18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

# Towards model integration and model-based decision support for environmental applications

**Struss, P.** [1]

[1] *Tech. Univ. Munich, Comp. Sci. Dept., Garching, Germany*
*Email: struss@in.tum.de*

**Abstract:** We argue that scientific results from different fields that can help decision making about environmental problems should be delivered in the form of executable and combinable models. For this purpose, such models have to be stated in a coherent modeling framework. For their integration, it is not essential (and even unnecessary) that they share a mathematical formalism. More fundamentally, they have to be stated at a conceptual level in order to identify and relate the objects and quantities that occur in the various model fragments. In order to be truly compositional, the models have to be formulated as independent model fragments that represent elementary processes, and they have to be stated in a context-independent way to enable their reuse for different purposes.

Computer-based decision support should be based on such models and generic algorithms for drawing inferences from these models. The basic steps in such a model-based decision support system are situation assessment, which is the task of generating a model that is compliant with the observations provided. Based on such a situation assessment, the next step is therapy proposal, which amounts to generating a model that combines the model of the current situation and models of human interventions and is compliant with the goals to be achieved by the remedy. Figure 1 displays this basic architecture. These steps can be formalized as instances of model revision in logic and realized by consistency-based diagnosis techniques.

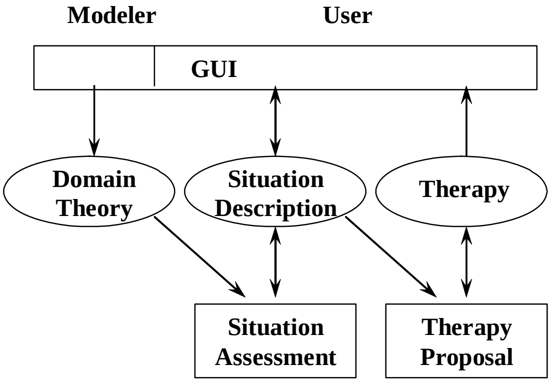


**Figure 1.** Generic architecture of a model-based decision support system.

As the modeling formalism, we propose what has been developed as process-oriented modeling in Artificial Intelligence. The elementary model fragments, called processes, contain an explicit representation of their preconditions, stated in conceptual terms by reference to objects, their existence, properties, and relations. Their effect part does not only relate quantities, but also create objects and relations. Such model fragments can be formalized as logical formulas of the form

$$\mathit{StructuralConditions} \wedge \mathit{QuantityConditions} \quad \Rightarrow \quad \mathit{StructuralEffects} \wedge \mathit{QuantityEffects}.$$

The logical foundation of this modeling formalism allows for the integration with the logic-based model revision algorithm. Since the effects of a process may imply (or negate) the preconditions of other proccesses, the composition of the model for a certain scenario based on a library of elementary processes can be performed by an automated reasoning process and is not dependent on modeling or domain experts.

As a consequence, this approach promises multiple benefits, the major ones being
- support for the integration of models from different research fields and sources,
- re-use of model fragments in different contexts and for different purposes,
- availability of expert knowledge (captured by the model library) for non-experts.

***Keywords:*** *Decision Support System (DSS), model-based reasoning, automated modeling*

## 1. INTRODUCTION

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. 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 and 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.

## 3. MODEL-BASED DECISION SUPPORT

Environmental decision support systems have to capture domain knowledge of experts in the field of ecology and/or environmental issues and facilities for enabling the users to state information about their particular problems, for generating answers and solutions based on this information and the domain knowledge, and for presenting them in a comprehensible manner.

We chose model-based systems to approach this ambitious goal. Some fundamental assumptions underlying our work are:

- The domain knowledge can be represented by a set of generic, independent, and, hence, re-usable model fragments (“processes”) that describe the relevant phenomena and are collected in a library (“domain theory”).
- The user faces two distinct tasks (Figure 1): situation assessment (understanding “What goes on?”) and therapy proposal (“What can be done?”).
- A proper answer to situation assessment is given by a model that can be composed from the library and “explains” the partial information about a situation that is available to the user, and, similarly,
- An adequate therapy can be found by constructing an extension of the situation model by feasible actions that satisfies a set of behavior goals.

The basis of our solution is a novel integration of logical theories and implementations of process-oriented modeling ([Forbus 84]) and consistency-base diagnosis ([Struss 08]) in a generalized diagnosis engine, called G+DE. In this paper, we will not restate the formal theories and technical details which have been described in [Heller-Struss 02], but just summarize the concepts and focus on discussing how a decision support system can be based on them.

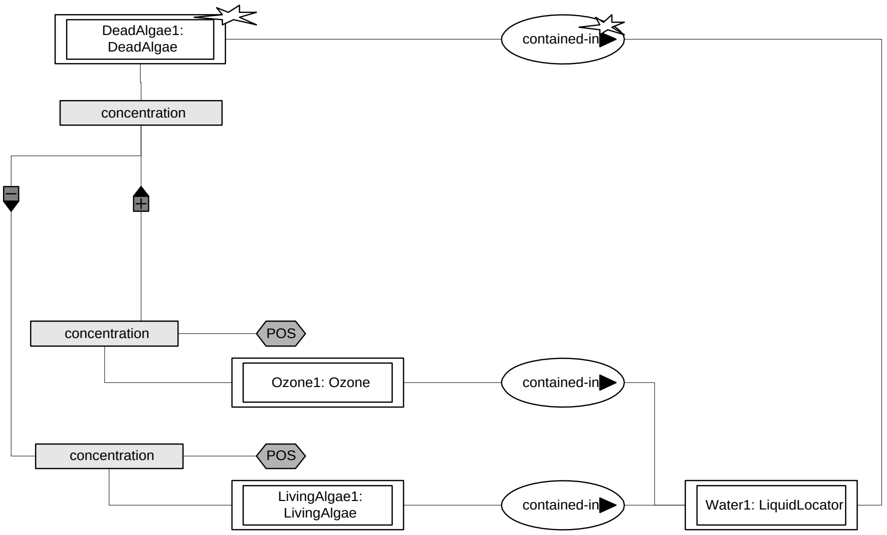
## 4. PROCESS-ORIENTED MODELING

The domain theory has to provide a vocabulary for behavior descriptions and the inferences that derive behavioral constituents from a structural description. It introduces

- *behavior constituent types*. These are physical phenomena which are considered to contribute to the behavior of the overall system. Examples are alcalinization, water transport or algal blooms. They occur deterministically under certain conditions, and their occurrence generates particular effects.

Applying the distinction between structural and quantity aspects to both conditions and effects, we obtain

- *structural conditions*: assertions about the existence of relations and objects (e. g. of sedimental iron)
- *quantity conditions*: statements about values of quantities (e. g. a low pH in a reservoir)
- *structural effects*: creation or possibly even elimination of objects and relations (e. g. the generation of dissolved iron from the sedimental one)
- *quantity effects*: can be expressed as restrictions on variables (e. g. the dissolved iron concentration rises with the sedimental iron concentration and lower pH). Here, we also allow for partially specified effects in the form of influences as in the Qualitative Process Theory of [Forbus 84].



Structural conditions:
Ozone contained-in Water
LivingAlgae contained-in Water
Quantity conditions:
concentration of Ozone greater than zero
concentration of LivingAlgae greater than zero
Structural effects:
DeadAlgae contained-in water
Quantity effects:
Positive influence on concentration of DeadAlgae
Negative influence on concentration of LivingAlgae

**Figure 2. A process "Ozone kills algae"**

The concept of influences is special and goes beyond classical mathematical modeling through equations or differential equations. Influences capture the contributions of a process to changes in the system. This concept reflects a requirement that arises from the compositional modeling scheme: we need to describe the effects locally w.r.t. a model element; but without knowing which other model elements will affect the same quantity, no definite constraint can be established for the influenced quantity. For instance, the evaporation rate influences the amount of a (liquid) water body negatively. But this may, nevertheless, grow due to other processes, such as precipitation. Influences can be positive or negative and act on a variable or its derivative. Roughly, an influence is a statement about the partial derivative of the influenced variable w.r.t to the influencing one. Hence, the abstract form of a behavior constituent type can then be written as

$$\mathit{StructuralConditions} \wedge \mathit{QuantityConditions} \quad \Rightarrow \quad \mathit{StructuralEffects} \wedge \mathit{Constraints} \wedge \mathit{Influences}$$

where “Constraints” refers to fixed relations or equations. More precisely, we state that for each constellation of objects satisfying the structural and quantity conditions, an instance of the behavior constituent occurs and imposes the respective effects on the constellation. Figure 2 shows a process in a graphical and a textual notation.

Based on this modeling formalism, we discuss the two steps of model-based decision support.

## 5. SITUATION ASSESSMENT

In a model-based system, the first step, situation assessment, is formally stated as follows: Given a set of observations (measurements, descriptions of geomorphological or biological features, results of visual inspection, etc.) and general domain knowledge, determine the relevant phenomena and active processes which cannot be or have not been observed directly. We might distinguish two different kinds of this task, although they frequently have to be solved together:

- *System Identification* which again has two aspects. One is *structure* identification, i.e. determining the (types of) constituents of the system and their interactions. Second, *parameter* identification, i.e.

determining constants that characterize the particular instances of such constituents. The starting point is a description of the entities that are directly observable (the topography, present species, etc.) and measurable quantities (salinity of the water, current, etc.). The goal is to establish a behavior model of the target system. Obviously, this is a reasoning task which requires scientific knowledge and expertise, since it is about recognizing the processes that are caused by or triggered by the visible configuration of entities, but happen behind them, often difficult or impossible to observe (e.g. deposit of sediments, chemical processes, etc.). Computer support to this task should satisfy a number of requirements. First, knowledge that has been gained in modeling similar systems and through generalization should be *re-used* directly where appropriate. Thus, we want to avoid having to build a model of each area from scratch. Second, the input to such a modeling system should be entered in a way that is natural to the user who may be a domain expert or a local semi-expert and whom we cannot force to use some mathematical formalism or computer language. Third, many of the available observations are inherently qualitative in nature, such as "trough-shaped area" or "increased degradation", but, nevertheless, carry crucial information that has to be taken into account. .

- *State Identification* interprets observations in order to infer internal states or tendencies in the system *at a particular time*. Obviously, this is based on, or combined with, system identification.

Both tasks can occur as "diagnostic" tasks if the given observations do not match the expectations or a given model.

## 6. THERAPY PROPOSAL

Given descriptions of a disturbed system (including the causes for the disturbance as discussed above) and of our goals, determine actions suited to re-establish a state that complies with our goals. In the simplest case, such actions are designed to shift a single quantity in the proper direction, e.g. increasing the flow from the trough to the sea by digging canals (which is what is actually being done). In general, the task can require planning a sequence of actions over time achieving a number of different sub-goals ("First, take steps to decrease salinity, then introduce species X to change the soil, finally, re-introduce species Y"). In addition to representing the ecosystem itself and the goals of management or conservation, computer support to this task needs to incorporate knowledge about the potential interventions, and, more specifically, this has to be done in a way that enables the analysis of their impact on the ecosystem.

Again; this can be regarded as a “diagnostic” task, applying revisions to a model that is inconsistent with some goals, i.e. what is considered healthy conditions, until consistency is achieved.

## 7. MODEL REVISION

It turns out that both tasks can be solved by the same algorithm, in switching the roles of the various ingredients. For instance, in situation assessment, we may start with a model of the normal state, but the given observations may contradict this model and force us to retract some normality assumptions to gain a picture of the actual situation.

Formally, the general revision process starts with a model that is inconsistent with the observations or some goals (both represented as sets of quantity specifications):

$MODEL_0 \cup OBSorGOALS \vdash \bot$.

Its result is a modified model (or several candidates) that removes the inconsistency with the observations (for situation assessment) or with the goals (for diagnosis and therapy):

$MODEL_1 \cup OBSorGOALS \nvdash \bot$.

Such a revision process is the core of model-based diagnosis systems that have been developed for technical systems and which currently enter industrial applications (see [Struss 08]). Their techniques also form the basis for solutions in our domain. The key to a focused proposal of model revisions is the following: the inconsistencies occur as conflicting values for variables derived from observations and the constraints of the behavior model. Techniques that record the dependencies of the constraints and values on specific elements of MODEL0 identify candidates for a revision. Guidance is usually given by some minimality criterion (w.r.t. sets, number, or probability of revisions).

Usually, there is not a unique proposed revision. Since the model-based system can explore the consequences of different hypotheses and, hence, their distinctions, it is able to propose measurements or tests that help to narrow down the set of candidates.

What distinguishes the two kinds of applying this revision algorithm is not only the interchange of observations and goals, but also what establishes the revisions of the model: in situation assessment, objects and conditions related to the physical world that may cause disturbances or unexpected phenomena. For therapy proposal, the model is extended by actions that may cause remedies. Therefore, actions have to be represented as processes, as well.

## 8. REPRESENTING ACTIONS

Actions may have certain physical preconditions for their applicability which can be stated in terms of structural conditions and quantity conditions. However, unlike ordinary processes, they do not become automatically active when these physical preconditions are satisfied. They have an additional precondition to become effective: some human intervention. Such an intervention can be an entire sequence of human activities (filling a container with some substance, connecting it to a treatment tank by a pipe, and opening a valve) which has to be taken into account when actually planning the work or estimating its costs. However, from the point of view of reasoning about adequate therapies, it suffices to regard them as atomic entities.

Hence, an easy way to integrate actions in the modeling formalism is to represent the human interventions as a special kind of objects, called action triggers, whose existence is a structural condition of the respective actions.

Since action triggers only depend on the decision and the respective activities by humans, they can never appear as structural effects of processes or other actions. Furthermore, we have to make sure that different action instances have different action trigger objects, even if they are instances of the same action type. Otherwise, several instances of an action type could be triggered by the same object. One way to achieve this within our modeling formalism without additional concepts is to guarantee that action triggers have a location that is unique to every instance of an action type. For instance, if a container has several connections to other tanks, the trigger of an *open_connection* action has to be specific to each opening rather than the container, because in the latter case, the opening of one connection would also trigger the opening of all others.

## 9. 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?

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 "evolution 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 84] Forbus, K., Qualitative Process Theory. Artificial Intelligence, 24(1-3), 1984.

[Heller-Struss 02] U.Heller, P.Struss: Consistency-Based Problem Solving for Environmental Decision Support. In: Computer-Aided Civil and Infrastructure Engineering 17 (2002) 79-92. ISSN 1093-9687

[Struss 08] P. Struss: Model-based Problem Solving. In: van Harmelen, F., Lifschitz, V., and Porter, B. (eds.). Handbook of Knowledge Representation, Elsevier, 2008