

Process-model-based Situation Assessment in a Case-study in Limnology

Jana von Wedel, Peter Struss

Technische Universität München, Comp. Sci. Dept.
{jana.vonwedel, struss}@mytum.de

Abstract

The paper describes work in progress the application of process-oriented modeling to situation assessment within a research project in limnology. In the field study, some 50 lakes in Southern Germany are monitored in order to explore and document the changes in the distribution of a particular aquatic plant, *Najas marina*, and understand their preconditions. The ultimate objective is using this native plant, which is dependent on favorable temperature conditions, as an indicator for predicting the possibility of the intrusion of thermophilic neophytes. We discuss how process-oriented modeling is used to perform situation assessment, i.e. in this case to explain the (non-)occurrence of *Najas marina* in a lake, and also discuss challenges to process-oriented modeling, especially concerning spatial and temporal representations and reasoning.

Introduction

The application of qualitative process-oriented modeling to ecological problems has been subject to a number of publications in the field. Often, this activity amounted to manually generating a model of a particular system as a basis for producing a simulation or an envisionment that coincided with some expected behavior (usually as one out of a zillion of other possible behaviors).

We believe that building models to confirm what has been the basis for constructing the model does not really exploit the potential of process-oriented modeling as a basis for automatically constructing a system model. This has been the motivation for previous work on a “diagnostic” use of model libraries ([Collins 1993], [Heller-Struss 02]): starting from a set of observations about a particular section of the real world, say an ecological system, try to perform automated **situation assessment**, which means answering the question “What is going on?”. We formalize this as the task as **constructing a model** that provides a causal account for what has been observed from a library of domain-specific processes.

We follow this approach accompanying a project conducted by the limnology department of our university that aims at understanding the factors affecting the distribution of an aquatic macrophyte in German lakes.

The next section will present situation assessment in a nutshell. Then we present the field study, its requirements on modeling and situation assessment, and how we addressed them in the model library. Finally, we highlight some of the challenges we encountered, which are of general interest to process-oriented modeling of ecological systems.

Situation Assessment Based on Process-oriented Modeling

It is performed by constructing **situations**, i.e. sets of assertions regarding existing objects, their relations, processes, and quantities that are consistent with the observations based on a domain library of processes. Following Qualitative Process Theory ([Forbus 84]) and our logical and computational reconstruction of it ([Heller 2001], [Heller-Struss 02], [Struss 11]), a process states that certain effects will be established whenever its preconditions are satisfied, i.e. an **implication**:

$$\text{StructuralConditions} \wedge \text{QuantityConditions} \\ \Rightarrow \text{StructuralEffects} \wedge \text{QuantityEffects},$$

where **StructuralConditions** and **StructuralEffects** assign existence to objects and structural relations, and **QuantityConditions** and **QuantityEffects** contain assignments of values (or ranges) to quantities. In addition, **QuantityEffects** include also **influences**, which capture the contribution to the process on the dynamics of the systems (which may rival with counteracting influences of other processes). We assume that there exists a **process library** representing the core of the domain knowledge.

Constructing situations involves starting from the given observations (which may be considered as facts or assumptions) and iteratively completing them in two directions:

- **Forward completion**: adding all implications of an intermediate result of the construction process and the process library, i.e. instantiating processes whose preconditions are satisfied and their effects. This establishes the causal impact of the intermediate

result, but it does not address the main goal, namely finding a reason for what has been observed.

- **Consistency check:** if the resulting situation is consistent, it is possible answer to situation assessment. Otherwise, it is incomplete causally upstream and requires

- **Backward completion:** this looks for process candidates whose effects yield changes in quantities and/or existence of objects and relations that are unexplained. If there is no such process, the search is cut off here. Since there can be several candidates, the search may branch.

The somewhat surprising fact that pure consistency-based reasoning yields an abductive result is due to two axioms:

- **influence resolution:** if its result yields an inconsistency, it can only be resolved by an additional process; a special case is a change in a quantity that is not influenced in the current model .

- **Existence default:** objects and relations do not exist unless they are given as observations, effects of active processes, or as introducibles, which are discussed below.

These closures are supported by closed-world assumptions that are associated with existence variables and quantities. In the Generalized Diagnosis Engine G+DE ([Heller 2001], [Heller-Struss 02]) consistency-based diagnosis ([de Kleer-Williams 1987], [Dressler-Struss 1994]) is performed to deliver (minimal) sets of assumptions that create an inconsistency. Revising such a closed-world assumption means searching for additional processes that provide an effect on the respective variable and, hence, performing backward completion in an informed and focused way.

The concept of **introducibles** is crucial for terminating the search: otherwise, repeated backward completion would usually ultimately result in an inconsistency, because some object remains unexplained. This reflects that each model library has a limited horizon of what can be explained. For instance, in our case study, the existence of a sediment in a lake does not have to be established by some process (as opposed to the existence of seeds in a lake). All objects that do not occur in a StructuralEffect of any process in the library will have to be marked as introducible. However, for a particular task, the model boundary, i.e. the scope of the explanation, may be tighter.

As a result, situation assessment, starting from initial assertions, includes causally downstream processes and their impact, but also causally upstream processes and their impact. It will usually deliver alternative results, and the user may have to pick the most plausible one.

Characterized in a more formal way, the **result of situation assessment** should be a **minimal** situation

- containing **all facts**
- otherwise only **introducibles** and
- **effects** of occurring processes
- being **closed** w.r.t. **effects**,
- in which a **maximal set of assumptions** holds.

The Case Study on Najas Marina

The Field Study

The environmental decision support system that is described in this paper was developed in the context of a research project headed by Dr. Arnulf Melzer and Dr. Uta Raeder of the limnology department of the Technische Universität München, the goal of which was to determine the vulnerability of German lakes to the invasion of foreign macrophytes as a result of climate change. Due to climate change higher water temperatures and an increased chance of flooding, which in turn can lead to a higher nutrient concentration in the water, are expected. These changes may allow some foreign species to establish themselves in German lakes, thereby changing their ecosystems in the long term.

The indigenous macrophyte *Najas marina* (Figure 1) shares many properties with the relevant non-native macrophytes, most prominently its dependency on sufficiently high water temperatures. *Najas marina* was for a long time endangered but has re-established itself in German lakes in recent years. Between 1988 and 2011, the number of Bavarian lakes in which the plant could be observed has more than doubled, as depicted in Figure 2. This dispersal area can be explained by a rise in water temperatures in the same time period.



Figure 1: The indigenous macrophyte *Najas marina*

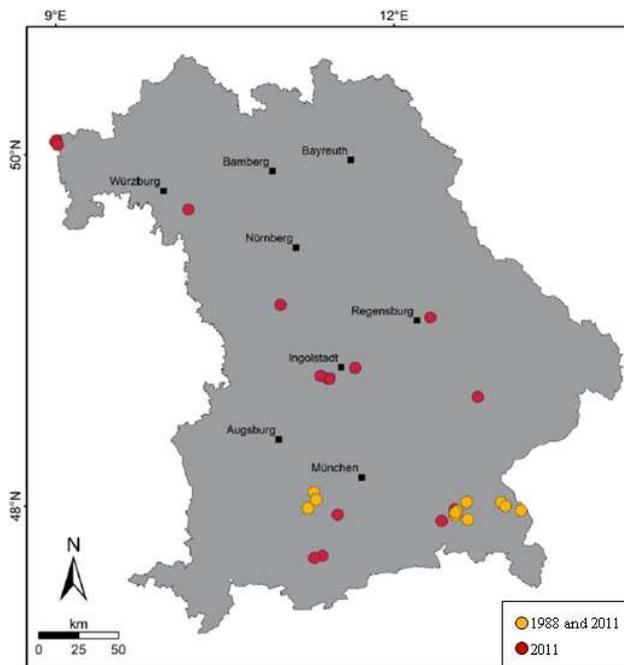


Figure 2: Occurrences of *Najas marina* in Bavaria, Germany in 1988 and 2011

Due to these facts it is thought that *Najas marina* can function as an indicator as to whether foreign thermophilic macrophytes are likely to establish themselves in a given lake as a result of climate change. In order to better understand the growth and spreading of *Najas marina*, as well as to allow for the prediction of future spread, 54 lakes in southern Germany were monitored with respect to the following factors:

Water temperature	Hardness
pH	Phosphorus content
Conductivity	Ammonium
Oxygen saturation	Nitrate
Oxygen concentration	soluble reactive phosphorus SRT
Secchi depth	genotype(s) of <i>Najas marina</i>

The conclusions that could be drawn from these measurements, as well as from additional experiments and research, were the basis for the process library we created that reflects the natural processes relevant to the growth and spread of *Najas marina*.

Relevant Impacts on *Najas Marina* Growth and Spread

Whether and if so, to which extent *Najas marina* can grow in a certain lake depends on a number of conditions that influence seed development, root establishment and growth. Since *Najas marina* is a thermophilic plant water temperature is one of the most important factors. In

general, seeds can never germinate in the same season they were created but have to go through a resting period with low water temperatures first and can only germinate after this period, if the water temperature reaches 15°C or higher. [Melzer-Raeder 2012] If seed coats are cracked, germination may also occur at slightly lower temperatures, around 12°C ([Agami-Waisel 1984], [van Vierssen 1982]). Seed coats may crack due to the seeds undergoing a period of very low temperature or due to exposure to an animal digestive system.

Water temperature also plays a crucial role in the later development stages of *Najas marina*. To develop blossoms and new seeds temperatures of 20°C or higher are required. [Melzer-Raeder 2012] Another important factor in the phase of root development and plant growth is the lake's sediment. The nutrient composition and concentration as well as the cohesive strength of the sediment must be favorable ([Hoffmann et al. 2013], [Handley-Davy 2002]).

Additional factors are among others the pH-level, which must be seven or higher in order for *Najas marina* to grow, Secchi depth and light intensity, water level and turbidity, electrolyte concentration as well as the presence of competing plants, e.g. *Ceratophyllum demersum* (hornwort) or *Myriophyllum spicatum* (Eurasian watermilfoil), dense populations of which may hinder the immigration of *Najas marina*.

An overview over the life cycle of *Najas marina*, annotated with conditions that must hold for the transition from one development stage to another, is given in Figure 3.

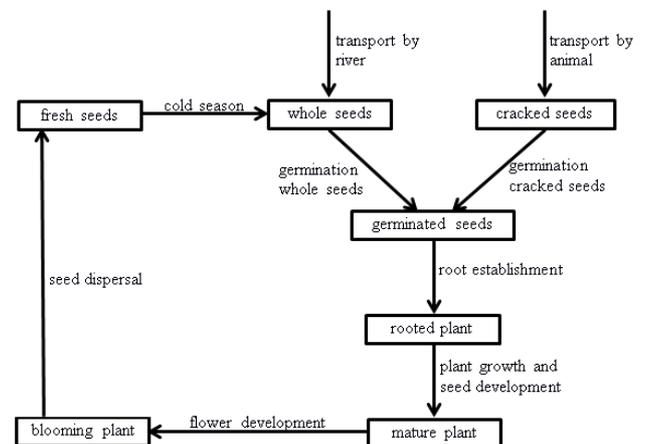


Figure 3: The life cycle of *Najas Marina*

Spatial Representation

In order to be able to correctly model the spread of *Najas marina* we must take into account the different ways in which seeds may be transported from one lake to another one, namely via water connections or by animals. For the

latter the beeline distances between lakes are relevant, as they determine whether water birds may be able to transport seeds from one lake to another. For streams connecting lakes, we need to model their direction, as seeds may only be transported in the same direction as that of the water flow.

Since the conditions relevant to the growth and spreading of *Najas marina* may vary greatly even within a lake it is further necessary to allow for the modelling of **compartments**, i.e. different sections of a lake with homogenous conditions. Compartments may be adjacent to one another, allowing for exchange processes, but also be contained within one another.

Figure 4 shows an example of 2D spatial relations between three lakes, each of which is subdivided into several compartments.

Furthermore, a compartment may consist of several vertical compartments, which, for instance, correspond to the different temperature layers of the lake (Figure 5).

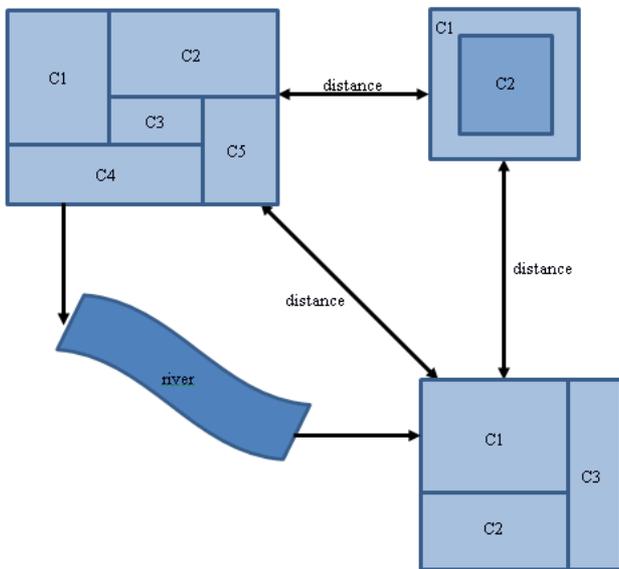


Figure 4: Spatial relations between lakes

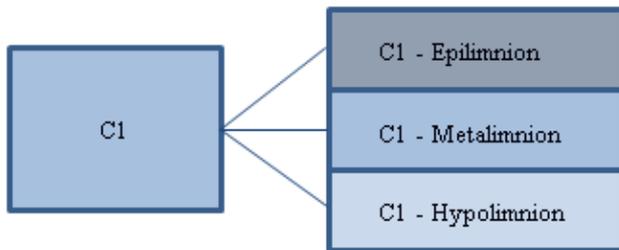


Figure 5: Vertical partitioning of a lake

Process Library

We explicitly model *Najas marina* at its different development stages, allowing us to check for plants in a certain development stage and to model the transitions between development stages as structural preconditions and effects, respectively. By associating quantities with structural elements we are further able to model processes which result in changes for some *Najas* individuals of a certain development stage, but not all of them. For example not all mature *Najas marina* seedlings will germinate once the conditions for germination are met, but only a certain percentage of them, in order to ensure the survival of the species in the lake in case unfavorable conditions lead to the death of germinated seeds or plants at a later point in the season.

Spatial relations between and within a lake are modeled using the relations ‘contained_in’ and ‘connected’, the latter of which indicates adjacency.

Table 1 lists some examples from our library, which contains roughly 25 process types.

Examples of Situation Assessment

Table 2 contains four simplified examples of information that might be available concerning a lake. The beeline distances between the lakes can be found in Table 3. In addition to the information that can be drawn from these tables it is known that a connection exists from lake C to lake D and that mallard ducks live by lakes A and D.

	Preconditions		Effects	
	Structural Conditions	Quantity Conditions	Structural Effects	Quantity Effects
Seed preparation	\exists seeds_A Type(seeds_A, fresh_NM_seeds) \exists lake_A Type(lake_A, lake) \exists hypolimnion_A Type(hypolimnion_A, hypolimnion) Contained_in(lake_A, hypolimnion_A) Contained_in(hypolimnion_A, seeds_A)	$0^\circ\text{C} < \text{Temp}(\text{hypolimnion_A}) \leq 4^\circ\text{C} (\geq 28 \text{ days})$ $\text{ph-value}(\text{lake_A}) \geq 7$	\exists seeds_effect Type(seeds_effect, whole_NM_seeds) Contained_in(hypolimnion_A, seeds_effect)	$1 - (\text{seeds_A_amount})$
Permination of whole seeds	\exists seeds_A Type(seeds_A, whole_NM_seeds) \exists lake_A Type(lake_A, lake) Contained_in(lake_A, seeds_A)	$\text{Temp}(\text{lake_A}) \geq 15^\circ\text{C}$ $\text{ph-value}(\text{lake_A}) \geq 7$	\exists seeds_effect (@ 1 day) Type(seeds_effect, germinated_NM_seeds) Contained_in(lake_A, seeds_effect)	$1 - (\text{seeds_A_amount})$ (@ 1 day)
Seed transport by birds	\exists lake_A Type(lake_A, lake) \exists lake_B Type(lake_B, lake) \exists seeds_origin Type(seeds_origin, whole_NM_seeds) Contained_in(lake_A, seeds_origin) \exists birds_A Type(birds_A, water_birds) live_by(birds_A, lake_A)	$\text{distance}(\text{lake_A}, \text{lake_B}) \leq \text{max_distance}(\text{birds_A})$	\exists seeds_dest Type(seeds_dest, cracked_NM_seeds) (@ 1 day) Contained_in(lake_B, seeds_dest)	$1 - (\text{seeds_orig_in_amount})$
Seed transport via connection (whole)	\exists lake_A Type(lake_A, lake) \exists river_A Type(river_A, river) \exists seeds_origin Type(seeds_origin, whole_NM_seeds) Contained_in(lake_A, seeds_origin) Connected(lake_A, river_A)		\exists seeds_dest Type(seeds_dest, whole_NM_seeds) Contained_in(river_A, seeds_dest)	$1 - (\text{seeds_orig_in_amount})$

Table 1: Examples from the process library

Lake	Avg temp spring	Avg temp summer	Size	ph-value	Secchi depth	Cohesive strength sediment	Najas marina
A	19.6°C	25.5°C	175ha	8.3	30cm	Low	yes
B	13.9°C	18.6°C	132ha	7.9	60cm	Medium	no
C	13.7°C	20.8°C	45ha	7.6	130cm	Medium	yes
D	17.4°C	22.9°C	86ha	8.1	80cm	high	no

Table 2: Relevant properties of lakes

	A	B	C	D
A	-	227km	72km	181km
B	227km	-	152km	48km
C	72km	152km	-	110km
D	181km	48km	110km	-

Table 3: Distances between lakes

Given this data, the following are three examples of questions that we might ask ourselves and that can be answered using situation assessment:

1. If the average air temperature were to rise by 2°C, would *Najas marina* likely occur in lake B?
2. How can the occurrence of *Najas marina* in lake C be explained? The temperature in spring is below the critical 15°C mark for germination.
3. How can it be explained that no *Najas marina* growth was observed in lake D? The critical temperatures for germination and plant development are met.

Situation assessment based on the library derives answers to these questions as follows.

- 1.1. Virtual observation: “Air temperature rise”
- 1.2. Effect: Lake temperature rises according to the size of the lake
- 1.3. Given the size of lake B this rise leads to a temperature <20°C
- 1.4. Plant growth process remains inactive, since its quantity conditions are not satisfied.

We can therefore conclude that *Najas marina* is not likely to occur. Given the size of the lake an increase of 2°C in air temperature would not lead to average summer temperatures above the critical mark of 20°C.

- 2.1. Observation: ∃ NM plants in lake C
- 2.2. Only process that has ∃ NM plants as structural effect: “plant growth”
- 2.3. Quantity conditions met according to observations; Structural condition: ∃ germinated NM seeds needs to be explained
- 2.4. Two processes with ∃ germinated NM seeds as structural effect:

2.4.1. “germination of whole seeds”

2.4.2. “germination of cracked seeds”

- 2.5. Quantity conditions for “germination of whole seeds” violated according to observations
- 2.6. Quantity conditions for “germination of cracked seeds” met according to observations; Structural condition: ∃ cracked NM seeds needs to be explained
- 2.7. Two processes with ∃ cracked NM seeds as structural effect
 - 2.7.1. “seed transport by birds”
 - 2.7.2. “seed transport via connection (cracked)”
- 2.8. Structural conditions for “seed transport via connection (cracked)” violated
- 2.9. All conditions for “seed transport by birds” met

This result indicates that mallard ducks transported seeds from lake A to lake C. Since seeds with a cracked coat can germinate at lower temperatures than whole seeds, the temperatures in spring were sufficient for those seeds to germinate.

- 3.1. Approach: Pretend that NM plants were observed and see whether an inconsistency is detected
- 3.2. Only process that has ∃ NM plants as structural effect: “plant growth”
- 3.3. An inconsistency is detected: quantity conditions are violated. Quantity condition “cohesive strength sediment ≤medium” is inconsistent with our observation of high cohesive strength of the sediment

We therefore conclude that due to the high cohesive strength of the sediment *Najas marina* plants are not able to root in lake D and any germinated seeds would therefore wither.

Issues and Challenges

In this case study, we face a number of requirements that raise questions both at a principled and a technical level, several of which still wait for a general answer. In the following, we discuss some of these challenges, related to temporal and spatial aspects and to some kind of non-determinism of processes.

Temporal Representation and Inferences

The above presentation of the domain problems and the samples from the library clearly indicates the importance of a non-trivial representation of time.

- **Timed preconditions:** conditions need to be satisfied for certain periods, as, for instance, in $0^{\circ}\text{C} < \text{Temp}(\text{hypolimnion_A}) \leq 4^{\circ}\text{C} (\geq 28 \text{ days})$.

Other preconditions, such as on pH, must hold during the same period, and reasoning about metric time, at least at the level of calendar days, is required.

- **Approximate temporal preconditions:** the preconditions like the above one are less crisp than they may appear at a first glance: it may be considered fulfilled if there were one or two days with temperatures a little higher. Moreover, there is no way to safely infer the precondition from observations obtained twice a week. One may be tempted to try to refine the precondition by introducing permissible short violations etc. However, we propose not to do so, because this would mix the conceptual level of the model with specificities of the frequency and nature of measurements (e.g. sensors or use of available meteorological data), but to encapsulate this in a **data interpretation** module, accepting that this will tend to be less declarative, but increase the reusability of the process type as a model fragment.

- **Repetitive temporal structures:** in the considered application (as in many other biological and ecological domains), there is a natural repetitive temporal structure imposed by seasonal changes, and one may want to model this explicitly. Another example of such a cyclic structures would be life histories of population, i.e. a sequence of juvenile, reproductive, and post-reproductive phases. Our current solution keeps this implicit, but could generate the pattern of seasons triggered by the observations.

Other aspects of the temporal representation are related to the spatial representation.

Spatial Representation

While the compartment-oriented approach described above appears to be adequate and matches domain concepts, for instance, in ecology and hydrology, there are a number of superficial and deep issues here:

- **Intersecting compartments:** compartments should be characterized by homogeneous conditions regarding the relevant physical quantities and, hence, introduce the local granularity for process instantiation (which is discussed in a little more detail below). However, since there are usually several quantities that induce a spatial partitioning into compartments, one has to intersect all of them to obtain the required compartments. As an example, consider that the entire lake has a certain pH; temperature introduces a vertical structure (as in Fig. 5), while depth introduces a horizontal partitioning of the lake. Unless a geometric description of these compartments is given, which would allow computing the relevant intersection of the partitions induced by

ranges of the various quantities, one has to enumerate the compartments resulting from intersection manually. A desirable feature would be that process instances can be located at the largest compartments where their preconditions hold and applied to all sub-compartments automatically. For instance, a process that depends on temperature only should be instantiated for a whole horizontal layer and automatically affect all sub-compartments.

- **Dynamic compartment structure:** There may be a need to keep the compartment topology dynamic: rivers may dry off dependent on precipitation; a flooding may extend the lake to include the reed belt, which triggers a process of washing nutrients into the lake, etc. This requires treating compartments as objects that may be modified, destroyed, or created, which may change the spatial relations among them. Actually, this is related to a fundamental modeling issue:

- **Compartments as spatial or physical objects:** is the compartment “Lake C” considered as a depression in the landscape that may contain water or as the water body that is contained in this depression? The implication of, say, the water evaporating completely would be dramatically different: the depression persists, the water body does not.

- **Locating objects in compartments:** there is a need to classify objects w.r.t. how they can be localized and distributed in sub-compartments. For instance, for a substance, such as phosphorus, dissolved in a lake, we want to conclude it appears in all sub-compartments, and we may express this by making its concentration a homogeneous quantity across the entire lake. In contrast, a barrel with a contaminated content known to have been dumped somewhere in the lake, could be hypothesized in several sub-compartments, but in any situation in only one of them. A single big catfish, however, could occur in different compartments, although not at exactly the same time. This suggest a classification of objects into, at least, distributed objects, fixed single objects, mobile objects.

- **Compartments and locating processes:** In providing homogeneous conditions regarding the preconditions of processes, they are the smallest entities to locate processes. We propose a fundamental classification of process types: on the one hand, processes occurring **within** a compartment (which may be called **transformation** processes), such as chemical processes, reproduction of a population in an area, etc.), and, on the other hand, processes occurring **at the interface** of two compartments (**exchange or transportation** processes). The latter ones would imply some

straightforward preservation laws (regarding mass, energy etc.), while the former ones may preserve quantities that are not explicit in the model (e.g. transforming chemical into heat energy).

Non-determinism in Processes

Modeling processes as implications means that they cause the effects in a deterministic way. This may be inappropriate, as illustrated by the seed transportation by birds. Even if the distance between lakes is short enough, it is not certain that water birds will actually move, and if they do, there is no guarantee they will carry seeds in their stomach.

In situation assessment, i.e. in its **abductive** use, this does not cause a problem, because the process is correctly identified as a possible cause for the occurrence of plants in the target lake. However, the process should not deliver certain **predictions**. A way to control this is within the current logic-based solution is to add a virtual precondition that represents the unknown implicit trigger that establishes the effect. In our view this is more appropriate than following the frequent suggestion to handle this problem by probabilities, which will usually have no empirical foundation.

Summary

The work on this project so far, has validated the utility of our approach to situation assessment, but is still in progress and incomplete regarding the issues discussed above that will require both more theoretical work and new technical solutions and will continue to do so. Based on a re-implementation of the Generalized Diagnosis Engine G+DE ([Heller 2001], [Heller-Struss 02]) we aim at first providing an interactive tool that allows a user to control the search in situation assessment, rather than a fully automated solution.

References

- Agami, M., and Waisel, Y., 1984: *Germination Of Najas Marina L.* Aquatic Botany, Vol. 19 (1984), p. 37-44
- J. Collins: Process-based Diagnosis: An Approach to Understanding Novel Failures. Doctoral thesis, Institute for the Learning Sciences, Northwestern University, 1993.
- J. de Kleer and B. Williams: Diagnosing Multiple Faults. Artificial Intelligence, 32(1), 1987.
- O. Dressler, P. Struss: Model-based Diagnosis with the Default-based Diagnosis Engine: Effective Control Strategies that Work in Practice. In: Proceedings of the 11th European Conference on Artificial Intelligence (ECAI-94), John Wiley, 1994.

K. Forbus: Qualitative Process Theory. Artificial Intelligence 24(1-3), 1984.

Handley, R. J., and Davy, A. J., 2002: *Seedling root establishment may limit Najas marina L. to sediments of low cohesive strength.* Aquatic Botany, Vol. 73 (2002), Issue 2, p. 129-136

Ulrich Heller: Process-oriented Consistency-based Diagnosis-Theory, Implementation and Applications. Dissertation, Fakultät fuer Informatik, Technische Universität München, 2001.

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

Hoffmann, M., Sacher, M. , Lehner, S., Raeder, U., and Melzer, A., 2013: *Influence of sediment on the growth of the invasive macrophyte Najas marina ssp. intermedia in lakes.* Limnologica, Vol. 43 (2013), p. 265– 271

Melzer, A., and Raeder, U., 2012: *Zwischenbericht zum Forschungsvorhaben: Die Invasion von Wasserpflanzen als Folge des Klimawandels – ein Vorhersagemodell für die Gefährdungsanfälligkeit stehender Gewässer in Bayern*

Struss, P. : A Conceptualization and General Architecture of Intelligent Decision Support Systems. In: Chan, F., Marinova, D. and Anderssen, R.S. (eds) MODSIM2011, 19th International Congress on Modelling and Simulation. pp. 2282-2288. ISBN: 978-0-9872143-1-7.

Van Vierssen, W., 1982: *Some Notes On The Germination Of Seeds Of Najas Marina L.* Aquatic Botany, Vol. 12 (1982), p. 201-203.