a model of this situation by human-controlled processes such that the extended model becomes consistent with a formulated goal or healthy development. All these inferences crucially depend on the mentioned properties of the models: explicitly represented concepts are the basis for automated reasoning about the model, and compositionality and context-independence allows for the automated composition and revision of models, without the intervention of a modeler. Furthermore, the building of model libraries itself can exploit the formal theories and techniques, e.g. by checking for inconsistencies among model fragments or between models and data, which supports debugging and modifying, or even discovering models.

2. 1 The Vision

What we are proposing is more than yet another application of the modeling technology developed in AI. Rather, it provides the starting point for a development that aims at a major qualitative step in the research on environmental problems and decision support. The grand vision is that **research** in relevant areas does no longer produce results only in terms of reports, scientific papers, and collections of empirical data. Instead, it **produces models**, more specifically model fragments as contributions to a large general library of phenomena that are relevant to a subset of environmental issues. This way, one cannot only read and understand the results obtained by other researchers (or from experience and traditional knowledge); these results are also incorporated in a set of new model fragments that are ready for being integrated in existing models, replacing refuted old models or establishing an alternative hypothesis. The minute they are published this way, the results would immediately be available to other researchers, to enhancing existing models and also to checking their validity in a different context. Comparison of rivaling hypotheses would become much easier, because alternative models could easily be generated by replacing well-identified model fragments. Wouldn't this not only ease model building itself, but speed up scientific progress in the relevant disciplines?

Of course, a lot work remains to be done. The field of model-based systems needs to solve a number of open theoretical and technical problems. Not only simplistic qualitative models need to be composed and exploited, but also numerical or semi-quantitative (e.g. interval) models. The integration with other modeling approaches, e.g. finite elements analysis, has to be established. In general, stronger spatial representations need to be incorporated. Also temporal reasoning has to be included, adding a dimension of complexity, e.g. in stepping from situation assessment to "evoluation assessment" and from remedy proposal to complex "remedy planning".

However, for environmental researchers, there is no use in simply waiting for the model-based technology to become more mature and powerful. It will never, if not exposed to and challenged by the application domain in real contexts. The research relevant to environmental systems has to contribute its own major share to the solution. An ontology has to be developed that comprises the basic concepts as a foundation for being able to state model fragments in a coherent way and for combining them. It has to be open and flexible enough for future modifications and revisions of the existing model fragments. A uniform modeling formalism with the necessary expressive power has to be designed or, perhaps, a set of specialized formalisms with appropriate interfaces between them. It is a huge research project in itself, and it is certainly overwhelming when attempted in a comprehensive way. However, for certain focused areas (water quality as an example) the time and the technology appear to be mature to start such an enterprise.

REFERENCES

Forbus, K.., Qualitative Modeling. van Harmelen, F., Lifschitz, V., and Porter, B. (eds.). *Handbook of Knowledge Representation*, Elsevier, 2008.

Heller, U., Struss, P., Consistency-Based Problem Solving for Environmental Decision Support. *Computer-Aided Civil and Infrastructure Engineering, 17*, 79-92, 2002.

Struss, P., Artificial Intelligence Methods for Environmental Decision Support. *e-Environment: Progress and Challenge*, ISBN: 970-36-0215-0, 1- 14, 2004

Struss, P., Model-based Problem Solving. van Harmelen, F., Lifschitz, V., and Porter, B. (eds.). *Handbook of Knowledge Representation*, Elsevier, ISBN-13: 978-0-444-52211-5, 395-465, 2008.