© *Acta hydrotechnica* **24/40 (2006)**, Ljubljana
ISSN 1581-0267

UDK/*UDC*: 519.61/.64:556.55 Prejeto/*Received*: 15. 1. 2008
Pregledni znanstveni prispevek – *Review scientific paper* Sprejeto/*Accepted*: 5. 8. 2008

# AVTOMATIZIRANO MODELIRANJE JEZER Z UPORABO PODATKOV IN EKSPERTNEGA ZNANJA: EVALVACIJA APLIKACIJ
# AUTOMATED MODELLING OF LAKES FROM DATA AND EXPERT KNOWLEDGE: EVALUATION OF APPLICATIONS

**Nataša ATANASOVA**

*Ekološki modeli jezer so uporabna orodja tako za boljše razumevanje delovanja ekosistema kot tudi za upravljanje z jezeri, oblikovanje okoljske politike ter testiranje in vrednotenje različnih inženirskih rešitev. Postavitev takega modela močno otežuje sama kompleksnost ekosistema. Zato je za povečanje relevantnosti in zanesljivosti bodočega modela ekosistema zaželeno uporabljati čimveč modelarskih pristopov. V tem prispevku podajamo oceno metode Lagramge, ki združuje dva osnovna pristopa k modeliranju, to sta učenje modelov iz podatkov (induktivni pristop) in gradnja modelov iz teoretičnega domenskega znanja (deduktivni pristop). Metoda podpira vnos domenskega znanja v proces učenja iz podatkov, pri čemer je domensko znanje shranjeno v obliki modelne knjižnice. Štiri aplikacije metode, tj. jezero Glumsø, Blejsko jezero, jezero Kasumigaura in jezero Greifensee, obsegajo različne naloge za Lagramgea, katerih rezultat so strukturno in po vrednosti parametrov specifični modeli opazovanih ekosistemov. Modele smo vrednotili v smislu njihove uporabnosti in natančnosti. Kljub nekaterim pomanjkljivostim evalvacija metode pokaže, da je ta uporabna za modeliranje kompleksnih domen. Uspešno jo lahko uporabimo tako za gradnjo modelov kot tudi za druga znanstvena dognanja, kot so prepoznavanje dinamičnih vzorcev obnašanja ekosistemov oziroma strukturne dinamike ekosistemov.*
**Ključne besede:** *jezera, avtomatizirano modeliranje, Lagramge, modelna knjižnica, konceptualno modeliranje, modeliranje iz podatkov.*

*Ecological models of lakes are useful tools for a better understanding of the ecosystem behaviour, lake management, policy making, as well as testing and accepting engineering solutions. Setting such model is a difficult task due to the complexity of these ecosystems. Therefore it is reasonable to use as many approaches as possible to construct a reliable model of the observed domain. In this paper the evaluation of an automated modelling method, called Lagramge, that combines the two basic approaches, i.e. data-driven (inductive) approach and knowledge-driven (deductive) approach, is given. The method supports the introduction of domain knowledge in the procedure of equation discovery from measured data, where the domain modelling knowledge is introduced in a form of modelling knowledge library. Four applications of the method, i.e. Lake Glumsø, Lake Bled, Lake Kasumigaura, and Greifensee, comprise different modelling tasks for Lagramge, each of them resulting in a specific model of the observed domains. The models are evaluated in terms of their descriptive power and their performance (goodness of fit to the measurements). Although faced with some constraints, the method can be successfully used in complex domains. It can be used successfully for model discovery as well as for other scientific discoveries, such as identifying dynamic patterns in the observed system, i.e. dynamic structure of the ecosystem.*
**Key words:** *lakes, automated modelling, Lagramge, modelling knowledge library, conceptual modelling, data-driven modelling.*

## 1. UVOD

Jezera so kompleksni in dinamični ekosistemi. V številnih primerih je njihovo

## 1. INTRODUCTION

Lakes are complex and dynamic ecosystems. In many cases their behaviour is obnašanje težko predvidljivo, kar botruje neuspešnemu upravljanju z jezeri. Ekološki modeli jezer so orodja, ki se zaenkrat večinoma uporabljajo v znanstvenih raziskavah, vendar pa se njihova uporabnost vedno znova pokaže tudi pri upravljanju, pri sprejemanju okoljske politike kot tudi pri preverbi in sprejemanju inženirskih rešitev. Veliko takih modelov je bilo razvitih in objavljenih v svetovni literaturi. Zajetna baza ekoloških modelov obstaja na: http://dino.wiz.uni-kassel.de/ecobas.html (Benz & Voigt, 1996; Benz & Knorrenschild, 1997). Modeli v bazi so dokumentirani pod enotnim sistemom, imenovanim ECOBAS (Benz & Hoch, 1997; Hoch *et al.*, 1998; Benz *et al.*, 2001).

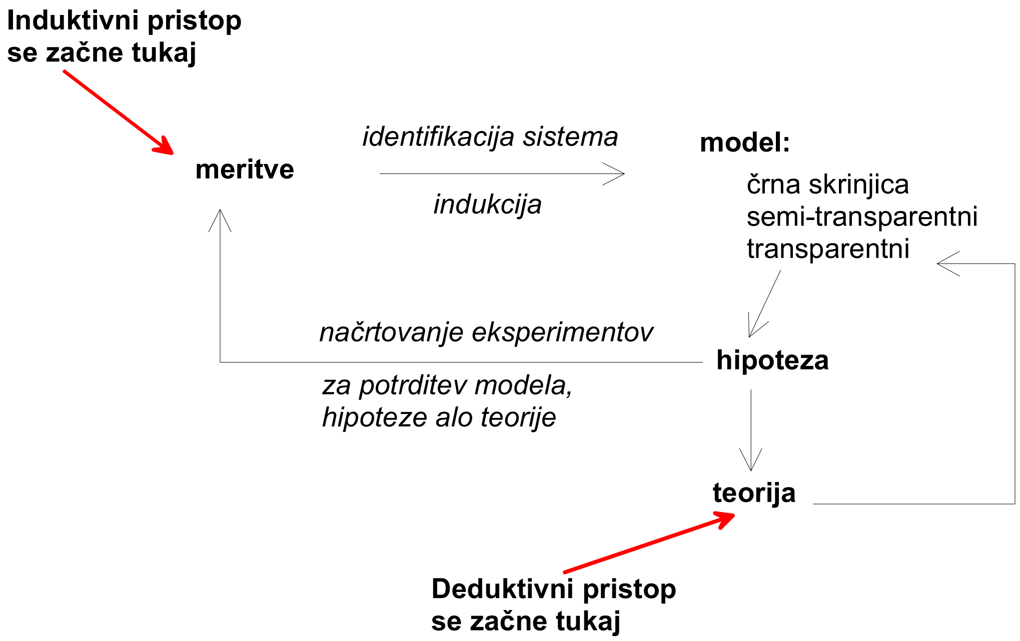
Kjub njihovi prisotnosti in popularnosti med raziskovalci ostaja naloga gradnje ekološkega modela velik izziv, saj smo pri tem dostikrat soočeni s problemom razumevanja delovanja nekaterih procesov v sistemu. Zato je uporaba čimveč pristopov k modeliranju zelo priporočljiva. Slika 1 prikazuje temeljne pristope k modeliranju (Kompare, 1995). Razvidna je konstantna prepletenost med osnovno teorijo, meritvami (opazovanji sistema) in modelom. V odvisnosti od začetne točke našega modelarskega postopka v tem diagramu ločimo med dvema osnovnima pristopoma, to sta pristop gradnje (učenje) modela iz podatkov (induktivni) in gradnja modela iz osnovne teorije (deduktivni pristop).

Pri teoretičnem (deduktivnem) pristopu so modeli zgrajeni iz osnovnih fizikalnih, kemijskih in bioloških principov. So torej transparentni in razumljivi raziskovalcem in domenskim ekspertom, zaradi česar so zelo priljubljeni (npr. Jørgensen & Bendoricchio, 2001; DeAngelis, 1992; Chapra, 1997). Kot pa smo že omenili, so ekosistemi zelo kompleksni in naše znanje o njihovem delovanju omejeno. Zato so tudi enačbe modelov prirejene našemu pomanjkljivemu znanju. To ima za posledico številne neznane parametre v modelih po eni strani, po drugi pa številne modelne strukture v literaturi, ki opisujejo isti proces. Z drugimi besedami: enolični model (lahko je ustrezen, a ne nujno 'pravi') za določen ekosistem ne obstaja. Torej je kvaliteta modelov močno odvisna od modelarjevih sposobnosti in izkušenj.

difficult to predict, which makes their management unsuccessful. Ecological models of lakes are highly appreciated tools, whose use is still primarily for scientific purposes, but they can also be well used in environmental management, policy making, as well as testing and accepting engineering solutions. Many such models have been developed and published in the literature. A comprehensive database of ecological models can be found at: http://dino.wiz.uni-kassel.de/ecobas.html (Benz & Voigt, 1996; Benz & Knorrenschild, 1997). The models in the database are documented under a unifying documentation system called ECOBAS (Benz & Hoch, 1997; Hoch *et al.*, 1998; Benz *et al.*, 2001).

Despite of their omnipresence, the task of establishing models is very demanding. Many times the modeller is faced with a problem of understanding the system in the first place. Therefore, using as many approaches to modelling as possible is strongly encouraged. Figure 1 represents the basic modelling paradigm (Kompare, 1995). It indicates a constant interaction between the basic theory, measurements (systems observations) and the model. Based on the starting point of our modelling procedure in this diagram, we distinguish between two basic approaches, i.e. data-driven (inductive) approach and knowledge-driven (deductive) approach.

Knowledge-based (deductive) approach models are constructed from basic physical, chemical and biological principles. Therefore they are transparent and clear to the domain experts, which results in their popularity (e.g., Jørgensen & Bendoricchio, 2001; DeAngelis, 1992; Chapra, 1997). However, as already stated, ecosystems are complex and our knowledge is not complete. Thus, the equations in mathematical models are adapted with regard to our incomplete knowledge. This results in many unknown parameters in our models on one hand, and in a set of possible mathematical model structures for a single process on the other hand. In other words, there is no single suitable (corresponding, but not necessarily correct!) model for a specific system. Thus, the quality of the obtained models greatly depends on the modeller's skills and experiences.




Slika 1. Pristopi k modeliranju, prirejeno po Kompare (1995).

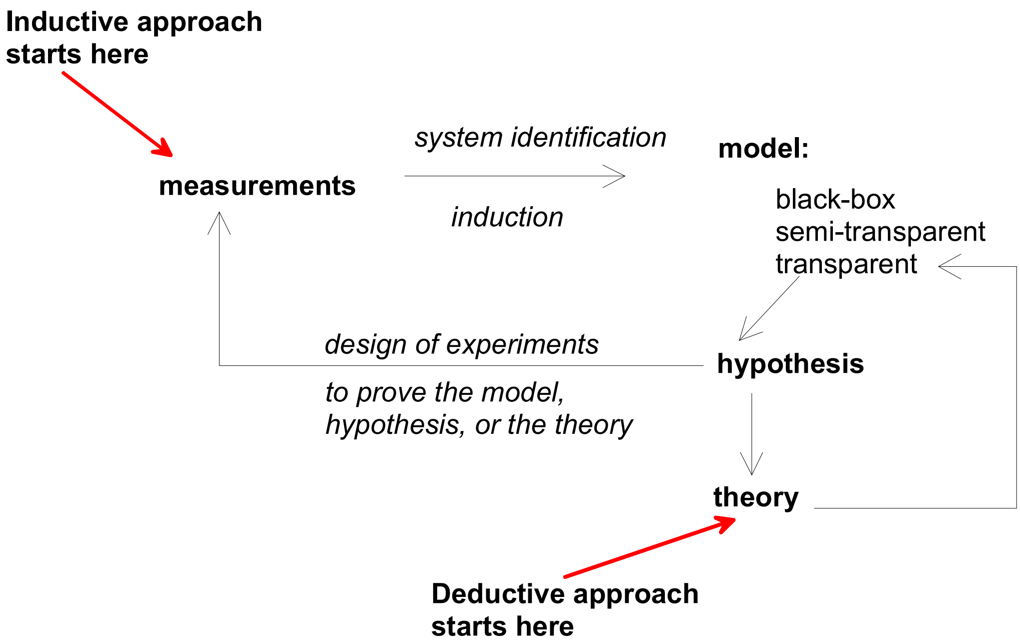



*Figure 1. Modelling paradigm, adopted from Kompare (1995).*

Nasprotno deduktivnemu pristopu je učenje modelov iz podatkov (opazovanj). Induktivni pristop vključuje statistične metode kot tudi metode s področja umetne inteligence. Ta pristop omogoča reševanje različnih ekoloških problemov brez vnosa domenskega znanja v proces gradnje modela. Model je zgrajen samo na podlagi meritev oziroma opazovanj.

In contrast to the knowledge-driven modelling, data-driven modelling is aimed at building models from observations. It comprises statistical methods, as well as methods that include data mining (artificial intelligence). This approach enables to tackle various (ecological) problems without the necessity to introduce any domain knowledge

Številne izmed omenjenih metod, še zlasti pa statistične metode, proizvajajo t. i. 'black-box' modele, ki jih eksperti ne morejo razložiti (ali pa zelo pomanjkljivo) ali opisati njihovega delovanja. Vendar pa podpodročje umetne inteligence, strojno učenje (ML), vsebuje algoritme, s katerimi lahko dobimo t. i. semitransparentne modele. Te lahko domenski eksperti delno razumejo in razložijo. Uspešne aplikacije različnih algoritmov strojnega učenja v ekologiji najdemo npr. v (Kompare, 1995; Kompare *et al.*, 2001; Todorovski *et al.*, 1998).

Zavedajoč se prednosti in pomanjkljivosti teoretičnega in induktivnega pristopa h gradnji modela sta Džeroski & Todorovski (2003) razvila metodo Lagramge, ki uporablja kombinacijo teh dveh oziroma združuje prednosti obeh pristopov: transparentnost, enostavnost in natančnost. Metoda podpira vnos domenskega znanja v postopek indukcije enačb iz merjenih podatkov. Domensko znanje vnašamo v obliki generičnih procesov. V tem smislu je Todorovski (2003) razvil formalizem za zapis domenskega znanja v knjižnico za modeliranje. Z uporabo formalizma je bila izdelana domenska knjižnica za modeliranje jezerskih ekosistemov (Atanasova *et al.*, 2006a). Znanje v knjižnici obsega modeliranje prehranjevalnih mrež (ali kroženja hranil) v jezerih z navadnimi diferencialnimi enačbami oziroma s principom masnih bilanc. Formalizirano je v obliki:

(1) taksonomije tipov spremenljivk,
(2) osnovnih generičnih procesov, ki opisujejo delovanje vodnih ekosistemov,
(3) alternativnih modelov (formulacije) osnovnih procesov in
(4) znanja, kako kombinirati te modele posameznih procesov v model celotnega ekosistema.

Z uporabo knjižnice in orodja za avtomatsko modeliranje Lagramge so bili zgrajeni modeli sledečih realnih domen: Blejsko jezero, Slovenija (Atanasova *et al.*, 2006b), jezero Kasumigaura, Japonska (Atanasova *et al.*, 2006c), jezero Glumsø, Danska (Atanasova *et al.*, 2007), Beneška laguna, Italija (Atanasova, 2005) in jezero

in the process of model construction. The measurements solely drive the procedure of model construction. As a result many of these methods, e.g. statistical methods, produce the so called 'black-box' models, which cannot be explained, or rather vaguely, by domain experts. Yet, a branch of artificial intelligence algorithms, i.e., machine learning (ML), tends to produce the so called semi-transparent models. These can be partly explained and understood by a domain expert. Successful applications of different machine learning techniques in ecology can be found for example in (Kompare, 1995; Kompare *et al.*, 2001; Todorovski *et al.*, 1998).

Knowing the benefits and the drawbacks of the knowledge-driven and data-driven approaches, respectively, Džeroski & Todorovski (2003) developed a method (Lagramge) that brings together the good properties of both approaches (deductive and inductive), i.e. transparency, simplicity and accuracy. The method supports the introduction of domain knowledge in the procedure of equation discovery from the measured data. Domain knowledge is introduced in a form of generic processes. In this manner Todorovski (2003) developed the formalism for encoding the domain knowledge into a modelling library. Using the developed formalism Atanasova *et al.* (2006a) elaborated a knowledge library for modelling of lake ecosystems. The knowledge in the library comprises modelling of the food web (or nutrient cycling) in a lake by following the mass conservation principle. It is formalized in terms of:

(1) taxonomy of variable types,
(2) basic processes that govern the behaviour of aquatic ecosystems,
(3) alternative models of the basic processes, and
(4) knowledge how to combine models of individual processes into a model of the entire ecosystem.

Using the library and the automated modelling method Lagramge models on several real-world domains, i.e., Lake Bled, Slovenia (Atanasova *et al.*, 2006b), Lake Kasumigaura, Japan (Atanasova *et al.*, 2006c), Lake Glumsø, Denmark (Atanasova *et al.*, 2007), Lagoon of Venice, Italy (Atanasova, Greifensee, Švica (Atanasova *et al.*, 2005).

Cilj tega prispevka je podati pregled teh aplikacij in na ta način ovrednotiti metodo avtomatskega modeliranja s strani uporabnosti v ekološkem modeliranju. Predstavljene so tudi smernice za nadaljnji razvoj orodja Lagramge.

## 2. AVTOMATIZIRANO MODELIRANJE Z METODO LAGRAMGE

### 2.1 POSTOPEK MODELIRANJA

Postopek modeliranja pri teoretičnem pristopu se praviloma začne z *identifikacijo problema* in *zbiranjem podatkov* (slika 2, levo).

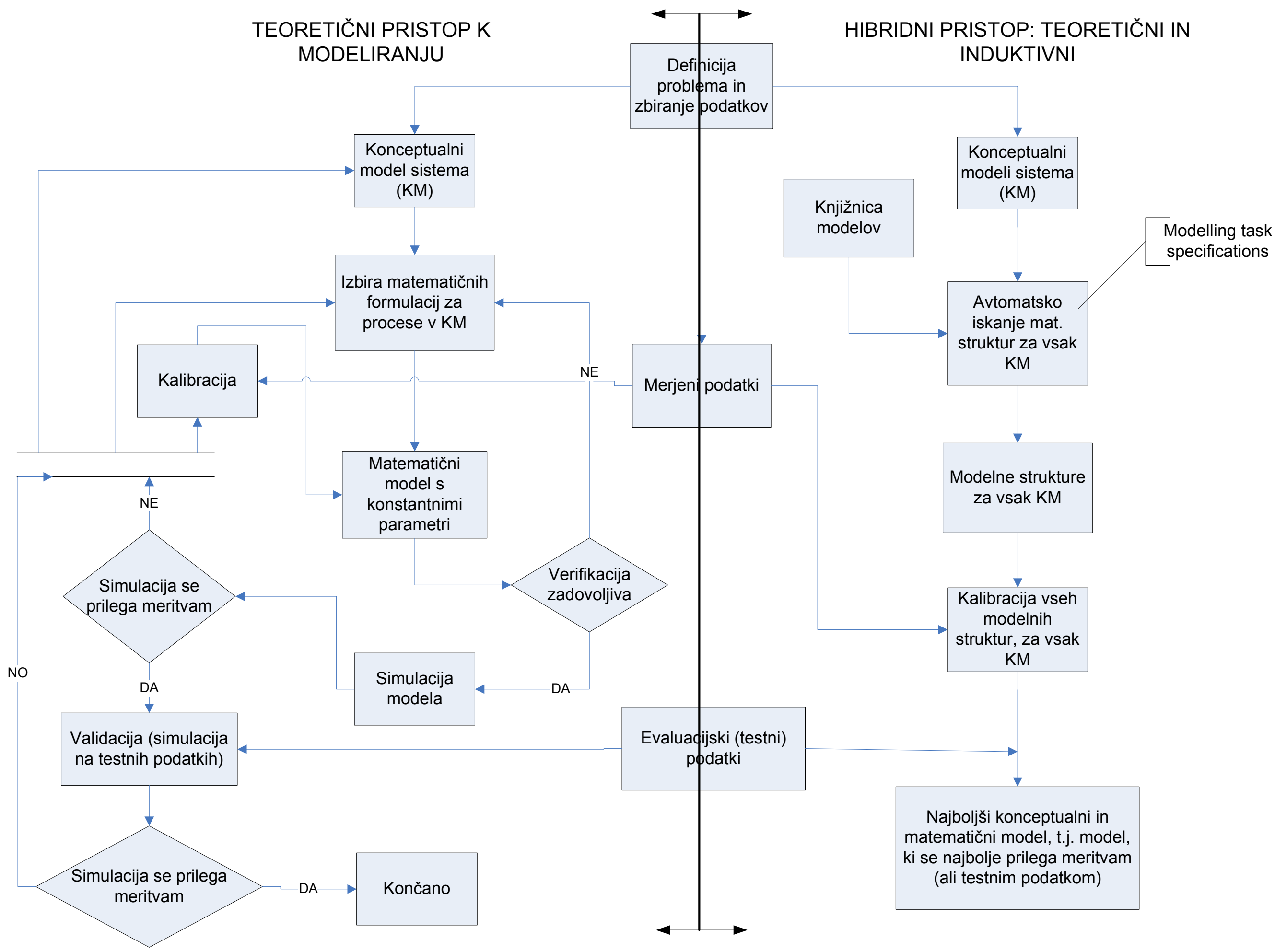
2005), and Greifensee, Switzerland (Atanasova *et al.*, 2005), were constructed.

The aim of this paper is to give an overview of these applications, and to evaluate the method in the light of its applicability in ecological modelling. Finally, future trends in the development of Lagramge are presented.

## 2. AUTOMATED MODELLING METHOD LAGRAMGE

### 2.1 MODELLING PROCEDURE

Typically the modelling procedure within the theoretical approach begins with *Problem identification* and *Data collection* (Figure 2, left).




Slika 2. Primerjava dveh pristopov k modeliranju. Levo: teoretični pristop, desno: hibridni pristop oziroma kombinacija teoretičnega pristopa in pristopa učenja modelov iz podatkov (induktivnega).

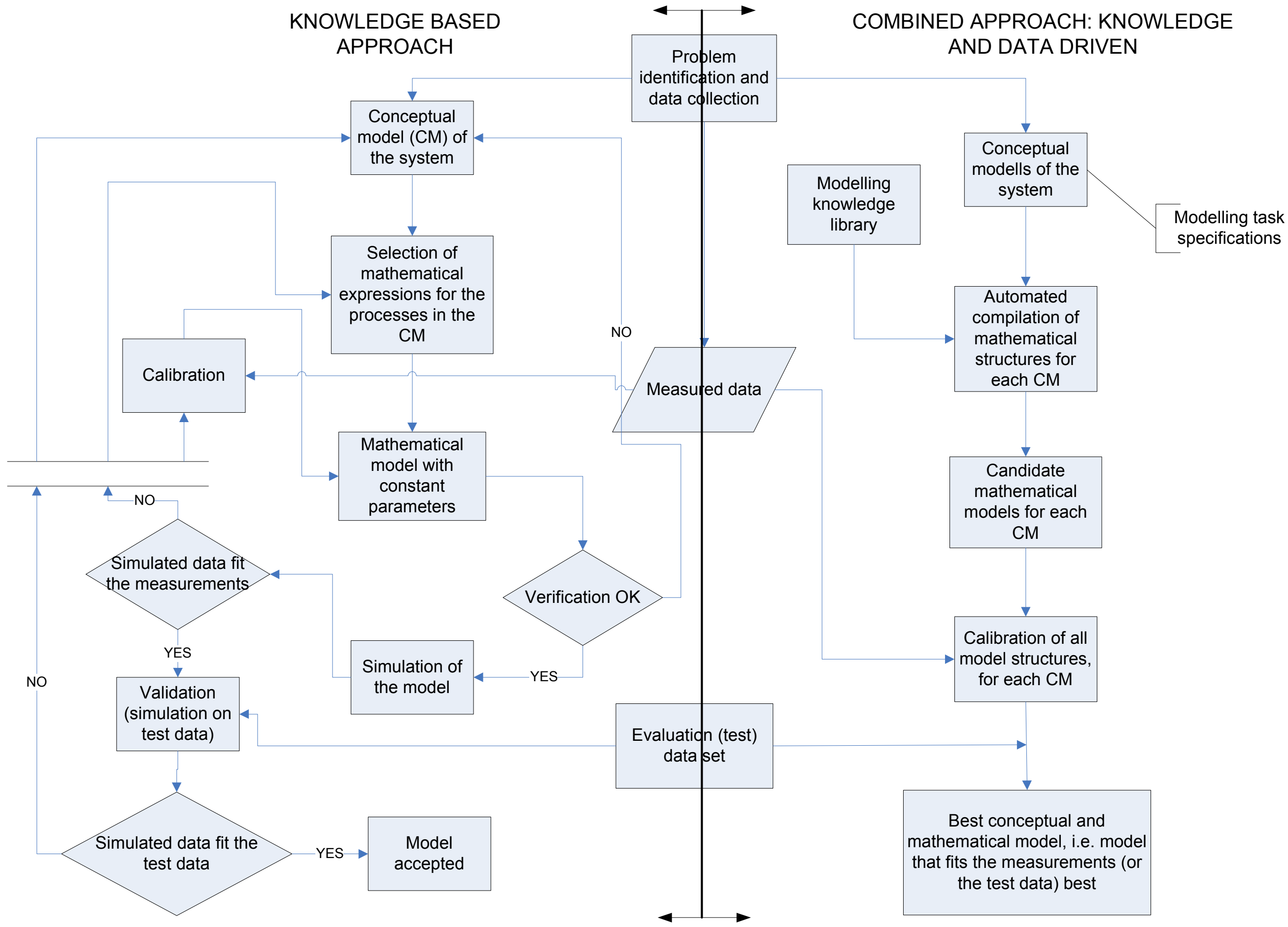



*Figure 2. Comparison of two approaches to process based modelling. Left: knowledge based approach, right: hybrid approach, i.e. combination of knowledge based approach and data driven approach.*

Naslednji korak je *konceptualno modeliranje*, kjer določimo poenostavljeno strukturo sistema z izbiro (1) relevantnih spremenljivk v sistemu in (2) relevantnih biogeokemijskih procesov, ki delujejo na izbrane spremenljivke. Tu velja opozoriti, da je izbira ustreznega konceptualnega modela zahtevna naloga, ki dostikrat zahteva preverbo različnih variant.

Ko izberemo konceptualni model, nastopi *matematično modeliranje*, kjer izbrane procese zapišemo z ustreznimi matematičnimi formulacijami. Tudi tu moramo opozoriti, da za en proces obstaja več (pravilnih) matematičnih izrazov. Rezultat tega koraka je matematični model sistema, sestavljen iz enačb s konstantnimi parametri. Vrednosti konstantnih parametrov se določijo bodisi z meritvami, privzamejo iz literature ali pa se

*Conceptual modelling* or system identification is the next step, where a simplified structure of the system is determined by selecting (1) the relevant variables in the system and (2) the relevant bio-geo-chemical processes that connect these variables. Note here that choosing a correct conceptual model of the observed system is a non-trivial task that many times requires the consideration of different possibilities and trying them out.

Once the conceptual model is selected, *Mathematical modelling* comes to play, where all processes are identified with suitable mathematical formulations. Here too, we should note that each process can be identified with more than one mathematical expression. The result of this step is a mathematical model of the system, composed of equations with constant parameters. The values of the umerijo na podane meritve. Naslednji trije koraki so namenjeni preverjanju in izboljšavam ekološkega modela.

*Verifikacija* je definirana kot "*prikaz, da je formalizem modela pravilen*", Rykiel (1996). V tem koraku pokažemo, da so kalkulacije z modelom smiselne, vhodi v model in računalniška koda pa pravilni. Vendar pa model, s katerim vršimo te kalkulacije, vsebuje parametre, katerih vrednost še ni popolnoma znana. Zato jih ne smemo smatrati kot popolnoma zanesljive (Oreskes *et al.*, 1994).

*Kalibracija* (identifikacija parametrov) je "*ocena in prilagajanje parametrov in konstant modela za izboljšanje prileganja simuliranih vrednosti z modelom meritvam*", Rykiel (1996). Kalibracija se lahko vrši kot del verifikacije ali validacije.

Zadnji korak, *validacija*, se nanaša na delovanje modela v domeni, za katero je zgrajen (Rykiel, 1996). Validacija primerja simulirane vrednosti z modelom z realnimi opazovanji v sistemu, ki niso bili uporabljeni med razvojem modela.

## 2.2 PRIMERJAVA TEORETIČNEGA IN LAGRAMGE PRISTOPA K MODELIRANJU

Pri teoretičnem pristopu k modeliranju se vsi koraki v postopku izvajajo iterativno. Potrebna je stalna interakcija modelarja, ki spreminja konceptualni in/ali matematični model in posledično v zanki izvaja preostale korake za vsako spremembo. Zanka se začne z *verifikacijo modela* (slika 2, levo). Po formulaciji matematičnega modela za izbrani konceptualni model izvedemo verifikacijo, tj. preverimo pravilnost formulacije in obnašanja modela. Če je preverba zadovoljiva, izvedemo simulacijo modela in pri tem primerjamo simulirane in merjene vrednosti spremenljivk. Če primerjava ni zadovoljiva, imamo tri možnosti za izboljšavo modela: kalibracijo, spremembo konceptualnega modela in spremembo matematičnega modela za podani konceptualni model. Vsaka možnost vodi v nov matematični model s konstantnimi parametri, ki ga ustrezno preverimo in simuliramo. Te postopke ponavljamo toliko časa, dokler ne dobimo zadovoljive primerjave

constant parameters are measured, estimated or adopted from the literature. The next three steps are for testing and improving ecological models.

*Verification* is defined as "*a demonstration that the modeling formalism is correct"*, Rykiel (1996). Here we determine that any calculations, inputs, or computer code are correct or true. However, the models that use these calculations are based on parameters which are not completely known so they cannot be taken as absolute truth (Oreskes *et al.*, 1994).

*Calibration* (parameter identification) is defined as "*the estimation and adjustment of model parameters and constants to improve the agreement between model output and a data set"*, Rykiel (1996). Calibration can occur as part of either verification or validation.

Finally, *Validation*, refers to model's performance (Rykiel, 1996). Validation compares simulated system output with real system observations using data not used in model development.

## 2.2 THEORETICAL VS. LAGRAMGE APPROACH

In theoretical approach all modelling steps are performed iteratively. The modeller constantly interacts by adopting/changing the conceptual and/or mathematical model and performing the rest of the modelling steps for each change. The loop starts with the *Verification of the model* (see Figure 2, left). After we formulate a mathematical model for the selected conceptual model, we perform verification, i.e. determine the correctness of the model formulation and behaviour. If the test is passed we simulate the model and compare the simulated data with the measurements. If this comparison is unsatisfactory, there are three options for improving the model – calibration, modification of the conceptual model, or modification of the mathematical formulation of the existing conceptual model. Each option leads to a new mathematical model with constant parameters, which is verified and simulated. This loop is repeated until satisfactory comparison between the simulated med simuliranimi vrednostmi v sistemu z modelom in merjenimi.

*Število ponavljanj korakov v postopku modeliranja določajo tri pomembna vprašanja: (1) ali je modelar izbral pravilen konceptualni model, (2) ali je modelar izbral pravilno matematično strukturo za izbrani konceptualni model in (3) ali so bili parametri modela optimalno določeni.*

Ta vprašanja lahko sistematično rešujemo z uporabo metodologije za avtomatizirano modeliranje, Lagramge, ki deluje na podlagi metode za odkrivanje enačb. V nasprotju s teoretičnim pristopom, ki zahteva veliko interakcij modelarja, Lagramge izvaja vse korake avtomatično, razen izbire konceptualnega modela. Kljub temu lahko Lagramgeu podamo več možnih konceptualnih modelov hkrati, za katere izvede odkrivanje ustreznega matematičnega modela (slika 2, desno).

Za dani konceptualni model torej Lagramge odkrije matematični model, strukturo in vrednost parametrov. To stori na osnovi (1) knjižnice znanja, kjer je formalno zapisano splošno znanje o modeliranju, (2) specifikacije sistema za modeliranje, ki ustreza konceptualnemu modelu ekosistema. Tu uporabnik opredeli pomembne spremenljivke in procese v opazovanem sistemu in (3) časovne serije merjenih podatkov opredeljenih spremenljivk.

Na podlagi podane specifikacije ekosistema in merjenih podatkov Lagramge izvede hevristično iskanje v knjižnici splošnega znanja o modeliranju. Rezultat tega iskanja je množica specifičnih matematičnih struktur modelov, primernih za modeliranje sistema, tj. Lagramge določi prostor potencialnih modelnih (matematičnih) struktur za dani konceptualni model. Hkrati Lagramge umerja parametre potencialnih matematičnih modelov glede na podane meritve, tj. izvaja nelinearno optimizacijo konstantnih parametrov vsake potencialne matematične strukture modela na podane meritve. Ti modeli se vrednotijo z dvema hevrističnima funkcijama. Prva je srednja vrednost vsote kvadratov napake, MSE (mean square error), ki meri odstopanje med merjenim podatkom in podatkom, dobljenim s simulacijo modela. Druga je funkcija najkrajše dolžine opisa, MDL (minimum description length) in upošteva kompleksnost dobljenega

and the measured data is achieved.

*There are three important issues, which determine the number of repetitions of the modelling procedure: (1) has the modeller selected the correct conceptual model, (2) has the modeller selected the correct mathematical structure for the selected concept, and (3) parameter estimation procedure.*

These issues can be addressed systematically using an automated modelling methodology, Lagramge, which is based on equation discovery method. In contrast to the knowledge based approach, Lagramge performs all of the modelling steps automatically, except for the conceptual modelling. Still, we can feed Lagramge with several conceptual models at once and perform the mathematical model development in a single iteration for each conceptual model (See Figure 2, right).

Thus, for a given conceptual model Lagramge discovers the mathematical model, i.e. structure and parameters' values. This is done based on (1) knowledge library, where the general modelling knowledge is encoded, (2) modelling task specification, which corresponds to the conceptual model of the system, where the user specifies important variables and processes that take place in the observed system, and (3) time series data of the specified variables.

After taking the modelling task specification and the measurements, Lagramge performs a heuristic search in the knowledge library. The result of this search is a list of specific mathematical model structures that can be used to model the processes specified in the task specification, i.e. a space of candidate model structures for a given conceptual model. At the same time Lagramge takes all candidate model structures and matches each of them to submitted data by fitting the values of the constant parameters, i.e. Lagramge performs non-linear optimisation of the constant models' parameters according to the measurements. These models can be evaluated by two heuristic functions. One is mean square error (MSE) – it measures the discrepancy between the measured data and the data obtained by simulating the model. The other is the minimum description length (MDL) modela, s tem da vnese preferenco za enostavnejše modele. Model z najmanjšo napako (MSE ali MDL) je najboljši matematični model za podani konceptualni model (specifikacija sistema) opazovanega sistema.

Uporaba orodja Lagramge za ekološko modeliranje je predstavljena skozi aplikacije Lagramgea na realnih domenah, in sicer natančneje na primeru jezera Glumsø (poglavje 3.1).

# 3. APLIKACIJE LAGRAMGEA: OPIS DOMEN, PODATKOV IN EKSPERIMENTOV

## 3.1 JEZERO GLUMSØ

Jezero Glums ø (Jørgensen *et al.*, 1986) se nahaja v sub-glacialni dolini na Danskem. Je plitvo jezero s povprečno globino blizu 2 m. Površina jezera Glumsø meri 266.000 m$^2$. Nekaj let se je v jezero stekala odpadna voda naselja s 3000 prebivalci, očiščena do druge stopnje, tj. biološko je bil odstranjen organski ogljik, ne pa tudi dušik in fosfor. Dodatna obremenitev z dušikom in fosforjem je prispevna površina jezera, ki meri 10,9 km$^2$ in je pretežno agrarnega značaja. Visoke obremenitve s hranili (dušik in fosfor) v odpadni vodi povzročajo hiperevtrofno stanje jezera. Jezero ne vsebuje podvodne vegetacije, najverjetneje zaradi slabe prosojnosti in pomanjkanja kisika.

Naša naloga je odkriti ustrezen model jezera z orodjem Lagramge. Kot že rečeno, Lagramge uporablja modelno knjižnico za sestavljanje matematično pravilnih modelov. Za uporabo znanja iz knjižnice uporabnik Lagramgea najprej sestavi specifikacijo sistema za modeliranje, nato pa poda še merjene podatke. Specifikacija sistema vključuje specifikacijo spremenljivk in ekoloških procesov, pomembnih za opazovani sistem, ki ga modeliramo. Specifikacija spremenljivk in procesov v jezeru Glumsø je podana v preglednici 1. V vrsticah od 1 do 6 so podane spremenljivke sistema, tj. deklaracija tipov spremenljivk: *ns* (topni anorganski dušik), *ps* (topni anorganski fosfor), *phyto* (fitoplankton, izražen kot Chl-a), *zoo* (zooplankton), *temp* (temperatura) in *light* (radiacija).

function that takes into account model complexity and introduces preference towards simpler models. The model with the lowest error is considered as the best model for the given conceptual model (task specification) and data set.

The use of Lagramge for modelling from the user's perspective is presented through the case studies, i.e. more detailed in the case of Lake Glumsø (section 3.1).

# 3. APPLICATIONS OF LAGRAMGE: DOMAINS, DATA, AND EXPERIMENTS

## 3.1 LAKE GLUMSØ

Lake Glumsø (Jørgensen *et al.*, 1986) is situated in a sub-glacial valley in Denmark. It is a shallow lake with an average depth of about 2 m. Its surface area measures 266,000 m$^2$. For several years, it was receiving waste water from a community with 3000 inhabitants, mechanically-biologically treated, i.e. without nutrients (phosphorus and nitrogen) removal. Additional load to the lake represents the lake's watershed (10,9 km$^2$) which is mainly agricultural with almost no industry. The high nitrogen and phosphorus concentrations in the treated waste water caused hypereutrophication. The lake has no submerged vegetation, probably due to the low transparency of the water and oxygen deficit at the bottom of the lake.

Our task is to find a suitable model of the lake with Lagramge. As stated previously Lagramge uses a modelling knowledge library for composing mathematically correct models. In order the knowledge to be used for building models the user needs to first specify the modelling task and next to provide data measurements. The task specification includes declaration of the variables and ecological processes relevant for the system to be modelled. The modelling task specification for Lake Glumsø is presented in Table 1. In lines from 1 to 6 the variable types are declared, i.e. *ns* (dissolved inorganic nitrogen), *ps* (dissolved inorganic phosphorus), *phyto* (phytoplankton, expressed as Chl-a), *zoo* (zooplankton), *temp* (temperature), and *light* (radiation).

Preglednica 1. Specifikacija modela za jezero Glumsø.
*Table 1. Modelling task specification for lake Glumsø.*

| | |
|---|---|
| 1: | variable Inorganic ns |
| 2: | variable Inorganic ps |
| 3: | system variable Primary_producer phyto |
| 4: | variable Animal zoo |
| 5: | variable Temperature temp |
| 6: | variable Light light |
| 7: | process PP_growth(phyto, {ps}, {temp}, {light}) p1 |
| 8: | process Feeds_on(zoo, {phyto}, {temp}) p3 |
| 9: | process Respiration_PP(phyto, {temp}, {ps}, {light}) resp0 |
| 10: | process Sedimentation(phyto, {temp}) sed0 |

Procesi so deklarirani v vrsticah od 7 do 10. Rast fitoplanktona je opisana v vrstici 7, kjer ime procesa *PP_growth* vsebuje pa štiri argumente. Prvi je ime spremenljivke stanja, ki označuje fitoplankton. Argumenti v zavitih oklepajih {}, tj. {ps}, {light} in {temp}, označujejo vplive na proces s strani hranil, svetlobe in temperature. Proces Feeds_on (vrstica 8) predstavlja (1) izgubo fitoplanktona zaradi pašnje zooplanktona in (2) rast zooplanktona (zoo). Izbirna argumenta tega procesa sta hrana (*phyto*) in temperatura (*temp*), kar pomeni, da na rast zooplanktona lahko (ali pa tudi ne) vplivata razpoložljiva hrana (ena ali več vrst fitoplanktona) in temperatura. Na podoben način sta definirana procesa respiracije fitoplanktona in sedimentacija (*Respiration_PP* in *Sedimentation*), vrstici 9 in 10.

Glede na podano specifikacijo sistema Lagramge preišče modelno knjižnico za ustrezne modele (formulacije) deklariranih procesov in jih ustrezno združi v model opredeljene sistemske spremenljivke, tj. model fitoplanktona. Spomnimo, da se v knjižnici nahajajo alternativni modeli posameznega procesa. Posledica tega je množica modelov fitoplanktona, ki se v naslednjem koraku kalibrirajo na podane meritve.

Podatkovni niz jezera Glumsø vsebuje dnevne meritve za obdobje od aprila 1973 do aprila 1974 ter od oktobra 1974 do oktobra 1975. Merjeni so podatki, uporabljeni v specifikaciji sistema: *temp, ns, ps, phyto* in *zoo*.

Cilj raziskave je bil odkriti in validirati

The processes are defined in lines from 7 to 10. Phytoplankton growth is described in line 7. The process name is *PP_growth* and it has four arguments. The first is the name of the phytoplankton state variable. The arguments in the {} brackets, i.e. {ps}, {light} and {temp}, define the influences and limitations of the process by nutrients, light and temperature, respectively. The process *Feeds_on* (line 8) stands for (1) predatory loss of phytoplankton and (2) growth of zooplankton (zoo). Optional arguments of this process are food (*phyto*) and temperature (*temp*), which means that the growth of zoo can be, or not, influenced by the food (none or many species) and temperature. Similarly, the rest of the processes in the system (*Respiration_PP,* and *Sedimentation*) are defined (see lines 9 and 10).

Having this specification Lagramge performs a search through the modelling knowledge library for suitable mathematical models of the specified processes and combines them in a model of the specified system variable, i.e. phytoplankton model. Note that there are alternative models for each of the processes. These result in a set of phytoplankton model structures, which are all calibrated against given measured data in the next step.

The data set for Lake Glumsø includes time series of the measurements for the variables in the task specification, i.e. *temp, ns, ps, phyto,* and *zoo*. In this case we used two years of daily measurements (provided by Jørgensen, 2004) from April 1973 to April 1974.

The experiment was aimed at discovering and model za napoved koncentracije fitoplanktona. Eksperiment vključuje odkrivanje enačbe na podatkih enega leta in validacijo na podatkih drugega leta (Atanasova *et al.*, 2007).

## 3.2 BLEJSKO JEZERO

Blejsko jezero je tipično subalpsko jezero ledeniško-tektonskega izvora. Zaseda površino 1,4 km$^2$ z maksimalno globino 30,1 m, povprečno pa 17,9 m (Sketelj & Rejic, 1958; Rismal, 1980; Remec-Rekar, 1995). Potopljeni greben deli jezero v dve kotanji – zahodno in vzhodno. Podatki o jezeru vsebujejo meritve (od 1987 do 2002) fizikalnih, kemijskih in bioloških parametrov. Vendar pa lahko kot dosledne in ustrezne podatke za indukcijo modelov z Lagramgeom obravnavamo le tiste od leta 1995 do 2002. Pogostost meritev znaša enkrat mesečno. Vzorčevanje poteka na dveh lokacijah (vzhodna in zahodna kotanja), in sicer vsaka dva metra od površine do dna. Podatki, ki smo jih uporabili v eksperimentih, so: temperatura, svetloba, pretoki in raztopljena hranila v pritokih, iztoki iz jezera, raztopljena anorganska hranila v jezeru, tj. fosfor, nitratni dušik in silicij, skupna biomasa fitoplanktona in vrsta zooplanktona *Daphnia hyalina*.

Izvedli smo tri eksperimente (Atanasova *et al.*, 2006b). Najprej smo odkrivali model, ki ustrezno opisuje dinamiko fitoplanktona skozi več let. Identifikacija modela je bila izvršena na podatkih od 1995 do 2001 (podatki za učenje). Leto 2002 je bilo uporabljeno za validacijo. Dobljeni model fitoplanktona se je slabo prilegal meritvam. Zato smo predpostavili, da se struktura jezera preveč spreminja iz leta v leto, da bi ga lahko opisali z enim modelom. To hipotezo smo preverili v drugem eksperimentu, tj. pri identifikaciji sistema ločeno za vsako leto. Za vsako leto smo torej dobili drug model. Vsi modeli so se dobro prilegali meritvam, razen modela za leto 1996. Za to leto smo v tretjem eksperimentu odkrivali model za opis osnovne prehranjevalne mreže, tj. fosfor–fitoplankton–zooplankton. Zaradi zahtevne nelinearne optimizacije, ki bi nastopila v tej nalogi, smo identifikacijo oziroma prostor možnih rešitev za ta model močno omejili (Atanasova *et al.*, 2006b).

validating a phytoplankton model. The discovery was performed on one-year data and the validation on the other (Atanasova *et al.*, 2007).

## 3.2 LAKE BLED

Lake Bled is a typical subalpine lake of glacial-tectonic origin. It occupies an area of 1.4 km$^2$ with a maximum depth of 30.1 m and an average depth of 17.9 m (Sketelj & Rejic, 1958; Rismal, 1980; Remec-Rekar, 1995). A sunken reef divides the lake into two basins – eastern and western. The data about the lake comprise long-term (from 1987 to 2002) measurements of physical, chemical and biological parameters, but only the data from 1995 to 2002 are consistent and suitable for model induction with Lagramge. The samples are taken at two deepest locations in the western and eastern basins, at every two metres from the surface to the bottom with a frequency of one month. The data used for modelling are as follows: temperature, light, inflow data (flow rates and nutrient concentrations), outflow from the lake, dissolved inorganic nutrients in the lake (phosphorus, nitrogen and silica), total phytoplankton biomass, and zooplankton species *Daphnia hyalina*.

Three experiments were performed on Lake Bled (Atanasova *et al.*, 2006b). The first was aimed at discovering a model that would describe the long-term phytoplankton behaviour of the lake. Data from 1995 to 2001 were used for model identification (model structure and parameters), and the data of 2002 to validate the model. Failing to get a very good fit to the long-term data, we conjectured that the lake dynamics changes from year to year. The second experiment was aimed at testing this hypothesis, so we applied Lagramge to build separate models for each year's data. In the final experiment, we aimed at discovering a model that includes three system variables (phosphorus, phytoplankton, and zooplankton) from one year's data (1996). Due to the complexity of space of candidate models and limited computational resources, a strong limitation and control of the search space was applied (Atanasova *et al.*, 2006b).

## 3.3 JEZERO KASUMIGAURA

Jezero Kasumigaura je plitvo jezero na Japonskem s povprečno globino 4 m. Prostornina jezera znaša 662 milijonov m$^3$, površina pa 220 km$^2$. Nahaja se v hiperevtrofnem stanju, ki ga najpogosteje povzroča cvetenje modro-zelenih alg poleti in spomladi z visokimi koncentracijami vrst *Microcystis* in *Oscillatoria*.

Podatki o jezeru vsebujejo meritve od 1986 do 1992. Merjeni podatki so temperatura vode, globalna radiacija, raztopljeni fosfor, nitrat, amonij, silicij, skupni fitoplankton, merjen kot klorofil-a (chl-a), vrste fitoplanktona, tj. *Microcystis, Oscillatoria, Scenedesmus* in *Synedra rumpens* in vrsta zooplanktona *Cladocera*. Podatki so bili uporabljeni kot linearno interpolirane vrednosti med dejanskimi meritvami. Domnevamo, da je bila pogostost meritev enkrat mesečno. Zooplankton je merjen le do leta 1989.

Z uporabo podatkov in ekspertnim znanjem pri postopku odkrivanja enačb smo izvedli naslednje eksperimente za odkrivanje modela za simulacijo Chl-a (Atanasova *et al.*, 2006c):

- Identifikacija dinamike fitoplanktona za vsako leto z odkrivanjem modelov na podatkih, pripravljenih ločeno za vsako leto. Pri tem eksperimentu smo želeli preveriti, ali se dinamika iz leta v leto ponavlja. Ali potrebujemo različne strukture modelov za vsako leto (od 1986 do 1992) ali pa zgolj različne parametre in enako strukturo? Vsakega izmed odkritih modelov, naučenega na podatkih enega leta, smo validirali na preostala leta. S tem smo preverili, ali morda obstaja generični model. Vpliv zooplanktona v tem eksperimentu ni bil upoštevan.
- Odkrivanje modela za Chl-a z učenjem iz celotnega podatkovnega niza, oziroma na podatkih od 1986 do 1991, leto 1992 smo uporabili za validacijo. S tem eksperimentom smo odgovorili na sledeča vprašanja: Ali dolžina učnega niza vpliva na tovrstno modeliranje in kako? Ali je bolje najti eno reprezentativno leto za učenje ali pa se učiti na celotnem podatkovnem nizu, čeprav lahko vsebuje veliko šuma? Vpliv zooplanktona ni bil

## 3.3 LAKE KASUMIGAURA

Lake Kasumigaura is a shallow lake in Japan with an average depth of 4 m. It has a volume of 662 million m$^3$ and a surface area of 220 km$^2$. The hypereutrophic state of the lake causes blue-green algal blooms in summer and autumn with frequently high abundances of *Microcystis* and *Oscillatoria.*

The lake's dataset comprises measurements from 1986 to 1992. The following data are measured: water temperature, global radiation, dissolved inorganic phosphorus, total phytoplankton, measured as chlorophyll-a (chl-a), phytoplankton species (*Microcystis, Oscillatoria, Scenedesmus* and *Snyedra rumpens*), and zooplankton species, *Cladocera*. The measurements were used as interpolated values between the actual measured values. The actual frequency of the measurements is monthly. Zooplankton was measured only until 1989.

Using the data and introducing the expert knowledge in the procedure of model discovery, the following experiments for discovery of Chl-a model were designed and conducted (Atanasova *et al.*, 2006c):

- Discover chl-a model for each year separately. This experiment focused on the question whether it is possible to find a generic model structure for all years from 1986 to 1992 and just optimise the parameter values for each year or if specific model structures for each year were required. We tested each year-specific model on the remaining years in order to find out whether there is a generic model for all the years measured. Algal grazing by zooplankton was not included in this experiment as zooplankton data were only available for the years 1986 to 1989.
- Discover one chl-a model by learning on the entire data set, i.e. from 1986 to 1991, while 1992 was used for validation. This experiment focused on the question whether and how the length of the training data set influences the model. Is it better to learn models on one-year data or to use the entire data set, although they tend to contain a lot of noise? Algal grazing by zooplankton was not included in this experiment as zooplankton data were only

upoštevan.

- V tretjem eksperimentu smo vključili pašnjo zooplanktona na fitoplankton, pri odkrivanju modela za fitoplankton (Chl-a). Za učenje smo uporabili podatkovni niz od 1986 do 1988, medtem ko smo model validirali na podatkih iz leta 1989.

### 3.4 GREIFENSEE

Jezero Greifensee se nahaja v Švici. Osnovne značilnosti jezera so: maksimalna globina jezera znaša 32 m, povprečna globina je 18 m, površina jezera meri 8,5 km$^2$, prostornina pa 148 milijonov m$^3$. Prispevno območje jezera meri 163 km$^2$. V 60. letih prejšnjega stoletja je bilo jezero v evtrofnem stanju s povprečno koncentracijo fosfatov prek 500 mg/m$^3$. Kasneje, v 70. letih prejšnjega stoletja, po sprejetju določenih ukrepov za izboljšanje kakovosti vode, si je jezero nekoliko opomoglo, a je še vedno v evtrofnem stanju. Podatkovni niz jezera vsebuje časovno serijo v obdobju 1988 do 1999. Merjeni so vhodni podatki v jezero ter kemijski, fizikalni in biološki parametri v samem jezeru.

*Vhodni podatki* vključujejo dnevne meritve dveh pritokov, Aabach-Mönchaltdorf in Aabach-Niederuster. Merjeno je sledeče: pretoki, temperatura, pH, kisik, anorganske oblike dušika in fosforja ter skupni dušik in fosfor.

Meteorološki podatki vključujejo dnevne meritve globalnega sončnega obsevanja (vir: Swiss Meteorological Institute – MeteoSchweiz). *Kemijski in fizikalni parametri* kvalitete vode jezera so merjeni enkrat na mesec, medtem ko so *biološki parametri* merjeni z mesečno do tedensko pogostostjo (vir: Swiss Federal Institute of Aquatic Science and Technology – Eawag).

Cilj eksperimentov je bil gradnja oziroma odkrivanje enostavnega ekološkega modela za napoved relevantnih spremenljivk stanja, ki opisujejo trofično stanje jezera, to so koncentracije raztopljenega fosforja in klorofila-a (Atanasova *et al.*, 2005).

available for the years 1986 to 1989.

- In this experiment we included algal grazing by zooplankton, while discovering a model for phytoplankton (Chl-a). For learning we used the data from the period from 1986 to 1988. 1989 was used for validation.

### 3.4 GREIFENSEE

Greifensee is located in Switzerland with a watershed area of 163 km$^2$, and maximal and average depths of 32 m and 18 m, respectively. The surface area of the lake measures 8.5 km$^2$, while the volume is 148 million m$^3$. In the 1960s the lake was highly eutrophic with average phosphate concentrations of over 500 mg/m$^3$. The lake begun to recover around the 1970s after some measures were taken to improve the water quality (but it is still eutrophic). The lake's data set comprises a time series data for the period from 1988 to 1999 of the input to the lake, and chemical physical, and biological data in the lake.

The *input data* include daily measurements of two river inflows, i.e. Aabach-Mönchaltdorf and Aabach-Niederuster. The measurements include the flow rates, temperature, pH, Oxygen, inorganic forms of nitrogen and phosphorus, and total nitrogen and phosphorous.

Meteorological data include global solar radiation obtained from the Swiss Meteorological Institute (MeteoSchweiz). *Chemical and Physical variables in the lake* include monthly measurements, whereas *biological data* in the lake are monthly to weekly measurements (obtained from the Swiss Federal Institute of Aquatic Science and Technology – Eawag).

The experiments were aimed at discovering a simple lake model for prediction of the relevant state variables that describe the trophic state of the lake, i.e. phosphorus and chlorophyll_a concentrations (Atanasova *et al.*, 2005).

# 4. EVALVACIJA EKSPERIMENTOV

Vrednotenje opisanih eksperimentov je povzeto v preglednici 2. Eksperimenti so prikazani z značilnostmi domene, razpoložljivimi podatkovnimi nizi, uporabnostjo in tipom modela, ki se je gradil za to domeno, in učinkovitostjo zgrajenih modelov. Vsi eksperimenti so bili izvedeni tako, da je bila omogočena validacija modelov, kar pomeni, da je bil en del vsakega podatkovnega niza uporabljen za učenje (tj. kalibracijo, optimizacijo) modelov, drugi pa za validacijo naučenih modelov. Modeli vsakega eksperimenta so bili ovrednoteni po dveh kriterijih za uspešnost. Prvih deset najuspešnejših modelov smo določili glede na napaki, vključeni v orodju Lagramge MSE in MDL (glej poglavje 2) (najboljše prileganje k meritvam). Najboljšega izmed teh deset modelov smo izbrali po simulaciji in ekspertovem vrednotenju simulacije. Opozoriti velja, da model z najmanjšo napako (MSE ali MDL) glede na meritve ni nujno najboljši model glede na ekspertovo oceno.

Poglejmo si najboljši modela jezera Glumsø (enačba (1)). Razvidno je, da je ta enačba popolnoma v skladu z ekspertnim znanjem o modeliranju dinamike fitoplanktona, kljub temu da je bila odkrita z orodjem za strojno učenje Lagramge.

Prikazani model fitoplanktona za jezero Glumsø vsebuje znane formulacije ekoloških procesov. Prvi člen enačbe predstavlja rast fitoplanktona, ki jo omejujejo hranila (Monodov model), temperatura (linearni vpliv) in svetloba (zvončasta krivulja). Sledeči členi opisujejo: respiracijo (kinetika 1. reda), sedimentacijo (reakcija 1. reda z vplivom temperature) ter pašnjo zooplanktona (zadnji člen v enačbi).

Simulacija modela je prikazana na sliki 3. Na levi strani je prikazana simulacija na učnih podatkih, tj. podatkih, na katerih je model skalibriran, na desni pa validacija na testnih podatkih (od aprila 1973 do aprila 1974).

# 4. EVALUATION OF THE EXPERIMENTS

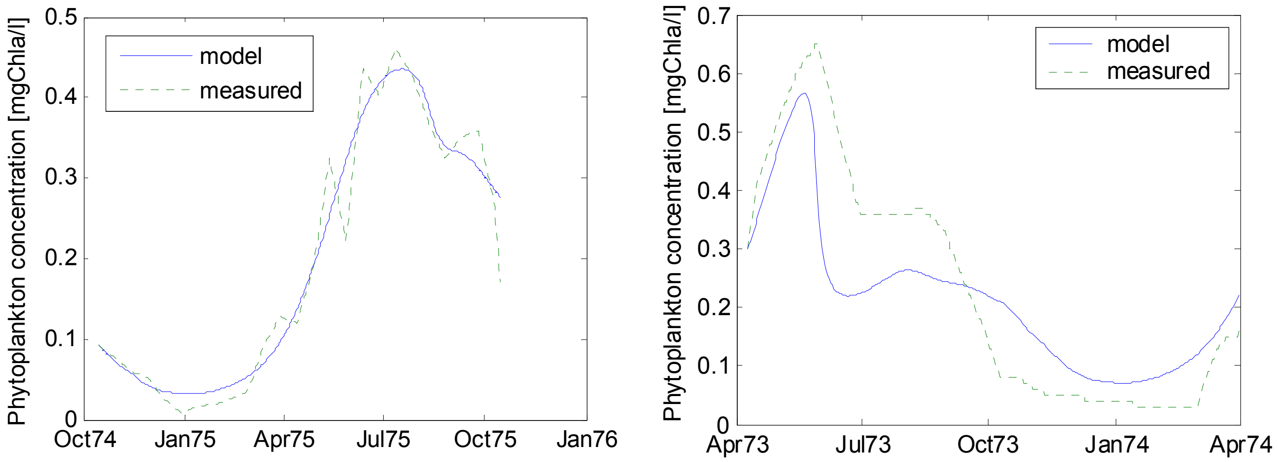
Evaluation of all described experiments is summarized in Table 2. The experiments are presented with domain characteristics, data sets available, applicability and type of the model that was built for that domain, and model performance. All experiments were constructed in a way to enable validation of the models, meaning that one portion of each data set was used for training (calibrating, optimizing) models, and another portion was left out for validation. The models from each experiment were evaluated by two error measures. To rank the first ten models (best fitted to the measurements) the error measure included in Lagramge, i.e. MSE and MDL (see section 2), was used. These best models are then simulated and the simulation is evaluated according to the visual perception of the expert. Note that models with lowest MSE are not necessarily the best according to the domain experts.

Let us here present the best model for Lake Glumsø (equation (1)). It is evident that this model is in-line with the expert knowledge about modelling of phytoplankton dynamics, in spite the fact that it was discovered with a ML tool (Lagramge).

The phytoplankton model is composed of known formulations for the ecological processes. The first term in the equation represents the phytoplankton growth, limited by the nutrients (Monod model), temperature (linear function), and light (photoinhibition curve). The following terms in the model stand for respiration (first order kinetics), sedimentation, which is influenced by temperature, and grazing by zooplankton.

Model performance is shown on the left hand side of Figure 3. The data set measured from April 1973 to April 1974 was used for model validation on testing data set. This is shown on the right hand side of Figure 3.

$$\frac{dphyto}{dt} = phyto \cdot 0.18 \cdot \frac{ps^2}{ps^2 + 0.0012} \cdot \frac{temp}{15.4} \cdot light \cdot e^{-\frac{(\frac{light}{116.6}+1)}{116.6}} - phyto \cdot 0.01 - phyto \cdot \frac{0.36}{2} \cdot \frac{temp-4}{18-4}$$
$$-zoo \cdot 0.14 \cdot 1.13^{(temp-19)} \cdot \frac{phyto^2}{phyto^2 + 0.44} \cdot phyto \cdot 0.007 \qquad (1)$$




Slika 3. Simulacija modela fitoplanktona za jezero Glumsø. Levo: simulacija na učnih podatkih oziroma podatkih, na katerih je model kalibriran. Desno: validacija modela na testnih podatkih.
*Figure 3. Phytoplankton model performance on Lake Glumsø data. Left: performance on the training set. Right: validation on unseen data.*

Vsak model smo ocenili z dvema ocenama, tj. z oceno natančnosti simulacije na učnem nizu in testnem nizu. Uporabljene ocene so: slabo (zelo slabo prileganje simulacije k meritvam), zadovoljivo (sprejemljivo prileganje k meritvam), dobro in zelo dobro. Na osnovi teh ocen pridemo do nekaj ugotovitev. Verjetno najbolj očitna je povezava med natančnostjo modelov in kvaliteto podatkov. Vzemimo za primer model jezera Glumsø, kjer podatkovni niz obsega dnevne meritve vseh parametrov. Model dosega veliko natančnost tako na učnem nizu podatkov kot na testnem (oziroma pri validaciji). Nadaljnje ugotovitve evalvacije aplikacij so podane v naslednjem poglavju.

There are two evaluation marks for each experiment model: performance on training data and performance on validation data. The marks used here are: poor (very bad or no fit to the measurements), fairly good (acceptable fit to the measurements), good (good fit to the measurements), and very good (nearly perfect fit to the measurements). Based on these marks several conclusions can be drawn. Probably the most obvious one is the connection between the goodness of the models and the quality of the domain data. In the case of Lake Glumsø a data set of frequent (daily) measurements was available, which lead to successful model induction and validation on unseen data. The evaluation results are further discussed in the following section.

ISSN 1581-0267

| JEZERO / REFERENCE EKSPERIMENTOV | KVALITETA PODATKOV | | EKSPERIMENT | CILJ EKSPERIMENTA | SPREMENLJIVKA STANJA | NEODVISNE SPREMENLJIVKE | UČNI NIZ PODATKOV | TESTNI NIZ; VALIDACIJA | EVALVACIJA EKSPERIMENTA | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ČAS MERITEV | POGOSTOST MERJENJ | | | | | | | UČINEK MODELA NA UČNEM NIZU | UČINEK MODELA NA TESTNEM NIZU |
| JEZERO GLUMSOE / Atanasova et al. 2007 | apr.73 do apr.74 in okt.74 do okt.75 | DNEVNO | EX 1 | odkriti in validirati model za dinamiko fitoplanktona | *Chl-a* | *temp, light, ps, no, zoo* | apr.73 do apr.74 | okt.74 do okt.75 | zelo dobro | zelo dobro |
| | | | EX 2 | odkriti in validirati model za dinamiko fitoplanktona | *Chl-a* | *temp, light, ps, no, zoo* | okt.74 do okt.75 | apr.73 do apr.74 | zelo dobro | zadovoljivo |
| JEZERO KASUMIGAURA / Atanasova et al., 2006c | 1986 - 1992; zooplankton: 1986-1989 | MESEČNO | EX 1 | pokazati dinamiko sistema, z odkrivanjem modelnih struktur ločeno za vsako leto merjenega obdobja; odkriti najbolj generično strukturo. | *Chl-a* | *temp, light, ps, no* | 1986, 1987, **1988**, 1989, 1990, 1991, 1992 | model iz 1988 je bil validiran na podatkih 1986, 1987, 1989, 1990, 1991, 1992 | zelo dobro | dobro |
| | | | EX 2 | odkriti in validirati model za dinamiko Chl-a za celotno obdobje meritev | *Chl-a* | *temp, light, ps, no, zoo* | 1986-1991 | 1992 | zadovoljivo | dobro |
| | | | EX 3 | izboljšati rezultate EX1 z upoštevanjem procesa pašnje zooplanktona | *Chl-a* | *temp, light, ps, no, zoo* | 1986-1988 | 1989 | dobro | dobro (toda kratko validacijsko obdobje) |
| BLEJSKO JEZERO / Atanasova et al., 2006b | 1995 - 2000 | MESEČNO | EX 1 | odkriti in validirati model fitoplanktona za celotno obdobje meritev | *phyto* | *temp, light, ps, no, zoo* | 1995 - 1999 | 2000 | zadovoljivo | slabo |
| | | | EX 2 | pokazati dinamiko sistema, z odkrivanjem modelnih struktur ločeno za vsako leto merjenega obdobja | *phyto* | *temp, light, ps, no, zoo* | 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002 | 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002 | zelo dobro | slabo |
| | | | EX 3 | odkriti model treh enačb: prehranjevalna mreža | *ps, phyto, zoo* | *temp, light, nutrients inflows* | 1996 | 1995 - 2000 | zelo dobro | slabo |
| GREIFENSEE / Atanasova et al., 2005 | 1988-1991 | TEDENSKO DO MESEČNO | EX1 | odkriti model treh enačb: prehranjevalna mreža | *ps, phyto, zoo* | *temp, light, nutrients inflows* | 1989 | 1988-1991 | dobro | dobro |

Preglednica 2. Evalvacija aplikacij Lagramgea. Oznake pomenijo sledeče: *temp* (temperatura vode), *light* (intenziteta svetlobe oz. radiacije), *ps* (anorganski raztopljeni fosfor), *no* (nitrat), *zoo* (koncentracija zooplanktona), *phyto* (koncentracija fitoplanktona), *chl-a* (koncentracija klorofila-a), *nutrients inflows* (dotoki hranil v jezero, tj. vrednosti pretokov in koncentracij).

| LAKE / REFERENCES TO EXPERIMENTS | DATA QUALITY | | EXPERIMENT | GOAL | STATE VARIABLE | INDEPENDANT VARIABLES | TRAINING DATA | VALIDATION DATA SET | EVALUATION OF THE EXPERIMENT | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MEASURED PERIOD | SAMPLING FREQUENCY | | | | | | | MODEL PERF. ON TRAINING DATA | MODEL PERF. ON VALIDATION DATA |
| LAKE GLUMSOE / Atanasova et al. 2007 | Apr.73 to Apr.74 and Oct.74 to Oct.75 | DAILY | EX 1 | discover and validate a phytoplankton model | *Chl-a* | *temp, light, ps, no, zoo* | Apr.73 to Apr.74 | Oct.74 to Oct.75 | very good | very good |
| | | | EX 2 | discover and validate a phytoplankton model | *Chl-a* | *temp, light, ps, no, zoo* | Oct.74 to Oct.75 | Apr.73 to Apr.74 | very good | fairly good |
| LAKE KASUMIGAURA / Atanasova et al., 2006c | 1986 - 1992; zooplankton: 1986-1989 | MONTHLY | EX 1 | evaluate systems dynamicity by discovering best model structure for each year of the data period: find most generic model | *Chl-a* | *temp, light, ps, no* | 1986, 1987, **1988**, 1989, 1990, 1991, 1992 | the model from 1988 was validated on 1986, 1987, 1989, 1990, 1991, 1992 | very good | good |
| | | | EX 2 | discover and validate a Chl-a model over the entire period | *Chl-a* | *temp, light, ps, no, zoo* | 1986-1991 | 1992 | fairly good | good |
| | | | EX 3 | improve the results of EX 1 by including grazing by zooplankton | *Chl-a* | *temp, light, ps, no, zoo* | 1986-1988 | 1989 | good | good (but short validation period) |
| LAKE BLED / Atanasova et al., 2006b | 1995 - 2000 | MONTHLY | EX 1 | discover and validate a phytoplankton model over the entire period | *phyto* | *temp, light, ps, no, zoo* | 1995 - 1999 | 2000 | fairy good | poor |
| | | | EX 2 | evaluate systems dynamicity by discovering best model structure for each year of the data period | *phyto* | *temp, light, ps, no, zoo* | 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002 | 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002 | very good | poor |
| | | | EX 3 | discover a food-web model of three equations | *ps, phyto, zoo* | *temp, light, nutrients inflows* | 1996 | 1995 - 2000 | very good | poor |
| GREIFENSEE / Atanasova et al., 2005 | 1988-1991 | WEEKLY TO MONTHLY | EX1 | discover a food-web model of three equations | *ps, phyto, zoo* | *temp, light, nutrients inflows* | 1989 | 1988-1991 | good | good |

*Table 2. Evaluation of Lagramge's applications. Marks: temp (water temperature), light (light intensity), ps (inorganic dissolved phosphorus), no (nitrate), zoo (zooplankton concentration), phyto (phytoplankton concentration), chl-a (chlorophill-a), nutrients inflows (inflows of nutrients into the lake, i.e. flow rates and concentrations).*

## 5. RAZPRAVA

V tem prispevku evalviramo metodo za avtomatizirano modeliranje jezer, ki predstavlja enotni okvir za pregledovanje ekoloških modelov kot tudi za njihovo avtomatsko odkrivanje (indukcijo) iz merjenih podatkov. To je omogočeno z zapisom obstoječega znanja o modeliranju v knjižnico generičnih spremenljivk, konstant in procesov. Po specifikaciji opazovanega ekosistema Lagramge z uporabo znanja iz knjižnice transformira to specifikacijo v možne modelne strukture, ki ta sistem opisujejo. V naslednjem koraku Lagramge optimizira te strukture tako, da jih umeri na podane meritve. Očitno je torej, da za razliko od ostalih orodij ML, kjer je kvaliteta modela večinoma odvisna od podatkov, ki so uporabljeni za gradnjo modela, v primeru Lagramgea na kvaliteto modelov vpliva več faktorjev: (1) domensko znanje, vsebovano v knjižnici, (2) kvaliteta in kvantiteta opazovanj sistema, (3) kompleksnost ekosistema in (4) specifikacija ekosistema v procesu indukcije modela.

Zato lahko rečemo, da je Lagramge ustrezen ne samo za indukcijo modelov iz podatkov, temveč tudi za odkrivanje/potrditev specifičnih znanstvenih dognanj v opazovanih domenah. Znova moramo opozoriti, da so modeli, odkriti z Lagramgeom, popolnoma v skladu s teoretičnim znanjem oziroma so matematično pravilni. Dosedanje aplikacije Lagramgea vključujejo indukcijo modelov v domeni modeliranja jezer. Nanašajo se na dve poglavitni nalogi: odkrivanje modela za dinamiko fitoplanktona in odkrivanje modela prehranjevalne mreže, hranilo–fitoplankton–zooplankton.

***Vpliv kvalitete podatkov.*** Kot velja za ostala orodja ML (in tudi ostala orodja za za modeliranje), je tudi uspeh Lagramgea močno odvisen od razpoložljivih podatkov (kvaliteta in kvantiteta). To se je pokazalo v praktično vseh aplikacijah Lagramgea. Jezero Glumsø, ki razpolaga z dnevno merjenimi podatki, je bilo identificirano z modelom, ki je najuspešneje prestal validacijo, tj. simulacijo na nevidenih podatkih. Res pa je tudi, da je podatkovni niz razmeroma kratek in je za potrditev ocene modela potreben dodatni niz meritev za validacijo. Od preostalih

## 5. DISCUSSION

In this paper we evaluate an approach to AM of lakes as a solid unifying framework for both handcrafting ecological models as well as their automated induction from measured data. This is enabled by encoding the existing modelling knowledge into a library of generic variables, constants, and processes. Given a specification of an observed system, Lagramge (by using the knowledge from the library) transforms the task specification into specific model structures for the observed system. The structures are later optimized (according to given measurements of the system variables). Unlike the other ML tools, where the model quality mostly depends on the data used for model building, in the case of Lagramge there are more issues that influence the resulting models: (1) the domain knowledge encoded in the library, (2) quality and quantity of the observed data (3) complexity of the ecosystem, and (4) expert knowledge introduced in the procedure of model induction, i.e. modelling task specification.

It is why Lagramge can be used not only for model induction from data, but rather for some specific scientific discoveries in the domains of interest. Note here that models discovered by Lagramge completely follow the background theoretical knowledge, i.e. they are mathematically correct. So far applications of automated modelling tool Lagramge include model discovery in the domain of lake modelling. More specifically, they include two major modelling tasks, (1) discovery of phytoplankton model and (2) discovery of food web models, i.e. nutrient–phytoplankton–zooplankton.

***Influence of data quality.*** As it is the case with any other modelling method (not just ML), the success of Lagramge is also dependant on the data quality and quantity. This is confirmed in all applications of Lagramge. Lake Glumsø, with a data set with daily measurements, was most successfully identified in terms of validation on unseen data. However the data set is rather short and further validation is necessary to confirm the model validity. Of the other cases discussed here, lake Greifensee possesses most complete and frequently measured data. The result is a obravnavanih domen poseduje jezero Greifensee najbolj popolne in najpogosteje merjene podatke. Tu je Lagramge induciral model, ki smo ga uspešno validirali na daljši podatkovni seriji, za obdobje štirih let. Poleg tega pa vsebuje omenjeni model tri enačbe, kar predstavlja mnogo težjo nalogo kot pri ostalih eksperimentih, pri katerih smo inducirali modele z eno enačbo. Validacija modela jezera Greifensee je pokazala dobro prileganje simuliranih podatkov k meritvam kot tudi dolgoročno stabilnost modela, ki je zelo pomembna za tako kompleksno domeno.

***Vpliv kompleksnosti ekosistema.*** Znano je, da se naravni ekosistemi prilagajajo spremembam v okolju oziroma spreminjajo svojo strukturo s časom (Jørgensen, 1999). Jezera imajo ponavljajoče letne vzorce obnašanja (npr. cvetenje alg spomladi) in je zato pričakovati, da bo določena struktura modela opisala letno obnašanje ekosistema. Vendar pa natančnejši vpogled pokaže, da se vzorci nekaterih jezer ne ponavljajo do te mere, da lahko sistem *zadovoljivo* opišemo z eno strukturo modela skozi daljše obdobje. Npr. cvetenje alg se lahko pojavlja vsako leto v približno istem obdobju, povzroča ga pa vsakič druga vrsta alg. To potrjujejo eksperimenti na Blejskem jezeru in jezeru Kasumigaura. V primeru Blejskega jezera Lagramge ni uspel najti dobrega modela fitoplanktona za simulacijo skozi daljše obdobje (pet let). Primerjava simuliranih in merjenih vrednosti koncentracije fitoplanktona je sicer nakazovala delno ujemanje (model je pokazal pravilne letne trende obnašanja), a je bil kljub temu precej nenatančen (Atanasova *et al.*, 2006b). Zato smo v naslednjem eksperimentu odkrili letne modele. Za vsako leto je Lagramge odkril model, ki se zelo dobro prilega meritvam. Vendar pa se modeli med seboj razlikujejo tako po strukturi kot po vrednosti parametrov. Prav tako je bila validacija modelov na nevidenih podatkih neuspešna. Strukturna dinamika Blejskega jezera se je pokazala tudi v tretjem eksperimentu, ko je Lagramge odkril model treh enačb na podatkih iz 1996. Kljub razmeroma natančni simulaciji na podatkih istega leta je bila validacija modela na preostanek podatkovnega niza neuspešna. Podobne rezultate lahko opazujemo pri jezeru Kasumigaura, kjer je bilo odkrivanje ‘letnih’

successfully identified and validated model on longer period data (four years). Moreover, Lagramge managed to successfully discover a model of three equations, which is a much more difficult modelling task compared to the rest of the experiments, where models of only one equation were discovered successfully. The validation shows good model behaviour in terms of degree of fit to the measurements as well as long term stability. The stability is especially important due to the complexity of the observed system.

***The influence of ecosystem’s complexity.*** It is well known that natural ecosystems tend to adapt to the changes in the environment, thus they change their structure in time (Jørgensen, 1999). Typically lakes have repeating yearly patterns (e.g. algal blooms in spring), and it is thus reasonable to expect that a single model structure can describe a year-to-year behaviour of the ecosystem. However, taking a closer look will show that the patterns in some ecosystems do not repeat to such extent, where we can *satisfactorily* apply a single model structure for describing long-term behaviour. For example, the algal blooms may happen at the same time as in previous years, but not necessarily by the same algae species. The experiments on lakes Bled and Kasumigaura confirm this. In the case of Lake Bled Lagramge failed to discover a long-term phytoplankton model of satisfactory accuracy. Comparing the simulations to the measured data, the model showed correct annual trends, but the accuracy (the amplitudes) was very low (Atanasova *et al.*, 2006b). In contrast, the models discovered by fitting structures on yearly data showed nearly perfect fit. However, comparing their structures, they all differ between each other (in structures as well as parameters). The structures could not be satisfactorily validated on unseen data. The dynamicity of the ecosystem was also shown in the third experiment, where a food-web model was discovered on the data from 1996. The model performed well on the training data (1996) but not on the rest of the data set. Similar results were obtained on lake Kasumigaura, where discovering yearly models was very successful. However, model validation was also more satisfactorily performed compared to Lake Bled.

modelov zelo uspešno. Za razliko od Blejskega jezera je bila tu validacija nekaterih modelov uspešnejša.

***Vpliv ekspertnega znanja***. Z vnosom ekspertnega znanja v specifikacijo opazovanega sistema določamo vplive oziroma procese, ki po našem mnenju vplivajo (ali bi lahko vplivali) na sistemske spremenljivke. Dodati ali 'pozabiti' na določen proces lahko bistveno vpliva na samo kakovost in strukturo modelov. V primeru jezera Kasumigaura smo odkrivali dva tipa modelov za skupni fitoplankton. Prvi je izključeval vpliv zooplanktona, drugi pa vseboval še ta vpliv, tj. proces pašnje zooplanktona na fitoplanktonu. V prvem primeru je odkriti model pokazal, da imajo hranila zanemarljiv vpliv na rast fitoplanktona. Ko smo v postopek odkrivanja modela vključili še proces pašnje, pa se je izkazalo, da imajo nekatera hranila večji vpliv. Torej se je struktura modela spremenila.

***Omejitve pri uporabi Lagramgea za indukcijo modelov:*** glavna omejitev metode je njena računska zahtevnost. Naj spomnimo, da pri vsakem poganjanju Lagramge izvaja nelinearno optimizacijo velikega števila modelnih struktur (odvisno od same specifikacije problema). To lahko vpliva na uporabnika tako, da daje prednost enostavnejšim specifikacijam sistema v izogib dolgotrajnemu iskanju kompleksnih modelov. Vsi obravnavani eksperimenti, razen jezera Greifensee in delno Blejskega jezera, so omejeni na odkrivanje enostavnega modela ene enačbe. Vendar se metoda razvija v smeri izkoriščanja dodatnih računskih moči, kot je na primer paralelno procesiranje. Tako lahko v bodoče realno pričakujemo, da bo ta ovira vsaj delno odpravljena.

Druga omejitev, ki jo je treba omeniti, je ta, da več ko vnašamo domenskega znanja (v smislu alternativnih formulacij generičnih procesov), večji je iskalni prostor, tj. število možnih struktur modela. To otežuje odkrivanje 'pravega' modela, saj so v tem primeru potrebni večji podatkovni nizi za identifikacijo modela izmed vseh možnih struktur. Torej je treba za identifikacijo najbolj relevantnih procesov v modelu sistema specifikacijo sistema prilagoditi realni situaciji, tj. našemu znanju in razpoložljivim meritvam.

***Influence of the expert knowledge.*** By introducing the expert knowledge in the modelling task we specify the influences and the processes that in expert's opinion take place (are the most important) in the observed system. Adding or 'forgetting' a specific process can greatly influence the structure and the success of the resulting models. In the case of lake Kasumigaura a phytoplankton model was identified first without the influence of the grazing process by zooplankton, and next by taking this influence into account. In the first case the models showed a negligible influence of nutrients on phytoplankton growth, whereas in the second, this influence was well captured in the equations. Thus, the structure of the ecosystems model has changed based on how we specify our background knowledge.

***Limitations in using Lagramge for model induction:*** The main constraint is the computational intensity of the method. Recall that Lagramge performs non-linear optimization of tens or hundreds (depending on the modelling task specification) model structures per one run. This can force the user to make more simple tasks and to avoid searching for complex model structures in the observed domain. This is also evident from the presented applications, where except for Greifensee and partly Lake Bled they were all limited to discovery of single phytoplankton equation. However, the method is developing towards exploitation of more computational power, i.e. parallel computing, and thus it is reasonable to expect that this obstacle will be at least partly solved.

Another limitation we should mention here is that the more background knowledge (in terms of alternative formulations of generic processes), the larger the search space, i.e., number of generated models, is. This, in turn, makes a discovery of the "right" model more difficult, in the sense that larger quantities of good quality data are needed to identify it from the remaining candidate models. Thus, one should be careful and design the task specification (and to an extent the knowledge library) so as to identify the most relevant processes and formulations for the task at hand.

# 6. ZAKLJUČEK

Metoda za AM, Lagramge, ki uporablja hibridni pristop k modeliranju, tj. učenje iz meritev in ekspertnega znanja, je bila uporabljena na realnih podatkih jezer. Aplikacije zajemajo gradnjo modelov za (1) identifikacijo sistema, (2) napoved evtrofikacije in (3) odkrivanje ali potrditev dinamičnih struktur ekosistemov. Za vsako domeno je bilo izvedenih nekaj eksperimentov. Vsi eksperimenti so bili evalvirani in s tem tudi sama uporabnost metode v ekološkem modeliranju. Kljub nekaterim omejitvam, kot je velika računska zahtevnost, se metoda lahko uspešno uporablja za identifikacijo kompleksnih domen. Uporabna je ne le za gradnjo modelov, temveč tudi za odkrivanje in potrditev drugih znanstvenih dognanj, kot je identifikacija dinamičnih vzorcev obnašanja opazovanih ekosistemov, tj. strukturne dinamike sistemov. Omejitve metode, omenjene v prejšnjem poglavju, narekujejo nadaljnji razvoj Lagramgea. Eden izmed glavnih ciljev nadaljnjega razvoja je omogočiti odkrivanje strukturno dinamičnih modelov in boljšo identifikacijo generičnih modelov. Razvoj grafičnega vmesnika je prav tako pomembno nadaljnje delo. Grafični vmesnik bi bistveno prispeval k popularizaciji oziroma širši uporabi metode med eksperti.

# 6. CONCLUSIONS

AM method (Lagramge), which uses hybrid approach to modelling, i.e. learning from data and expert knowledge, was applied to several real world lakes. The applications comprise models construction for (1) system identification, (2) predictions of eutrophication, and (3) discovery or confirmation of dynamic ecosystems structures. Thus, several experiments with Lagramge were conducted to each domain. All experiments were evaluated to investigate the applicability of the method in ecological modelling. Although faced with some constraints, such as the requirement for intensive computational power, the method can be successfully used in complex domains. It can be used successfully for model discovery as well as for other scientific discoveries, such as identifying dynamic patterns in the observed system, i.e. dynamic structure of the ecosystem. The limitations listed in the previous section direct the focus on the further development of Lagramge. One of the major goals is enabling the discovery of structurally dynamic models and easier identification of generic models. In order to spread this approach of automated modelling among experts further work is aimed to developing a graphical user interface for model construction.

# VIRI – REFERENCES

Atanasova, N., Mieleitner, J., Džeroski, S., Todorovski, L., Kompare, B. (2005). Development of a lake model using data and expert knowledge – case study : Greifensee. In: Markič, O. (ur.), Gams, M. (ed.), Kordeš, U. (ed.), Heričko, M. (ur.), Mladenić, D. (ur.), Grobelnik, M. (ur.), Rozman, I. (ur.), Rajkovič, V. (ur.), Urbančič, T. (ur.), Bernik, M. (ur.), Bohanec, M. (ur.). *Zbornik 8. mednarodne multikonference Informacijska družba IS 2005, 11. do 17. oktober 2005*, (Informacijska družba). Ljubljana: Institut "Jožef Stefan", 216–219.

Atanasova, N., Todorovski, L., Džeroski, S., Kompare, B. (2006a). Constructing a library of domain knowledge for automated modelling of aquatic ecosystems. *Ecological Modeling* **194**(1–3), 14–36.

Atanasova, N., Todorovski, L., Džeroski, S., Remec-Rekar, Š., Recknagel, F., Kompare, B. (2006b). Automated modelling of a food web in lake Bled using measured data and a library of domain knowledge. *Ecological Modeling* **194**(1–3), 37–48.

Atanasova, N., Todorovski, L., Džeroski, S., Recknagel, F., Kompare, B. (2006c). Computational Assemblage of Ordinary Differential Equations for Chlorophyll-*a* Using a Lake Process

Equation Library and Measured Data of Lake Kasumigaura. In: Recknagel, F. (Ed.): Ecological Informatics, 2nd ed., Springer-Verlag. Berlin, New York, 1–485.

Atanasova, N., Todorovski, L., Džeroski, S., Kompare, B. (2007). Application of automated model discovery from data and expert knowledge to a real-world domain: Lake Glumsø. *Ecol. Modeling* 2008, vol. 212, no. 1/2, str. 92-98.

Benz, J., Hoch, R. (1997). Ein Modelldokumentationssystem. ASIM Simulationstechnik, 11. Symposium in Dortmund, 232 p.

Benz, J., Knorrenschild, M. (1997). Call for a common model documentation etiquette. *Ecological Modeling* **97**(1–2), 141–143.

Benz, J., Voigt, K. (1996). Aufbau eines Systems zur strukturierten Suche von Informationsquellen für den Umweltschutz im Internet. Informatik für den Umweltschutz, 10. Symposium, 241 p.

Benz, J., Hoch, R., Legovic, T. (2001). ECOBAS – modelling and documentation. *Ecological Modelling* **138**(1–3), 3–15.

Chapra, S. C. (1997): *Surface Water-Quality Modeling*. McGraw-Hill, New York, N.Y., 1997.

DeAngelis, D. L. (1992). *Dynamics of Nutrient Cycling and Food Webs*. Chapmann & Hall, 2-6 Boundary Row, London SE1 8HN, UK.

Džeroski, S., Todorovski, L. (2003). Learning population dynamics models from data and domain knowledge. *Ecological Modelling* **170**, 2–3, 129–140.

Hoch, R., Gabele, T., Benz, J. (1998). Towards a standard for documentation of mathematical models in ecology. *Ecological Modelling* **113**, 1–3, 3–12.

Jørgensen, S. E., Bendoricchio, G. (2001). *Fundamentals of Ecological Modelling*. Elsevier Science B.V. Ltd, The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK.

Jørgensen, S. E. (1999). State-of-the-art of ecological modelling with emphasis on development of structural dynamic models. *Ecological Modelling* **120**(2–3), 75–96.

Kompare, B. (1995). The Use of Artificial Intelligence in Ecological Modelling: *Ljubljana, FGG; Royal Danish School of Pharmacy*, FGG, Ljubljana; Royal Danish School of Pharmacy, Copenhagen, Ljubljana, Copenhagen.

Kompare, B., Todorovski, L., Džerovski, S. (2001). Modelling and prediction of phytoplankton growth with equation discovery: case study – Lake Glumsø, Denmark. *Verh. Internat. Verein. Limnol.* **27**, 3626–3631.

Oreskes, N., Shrader-Frechette, K., Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science* **263**, 641–646.

Rismal, M. (1980). Presoja posameznih metod za sanacijo Blejskega jezera (Assessment of different methods for Lake Bled remediation). *Gradbeni vestnik* **29**(2–3), 34–46 (in Slovenian).

Remec-Rekar, S. (1995). Življenjska strategija in absorbcija fosforja pri nekaterih fitoplanktonskih vrstah Blejskega jezera-123 (Life strategy and phosphorus absorption in some phytoplankton species in Lake Bled), University of Ljubljana, Ljubljana (in Slovenian).

Rykiel, E. J., Jr. (1996). Testing ecological models: the meaning of validation. *Ecological Modeling* **90**, 229–244.

Sketelj, J., Rejic, M. (1958). Preliminary account on the examination of Lake Bled. *Gradbeni vestnik* **61**–**64**.

Todorovski, L. (2003). Using Domain Knowledge for Automated Modeling of Dynamic Systems with Equation Discovery: *Fakulteta za računalništvo in informatiko*, University of Ljubljana, Ljubljana, Slovenia.

Todorovski, L., Džeroski, S., Kompare, B. (1998). Modelling and prediction of phytoplankton growth with equation discovery. *Ecological Modelling* **113(**1–3), 71–81.

## Naslov avtorice – Author's Address

*asist. dr.* Nataša Atanasova
Univerza v Ljubljani – University of Ljubljana
Fakulteta za gradbeništvo in geodezijo – Faculty of Civil and Geodetic Engineering
Jamova 2, SI-1000 Ljubljana, Slovenia
E-mail: natanaso@fgg.uni-lj.si