Hydroinformatics techniques in eco-environmental modelling and management

Qiuwen Chen, Yenory Morales-Chaves, Hong Li and Arthur E. Mynett

ABSTRACT

The high complexity of aquatic ecosystems and the multiple processes involved, make the development of ecohydraulics and eco-environmental models a challenging subject. Conventionally, computer-based models use a mathematical formulation for the processes involved which are then solved by numerical methods. These models are often derived based on the assumption of spatial homogeneity and conservation principles of mass, momentum and energy. Development of these models often demands a clear understanding of the processes involved. However, the above assumptions are easily violated when spatial heterogeneity, individual species behaviour and local interactions play a significant role in the system dynamics. In particular for eco-environmental systems, knowledge on local interactions that determine the overall system behaviour is not always available. Although the rapid advances of data-driven techniques have recently made great contributions to water-environment related research, data on ecosystems are often quite limited, which restricts the application of data mining methods to eco-environmental system modelling. In addition, no modelling – also not black-box modelling – can be undertaken without having at least some understanding of the basic processes and mechanisms involved. It is always advisable to start exploring any dataset using conventional statistical techniques, as elaborated in this paper for a case study on Western Xiamen Bay, China. Neural network trimming was then used to establish the dominant factors; it was shown that a relatively simple ANN model was quite capable of capturing the essential features, provided the right input parameters are chosen. Examples of integrated approaches to ecohydraulics modelling coupling formulations with cellular automata and physical equations with fuzzy rules are presented for applications on eutrophication modelling of Taihu Lake in China, competitive growths and colonization of two underwater macrophytes in Lake Veluwe in The Netherlands, and forecasting of algal blooms in the Dutch coastal waters on the North Sea. A mussel dynamics model developed for the Upper Mississippi River in the USA demonstrates the feasibility of individual based modelling in ecosystem dynamics. Numerical models are quite capable of simulating the abiotic aquatic environment, including complicated fluid flow and transport mechanisms. However, when it comes to simulating the biotic and ecosystem dynamics, the interaction of individual species with their environment, as well as the interactions amongst species, has to be taken into account. The future of ecohydraulics and eco-environmental modelling thus seems to lie in the integration of different paradigms and techniques, which is the core content of the hydroinformatics discipline.

Key words | Data-model integration, eco-environmental modelling, ecohydraulics, environmental hydroinformatics, environmental systems modelling, harmful algal bloom prediction, individual-based modelling, mussel dynamics modelling

Qiuwen Chen
State Key Laboratory of systems ecology,
Research Centre for Eco-Environmental Sciences,
Chinese Academy of Science,
Beijing, 100085,
China

Yenory Morales-Chaves
School of Geography, Earth and Environmental
 Sciences,
University of Birmingham,
Edgbaston, Birmingham, B15 2TT,
UK

Hong Li
Arthur E. Mynett
Unesco-IHE Institute for Water Education,
P.O.Box 3015, 2601 DA, Delft,
Netherlands

Delft University of Technology,
Faculty CiTG, P.O. Box 5048, 2600 GA, Delft,
Netherlands

WL Delft Hydraulics,
P.O. Box 177, 2600 MH,
Netherlands

INTRODUCTION

Given the complexity of aquatic ecosystems, computer-based tools can become very useful for organizing knowledge and information about a system of interest, and support the analysis and discussion among experts in different disciplines. Since hydroinformatics is concerned with the application of information and communication technologies for the planning, management and conservation of the aquatic environment, the use of hydroinformatics technologies in areas of ecohydraulics and eco-environmental applications arise quite naturally (Mynett 2002).

Due to lacking mathematical formulations and limited available ecological survey data, most of the conventional ecohydraulics models are aggregated and conceptually based. For example, in the Lotka-Volterra model, populations (May 1975) are expressed in terms of biomass instead of the number of individual species. In the Michaelis-Menten growth model, the Monod curve is taken as the basic concept (Jorgensen 1994). These models are usually developed from Newton's 2nd law of motion and the 1st and 2nd laws of thermodynamic, employing partial differential equations (PDFs) to describe the physical processes (Abbott & Minns 1998). The variables in the models are discretised approximations of continuous functions in time and space, and the formulations strictly follow the conservation principles of mass, momentum and energy. Such modelling paradigms and techniques have been playing important roles in the progress of ecological research and are still fundamental tools (Jorgensen 1994).

However, these aggregated models mostly fail to take into account the effects of individual species differences, spatial heterogeneity or local interactions. These properties sometimes are crucial to ecosystem dynamics (DeAngelis & Gross 1992; Chen 2004). With respect to the underlying processes, it has been widely recognised that knowledge on the mechanisms of ecosystem dynamics are usually limited due to the high complexity of the system and restrictions on availability of laboratory experiments (Recknagel *et al.* 1994; Recknagel 1997; Lee *et al.* 2003). In addition, understanding is often qualitative rather than quantitative and difficult to be formulated in terms of PDFs (Chen 2004).

Following the rapid progress of high performance computing and advanced survey technologies, other modelling paradigms and techniques are being developed and applied to ecohydraulics modelling. These paradigms include cellular automata, individual based and box based schemes, fuzzy rule-based systems, evolutionary algorithms, and machine learning methods. Compared to conventional modelling paradigms, they are usually discrete in time, space and model variables. They take each spatial unit or individual species as the object to investigate the evolutions in time and the motions in space in order to obtain the global spatio-temporal patterns of the system. These techniques usually employ empirical knowledge as a reference, and discover embedded rules from the collected data which are then used to supplement the insufficiency of available understanding (Weiss & Indurkhya 1998; Witten & Frank 2000).

Hydroinformatics technologies can bring innovative approaches to ecohydraulics by integrating continuous formulations with discrete paradigms, and physical equations with data mining techniques. Integrated ecohydraulics modelling is seen to become a promising subject of environmental Hydroinformatics. Chen *et al.* (2002) used CA to simulate the competitive growths and colonization of two underwater macrophytes and explained the resulted ecosystem succession in Lake Veluwe, the Netherlands. Morales-Chaves (2004) investigated the growth and spreading of Zebra mussels in the Mississippi River by individual based modelling. Salski and Sperlbaum (1991) and Chen and Mynett (2003) successfully applied fuzzy logic to model eutrophication, and Recknagel (1997) and Li *et al.* (2005) used artificial neural networks (ANN) to forecast algal blooms. Baptist (2005) derived the vegetation-induced roughness formula from experiment data by genetic programming (GP). Meanwhile, applications of chaos theory in ecohydraulics modelling are under rapid development as well (Jorgensen 1994).

This paper demonstrates a range of hydroinformatics techniques for a number of aquatic ecosystem cases: harmful algal bloom prediction for Xiamen Bay, China, by combining data reduction techniques with artificial neural networks; eutrophication modelling of Taihu Lake, China by an integrated numerical-fuzzy method; algal bloom forecasting in the Dutch coastal waters by fuzzy cellular

automata; mussel bed evolution for the Upper Mississippi River by individual based modelling.

It can be observed that environmental hydroinformatics techniques can prove quite valuable for the design of river restoration measures and the development of adaptive river basin management procedures; combining mathematical models with in-situ or remote sensing data acquisition systems with graphical display techniques for effective communication of decision can be considered indispensable tools for successful operation and decision making in water resources management.

DOMINANT FEATURE EXTRACTION FROM DATA BY ANN TRIMMING PROCEDURE

Harmful Algal Blooms (HABs) are quite complex phenomena involving many physical, chemical and biological processes. Due to its highly nonlinear behaviour it is difficult to select the vital factors needed for HAB prediction. The causes of algal proliferation could be numerous and may include meteorological, hydrological, hydrodynamic, biological, and ecological factors. Only if there is a clear understanding of how these processes interact to cause HAB occurrence, can reliable models be developed that can identify systems and areas potentially susceptible to bloom events. However, at present the data and information, as well as the understanding of the mechanisms involved are still limited. Hence it is difficult to speculate on bloom occurrence and even more difficult to predict one.

Western Xiamen Bay is a semi-enclosed eutrophic bay (Qi *et al.* 2003) with a total area of 53 km^2, located in the South Eastern coast of China. Due to pollution, many HABs have occurred in the past (Qi *et al.* 2003) and mainly from April to June. The main reasons for the occurrence of HABs in Western Xiamen Bay are its semi-enclosed geography and eutrophication. Since the 1980's, the rapid development of industry and aquaculture, as well as the increase of population, which accelerate the eutrophication of this Bay, has increased the frequency of HABs. Most of the pollution loads are from the Jiulongjiang River, and the second largest source of pollution is the industry waste load. The annual average water temperature is 22 °C. Nitrogen and phosphorus are the main nutrients in this area in which

phosphorus was recorded as a limiting factor (Qi *et al.* 2003). Because of the subtropical climate, there are abundant algal species, which combines different types of algae with different favourite water temperature. Zhang (1993) reported that 110 algal species have been identified in Western Xiamen Bay, of which, 93 belonged to diatoms (84.5%), 13 dinoflagellates (11.8%), 2 blue-green algae and 2 Chrysophyta

Since there are so many factors to be considered, it is very important to first analyze whatever data and information is available and try to extract the dominant features in order to set up any predictive model. Conventional statistical methods can be used for multi-variable analysis and data set reduction, e.g. correlation analysis, principal component analysis (PCA), independent component analysis, etc. Recently, also data-driven models such as Artificial Neural Networks (ANNs), have been applied for HAB prediction (e.g. Lee *et al.* 2003; Maier & Dandy 1997). But in order to set up any Multi-Layer-Perceptron ANN, it is necessary to identify (i) what are the cause-effect relations and (ii) what are the parameters that dominate the process. Clearly, no modelling – not even black-box modelling – can be done without having at least some understanding of the basic processes and mechanisms involved. Precisely for this reason, it is always advisable to start exploring any dataset using conventional statistical techniques.

Statistical analysis

The input data were obtained from the HAB monitoring program carried out in Xiamen in 2003. Data from 4 stations (Figure 1) was collected, measuring physical, chemical and biological parameters. Since historical surveys (Qi *et al.* 2003) indicated that high Chlorophyll a (Chl-a) concentrations can indicate algal abundance as well as the occurrence of HABs, Chl-a was chosen as the indicator for predicting possible occurrence of HABs one week ahead. The ratio of total inorganic nitrogen and total inorganic phosphorous (TIN/TIP) was seen to vary from 7 to 513 while the average TIN/TIP ratio was about 33 (viz. much higher than the Redfield ratio 16), which indicates that phosphorus may be the limiting factor for HAB in this area in 2003.

A cross-correlation analysis was carried out to assess whether there is a similar tendency at the four different locations. In this case, stations 1, 2, and 3 were found to have a

Figure 1 | Monitoring stations.

high correlation (Table 1), but station 4 had quite a low correlation with the other three stations. Stronger hydrodynamics and different pollution sources could be the reasons for this. Therefore, only the data from stations 1, 2 and 3 were selected for model development and testing.

The correlation coefficients between all measured parameters and the Chl-a concentration one week ahead (Chl-a; $t + dt$) were in general low (Table 2). Rainfall (R) and Chl-a gave higher positive correlation with (Chl-a; $t + dt$), (dt is one week, *Tran* is transparency, *Sal* is salinity and *Irr* is irradiance). Dissolved Oxygen (DO), pH, water temperature (Tw) and Chemical Oxygen Demand (COD) gave high correlation with Chl-a at the time of measurement (Chl-a; t). The correlations between the nutrients and Chl-a are relatively high and negative as well, because of the nutrient uptake by algae.

A common statistical method for data set reduction is Principal Component Analysis (PCA) identifying patterns in the data. PCA has been widely applied in data reduction (e.g. Legendre & Legendre 1998; Park & Park 2000; Chen & Mynett 2003). Those variables, which are not contributing much to the variance of the components, can be eliminated from further consideration (Haan 1977). Those components

Table 1 | Correlation analysis between different stations (st1, st2, st3, st4)

Chl-a	St1	St2	St3	St4	TIN	St1	St2	St3	St4
St1	1				St1	1			
St2	0.85	1			St2	0.86	1		
St3	0.93	0.66	1		St3	0.86	0.89	1	
St4	0.07	− 0.10	0.27	1	St4	0.17	0.22	0.12	1
TIP					**Salinity**				
St1	1				St1	1			
St2	0.89	1			St2	0.91	1		
St3	0.90	0.91	1		St3	0.97	0.92	1	
St4	0.26	0.29	0.13	1	St4	0.64	0.66	0.64	1

Table 2 | Correlation analysis in the data of st1, st2, and st3

	Tw	WaC	Tran	pH	Sal	COD	DO	TIP	Irr	TIN	R	Chl-a	Chl-a(t + dt)
Chl-a	0.43	0.63	− 0.62	0.79	− 0.31	0.79	0.91	− 0.66	− 0.12	− 0.53	− 0.08	1	
Chl-a(t + dt)	0.20	0.27	− 0.08	0.12	− 0.26	0.25	0.16	− 0.15	− 0.43	0.19	0.44	0.29	1

with higher than average variance contribution are selected as the principal components (Weiss & Indurkhya 1998; Chen & Mynett 2003).

Based on PCA, three components were found to give higher than average contributions, representing almost 80% of the total loading from the whole data set. The results from the factor loadings to these three main components indicate that the significant factors may include pH, DO, Chl-a, R, TIN and Tw, which were selected as the input variables for HAB prediction.

Since phosphorus was recorded as the limiting factor in this area (Qi *et al.* 2003), it could also be one of the factors to be included in the model.

Data-driven modelling by ANNs

Data-driven modelling is nowadays quite popular because of its capability to detect trends from observed data and to provide fast predictions. Artificial Neural Networks (ANNs) are one of the well-established technologies in machine learning, and a mainstream technology for data-driven modelling (Solomatine 2004). It was inspired by neuro-science but did not attempt to be biologically realistic in detail. It combines simple processing elements (called neurons, units, or nodes), and the learning process in ANN is typically one of changing the strength of connections (weights) between the neurons (Figure 2).

ANN is one of the commonly used data-driven models in HAB prediction. Since it is able to map inputs to outputs even when the relationships between them are not completely clear (Murray 1993), ANNs have been used for input variable sensitivity analysis (Maier & Dandy 1997) and input variable selection by using network-trimming process (Lee *et al.* 2003). In this study, ANNs are used for extracting the dominant factors that determine HAB occurrence as well as
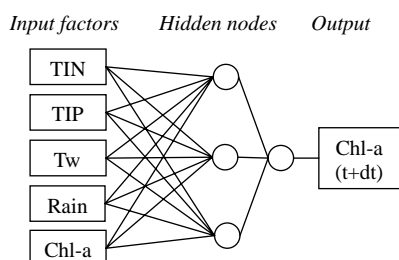

Figure 2 │ The example of ANN structure.

for HAB prediction in Western Xiamen Bay. The software used in this study is Weka, which is developed at the University of Waikato in New Zealand.

DOMINANT FEATURE EXTRACTION BY ANN TRIMMING PROCEDURE

Based on the results from correlation analysis and PCA, as well as expert knowledge, 8 scenarios with different input variables were set up (Table 3). The time step for prediction *dt* was chosen to be one week. The data from stations 2 and 3 were selected for training because they contain all extreme values of input and output variables. The data from station 1 were selected for testing. For scenario 1 (S1), the input variables are selected based on the correlation analysis and PCA results. The input variables for other scenarios are selected by removing one of the input variables at a time. This procedure represents a network trimming process, starting from the most complicated network (S1) and reducing towards the dominant features. The resulting errors of training and testing for the 8 scenarios are shown in Table 4 and Table 5. In addition, the testing results are shown in Figure 3.

The training results of the 8 scenarios were all quite successful in capturing the peak values of Chl-a. This proves that MLP-ANNs have a high learning ability for the given training data. However, in the testing results, scenarios 1 and

Table 3 │ Scenarios in using ANN for input variable sensitivity analysis

Scenarios	Input variables
S1	TIN, TIP, Tw, pH, R, Chl-a, DO
S2	TIN, TIP, Tw, R, Chl-a
S3	TIP, Tw, R, Chl-a
S4	TIN, Tw, R, Chl-a
S5	TIN, TIP, R, Chl-a
S6	TIN, TIP, Tw, Chl-a
S7	TIN, TIP, Tw, R
S8	Tw, R, Chl-a

Table 4 | Training errors in 8 scenarios applying MLP model using Weka

	S1	S2	S3	S4	S5	S6	S7	S8
Correlation coefficient	0.9837	0.9761	0.9525	0.9608	0.9234	0.873	0.849	0.9546
Mean absolute error	4.9431	8.472	9.9423	8.1605	9.6128	15.1644	9.4338	6.7367
Root mean squared error	6.579	10.386	12.4575	10.7034	11.3748	17.7621	14.8204	8.4752
Relative absolute error	28.94%	49.61%	58.22%	47.78%	56.29%	88.79%	55.24%	39.45%
Root relative squared error	23.71%	37.43%	44.90%	38.58%	41.00%	64.02%	53.42%	30.55%

6 gave very inaccurate results. Compared to scenario 2, the time delay of the occurrence of the peak value in scenario 1 is caused by pH and DO. From the correlation analysis, the highest correlation between DO, pH and Chl-a is when the time lag is 0. Algal photosynthesis uses CO2 and releases oxygen, and then leads to the increase of pH and DO values. Therefore, higher pH and DO may be the consequences of fast algal growth. Scenario 6 does not consider rainfall and is seen to give inaccurate results, which shows that rainfall has a vital influence on the ANN performance in this case.

Scenario 4, 5, 7, and 8 present relatively lower errors: this means that the input variables could be selected from any of these four scenarios. These scenarios have four or less input variables. Scenario 4 does not include TIP and nevertheless it gives good result, which indicates that TIP may not be a limiting factor in this case. S5 does not include water temperature, and S7 does not include Chl-a at time *t*; this may indicate that Tw and Chl-a may have low influence

in the network. It also shows the difficulties in input variable selection.

Scenario 8 with only 3 input variables is seen to have the best performance amongst the testing of all 8 scenarios. This means on the one hand that the results do not show any advantage of using more (than these three) environmental factors as network inputs. A simple network already leads to quite good result. Because of this, it means that rainfall, water temperature and chlorophyll-a are the dominant factors for one-week ahead prediction of the chlorophyll-a concentration.

For the Xiamen case, the ANN results showed the importance of rainfall in HAB prediction. This can be traced back to historical HABs (Chen *et al.* 1993). Rainfall is a very important nutrient supply carrier and indirect indicator of nutrient enrichment, especially in the period from April to June in Western Xiamen Bay. Also, rainfall may lead to salinity reduction, which may benefit the growth of some algae species.

Table 5 | Testing errors in 8 Scenarios applying MLP model using Weka

|	S1	S2	S3	S4	S5	S6	S7	S8
Correlation coefficient	0.5247	0.9367	0.8313	0.9413	0.9271	0.2189	0.8707	0.9328
Mean absolute error	11.023	11.0495	12.3508	8.4786	8.4742	20.5385	7.8753	7.032
Root mean squared error	23.25	13.1707	15.1474	10.1755	9.7253	31.688	11.2492	9.559
Relative absolute error	76.36%	76.54%	85.56%	58.73%	58.70%	142.28%	54.55%	48.71%
Root relative squared error	103%	58.33%	67.08%	45.06%	43.07%	140.33%	49.82%	42.33%
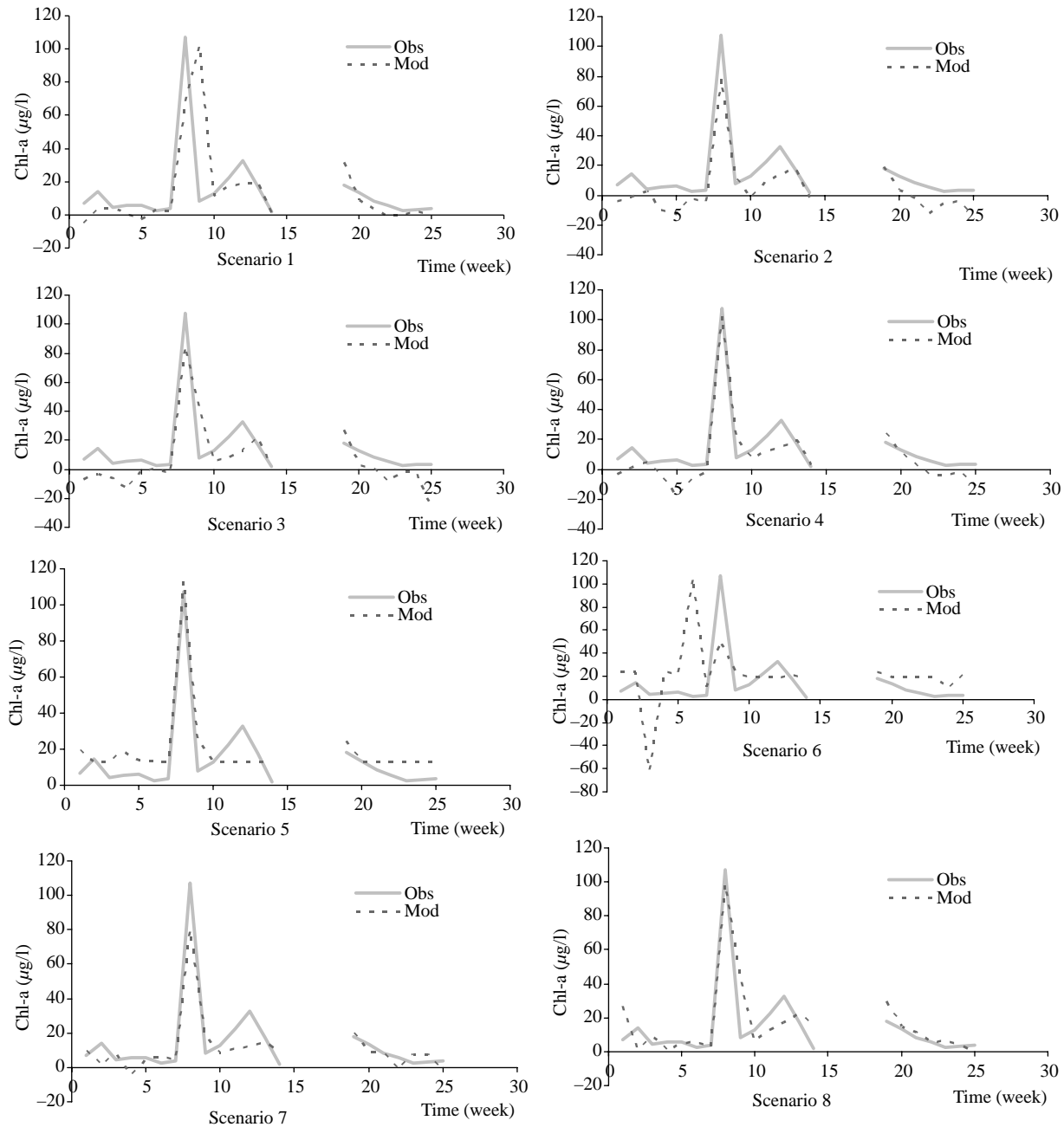
Figure 3 │ The testing results using ANN in 8 scenarios.

On a more general level, data pre-processing using statistical analysis is seen to be important not only for input variable analysis and input dimension reduction, but also for achieving a better understanding of the relationship between environmental factors and HABs in this research area. ANNs prove suitable tools for main factor selection for HABs and HAB prediction in this case.

The results of this study clearly demonstrate the power of applying hydroinformatics techniques to eco-environmental data-analysis and prediction; also, it shows

the importance of coupling hydrological processes, nutrient transport processes and algae growth processes in HAB prediction for this study area.

COMBINED NUMERICAL - FUZZY LOGIC MODELLING OF EUTROPHICATION IN TAIHU LAKE

Taihu Lake which is situated in the Yangtze Delta is the third biggest freshwater lake in China, and plays an important role in the regional development. In recent years, the lake exhibits serious problems of eutrophication, and blue-green algae blooms are observed frequently in most areas

To investigate the water quality changes, monitoring stations have been installed at 79 cross sections in the tributaries and 22 points in the lake. Besides, modelling studies focusing on hydrodynamic and water quality processes in Taihu Lake have been developed as well. Although previous models were able to provide spatial information on algae concentrations under certain wind fields by simulating the flow patterns, they could not model the temporal variations of the chemical and biological substances. Moreover, integration of ecological and hydro-dynamic processes is certainly far from established. In order to improve understanding of the detailed eutrophication features in Taihu Lake, a 3D water quality model was developed which takes into account release of sediment and unsteady meteorological forcing.

The hydrodynamics are conveniently described by the shallow water equations under σ-coordinates formulation, while the advection-diffusion processes are described by the well known transport equation. Nutrient variations, phyto-plankton growth, zooplankton predation and sediment disorption are all taken into account. A detailed description of the 3D model can be found in Mao (2005). To numerically solve the governing equations, all state variables are arranged on staggered "C" grids, which defines water level ζ and biochemical variables at the centre of any grid cell, and puts u at points to the east and west of the ζ points, and uses v at points to the north and south of the ζ points. An explicit finite-difference scheme using frontal-difference in temporal domain and central-difference in spatial domain is applied. Some of the modelling results are presented in Figure 1 for stations situated in Meiliang Bay where algae blooms often originate. It can be observed that the modelled temporal variation does not always agree well with the observations (Figure 4).

From the model studies, wind was found to be the main driving force for the currents within Taihu Lake, and Meiliang Bay turns out to be most susceptible to algal blooms. Algal blooms in Taihu Lake usually take place in summer with duration of about two months, except in Meiliang bay which is not so regular because of the complicated boundary conditions. The temporal behaviour of chlorophyll a concentration is similar to that of solar radiation and water temperature throughout the year, which indicates that the model is sensitive to solar irradiance.

Given the fact that biological processes of algal blooms are complicated and ambiguous, and that the results of the 3D numerical model and the observations have obvious difference, a fuzzy logic (FL) model was developed as a
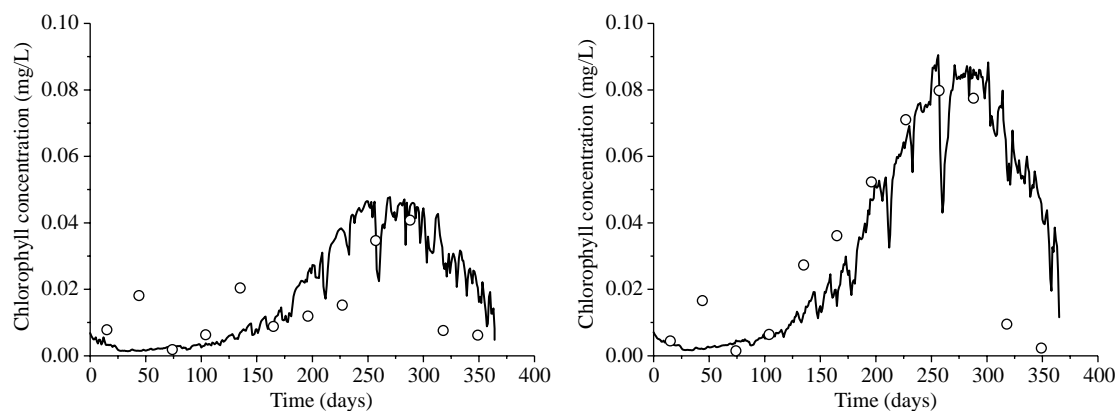

Figure 4 | Simulated (—) and measured (o) chlorophyll a concentration for Meiliang Bay.

Table 6 | Number of cases of incorrect fuzzy prediction

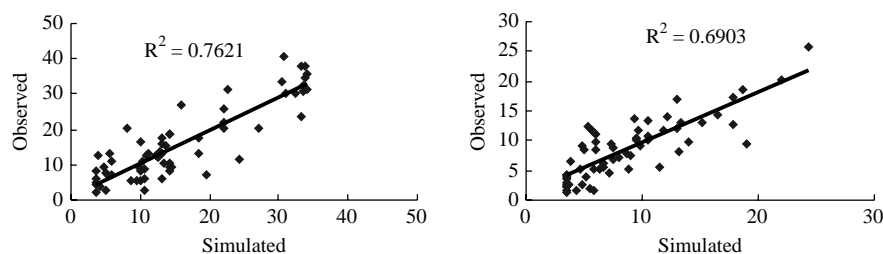| Station	L/M	L/H	M/L	M/H	H/L	H/M	Σ
S12	7	0	6	2	0	2	17
S17	4	0	10	0	0	0	14

L: low; M: middle; H: high; L/M: observation low / model middle. The same is to others.

complementary method to compute Chlorophyll a (Chl-a) concentrations. The model incorporates data mining techniques and heuristic knowledge. Principal component analysis (PCA) was used to identify the major abiotic factors and to reduce dimensionality. Self organising feature maps (SOFMs) and empirical knowledge were applied jointly to construct membership functions and induce inference rules (Chen & Mynett 2003). By combining rule-based learning (Chen 2004) and by capturing empirical knowledge in computer-based formulations, a total of 60 rules could be formulated.

The constructed FL model was tested at two sampling sites (noted S12 and S17) (Table 6) using monthly observed values for $Chla_{t-1}$ and TIN_{t-1}, TIP_{t-1} as inputs. The model outputs as summarised in Table 1 are seen to be in general qualitatively agreement with the field observations, especially for the cases of high concentration. The incorrect fuzzy predictions were mainly in the cases of low or middle concentration Figure 5.

In order to carry out a more quantitative evaluation, defuzzified outputs were calculated and plotted together with the observations. The R^2 is also computed and is given in the figures and can be considered quite reasonable for environmental studies.

Despite its drawbacks, the model was considered acceptable for qualitative prediction (without defuzzification).

The results are also promising even for quantitative prediction (with defuzzification), especially if the membership functions and inference rules are improved further by incorporating some optimisation and sensitivity analysis techniques.

DETERMINISTIC CELLULAR AUTOMATA FOR MODELLING MACROPHYTE DYNAMICS

There are many systems, in particular ecosystems where the global dynamics evolve from local interactions between species and spatial heterogeneity of environmental factors. In such cases, spatially-explicit biological models can be coupled with global environmental factors to simulate the ecosystem dynamics (Chen *et al.* 2002), as shown in Figure 6. A CA based model was developed to simulate the competition and succession of two macrophytes species in the Lake Veluwe, the Netherlands.

Description of study area

Lake Veluwe is an isolated part of the larger Lake IJssel in the centre of the Netherlands. The water surface is around 3300 ha, with an averaged depth of 1.4 m. It was formed by the construction of dams in the Southeast part of Lake IJssel in 1952 (Figure 7).

According to long-term documentation, the submerged vegetation of the lake has experienced considerable change after its formation, due to changes in nutrient loading. Before 1968, the water in the lake was clear, with diverse macrophytes vegetation. Due to discharge of wastewater from some small cities, the lake became eutrophicated and blue-green algae became dominant. Some restoration measures were taken in the late 1970s, which resulted in the increase of
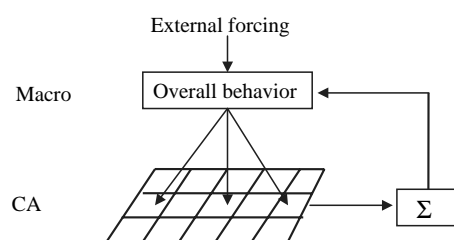

Figure 5 | Scatter plot of measured vs. simulated Chl-a concentrations at site S12 and S17.

Figure 6 │ Diagram of CA model coupling with other models.

Potamogeton pectinatus (*P. pectinatus*). The increase of *P. pectinatus* provided the precondition for the return of *Chara aspera* (*C. aspera*). After 1990, *C. aspera* colonised steadily and replaced the dominance of *P. pectinatus*.

From an ecological point of view, it seemed that *P. pectinatus* would outcompete *C. aspera* in the Lake Veluwe. However, *C. aspera* outcompeted *P. pectinatus* and replaced it gradually. Analysis of long-term observations indicated a self-reinforcing ability of *C. aspera* during eutrophication. *C. aspera* returned at a lower phosphorus level (0.1 mg/l) than the level at the time of its disappearance (0.3 mg/l), a phenomenon known as hysteresis; therefore it can be concluded that phosphorus is not a key factor in this case. It is supposed that the competition of

dissolved inorganic carbon HCO_3^{-1} and competition of light are the two main factors of the succession. However, the replacement process is still unclear, from which emerged the demand of model simulation. Considering local competition and colonisation between the two species, a Cellular Automata approach was selected to simulate the competition on light and HCO_3^{-1}, and try to explain the essential features of the replacement process.

Model development

In this CA model, deterministic evolution rules are developed for each of the two species, obtained from laboratory and field experiments. The model is designed to contain two partly interacting parallel submodels, one for *P. pectinatus* and the other for *C. aspera*. The processes considered in each submodel include shading, attenuation, HCO_3^{-1} competition, photosynthesis, respiration, morality and spreading. A conceptual framework of the model is presented in Figure 8, where solid lines refer to mass or energy flow, and dash lines indicate related processes. The local interactions between the two species are indicated by the two-directional dashed lines, for instance
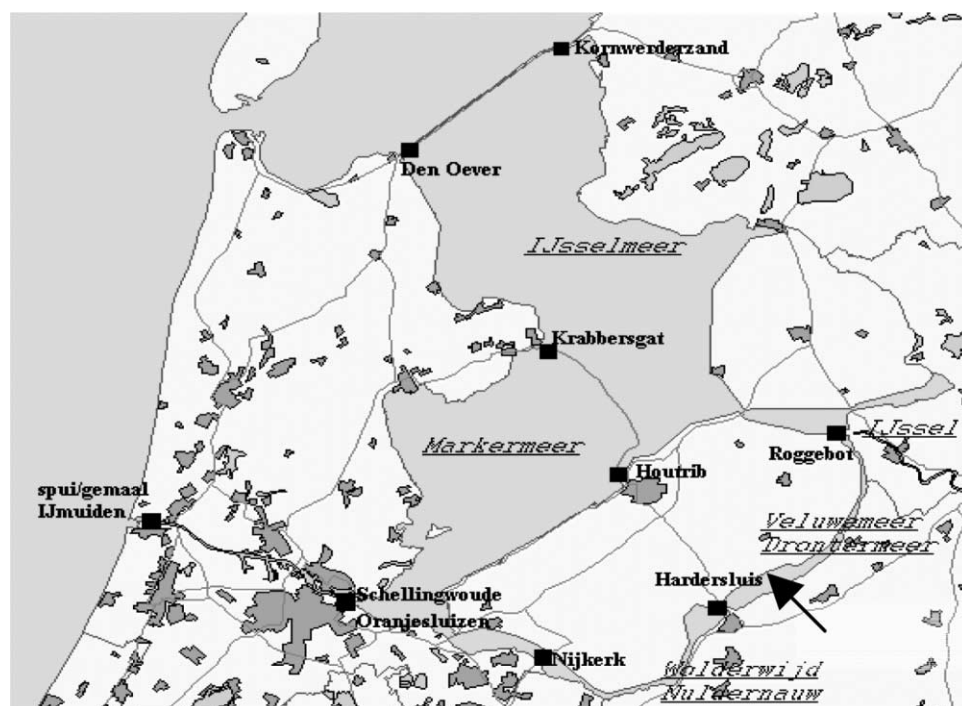

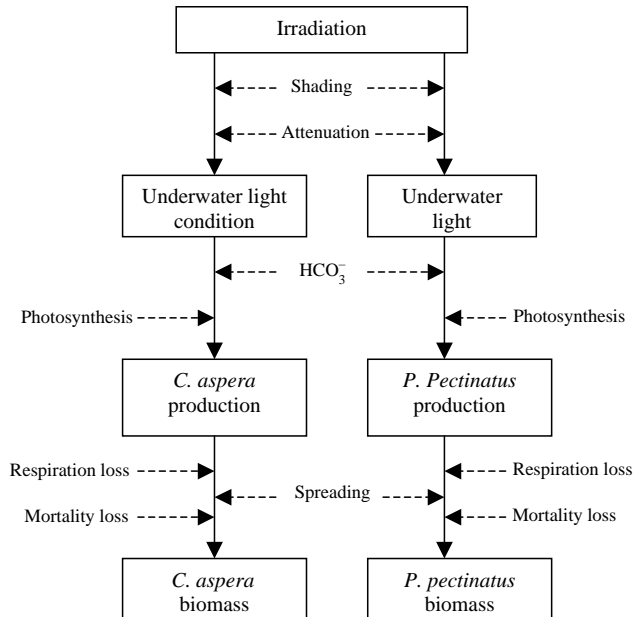**Figure 7** │ The study area Lake Veluwe.

Model development

Figure 8 │ Conceptual framework of the model.

"shading". General aspects of the model include: (1) germination of *P. pectinatus* and *C. aspera* from propagules; (2) initialisation with exponential growth rate; (3) growth and spreading; (4) production of propagules. Detailed model descriptions can be found in Chen *et al.* (2002).

Results and discussion

The results of the model are presented in two ways: by visualisation of the growing and spreading patterns of the two species (Figure 9), and by time series of biomass density averaged over sampled cells.

As shown in Figure 9, the colonisation is from the Northeast to the Southwest, and it is faster in longitudinal direction than in transverse direction. Besides, the colonisation of *C. aspera* is faster than that of *P. pectinatus*. Several simulation scenarios are carried out to test the governing factors. The result showed that the light intensity and HCO_3^{-1} are two major factors to the competitive growths of *C. aspera* and *P. pectinatus* in Lake Veluwe. Thus, shading and competition of HCO_3^{-1} become two important processes. The scarcity of HCO_3^{-1} has a great negative effect on the growth of *P. pectinatus*, while it has an indirect positive effect on the growth of *C. aspera*, which is advantageous to the replacement of *P. pectinatus* by *C. aspera*. These results are compatible with the field observations of Marcel (1999), who explored an individual based model to study the effects of changing environment conditions on the dynamics between *C. aspera* and *P. pectinatus*.

FUZZY RULE BASED CELLULAR AUTOMATA FOR MODELLING ALGAL BLOOMS

In the CA model of Lake Veluwe, deterministic rules were applied obtained from lab experiments and calibrated by field data. However, often such deterministic rules are not available because detailed mechanisms and their statistical properties remain unclear. In such cases, rule-based techniques can be used as an alternative. The following case study demonstrates the application of such Rule-based
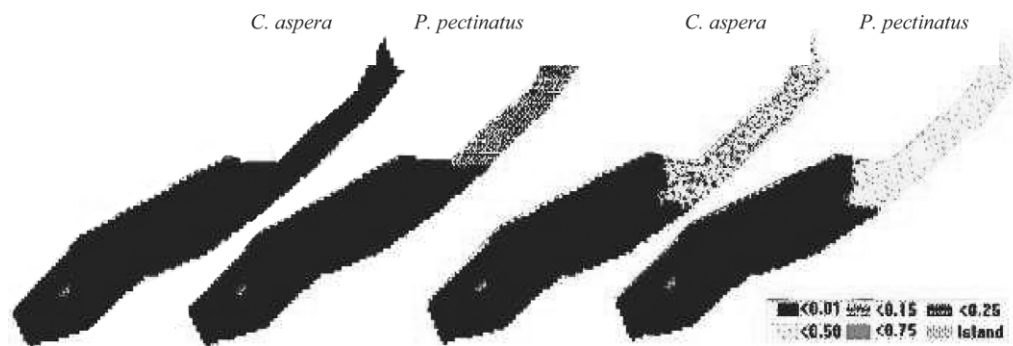

Figure 9 │ Germination of both species (*P. pectinatus* germinates earlier than *C. aspera*) at the left, and colonisation pattern by the end of the second year (*P. pectinatus* spreads slower than *C. aspera*) at the right.
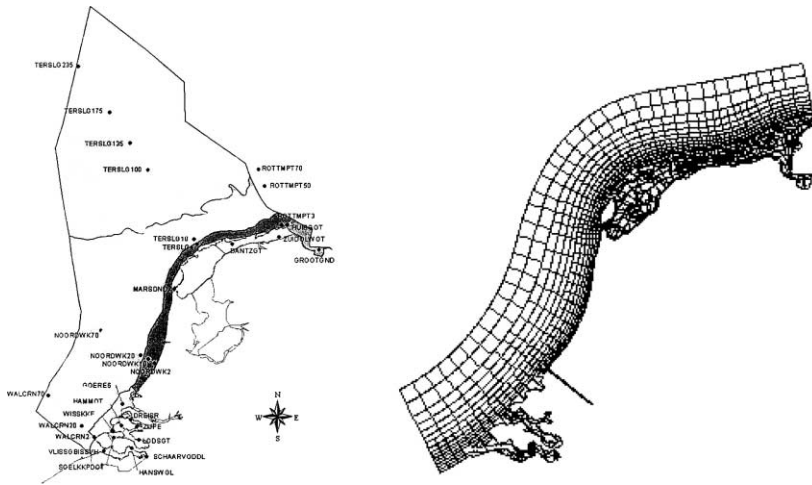
Figure 10 | Study area and monitoring stations in the Dutch coast and the computation grid.

Cellular Automaton (RbCA) to model algal blooming along the Dutch coast.

Description of the study area

The study focuses on the near shore area of the Dutch coast (Figure 10). The water depth is between 0 and 30 m, and water temperature varies from 5 to 22 °C, while the irradiance is between 132 ~ 1700 Whm-2day-1. The concentrations of inorganic nitrogen and phosphorus are between 0.007 ~ 1.246 mg/l and 0 ~ 0.073 mg/l respectively. The biomass concentration (in Chlorophyll a) is from 0.1 ~ 90.2 μg/l. The discharge from the River Rhine at the Maassluis station is between −2744 ~ 4649 m3/s, with a mean of 3182 m3/s. The water is usually well mixed except for temporary weak stratification caused by salinity. The RCA model is to forecast algal bloom (defined by chlorophyll a $\geq$ 30 μg/l) based on the monitored irradiance data and the nutrient concentrations data computed by Delft3D-WAQ of WL|Delft Hydraulics.

Model development

A curvilinear grid (Figure 10) was used in the model and the calculation of nitrate and phosphate concentrations was realized through the processes library configuration tool (PLCT) of the Delft3D-WAQ (Chen 2004). The boundary conditions are provided by the monitored data from the stations (Figure 10) in a way of block function, and the initial conditions were configured through linear interpolation of the monitored data.

The rule-based model developed by Chen & Mynett (2004) was introduced to predict algal biomass on the basis of the calculated nutrient concentrations from Delft3D-WAQ. The membership functions of nitrate and Chlorophyll a (Chl-a) concentrations are shown in Figure 11. The other variables include mean water column irradiance, water temperature and ortho-phosphate concentration. Chlorophyll_a concentration at the last time step (Chl-a$_{t-1}$) is also used as the model input.

The time step (Δt) for the hydrodynamic computations was set at 5 minutes, and the simulation completed a full
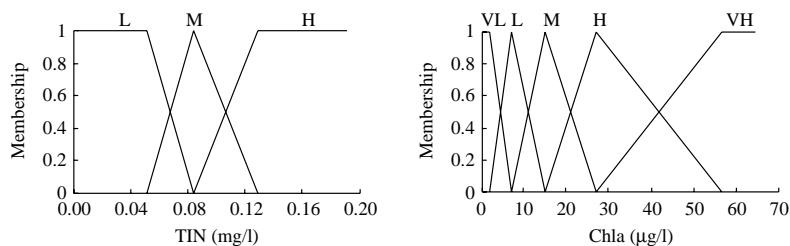

Figure 11 | Membership function of model variable and output (left: total inorganic nitrogen, right: chlorophyll-a).

tidal cycle, which was then repeatedly used for a year. The Δt was aggregated into 7 days in the rule-based model for algal biomass estimation.

The CA module was directly implemented on the curvilinear grid, which of course did not strictly follow the original definition of CA where square grids are used. However, this approximation can be acceptable as the geometry of the cells does not have much variation in the nearest neighbours. The Moore neighbourhood configuration (Wolfram 1984) was applied in the CA model and the local evolution rules were formulated in general as:

$$S_{i,j}^{t+1} = f\left({}^*S_{i,j}^{t+1}, \sum {}^*S_N^{t+1} \right) \tag{1}$$

where $S_{i,j}^{t+1}$ is the state of cell (i, j) at time step $t + 1$, ${}^*S_{i,j}^{t+1}$ is the state of the cell (i, j) at time step $t + 1$ which is preliminarily estimated without local interactions, $\sum {}^*S_N^{t+1}$ is the preliminarily estimated states of the eight neighbouring cells, and f are local evolution rules. In this study, the state S takes a value of the set $S_{\mathrm{Chl-a}} \in (\mathrm{L, M, H})$. Supposing ${}^*S_{i,j}^{t+1} = p$, the rules f are defined as:

$$S_{i,j}^{t+1} = \begin{cases} p & \text{if} > 3\,\text{neighbours}\, {}^*S_{neighbour}^{t+1} = p \\ 0.5(p+q) & \text{if} > 3\,\text{neighbours hold}\, {}^*S_{neighbour}^{t+1} = q \end{cases}$$

$$(p, q \in S_{\mathrm{Chl-a}}) \tag{2}$$

Model results

Some of the modelled results of Chlorophyll_α concentrations in the year 1995 are displayed in Figure 12 which presents the output at peak-bloom period. In spatial pattern, the algal blooms occur at the near shore area. The reason is that the residual flow of the river Rhine discharge is from the South to the North, following the coastline due to the effects of Coriolis force, so the nutrient concentrations are higher along the coast. It is also seen that the blooms are more severe near the Noordwijk transect and Wadden Sea area because of the discharge from the land. By examining the observations in 1995, the first peak bloom at station Noordwijk 10 (Figure 10) appeared on May 3rd with the Chlorophyll_α concentrations of 58.2 µg/l. The modelled bloom timing (28th, April) and intensity (48 µg/l) are quite close to these observations. It remains difficult to quantitatively evaluate the modelled spatial patterns, but this could become possible in future when using satellite images for comparison.

The development and application of Rule-based Cellular Automata is still at the initial stage, and the advantage of the method still requires proper evaluation. However, preliminary research outputs indicate that cellular automata are capable of capturing enhanced patchiness dynamics.
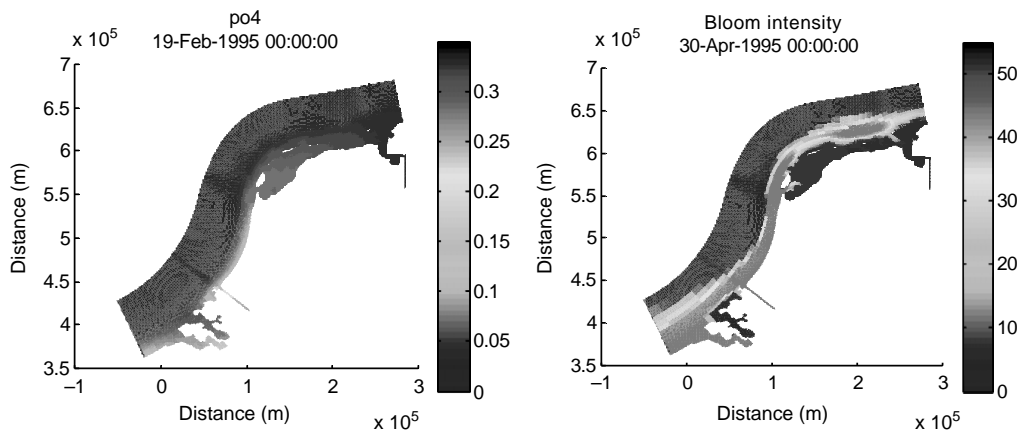

Figure 12 │ Model results of phosphate concentrations before algal bloom by Delft3D-WAQ and algal biomass concentrations by RCA.

INDIVIDUAL BASED MODELLING OF ECOSYSTEM DYNAMICS

Computer simulations using knowledge-based systems are rapidly gaining popularity in the field of ecosystem modelling. In particular when species interactions have to be taken into account, local conditions and local interactions play a dominant role.

Basically three model types can be distinguished for modelling population dynamics: traditional models based on differential equations (e.g. Lotka-Volterra models), spatially explicit models such as Cellular Automata (Mynett & Chen 2004), and individual-based models (DeAngelis & Gross 1992). Clearly, assumptions and simplifications inherent to each modelling paradigm determine its applicability for any given problem. In general, individual-based modelling (IBM) provides more flexibility to account for population and environmental heterogeneity both in space and time (Huston *et al.* 1988; Dunning *et al.* 1995). Within the modelling spectrum IBM is often seen as the most detailed model representation that can often be simplified to arrive at a cellular automaton or a differential equation model (Law & Dieckmann 1998; Cronhjort 2001). The main limitations of the IBM approach are related to input data requirements and computation time, though the last one becomes less significant as computer technology advances.

A recent example of the application of individual-based modelling is the Mussel Dynamics Model (MDM, Figure 13) developed for analyzing the dynamic interactions between freshwater mussels and their environment, and the interaction between native and invasive mussel species (Morales-Chaves 2004; Morales *et al.* 2006*a*). The model takes input data on river hydrodynamics, water quality, and distribution of available fish (required for survival of the parasitic life stage), and computes the population response to environmental conditions based on various functional processes. The life cycle of each species of mussels can be divided into as many stages as deemed necessary to give an appropriate representation of the population. Expert knowledge is integrated into a habitat suitability model that is used to estimate mussels' mortality. Larvae and young juveniles dropping from the host fish are suspended in the flow and experience hydrodynamic transport (dispersion). If habitat conditions deteriorate, juveniles and adults can move in search of more suitable habitats, though this commonly occurs at a very slow rate. As juveniles settle in the river bed, a basic bioenergetics model is used to compute mussels' growth and when adulthood is reached, reproduction is activated. The growth of individual mussels depends on food availability and mussels' ordering for feeding is the mechanism used.

The core of MDM simulation is an individual-based configuration model (DeAngelis & Gross 1992) that applies
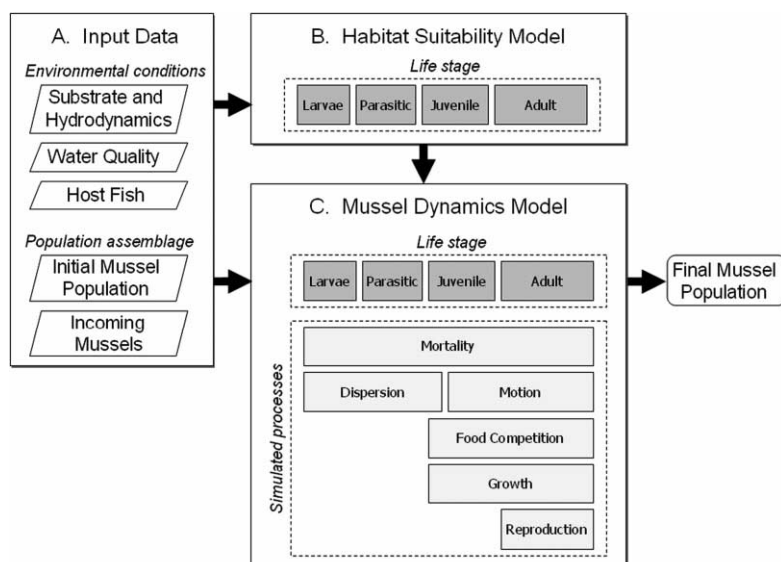

Figure 13 │ Diagram of the Mussel Dynamics Model (MDM).

a dynamic (time dependent) approach in a spatially distributed (two-dimensional) domain. The model has been designed to provide flexibility for simulating mussel dynamics in as much detail as desired: environmental conditions can be generated by sophisticated or simple models; any number of water quality parameters can be considered; the population can be separated into various species and life stages; for each life stage, the functional processes to be simulated can be selected. All these options make MDM a suitable tool for carrying out sensitivity analysis and hypothesis testing, as elaborated hereafter.

Investigating the effect of flow rate on mussels' distribution

Freshwater mussels spend most of their lives partially or completely buried in the sediments of rivers, therefore nearby bed currents, mean velocities, water depths, and substrate stability may affect their distribution (Holland-Bartels 1990; Layzer & Madison 1995; Strayer 1999a). Different species may have different tolerances to these parameters and the frequency and magnitude of extreme

flow events dictates which species can survive in a given habitat (Di Maio & Corkum 1995). In addition, flow conditions during larvae release and juvenile settlement may limit recruitment in established mussel beds (Hardison & Layzer 2001; Hastie et al. 2001).

In MDM the effect of flow conditions on mussel dynamics is simulated in two steps: (i) in the habitat suitability model hydrodynamic variables are used to determine the locations where mussels can survive at different life stages; (ii) in the dispersion routine a Lagrangian particle tracking mechanism is used to determine the position of larvae and young juveniles travelling with the flow. Simulations are made following the mean yearly hydrograph observed in the river reach so that flow extremes and seasonal flow variability can be taken into account.

To test MDM predictive capabilities, a 10-km reach of Navigation Pool 16 in the Upper Mississippi River (UMR) was used as a case study (Morales-Chaves 2004). An initial population of 700,000 native mussel larvae was homogeneously distributed in the 15 km^2 domain and followed throughout their life cycle. Suitable areas for mussel survival (Figure 14A) were identified based on water depth, mean
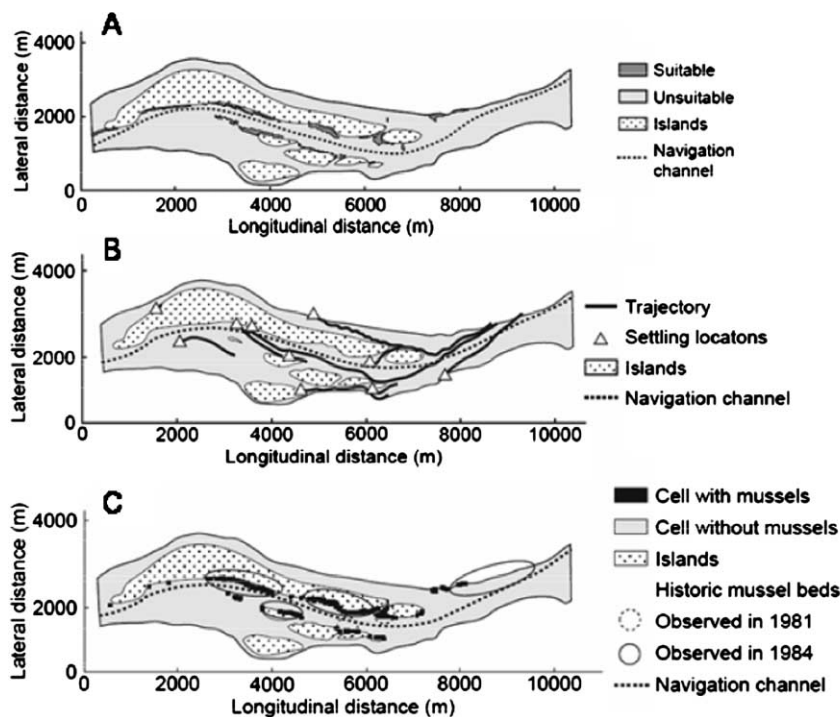

Figure 14 | Application of MDM to spatial distribution analysis of freshwater mussel communities (Morales *et al.* 2006a): (A) Distribution of suitable habitats. (B) Trajectories of 10 juvenile mussels traveling with the flow. (C) Simulated mussel accumulations and historic mussel beds.

velocity, substrate type, and substrate stability (Morales *et al.* 2006*b*). Dispersion of juveniles with the flow determined the patterns of colonization of suitable habitats (Figure 14B). Those mussels settling in suitable habitats survived and formed mussel accumulations that coincided with the location of historic mussel beds (Figure 14C).

The sensitivity of the results to changes in flow rates is discussed in Morales *et al.* (2006*a*, 2006*b*) and summarized in Table 7. As the flow rate increases, the number of suitable areas for mussels to inhabit decreases due to active sediment motion at medium to high flows. The mean flow condition greatly overestimate the real habitat suitability; for this reason the 5% flow exceedance probability was used which provided an appropriate estimate of the location of flow refuges and the spatial distribution of suitable habitats for mussel survival. The number of individuals settling inside the boundaries of the domain also decreases with increasing discharge. This indicates the significant effect that high flows during the spawning season could have on the recruitment of young individuals in the study area.

Food competition between native and invasive species

Native freshwater mussels in the UMR are members of the Family Unionidae and within this group food competition is thought to be negligible (Bauer *et al.* 1991), although this may depend on population density (Kat 1982). The situation has changed drastically with the arrival of the zebra mussels, an invasive species that smothers native mussels and competes with them for food supplies (McMahon & Bogan 2001). In comparison with native mussels, zebra mussels densities are generally orders of magnitude higher (Strayer 1999*b*), and filtration rates can be more than 10 times those of unionids (Strayer *et al.* 1999). Even at low

Table 7 | Sensitivity of the modeling results to flow rate

Flow rate (m³/s)	Exceedance probability	Relative extent of suitable areas	Relative number of settling juveniles
566	99%	1	1
2039	35% (mean flow)	0.14	0.71
3965	5%	0.05	0.56

infestation intensities, the loss of food resources due to zebra mussel invasion has been observed to cause a sharp decline in the native mussel population.

MDM was used to estimate the effect of food competition between zebra mussels and native mussels (Morales-Chaves 2004). Chlorophyll α records from the Long Term Resource Monitoring Program (USGS 2003) were used to estimate a base food concentration of 0.9 mg C/L. The amount of food available for mussels feeding was computed by multiplying the base food concentration by the volume of the cell and by a factor ff indicating the percentage of the water column to which mussels have access. Because zebra mussels often smother unionids, securing first access to inhalant waters (Strayer *et al.* 1999), in the simulations zebra mussels were given priority for feeding over unionids.

First, simulations were made for unionids alone for a long enough period to obtain a developed or stable population (60 y for ff = 100%, 142 y for ff = 10%, 115 y for ff = 1%). Then zebra mussels were introduced in the system applying a fixed upstream boundary condition of 200 larvae/L (based on measurements by USACE, personal communication) entering the domain during every day of the reproductive season (July to September). Because the aim of the simulations was to investigate coexistence of unionids and zebra mussels, only zebra mussels settling in native mussel beds (Figure 14C) were considered.

Zebra mussel population growth (Figure 15A) was determined by the concentration of incoming larvae, hydrodynamic conditions, and food availability (Morales-Chaves 2004). The base food concentration was enough to reach the maximum population size dictated by the fix immigration rate. But for ff = 10% and ff = 1% food shortage considerably reduced the population size that was attained. After 10 y the population size remained stable in the three scenarios and densities varied between 25 and 300 zebra mussels/m², which corresponds with ranges reported by Hart *et al.* (2001) for moderate to high zebra mussel infestation in an upstream reach of the Upper Mississippi River.

In competition with native mussels, zebra mussels took up most of the resources available causing a decline in the native mussel population (Figure15B). For ff = 100%, a slow recovery of the native mussels occurs after the zebra mussels population stabilizes. But for ff = 10% and ff = 1% the native mussels population size drops sharply and after
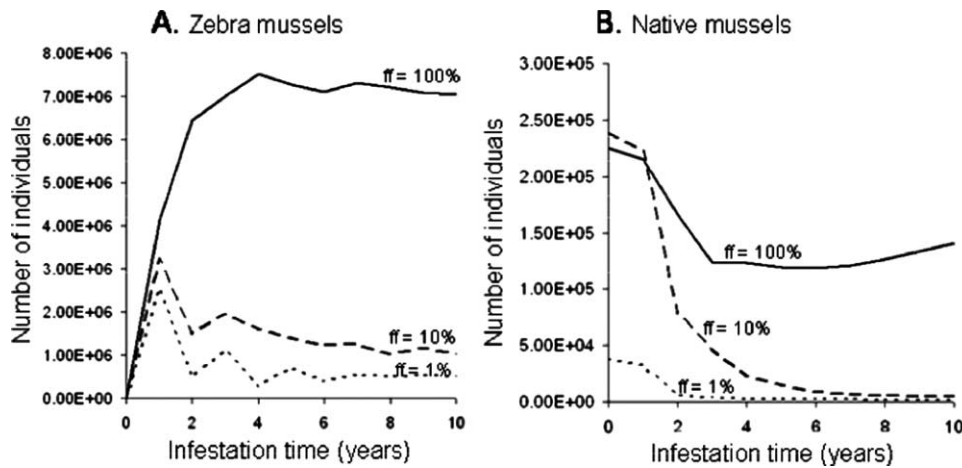
Figure 15 | Population growth as a function of percentage of food in the water column readily available for mussels feeding (ff). (A) Zebra mussels. (B) Native mussels infested by zebra mussels. The food concentration was 0.9 mg C algae/l and priority was given to zebra mussels over native mussels for feeding.

45 y simulation (~one unionids life cycle) it has decreased 99%. Annual survival rates of native mussels were between 0.75 and 0.90, similar to those reported by Hart *et al.* (2001) for moderate to high infestation densities. Even for the most optimistic scenario where mussels had access to all food in the water column (ff = 100%), the population size of the native mussels had to decrease to satisfy zebra mussels demand. When unionids were simulated alone, population densities varied between 2–98 mussels/m^2, but after 10 y simulation with zebra mussels, the maximum density of unionids had dropped to 17.8 mussels/m^2 and some former native mussel beds had disappeared (Morales-Chaves 2004). The situation worsened for lower food availability. The results suggest that coexistence of the two families of species might be possible, nevertheless it must be bear in mind that the sharp decline in native mussel densities at low food availabilities may produce extirpation due to unsuccessful egg fertilization (Downing *et al.* 1993).

In general, the relevance of food competition as a population structuring mechanism depends on the relative importance of mussels grazing compared to food resources transport and availability (Strayer *et al.* 1999). In the past, food availability allowed the development of rich native mussel communities in the UMR, some of which were extensively exploited with the button industry from 1850's to 1930's (Anthony & Downing 2001). Prior to the zebra mussel invasion, densities >100 mussels/m^2 had been observed in some areas (Whitney *et al.* 1997). The results of our analysis indicate that the situation is likely to change.

If the current pattern of zebra mussel colonization continues and the population becomes permanently established along the UMR a significant shift on resources allocation can be expected, as it has occurred in other freshwater systems in North America (Ricciardi *et al.* 1996; Strayer *et al.* 1999).

MDM as a decision support tool

MDM is a typical example of an application in the area of environmental hydroinformatics, where knowledge and information from various disciplines is integrated and the application of alternative modeling paradigms is explored. The individual-based model presented in this paper proved to be very effective for simulating mussel population dynamics. By integrating individual responses it was possible to observe a range of dynamic variations at the population level, information that was previously unavailable.

The overall agreement between simulation results and observations indicate that the MDM model successfully captured the essential mechanisms of mussel dynamics. The model can be a useful tool to estimate the potential effect of different stressors on long-term dynamics of freshwater mussel communities and consequently improve the current understanding of cause-effect relationships in such a complex system. The model can be used to evaluate the potential effect of different management practices, information that is considered to be extremely valuable for decision makers

since it can support the development of appropriate strategies for conservation of natural ecosystems.

CONCLUSIONS

Although discrete paradigm and rule-based techniques are becoming more widely applied in eco-environmental modelling, and already demonstrates certain advantages in some cases as shown by the examples in the paper, this does not mean that they can take over conventional methods that are often referred to as physically-based formulations using partial differential equations. The key issue nowadays is to select proper methods from a variety of available tools depending on the problems (Mynett 2002; Chen & Ouyang 2005) dealt with. In general, the selection of a paradigm or method depends on the research objectives, understanding of the problem of study, and availability of data.

If knowledge on the mechanisms involved is enough and data is limited, conceptual models can be a proper choice. If knowledge is limited but enough data is available, data driven model can be a right method. When only limited data and/or limited knowledge are available, rule-based methods taking experts experiences as the reference can provide a suitable alternative (Duel *et al.* 2002; Lee *et al.* 2002).

It is becoming increasingly recognised that eco-compatible adaptive water management strategies can best be developed and implemented by making use of advances in the development of hydroinformatics tools and technologies. Evolutionary algorithms and neuro-fuzzy computing prove extremely valuable for the design of river restoration measures and adaptive river basin management procedures, as well as for habitat suitability assessment. Mathematical models in combination with data acquisition systems and advanced graphical display techniques for effective communication are becoming indispensable prerequisites for successful water management. In the field of ecosystem modelling, the potential of computer-based simulation techniques in combination with expert knowledge from a wide range of disciplines ranging from the hydro-sciences to biological/ecological/environmental systems is gaining recognition. In fact, eco-compatible adaptive water management strategies could greatly benefit by making use of

advances in hydroinformatics tools for ecohydraulics modelling (Mynett 2004).

REFERENCES

Abbott, M. B. & Minns, A. W. 1998 *Computational Hydraulics*, 2nd Edition. Ashgate, Aldershot, UK & Brookfield, USA.

Anthony, J. L. & Downing, J. A. 2001 Exploitation trajectory of a declining fauna: a century of freshwater mussel fisheries in North America. *Can. J. Fish. Aquat. Sci.* **vol. 58**, 2071–2090.

Baptist, M. J. *Modelling floodplain biogeomorphology*. Ph.D thesis, Delft University of Technology.

Bauer, G., Hochwald, S. & Silkenat, W. 1991 Spatial distribution of freshwater mussels: the role of host fish and metabolic rate. *Freshwat. Biol.* **26**, 377–386.

Chen, Q. & Mynett, A. E. 2003 Integration of data mining techniques with heuristic knowledge in a fuzzy logic modelling of eutrophication in Taihu Lake. *Ecological Modelling* **162**, 55–67.

Chen, Q. & Mynett, A. E. 2004 A robust fuzzy logic approach to modelling algal biomass. *Journal of Hydraulic Research* **42**, 303–309.

Chen, Q. & Ouyang, Z. 2005 Integrated Ecohydraulics Model and the Application. *Journal of Hydraulic Engineering* **36** (11), 1273–1279.

Chen, Q. 2004 *Cellular Automata and Artificial Intelligence in Ecohydraulics Modelling*. Taylor & Francis Group plc. London UK.

Chen, Q., Mynett, A. E. & Minns, A. W. 2002 Application of cellular automata to modelling competitive growth of two

underwater species C. aspera and P. pectinatus in Lake Veluwe. *Ecological Modelling* **147**, 253–265.

Chen, Q. H., Zeng, Z., Zhang, S. & Xu, K. 1987 Report on the Red tide occurred in Xiamen Harbour in 1987. In *Collected papers on red tidesurvey and study in Xiamen Harbour*. Third Institute of Oceanography State Oceanic Administration, China Ocean Press. Beijing.

Cronhjort, M. B. 2001 The interplay between reaction and diffusion. In *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (Dieckmann, U., Law, R. & Metz, J. (Eds)), pp. 151–170. Cambridge University Press, UK.

1992 *Individual-based Models and Approaches in Ecology* (ed. in D. L. DeAngelis & L. J. Gross). Chapman and Hall. New York.

Di Maio, J. & Corkum, L. D. 1995 Relationship between the spatial distribution of freshwater mussels (Bivalvia: Unionidae) and the hydrological variability of rivers. *Can. J. Zool.* **73**, 663–671.

Downing, J. A., Rochon, Y., Perusse, M. & Harvey, H. 1993 Spatial aggregation, body size, and reproductive success in the freshwater mussel *Elliptio complanata*. *J. North Am. Benthol. Soc.* **12**, 148–156.

Duel, H., van der Lee, G. E. M. & Penning, W. E. 2002 Habitat Modelling of Rivers and Lakes in the Netherlands: An Ecosystem Approach. *Canadian Water Resources Journal* **28**, 231–248.

Dunning, J. B., Steward, D. J., Danielson, B. J., Noon, B. R., Root, T. L., Lamberson, R. H. & Stevens, E. E. 1995 Spatially explicit population models: current forms and future uses. *Ecological Applications* **5** (1), 3–11.

Haan, C. T. 1977 *Statistical Methods in Hydrology*. Iowa State University Press.

Hardison, B. S. & Layzer, J. B. 2001 Relations between complex hydraulics and the localized distribution of mussels in three regulated rivers. *Regulated Rivers: Research & Management* **17**, 77–84.

Hart, R. A., Grier, J. W., Miller, A. C. & Davis, M. 2001 Empirically derived survival rates of a native mussel, Amblema plicata, in the Mississippi and Otter Tail Rivers, Minessota. *Am. Midl. Nat.* **146** (2), 254–263.

Hastie, L. C., Boon, P. J., Young, M. R. & Way, S. 2001 The effects of a major flood on an endangered freshwater mussel population. *Biological Conservation* **98**, 107–115.

Holland-Bartels, L. E. 1990 Physical factors and their influence on the mussel fauna of a main channel border habitat of the Upper Mississippi River. *J. North Am. Benthol. Soc.* **9**, 327–335.

Huston, M., DeAngelis, D. & Post, W. 1988 New computer models unify ecological theory. *BioScience* **38** (10), 682–691.

Jørgensen, S. E. 1994 *Fundamentals of Ecological Modelling*, 2nd Edition. Elsevier Science BV, The Netherlands.

Kat, P. W. 1982 Effects of population density and substratum type on growth and migration of Elliptio complanata (Bivalvia: Unionidae). *Malacol. Rev.* **15**, 119–127.

Law, R. and Dieckmann, U. 1998 On scaling up from individual-based processes to macroscopic ecological dynamics in spatially extended communities, International Institute for Applied Systems Analysis, Interim Report IR-98-046, 28 pp.

Layzer, J. B. & Madison, L. M. 1995 Microhabitat use by freshwater mussels and recommendations for determining their instream flow needs. *Regulated Rivers: Research & Management* **10**, 329–345.

Lee, G. E. M., Aarts, H. P. A. & Boogaard, H. F. P. 2002 Analysis of uncertainty in expert rules and input data for habitat suitability models. In *Proceedings of the 4th Ecohydraulics Conference and Environmental Flows* (ed. in J. King). Cape Town, South Africa.

Lee, J. H. W., Huang, Y., Dickman, M. & Jayawardena, A. W. 2003 Neural network modelling of coastal algal blooms. *Ecological Modelling* **159**, 179–201.

Legendre, P. & Legendre, L. 1998 *Numerical Ecology*, 2nd English Ed.. Elsevier. Amsterdam, the Netherlands.

Li, H., Mynett, A. E., Huang, B. Q. & Hong, H. S. 2005 *Main factor selection in Harmful Algal Bloom Prediction with a case study for Western Xiamen Bay*. IAHS International Symposium on Methodology in Hydrology. Nanjing, China.

Maier, H. R. & Dandy, G. C. 1997 Modelling cyanobacteria (blue-green algae) in the River Murray using artificial neural networks. *Mathematics and Computers in Simulation* **43**, 377–386.

Mao, J. Q. 2005 *Study on Assessment Methods and Numerical Simulation of Eutrophication in Lakes and Reservoirs* PhD theis, Tsinghua University.

Marcel, S. 1999 *Charophyte Colonisation in Shallow Lakes*, PhD Thesis, University of Amsterdam, The Netherlands.

May, R. M. 1975 Biological populations obeying difference equations: stable points, stable cycles and chaos. *Journal of Theoretical Biology* **49**, 511–524.

McMahon, R. F. & Bogan, A. E. 2001 Mollusca: Bivalvia. In *Ecology and classification of North American freshwater invertebrates* (ed. in Thorp, J. H. & Covich, A. P.), pp. 331–429. Academic Press.

Morales-Chaves, Y. 2004 *Analysis of mussel population dynamics in the Mississippi River* (**228** Ph.D. Thesis, Civil and Environmental Engineering, University of Iowa, Iowa City, Iowa.

Morales, Y., Weber, L. J., Mynett, A. E. & Newton, T. J. 2006a Mussel Dynamics Model: a hydroinformatics tool for analyzing the effects of different stressors on the dynamics of freshwater mussel communities. *Ecological Modelling* (in press).

Morales, Y., Weber, L. J., Mynett, A. E. & Newton, T. J. 2006b Effects of substrate and hydrodynamic conditions on the formation of mussel beds in a large river. *J. N. Am. Benthol. Soc.* **25** (3), 664–676. (in press).

Murray, S. 1993 *Neural Networks For Statistical Modelling*. Van Nostrand Reinhold Press. New York.

Mynett, A. E. 2002 Environmental Hydroinformatics: the way ahead. In *Proc. 5th Int. Conference on Hydroinformatics*, pp. 31–36. Cardiff, UK, IWA Publishing. London.

Mynett, A. E. and Chen, Q. 2004 Cellular Automata in Ecological and Ecohydraulics Modelling, Proc Int. Conf. on Cellular Automata, University of Amsterdam, Springer Verlag Publishing, October.

Mynett, A. E. 2004 Hydroinformatics tools for ecohydraulics modeling. In *Proc. 6th Int. Conf. on Hydroinformatics* (ed. in Liong, V., Phoon, K. K. & Babovic, S. Y.),. World Scientific Publishing Company. Singapore, June.

Park, G. S. & Park, S. Y. 2000 Long-term trends and temporal heterogeneity of water quality in tidally mixed estuarine waters. *Marine Pollution Bulletin* **40**, 1201–1209.

Qi, Y. Z., et al. 2003 Coastal Red tide in China. In *Supported by Chinese National Natural Science Foundation*. China Science Press. Beijing.

Recknagel, F. 1997 ANNA-Artificial neural network model for predicting species abundance and succession of blue-green algae. *Hydrobiologia* **349**, 47–57.

Recknagel, F., Petzoldt, T., Jaeke, O. & Krusche, F. 1994 Hybrid expert system DELAQUA – a toolkit for water quality control of lakes and reservoirs. *Ecological Modelling* **71**, 17–36.

Ricciardi, A., Whoriskey, F. G. & Rasmussen, J. B. 1996 Impact of the Dreissena invasion on native unionid bivalves in the upper St. Lawrence River. *Can. J. Fish. Aquat. Sci.* **53**, 1434–1444.

Salski, A. & Sperlbaum, C. 1991 Fuzzy logic approach to modelling in ecosystem research. In *Uncertainty in Knowledge Bases. LNCS 521* (Bouchon-Meunier, B., Yager, R. R. & Zadeh, R. R. (Eds)), pp. 521. Springer.

Solomatine, D. P. 2004. Data-driven modelling: Machine learning, data mining and knowledge discovery, UNESCO-IHE Institute for Water Education lecture notes, HH482/04/1, pp. 65.

Strayer, D. L. 1999a Effects of alien species on freshwater mollusks in North America. *J. North Am. Benthol. Soc.* **18**, 74–98.

Strayer, D. L. 1999b Use of flow refuges by unionid mussels in rivers. *J. North Am. Benthol. Soc.* **18**, 468–476.

Strayer, D. L., Caraco, N. F., Cole, J. J., Findlay, S. & Pace, M. L. 1999 Transformation of freshwater ecosystems by bivalves: a case study of zebra mussels in the Hudson River. *Bioscience* **49** (1), 19–27.

USGS (US Geological Survey) 2003 *Long term resource monitoring program*. Upper Midwest Environmental Sciences Center. La Crosse, Wisconsin, (Available from: http://www.umesc. usgs.gov/ltrmp.html).

Weiss, S. M. & Indurkhya, N. 1998 Predictive Data Mining. In *A Practical Guide*. Morgan Kaufmann Publishers, Inc, USA.

Whitney, S. D., Blodget, K. D., and Sparks, R. E. 1997. A comprehensive evaluation of three mussel beds in reach 15 of the Upper Mississippi River, Long Term Resource Monitoring Program, U.S. Geological Survey, Environmental Management Technical Center, Reprint, 97–R022.

Witten, I. H. & Frank, E. 2000 Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.

Wolfram, S. 1984 Universality and complexity in cellular automata. *Physica* **10D**, 1–35.

Zhang, S. 1993 Ecological features of phytoplankton from red tide occurring area in Xiamen Harbour. In *Collected papers on red tide survey and study in Xiamen Harbour (in Chinese)*, pp. 29–36. China Ocean Press. Beijing, Third Institute of Oceanography State Oceanic Administration.