This conference is supported by
the EU 5th Framework Programme
through the PAEQANN project
(contract EVK1-CT1999-00026)

3rd Conference of the International Society for Ecological Informatics

Abstract book

Villa Grazioli Park Hotel
Grottaferrata (Rome), Italy
26-30 August 2002

http://www.isei3.org

University of Bari

University of Rome
"Tor Vergata"

econ S.R.L.

ICRAM

CoNISMa

Organising Committee

M. Scardi, Univ. of Rome "Tor Vergata", Italy
F. Recknagel, Univ. of Adelaide, Australia
S. Lek, Univ. of Toulouse, France

International Scientific Committee

M. Scardi, Univ. of Rome "Tor Vergata", Italy, Chair
L. Boddy, Univ. of Cardiff, UK
D. Chen, Internat. Pacific Halibut Commission, USA
T.-S. Chon, Univ. of Pusan, South Korea
D. D'Angelo Morrall, Procter & Gamble Co., USA
M. Dreyfus Leon, Univ. Aut. de Baja California, Mexico
S. Dzeroski, Jozef Stefan Institute, Slovenia
G. M. Foody, Univ. of Wales, UK
S. E. Jorgensen, Univ. of Copenhagen, Denmark
S. Lek, Univ. Paul Sabatier Toulouse, France
R. I. McKay, Univ. of New South Wales, Australia
F. Recknagel, Univ. of Adelaide, Australia
W. J. Walley, Staffordshire Univ., UK
H. Werner, Univ. of Kassel, Germany
P. A. Whigham, Univ. of Otago, New Zealand
X. Yao, Univ. of Birmingham, UK

Local Organising Committee

M. Scardi, Univ. of Rome "Tor Vergata", Italy
E. Fresi, Univ. of Rome "Tor Vergata", Italy
S. Cataudella, Univ. of Rome "Tor Vergata", Italy

List of participants

Adriaenssens	Veronique	BE	Veronique.Adriaenssens@rug.ac.be
Aguilar Ibarra	Alonso	FR	aguilar@ensat.fr
Akkermans	Wies	NL	l.m.w.akkermans@plant.wag-ur.nl
Anishchenko	Irina	UA	ira_anishchenko@hotmail.com
Bouharati	Saddek	DZ	sbouharati@yahoo.fr
Bredeweg	Bert	NL	bert@swi.psy.uva.nl
Bretin	Luc-Patrick	FR	luc-patrick.bretin@ulp.u-strasbg.fr
Cataudella	Stefano	IT	cataudel@uniroma2.it
Chen	Din	US	din@iphc.washington.edu
Chon	Tae-Soo	KR	tschon@pusan.ac.kr
Compin	Arthur	FR	compin@cict.fr
Coste	Michel	FR	michel.coste@bordeaux.cemagref.fr
Delmas	François	FR	francois.delmas@ bordeaux.cemagref.fr
de Thoisy	Benoît	FR	kwata@nplus.gf
Dedecker	Andy	BE	andy.dedecker@rug.ac.be
D'heygere	Tom	BE	tom_dheygere@hotmail.com
Di Dato	Paola	IT	pdidato@mclink.it
Dreyfus Leon	Michel	MX	dreyfus@cicese.mx
Dubois	Marc-André	FR	mad@cea.fr
Fielding	Alan	UK	a.fielding@mmu.ac.uk
Foody	Giles	GB	g.m.foody@soton.ac.uk
Fresi	Eugenio	IT	fresi@uniroma2.it
Frolova	Ludmila	RU	Lucy.Frolova@ksu.ru
Gabriels	Wim	BE	wim.gabriels@rug.ac.be
Gevrey	Muriel	FR	gevrey@cict.fr
Goethals	Peter	BE	peter.goethals@rug.ac.be
Gosselain	Véronique	BE	Veronique.Gosselain@fundp.ac.be
Holmberg	Maria	FI	maria.holmberg@vyh.fi
Horrigan	Nelli	AU	nelli.horrigan@adelaide.edu.au
Hraber	Peter	US	pth@santafe.edu
Jeong	Kwang-Seuk	KR	pow5150@hananet.net
Jorgensen	Sven	DK	sej@dfh.dk
Joy	Mike	NZ	mikejoy@clear.net.nz
Kaski	Samuel	FI	Samuel.Kaski@hut.fi
Kim	Cheol-Ki	KR	kck@harmony.cs.pusan.ac.kr
Knoflacher	Helmut Markus	AT	markus.knoflacher@arcs.ac.at
Koh	Sung-Cheol	KR	skoh@hanara.kmaritime.ac.kr
Koroleva	Tatiana	RU	Tata.Koroleva@ksu.ru
Lange	Holger	DE	holger.lange@bitoek.uni-bayreuth.de
Le Duc	Mike	GB	mled@liv.ac.uk

Lek	Sovan	FR	lek@cict.fr
Maamri	Abdellatif	MA	maamri@sciences.univ-oujda.ac.ma
Mancini	Laura	IT	lmancini@iss.it
May	Michael	DE	michael.may@ais.fhg.de
Mayer	Audrey	US	mayer.audrey@epa.gov
Moreau	Jacques	FR	moreau@ensat.fr
Naumenko	M.A.	RU	naumenko@limno.org.ru
Nepomnyashchikh	Valentin	RU	nepom@ibiw.yaroslavl.ru
Neumann	Michael	DE	m.neumann@uni-jena.de
Nijboer	Rebi	NL	r.c.nijboer@alterra.wag-ur.nl
O'Connor	Mark	GB	mo3@staffs.ac.uk
Ozesmi	Stacy	TR	stacy@erciyes.edu.tr
Ozesmi	Uygar	TR	uozesmi@erciyes.edu.tr
Park	Young-Seuk	FR	park@cict.fr
Parrott	Lael	CA	lael.parrott@UMontreal.CA
Podgornyj	Konstantin	RU	kap@ibiw.yaroslavl.ru
Ragia	Lemonia	DE	lemonia.ragia@ais.fhg.de
Recknagel	Friedrich	AU	friedrich.recknagel@adelaide.edu.au
Rimet	Frederic	LU	rimet@crpgl.lu
Russo	Tommaso	IT	tommy.russo@libero.it
Salski	Arkadiusz	DE	asa@email.uni-kiel.de
Scardi	Michele	IT	mscardi@mclink.it
Shanmuganathan	Subana	NZ	subana.shanmuganathan@aut.ac.nz
Siligardi	Maurizio	IT	msilig@tin.it
Soyupak	Selçuk	TR	selcuk_soyupak@atilim.edu.tr
Sperr	Tobias	DE	tosp@uni-oldenburg.de
Spezzano	Giandomenico	IT	spezzano@isi.cs.cnr.it
Stockwell	David	US	davids@sdsc.edu
Talib	Anita	MY	anita@usm.my
Tancioni	Lorenzo	IT	tancioni@uniroma2.it
Teriokhin	Anatoli	RU	ter@mat.bio.msu.su
Tison	Juliette	FR	juliette.tison@bordeaux.cemagref.fr
Trigg	David	GB	D.J.Trigg@staffs.ac.uk
Tsichritzis	Dennis	DE	dennis.tsichritzis@zv.fraunhofer.de
van Nes	Egbert	NL	Egbert.vanNes@aqec.wkao.wau.nl
Verdonschot	Piet	NL	P.F.M.Verdonschot@alterra.wag-ur.nl
Villanueva	Maria Concepcion	FR	ching@ensat.fr
Werner	Heinrich	DE	werner@neuro.informatik.uni-kassel.de
Whigham	Peter	NZ	pwhigham@infoscience.otago.ac.nz

Author index

River quality assessment based on fuzzy logic

Veronique Adriaenssens, Peter L.M. Goethals & Niels De Pauw

Laboratory of Environmental Toxicology and Aquatic Ecology,
Ghent University, J. Plateaustraat 22, B-9000 Ghent, Belgium.
E-mail: Veronique.Adriaenssens@rug.ac.be

Fuzzy logic is a soft computing technique that makes use of fuzzy sets and fuzzy linguistic rules to incorporate the uncertainty into a model. Fuzzy logic can be very useful within the domain of ecology. Many natural systems are inherently non-linear and cannot be represented by conventional linear models. Ecological expert knowledge is also often vague and uncertain to be expressed by mathematical functions. Fuzzy logic can process this imprecise and uncertain information. Fuzzy rule-based systems can also use knowledge of experts in the given field of interest because of the ability to work with linguistic variables. There are several applications for which the fuzzy models were designed in ecosystem management. Most of these developed models are used for assessment of ecosystem integrity or sustainability.

In this paper, the fuzzy logic technique will be evaluated for his use into river quality assessment. This will be done by means of a case study on the Zwalm river basin (Flanders, Belgium). An ecological dataset of 120 samples at 60 sites, collected over a two-year period (2000-2001) is used for this purpose. Fifteen structural, physical and chemical environmental variables such as temperature, pH, percentage of dissolved oxygen, water depth, stream velocity, presence/absence of hollow beds, deep/shallow variation… were measured at each site, as well as the abundance of the aquatic macroinvertebrate taxa. After input variable selection, input variables were fuzzificated into fuzzy sets, dependent on their degree of uncertainty. A fuzzy rule base system was developed using an expert knowledge database. The output of the fuzzy river assessment system was then compared with assessment systems used by The Flemish Environmental Agency. The robustness and sensitivity of the fuzzy assessment model for different water quality situations (range: very bad to very good) was analyzed.

Finding fish species patterns in the Garonne basin (France) with a self-organising map

Alonso Aguilar Ibarra [1], Young-Seuk Park[2], Puy Lim [1], Sovan Lek [2]

[1] Ecole Nationale Supérieure Agronomique, 1, ave de l'Agrobiopole, BP 107 Auzeville-Tolosane, 31326, France. E-mail: aguilar@ensat.fr
[2] CESAC, UMR 5576, CNRS-Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 4, France. E-mail: park@cict.fr

We present patterns of fish distribution in South Western France. This study is focused on defining ichthyoregions in the Garonne basin by using a Kohonen's self-organising map (SOM). The SOM is a non-supervised artificial neural network which allows to extract features in presented patterns, without being told what outputs (i.e. guilds) associated with the input data are desired. We used a two dimensional (4x3) output map.
Our fish community dataset was collected in sampling sites spread over the Garonne basin between 1986 and 1996. Five environmental variables (altitude, distance from the source, mean annual water temperature, mean annual water flow, surface of catchment area), and fish species richness were measured at each site. The SOM was trained with several datasets (native species, introduced species, migratory species, and total species). The trained SOMs were able to identify fish assemblage patterns, from which we recognised a gradient of fish species longitudinal succession, and in accordance to environmental features. This pattern was more evident for native species than for introduced species. In the latter case, human-induced introductions may have played a role in their distribution. Migratory species were rather associated to the larger rivers of the basin. After our analysis, we may infer the existence of four ichthyoregions in the Garonne basin following a species gradient: an upstream headwater region, an upstream transition region, a downstream transition region, and a downstream plains region. We conclude that our study would help to understand landscape ecology in the area, and would represent a tool for aquatic management and assessment agencies.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Prediction of class membership by means of Support Vector Machines

Akkermans W., Verdonschot P.F.M. and Nijboer R.C.

Alterra, Green World Research, team of Freshwater Ecology, Wageningen, The Netherlands.
E-mail: L.M.W.Akkermans@plant.wag-ur.nl

Support Vector Machines (SVMs) are nonlinear predictors that can be used both for classification and regression. In this paper the emphasis is on classification.
In the SVM framework class boundaries are defined by so-called 'boundary observations' or support vectors. Hence support vectors are data points lying in some sense 'between' two classes. The search for these boundary observations is the main task of a Support Vector Machine.
It appears that SVMs can be expressed as quadratic programming problems, which implies that they have a well-behaved loss function and clearly defined stop criterion. This can be considered as an advantage of this method over Neural Networks.
In this paper a data set of freshwater sites is considered, which has been classified into cenotypes on the basis of macro-invertebrate species abundance. An SVM is fitted to these data in order to predict the cenotypes using only environment variables. The performance of the classifier is assessed by means of leave-one-out crossvalidation.
Finally, a method proposed by Platt (1999) in the context of 2-class classification is adapted to multi-class classification and used to obtain estimates of posterior probabilities of class membership for each observation.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Creation and use of a database of Dinophyta of Ukraine

Anishchenko I., Krakhmalnyy A.

Institute of Botany, National Academy of Sciences of Ukraine
2 Thereshchenkivska St., 1601 Kyiv – UA.
E-mail: ira_anishchenko@hotmail.com

The problem of preservation and global inventory of all objects of the plant world is very important now. Data resulting from our work show, that Dinoflagellates developing in the various Ukrainian reservoirs (from small puddles to large rivers, lakes and estuaries), however the frequency of their occurrence is very low (1-5% from the total quantity of samples). At present the species of Dinoflagellates with wide amplitude of adaptation to the different ecological conditions are found in the continental reservoirs of Ukraine. They are following: Ceratium hirundinella (O.F.Muller) Berg, Peridiniopsis quadridens (Stein) Bourr., Peridinium cinctum (J.F.Muller) Ehr., P. umbonatum Stein (rivers, freshwater lakes, ponds) and Prorocentrum cordatum Dodge, P. micans Ehr. (estuaries, saltwater lakes). However, the most Dinoflagellates characterise by limited amplitude of the adaptation to the environment, that explains low frequency of their occurrence. The problem of creating and using a database of Dinophyta in Ukraine is connected to the need to inventory the biodiversity of all Dinoflagellate species - one of the most sensitive group of vegetative organisms and indicative of clean water. Such systems considerably increases an performance of the specialist phycologist, and also allows the conduct of otherwise labour-intensive analyses. Based on an analysis of a data domain we designed a conceptual pattern, which includes both a taxonomic information (accepted name: family, genus, species, synonyms) and references (place of growth, ecological characteristics of biotope, literature references, geography of distribution). Users of such system may to update a database by new species, replace of a main species by a synonym and reverse procedure, introduce nomenclatural changes, work with original and bibliographic information, prepare a systematic structure, make checklists, which can be used in natural reservations and landscape parks of the Ukraine. Moreover, this system allows to use an electronic mycroscope's photos of dinoflagellate species.

Detection of microbiological pollution in fresh water by Fuzzy logic method

Bouharati S.[1 2], Harzallah D.[1], Benmahamed K.[2], Abdesalem M.[3], Hachem A.[3]

[1]Biology Dept., Faculty of Sciences, UFA Setif Algeria. E-mail: sbouharati@yahoo.fr
[2]Intelligent systems laboratory, Electronic dept., Faculty of Ingineering, UFA Setfi Algeria.
[3]Electrical engineering Dept., Faculty of Ingineering, Assiut University, Egypt.

Even water which looks clear and pure may be sufficient contaminated with pathogenic microorganisms to be a health hazard. Some means are necessary to ensure that drinking water is safe. One of the main tasks of water microbiology is the development of laboratory methods which can be used to detect the microbiological contaminants that may be present in drinking water. To check each drinking water supply for these contaminants would be a difficult and time-consuming job. The coliform goup of organisms are suitable as indicators because they are common inhibitants of the intestinal tract. Althougt, some of the quantitative and qualitative methods used up to now to detect colirom bacteria are so effective, the confirmed presence of these bacteria is detectible following at least 24-hour incubation time. In this study, a new approch for the detec tion of colifom bacteria in fresh water is proposed. This is done by the use of the Fuzzy logic method. The Fuzzy logic method is based on the variations of the physical and chemical parameters occured during bacterial growth in fuzzy environment. The treatment of data is done by Fuzzy algorithm which ends by a Fuzzy program. The response will be in linguistic and numerique expression of microbiological contaminants of fresh water.

A fuzzy logic model for fish recruitment forecast

D. G. Chen

International Pacific Halibut Commission
University of Washington, P. O. Box 95009, Seattle WA 98145-2009, USA
E-mail: din@iphc.washington.edu

A fuzzy logic model is proposed for fish recruitment forecast using fish stock-recruitment data incorporating environmental information. Data from the Pacific halibut stock with the Pacific Decadal Oscillation index as the environmental variable is used for the model development. A comparison is conducted to the traditional stock-recruitment analysis and neural network model. It is demonstrated that the fuzzy logic model outperforms the traditional stock-recruitment analysis and the neural network model in model fitting and recruitment forecasting as measured by several diagnostic criteria.

A comparison of neural network and fuzzy logic models for estimating seasonal pseudo steady state chlorophyll-a concentrations in reservoirs

D. G. Chen[1] and Selçuk Soyupak[2]

[1]International Pacific Halibut Commission, University of Washington, P. O. Box 95009, Seattle WA 98145-2009, USA. E-mail: din@iphc.washington.edu
[2]Civil Engineering Department, Atilim University, Kizilcasar Koyu, İncek , 06836, Ankara, Turkey. E-mail: selcuk_soyupak@atilim.edu.tr

Fuzzy logic and neural network models were developed to estimate pseudo steady state chlorophyll-a concentrations in a very large and deep dam reservoir that exhibits high spatial and temporal variability. The estimation power of the developed fuzzy logic model was tested by comparing its performance with that from neural network modeling approach. The utilized data set include chlorophyll-a concentrations as an indicator of primary productivity as well as several other water quality variables such as PO_4 phosphorus, NO_3 nitrogen, alkalinity, suspended solids concentration, pH, water temperature, electrical conductivity, dissolved oxygen concentration and Secchi depth as independent environmental variables. Considering the existence of non-significant functional relationships between some of the water quality variables and the chlorophyll-a concentrations, an initial analysis was conducted to select the most important variables that can be used in estimating the chlorophyll-a concentrations within the studied water body. Following the outcomes from this initial analysis, the fuzzy logic model and neural network models were developed to estimate the chlorophyll-a concentrations and the advantages of the fuzzy logic model were demonstrated in model fitting over neural network modelling approach.

Self-Organizing Mapping on Response Behavior of Indicator Species Exposed to Toxic Chemicals for Developing Automatic Bio-monitoring Systems in Aquatic Environment

Tae-Soo Chon*, Inn-Sil Kwak, Mi-Young Song, Chang Woo Ji, Cheol Ki Kim[1], Eui-Young Cha[1], Sung Cheol Koh[2], Jong-Sang Kim[3], Joo-Baek Leem[4] and Sung-Kyu Lee[5]

**Division of Biological Sciences, Pusan National University, 30 Changjeon-dong, Keumjeong-ku, Pusan 609-735, Korea E-mail: tschon@pusan.ac.kr*
[1]Division of Computer Science, Pusan National University, 30 Changjeon-dong, Keumjeong-ku, 609-735, Pusan KORE
[2]Department of Marine Environmental Engineering, Korea Maritime University, Pusan 606-791, Korea
[3]Department of Animal Science and Biotechnology, Kyungpook National University, Taegu 702-701, Korea
[4]Korean Inter-University Institute of Ocean Science, Pusan 608-737, Korea
[5]Toxicology Research Center, Korea Research Institute of Chemical Technology, Taejon 305-600, Korea

Behavioral responses of indicator species have been reported to be sensitive to sub-lethal doses of toxic chemicals, and have been used for *in situ* biomonitoring tool to detect toxic chemicals in environment. Due to complexity and non-linearity in behavioral data, however, not much research has been conducted on computationally characterizing response behaviors of indicator species. Artificial neural networks were utilized to train data for the movement tracks to recognize the patterns of response behaviors. The indicator species such as medaka fish were treated with insecticides (e.g., carbofuran) in semi-natural conditions, and characteristic patterns of the movement tracks of test specimens were observed continuously in individuals through the automatic image recognition system before and after treatments for 4 – 5 days. The data for the locomotive tracks in a short time period (e.g., 1 min) were investigated in a sequence, and patterns characterizing behaviors before and after treatments were selected for training with the Kohonen network. Subsequently the self-organizing mapping was produced to pattern new input data of the movement tracks. As new input data were given to the trained map successively in one-minute sequence, the network was able to pattern continuously and automatically the sequence of one-minute input data. This study indicated that complex movement data could be patterned by the self-organizing mapping, and the training with artificial neural networks is useful for developing an automatic bio-monitoring system for detecting presence of toxic chemicals in environment.

Predicting the functional structure of macroinvertebrate communities in the Adour Garonne stream system (France)

Compin A., Park Y.S., Céréghino R., & Lek S.

Centre d'Ecologie des Système Aquatiques Continentaux, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse cedex, France. E-mail : compin@cict.fr

The aim of this work was to predict the composition of macroinvertebrate functional feeding groups (grazers, scrapers…) with three groups of environmental variables (typological, chemical, and climatic), using artificial neural networks. The data (invertebrates, variables) were recorded in the Adour-Garonne stream system (116000 km², South-Western France). The modelling procedure was conducted following two steps. First, these data were computed with a Self-Organising Map (SOM), an unsupervised artificial neural network, which helped to classify the sampling sites according to community structure and to interpret relationships between biological attributes and environmental variables. Then, a backpropagation algorithm (BP), a supervised neural network, was applied to predict the composition of functional feeding groups using environmental variables. From the database, two third of sampling sites were used to train the BP, while remaining sites were applied to test the model. Sensitivity analyses were carried out to assess the influence of environmental variables on each functional groups. The results of the model showed high predictabilities for each functional group and high variation of functional groups with different habitats and resources. The sensitivity analyses showed that different functional groups responded differently to environmental variables. Predicting the functional structure of macroinvertebrate communities with environmental variables is thus a valuable tool for the assessment of disturbance in a given area: by knowing what the community functional structure should be like under theoretically undisturbed conditions in a given area, we can provide explicit spatial schemes which may be useful to further investigation, but we primarily can determine the degree to which human activities have altered it.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Optimisation of predictive decision tree and neural network ecosystem models with genetic algorithms

Tom D'heygere, Peter Goethals & Niels De Pauw

Laboratory of Environmental Toxicology and Aquatic Ecology,
Ghent University, J. Plateaustraat 22, B-9000 Gent, Belgium
E-mail: tom_dheygere@hotmail.com

The selection of appropriate input variables is an important aspect of modelling. Numerous variables can be involved in modelling and most variables cannot be omitted without a significant information loss. Rigorous methods are thus needed for detecting which variables are relevant. Appropriate selection of input variables is not only important for modelling objectives as such, but also to ensure reliable decision support in river management and policy making. The selected variables can therefore also be considered as the critical variables in river restoration and conservation management programmes.

Genetic algorithms are effective in detecting the importance of input variable combinations for ecological models. An input variable selection scheme based on genetics algorithms has been developed for both decision tree and artificial neural network (ANN) models. The decision tree models are based on the J48 algorithm which is a java re-implementation of C4.5, one of the most well known and widely-used decision tree induction systems. The ANN models are feed-forward networks trained with the back-propagation algorithm.

In this paper, the results of the optimised decision tree and neural network models are compared. Therefore, a database of 360 samples of macroinvertebrate communities in unnavigable watercourses throughout Flanders was applied. Along with the abundances of 92 taxa of macroinvertebrates, the database contained a number of abiotic variables. It was a combination of physical-chemical (temperature, pH, dissolved oxygen concentration and total organic carbon, Kjeldahl nitrogen, total phosphorus concentrations of the sediment), eco-toxicological and structural variables (such as flow velocity and water depth). A comparison was made between the variables that were selected in both model types and their predictive performance on the basis of a validation dataset. Also the ecological relevance of the predictions and the usefulness of the models for decision support in water management was assessed.

Development of a Decision Support System for integrated water management in the Zwalm river basin, Belgium

Tom D'heygere, Veronique Adriaenssens, Andy Dedecker, Wim Gabriels, Peter Goethals & Niels De Pauw

Laboratory of Environmental Toxicology and Aquatic Ecology,
Ghent University, J. Plateaustraat 22, B-9000 Gent, Belgium
E-mail: tom_dheygere@hotmail.com

The Zwalm catchment is part of the Upper-Scheldt river basin and is characterised by an irregular flow regime. Floodings in urbanised areas led to the construction of weirs for water quantity control. Due to these weirs (in combination with numerous other structural disturbances) aquatic habitats are degraded and species migration is obstructed. In addition, several perturbations caused by domestic, industrial and agricultural activities have a serious impact on the river ecosystem quality. To achieve an optimal exploitation of the water system, all stakeholders such as nature conservationists, farmers, drinking water production companies, water quantity managers, … have to be involved in the decision making process. However, due to the enormous amount of information and the complex behaviour of the processes in a river basin, there is a high need for data management tools and models to perform simulations. A Decision Support System (DSS) is being developed to improve the reliability and efficiency of management decisions on the water use by the different stakeholders. The DSS includes information about aquatic species (macroinvertebrates, fish), water quality, habitat quality, stream characteristics, human activities,… Ecosystem models based on artificial neural networks, decision trees and fuzzy logic as well as a migration model based on calculations with overlays in a Geographical Information System (GIS) are applied to predict the response of key features of the riverine environment to proposed management actions. In this manner, the most sustainable actions can be selected, taking the goals of all stakeholders into consideration. The DSS also allows to estimate the time needed to observe the effect of restoration actions, what is often important to convince all stakeholders of the effectivity of particular management decisions. In this manner, the DSS is an interesting tool to convince water managers and stakeholders to take specific actions and to elaborate an efficient planning of the restoration activities.

Line transects: attempts to optimize sampling efforts with the use of neural networks

Benoît de Thoisy [1], Marc A. Dubois [2], Sébastien Brosse [3]

[1] Institut Pasteur, Cayenne, F97300, France, E-mail: kwata@nplus.gf
[2] SPEC, DSM, CEA Saclay – Orme des Merisiers F91191 Gif sur Yvette cedex France
[3] UMR 5576 CESAC-CNRS, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse cedex 4, France

Line transect sampling is a method commonly used to census large mammals and birds in neotropical rainforests. Despite numerous theoretical developments and field test applications, the effect on census reliability of a major parameter, namely the sampling effort (i.e., the total transect length), has been scarcely investigated. The aim of this study is to analyse the useful data acquisition rate during the sampling, with the use of backpropagated neural networks. We pooled data obtained on 12 transects conducted in French Guianan forests, surveying 12 areas along 100 km each. These areas were facing different human disturbance levels, ranging from no disturbance to heavy hunting pressure. Using information provided by the neural network model, we show that 1) we can optimise the sampling effort to reach a reliable assessment of the fauna composition and abundance 2) we can estimate indicators of the hunting pressure.

Sensitivity and robustness of predictive neural network ecosystem models for simulations of 'extreme' management scenarios

Andy Dedecker, Peter Goethals & Niels De Pauw

Laboratory of Environmental Toxicology and Aquatic Ecology,
Ghent University, J. Plateaustraat 22, B-9000 Ghent, Belgium
E-mail: andydedecker@hotmail.com

Back-propagation artificial neural network algorithms were used to induce predictive ecosystem models on a dataset collected at the Zwalm river basin, Flanders (Belgium). This dataset consisted of 120 samples, collected over a two-year period (2000-2001). Fifteen structural, physical and chemical environmental variables such as temperature, pH, percentage of dissolved oxygen, water depth, stream velocity, presence/absence of hollow beds, deep/shallow variation… were measured at each site, as well as the abundance of the aquatic macroinvertebrate taxa.
In a preliminary study, different neural networks were developed and optimized to obtain the best model configuration for the prediction of the presence/absence of macroinvertebrate taxa. The best performing number of hidden layers and neurons, transfer functions in the hidden and output layer and training algorithms were searched for based on a validation set that consisted of measurements similar to the model training dataset. In this way, the model performance could only be optimized and assessed for predictions of river conditions that were similar to those in the collected data. The river conditions at most sites in the Zwalm catchment do not meet the desired 'good ecological status' as described in the European Water Framework Directive. Therefore the collected data and induced models are less useful as such for predictions in stream restoration management. For this reason, most relevant predictions can be classified as 'extreme' simulations for the induced neural network models and are therefore unreliable for practical decision support purposes.
The aim of this paper is to test the sensitivity and robustness of the neural network models for these 'extreme' values and to elaborate model training methodologies allowing the development of models competent for simulations of 'extreme' management scenarios. If these 'extreme' values are not present in the collected dataset, the use of ecological expert knowledge is recommended. To import this expert knowledge in the artificial neural network models, a virtual dataset, containing 'extreme' values, was created. Artificial neural networks trained with and without this virtual dataset were validated on the original dataset, merely consisting of measurement data and also assessed for practical applications in simulations for stream restoration management. In this way, the sensitivity and robustness of both types of models was assessed from a theoretical and practical point of view. Strengths and weaknesses of the different methods are discussed.

A neural network approach to the prediction of the benthic macroinvertebrate fauna composition in rivers

Paola Di Dato[1], Laura Mancini[2], Lorenzo Tancioni[1] and Michele Scardi[1]

[1] Department of Biology, University of Roma "Tor Vergata", Via della Ricerca Scientifica, 00133 Roma, Italy, E-mail: pdidato@mclink.it
[2] Italian National Institute of Health, Viale Regina Elena 299, 00161 Roma, Italy

Predicting the composition of benthic macroinvertebrate fauna in rivers is not a trivial task, both because of the number of species to be modeled and because of the complexity of the biotic and abiotic relationships that determine their distribution.
However, the composition of the benthic macroinvertebrate fauna usually provide very useful insights into the ecological status of a river and this is the reason why it has been routinely used since many years for computing biotic indices.
Several modeling approaches based on neural networks have been tested on a data set from the Latium rivers (Central Italy). This data set included 153 records for 65 taxa and 10 environmental variables (namely: elevation, distance from source, watershed drainage area, gradient and 6 variables for sediment structure, as percentages of different granulometric classes, from boulders to silt and clay).
In order to reduce the effect of quantitative sampling errors, all the models have been developed on the basis of the minimum level of faunistic information, i.e. presence/absence. Therefore, they allow to predict the probability of presence for each taxon (that obviously can be transformed into a binary value) given a set of environmental variables.
The benthic macroinvertebrate species to be modeled have been selected and/or grouped in different ways in order to attain the best accuracy in the predictions and in order to preserve the maximum amount of information with respect to the whole data set. In particular, the role of the frequency of species occurrence will be discussed and the results provided by the different models will be compared.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

An artificial neural network approach to model fishermen search decisions and information exchange between fishing vessels

Michel Dreyfus-León[1,2], Daniel Gaertner [3]

[1]*Instituto Nacional de la Pesca-PNAAPD México. E-mail: dreyfus@cicese.mx*
[2]*Universidad Autónoma de Baja California, México.*
[3]*Istitut de la Recherche pour le Developement, France.*

A simulated fishery is modeled where 20 purse seine vessels have to decide where to search for fish. The simulated world consist of 25 areas and a port for all vessels. Schools of fish are all identical and recruitment as well as natural mortality are included in the model. Fish schools move randomly at a speed reported for tuna schools in nature.
Three scenarios are considered varying in variability: One with only random changes in recruitment within the six fishing grounds considered, another with shifts in recruitment to new areas and the last one a combination of the previous two.
All artificial vessels present characteristics of real purse seiners, speed, 12 hour daily search period, duration of sets and trips similar to real vessels. They are represented by an artificial neural network that has already learned with backpropagation technique and reinforcement learning to make decisions in this world. Decisions are represented by four output neurons that predict an economic benefit if the decision to continue the search in an area is taken, to move to a near by fishing ground, mid distance or farther away area. Input information consist of knowledge of the duration of the trip, presence of other vessels in the area, knowledge of performance in the area, how good is the area for fishing based on previous experience, knowledge of quality of other areas near to vessel position, at mid distance or far away.
In real tuna purse seine fisheries, some vessels belong to a "group" and exchange information related to the fishing activity. The effects of this interchange has not been evaluated. In this model some of the vessels belong to a "group " while the rest depend only on there own experience. Varying the "group" size from two to twelve in the three scenarios let us see the advantage of information sharing and depending on the scenario, the possibility of competition within the "group". In general they perform better due to the sampling of several areas at the same time and this generates a more accurate knowledge of the world.

Patterning forest structures from high resolution LAI transects using Kohonen neural networks

Marc A. Dubois [1], Laurent Cournac [2], Sébastien Brosse [3], Young Seuk Park [3]

[1] SPEC, DSM, CEA Saclay – Orme des Merisiers F91191 Gif sur Yvette cedex France, E-mail: mad@cea.fr
[2] DSV, CEA Cadarache F13191 France
[3] UMR 5576 CESAC, Université Paul Sabatier 118 Route de Narbonne 31062 Toulouse France

Leaf Area Index (LAI) have been measured every meter along many transects in several tropical forests from South America and Central Africa, using a cheap, easy and reliable method. These measures contain information on the forest canopy structure. A sliding fast Fourier transform (SFFT) was performed and then averaged on each LAI transect. The obtained spectra were ponderated for each transect by its mean LAI value. We show that Kohonen neural networks applied to such signals allow patterning of different forest structures. This semi automatic forest characterisation could be used to monitor forestry resources on extended scales.

Neural network modeling of diatom community structure in the Loire river basin

Luc Ector[1], Frederic Rimet[1], Paola Di Dato[2] and Michele Scardi[2]

[1] Centre de Recherche Public - Gabriel Lippmann, CREBS, 162a Avenue de la Faïencerie, L-1511 Luxembourg, E-mail: ector@crpgl.lu
[2] Dept. of Biology, University of Rome "Tor Vergata", Via della Ricerca Scientifica, 00133 Roma, Italy

Neural network models provided interesting results in several applications where community structure had to be predicted on the basis of a set of environmental variables. Such a goal, however, can be more difficult to attain when the number of species to be modeled is very large. This problem is very common when benthic diatom are taken into account, as several hundreds of species can be present in a data set. The data set for the Loire river basin is an example of this situation, as in its 641 samples more than 900 diatom species have been found.
In this framework, different criteria for reducing the complexity of the model have been tested. They have been based on the selection of species subsets (1) representing the ecological relationships between sampling sites with a minimal loss of information and (2) including only those species that could be actually modeled on the basis of the available environmental information.
The results of these different modeling approaches will be presented and the implications of the selection of species to be modeled on the basis of their occurrence in the data set will be discussed.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Accuracy, Utility and Costs

Alan Fielding

Department of Biological Sciences, Manchester Metropolitan University; Manchester, UK
Email: a.fielding@mmu.ac.uk

I will be reviewing the current use of machine learning methods across the ecological disciplines, paying particular attention to their predictive accuracy, utility (ease of use, interpretation, scope) and the incorporation of costs. As part of this process I will compare three techniques that have had similar growth patterns over the last 10 years: artificial neural networks, decision and regression trees and generalized additive models.
Although machine learning applications have expanded there is evidence that they are restricted by habitat, journal and geography. What evidence is there that machine learning methods are 'better' than alternative methods and what strategies can we adopt to promote their use by colleagues who are currently unaware of their utility? Simply demonstrating that they are 'better' is probably insufficient because we will be pushing against a number of well established paradigms and considerable inertia.

Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks

Giles M. Foody[1], Mark E. Cutler[2]

[1]Department of Geography, University of Southampton, SO17 1BJ UK. E-mail: g.m.foody@soton.ac.uk
[2]Department of Geomatics, University of Newcastle, UK

The impact of logging on tree species diversity is an important issue in sustainable development. Knowledge of the impacts of logging on forest diversity may help in biodiversity conservation, the development of reduced impact logging methods and the formulation and assessment of environmental management policies. At the scale of the landscape, the typical scale used in forest management, it is difficult to acquire repetitively accurate data on tree species diversity. Satellite remote sensing offers the ability to observe forests at a range of spatial and temporal scales that are compatible with forest management activities. Here we evaluate the potential of remote sensing for the acquisition of information on tropical forest tree species diversity. The test site is a region centred on the Danum Valley Field Centre in Sabah, Borneo. At this site are forests that have been logged by various methods over time providing a range of forests of differing species diversity. A Landsat TM image of this site was rigorously pre-processed and used to evaluate the potential of remote sensing as a source of data on tree species diversity. Here attention is focused on the diversity of commercially valuable trees, as this is fundamental to sustainable development actions. Data on species diversity acquired from a sample of 52 plots surveyed in the field near the time of the Landsat TM image acquisition were used. A relative evaluation of different methods of extracting diversity information from the imagery was undertaken. A range of issues connected with tree diversity were addressed but we focus especially on the major variable of tree species richness. The methods used to extract information from the remotely sensed data comprise conventional approaches such as the use of vegetation indices and multiple regression analyses together with a series of analyses based on neural networks. A variety of neural networks are investigated, including multi-layer perceptron, radial basis function and generalized regression neural networks. In general, the neural network based approaches were found to provide the most accurate estimates of tree species richness. For example, a basic multi-layer perceptron derived estimates of species richness that were strongly and significantly correlated with field observation (r>0.65). Species richness is, however, only one component of biodiversity. The paper concludes with a brief discussion on the potential to more fully extract biodiversity information from the imagery with the use of a further neural network. Here self-organizing feature map (SOM) networks were used to ordinate the species and remotely sensed data sets. The partitioning of the SOM space depicting the ground data on tree species composition by the SOM derived from the imagery revealed a high degree of correspondence (~80%) between the SOM representations indicating a great potential for biodiversity mapping from remote sensing. This is illustrated with the production of a map depicting the variation in species richness and composition for the site predicted from the neural networks.

Fuzzy Expert System Of Water Quality Management

Ludmila Frolova

Kazan State University
Kremlevskaya St.18, 420008 Kazan, Russia
E-mail: Lucy.Frolova@ksu.ru

At present, the majority of water reservoirs test strong antropogeneous influence and become dangerous for the environment. Result of anthropogeneous influence is deterioration of water quality. So, the treatment of water is very important for life and health of the people.
As know, the quality of water in reservoirs depends from its measured biotic and abiotic parameters. Also, the degree of clearing of water depends from the requirements showed to use of a reservoir: for drinking water supply, fishing, rest and agriculture. The above than requirement to quality of water, the more efforts it is required for its treatment.
The fuzzy sets are used as a management model for construction of an integrated index, on which meaning it is possible to estimate the water quality in reservoir and to offer technologies on its treatment. The higher is the quality of water, then it is offered less ways for water treatment and on the contrary. The meaning of an index is calculated in a range [0,1] and has the following interpretation: if the meaning of an index more less 0,4 - condition of water system unstable also requires external intervention, if the meaning of an index is more 0,6 - condition of system aspires to sustainable development. Fuzzy logic is a part of expert system pattern on personal computer.
So, for execution of any program of water reservoirs treatment it is necessary to take the following steps: monitoring and collect the data of analysis; evaluate the data obtained (the limit of measurement error is 20-40%); using the fuzzy model for choose managing influence (optimal technologies of water reservoir treatment); analyse the last results and making-decision to reach sustainable development water ecosystem.
Thus, the integrated index allows to define a general state water ecosystem and necessary managing influences, which direct a condition water ecosystem to the party of achievement of its sustainable development.

Input variables selection of artificial neural networks predicting aquatic macrobenthos communities in Flanders (Belgium)

Wim Gabriels, Peter LM Goethals & Niels De Pauw

Laboratory of Environmental Toxicology and Aquatic Ecology,
Ghent University, J. Plateaustraat 22, B-9000 Gent, Belgium
E-mail: wim.gabriels@rug.ac.be

In this study, a data set of 360 sediment samples from unnavigable watercourses in Flanders (Belgium) was used to develop artificial neural network models predicting macrobenthos communities. For each sampling point, abundances of the present macroinvertebrate taxa as well as a large number of abiotic variables were available. The abiotic variables were used as input variables for the artificial neural networks predicting the macrobenthos community. The impact of the input variables on model performance was assessed in order to select a limited number of variables. For this purpose, the variables, which were least important for the predictive performance of the models, were eliminated via an iterative process. In this manner, simplified models with a good predictive power were developed for application in river assessment and management.

A new approach to determine the significance of the two-way interaction in an artificial neural network model

Gevrey M. [1], Dimopoulos Y.[2], Lek S.[1]

[1] CESAC, UMR 5576, CNRS-Univ. Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 4, France. E-mail: gevrey@cict.fr
[2] Technological Educational Institute of Kalamata, Dep. of Health & Welfare Unit Administration, Antikalamos, 24100 Kalamata, Greece

The major shortcoming of artificial neural networks models, called the "Black-Box" model, is the difficulty to interpret the knowledge contained in the trained network. Several methods, commonly called sensitivity analysis, have been previously proposed in order to overcome this disadvantage. These methods allow to identify the relative importance of the input variables on the output. However, the change of the output is not the result of only one factor but of a pair of input acting in concert. Improving an existing method called "PaD", using PArtial Derivatives, previously implemented to study the contribution of each input variables independently, the "two-way" interactions have been studied ("PaD2" method).
This new method was tested on the prediction of the density of brown trout spawning redds using habitat characteristics. Sampling was done at twenty-nine stations, distributed on 6 Pyrenean rivers, subdivided into 205 morphodynamic units. Ten environmental features have been measured (i.e. Wetted width, Area with suitable spawning gravel for trout per linear meter of river, Surface velocity, Water gradient, Flow/width, Mean depth, Standard deviation of depth, Bottom velocity, Standard deviation of bottom velocity, Mean speed/mean depth). We proposed first to use a backpropagation artificial neural network to predict the density of trout redds per linear meter of streambed according to the 10 variables and then to apply the PaD2 method to study the two-way interactions.
In the previous study using the one-way interaction, two input variables were found to be the most important and the others had not significantly different contributions. Using the two-way interactions, the most important interaction are the relationships between the two most important variables of the previous study but also the relationship between one of the most important variable of the previous study and one of the other which were not significantly differenciated. This new method appears interesting to determine all the causes leading to a change of the output but also to better differenciate all the contributions. It would be possible to use its in management programs.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Use of artificial neural networks and diatom assemblages to predict rivers water quality

Gevrey M.[1], Rimet F.[2], Park Y.S.[1], Giraudel J.L.[1], Ector L.[2], Lek S.[1]

1 CNRS-UMR 5576, CESAC – Bât. 4R3, Univ. Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex, France, E-mail: gevrey@cict.fr
2 Centre de Recherche Public - Gabriel Lippmann, CREBS, 162a avenue de la Faïencerie, L-1511 Luxembourg, Grand Duchy of Luxembourg

The aim of this study is to use epilithic diatom communities as a basis for water quality assessments. The available data from Luxembourg brooks including the species composition and the environmental features constitute a large data matrix. Usually, the common statistical tools oblige to restrict the description of a complex community structure by one single attribute such as species richness or diversity, neglecting thus the fact that much valuable information is lost. The present study develops method by combining two procedures considering a community matrix to define and predict the community structure.
Two types of artificial neural networks were associated: (i) the self organizing mapping (SOM) to give a classification of the samples according to their species similarities that is to obtain as much assemblages as SOM cells of the map; (ii) the backpropagation networks (BPN) to predict these assemblages by the way of the coordinates of each sample in the SOM map according to the environmental characteristics.
The diatom species composition was used to train the SOM to obtain 12 representative diatom communities, relating to different species composition, which correspond to different water environmental characteristics. Then, the BPN was applied to predict these communities using at the input the environmental characteristics of each sample and at the output the spatial coordinates (X and Y) of the cell centres of the SOM map identified as diatoms communities. A comparison of the observed and estimated samples positions were done to evaluate the performance of the BPN, using the correlation coefficients for all the data and using for each cell two indices: (i) the mean square errors (MSE), (ii) the proportion of well-predicted samples. Correlation coefficients were 0.93 for X and 0.94 for Y. The MSE values of 12-cells of the map varied from 0.47 to 1.77 and the proportion of well-predicted samples from 37.5% to 92.86%. In conclusion, the diatom communities correspond to a known state of the environment. The sensitivity of the communities to a change of the environment highlights the importance to study the community prediction for water quality assessment. The diatom communities prediction approach had the potential to evaluate the actual state of the ecosystem, allowing to develop monitoring tools considering the structure of the community.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

The structuring index: a tool for analysing self-organizing maps

Giraudel J.L., Lek S.

CESAC, UMR 5576, CNRS-Univ. Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex, France.
E-mail: giraudel@montesquieu.u-bordeaux.fr

The Kohonen Self-Organizing Map (SOM) is one of the most well-known neural network with unsupervised learning rules; it performs a topology-preserving projection of the data space onto a regular two-dimensional space. So, SOM is a powerful tool for data mining and SOM can be recommended for studying large, high-dimensional data sets.
After the learning, a lot of methods can be used to visualise the SOM. Showing the values of one variable in each map unit, the component plane visualisation allows to show the part of each variable in the total organisation of the map but also in some areas of this one. However, with high-dimensional data, a lot of maps are to be considered and the task can become fastidious and even quite impossible to be achieved. The purpose of this paper is to propose a computational method to determine the most relevant variables for structuring the obtained map.
The Structuring Index (SI) is computed for each variable of the dataset, using the components of the codebook vectors available after the learning process: the higher the value of the SI, the more relevant the variable. Only keeping the variables with highest values of the SI, a new map is obtained, very similar with the previous one. By this way, SI can provide an efficient tool for pre-processing data, for instance in order to reduce the number of variables used in the input layer of a multilayer artificial neural network.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Development and assessment of fuzzy logic models predicting aquatic macroinvertebrate taxa in the Zwalm catchment

Peter L.M. Goethals, Veronique Adriaenssens, Bernard De Baets & Niels De Pauw

Laboratory of Environmental Toxicology and Aquatic Ecology,
Ghent University, J. Plateaustraat 22, B-9000 Gent, Belgium
E-mail: peter.goethals@rug.ac.be

Over a two year period, 120 samples were collected at 60 different sites in the Zwalm catchment. Fifteen environmental variables such as temperature, percentage of dissolved oxygen, water depth, stream velocity, presence/absence of hollow beds, … were measured at each site, as well as the abundance of the aquatic macroinvertebrate taxa. Fuzzy predictive models were developed based on expert knowledge from literature about the habitat preferences of the collected 52 macroinvertebrate taxa. The models were validated and optimized using the measurement data collected at the 60 sites. The major aim of these models was to obtain useful predictions that can be used for decision support on mitigation strategies of the disturbed stream ecosystems in the Zwalm catchment.

Identification and prediction of diatom assemblages in rivers accross a range of environmental conditions in Europe: case study of Belgium

Gosselain V. [1], Campeau S. [1], Gevrey M. [2] and Fauville C. [1]

[1] Laboratory of Freshwater Ecology (LFE), FUNDP, Namur, Belgium
Email: Veronique.Gosselain@fundp.ac.be
[2] CNRS-UMR 5576, CESAC – Bât. 4R3, Univ. Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex, France

Self Organizing Maps and Back Propagation are used to identify and predict benthic algal communities in Wallonian rivers (Belgium). Classical methods are useful tools to assess the quality of data matrices and to identify environmental variables driving the algal community changes. A first attempt is made to identify the contribution of environmental variables to the algal community changes. The relevance of the expression of records as relative biomass, compared to relative abundances, is also discussed.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Phytoplankton primary production in Chesapeake Bay: a comparison between neural networks and other models

Lawrence W. Harding., Jr.[1,2] and Michele Scardi[3]

[1] Horn Point Laboratory, University of Maryland, Center for Environmental Science, Box 775, Cambridge, Maryland 21613, U.S.A., E-mail: larry@kestrel.umd.edu
[2] Maryland Sea Grant College, 0112 Skinner Hall, University of Maryland, College Park, Maryland 20742, U.S.A.
[3] Dept. of Biology, University of Rome "Tor Vergata", Via della Ricerca Scientifica, 00133 Roma, Italy, E-mail: mscardi@mclink.it

An extensive data set about phytoplankton primary productivity has been collected since 1982 on a series of cruises in Chesapeake Bay, the largest bay on the Atlantic coast of the United States. On the basis of this data set several models have been developed to obtain estimates for phytoplankton primary production using environmental variables (e.g. light, temperature, etc.) and phytoplankton biomass as predictive variables.
Both conventional techniques and artificial neural networks have been used in order to achieve this goal. Even though the results provided by these approaches are comparable in terms of accuracy, the underlying modeling strategies are quite different. In fact, even when an empirical paradigm is adopted, conventional models are inspired by some form of knowledge about the processes to be modeled. On the other hand, neural network models do not require previous knowledge and (in theory) they are able to reproduce whichever function to an arbitrary degree of accuracy.
The role of this difference in the modeling activities will be pointed out, and several comparisons among models will be presented. Finally, some perspectives in the optimization of neural network models of phytoplankton primary production will be outlined.

An application of artificial neural networks to carbon fluxes in three boreal streams

Maria Holmberg[1], Martin Forsius[1], Michael Starr[2]

[1]Finnish Environment Institute, P.O. Box 140, FIN-00251 Helsinki, Finland
E-mail: maria.holmberg@vyh.fi
[2]Finnish Forest Research Institute, P.O. Box 18, FIN-01301 Vantaa, Finland

We have used an artificial neural network to model daily total organic carbon (TOC) concentrations and fluxes in streamwater from climate variables. The streams drain two forested catchments, which belong to the network of Integrated Monitoring sites, located in southern and eastern Finland. In the period 1990 to 1999, observed TOC concentrations were in the range 2 – 60 mg CL^{-1} with mean runoff between 10 and 15 $Ls^{-1}km^{-2}$ for the three streams.

The processes that control runoff water chemistry and changes in response to climate change are complex and their full range of operation is difficult to grasp with process-oriented models. Although such models exist and become more elaborate, few models encompass all relevant processes and are difficult to apply to sites other than those they were developed within.

An artificial neural network consisting of 8 input variables, one hidden layer with 5 nodes and one output variable was trained with the back-propagation algorithm to estimate the daily export load of TOC in streamwater. Daily air temperature and precipitation observations were used as input variables, together with average values of these variables calculated for the previous 3, 10 and 30 days. Although the simulation did not catch all extreme values, it reproduced most of the dynamics in the observations. The artificial neural network approach appears to offer a useful method for creating black-box models of water quality and carbon fluxes in cases where the involved processes are too complex to simulate directly.

Knowledge discovery in two Australian stream systems by means of self-organizing maps and evolved rules

Nelli Horrigan, Jason Bobbin and Friedrich Recknagel

Adelaide University, Department of Soil & Water
PMB 1, 5064 Glen Osmond, Australia.
E-mail: nelli.horrigan@adelaide.edu.au

In this paper Kohonen neural networks and evolutionary algorithms were used for the classification and rule-based prediction of stream macroinvertebrate assemblages in statewide stream databases of Queensland and Victoria. The Queensland database contained presence and absence data of 157 macroinvertebrate families at 2054 sampling sites, and 39 environmental predictor variables. The Victoria database contained abundance data of 128 macroinvertebrate families at 675 sampling sites, and 20 environmental predictor variables. Firstly SOM were applied to localize ecological regions in the two stream systems identified by distinctive clusters of macroinvertebrate communities. Secondly we applied evolutionary algorithms in order to describe the ecological regions by predictive rule sets. As a result each macroinvertebrate cluster discovered by SOM within the databases is characterized by a specific rule set. Each rule set reveals specific environmental conditions determining the spatial occurrence of macroinvertebrate assemblages in the landscapes, which allow underlying hypotheses to be examined. While rule sets for the Queensland streams relate to presence and absence of macroinvertebrates, rule sets for Victoria streams predict macroinvertebrate abundance. Practical implications of the approach are discussed in the context of stream assessment and management.

Information-based models of complex ecological processes

Peter T. Hraber

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501 USA
Email: pth@santafe.edu

An essential technique to understand and predict the dynamics of natural systems is the construction of models. Scientific models extract information from data to infer relationships between empirical patterns and the processes that produce them. Though the mechanisms by which different models process information vary according to the model category and application domain, similar challenges confront all such activities.

This talk will compare alternative approaches to modelling complex ecological processes. The focus will be on how each modeling approach processes information to establish a representation of the regularities therein. Examples will be drawn from a range of ecological applications, including agent-based models of community assembly, mutual information analyses of coevolving organisms, and application of information-based model selection to a study of climate change effects. Contributions and limitations of each approach will be discussed in reference to the state of the art of constructing and validating scientific models.

Fish diversity patterns in rivers of the Garonne basin (France)

A.A. Ibarra[1], P. Lim[1], A. Belaud[1], J. Moreau[1], F. Dauba[1], Y.-S. Park[2], M. Gevrey[2], S. Lek[2]

[1] Ecole Nationale Supérieure Agronomique, 1, ave de l'Agrobiopole, BP 107 Auzeville-Tolosane, 31326, France. E-mail: plim@ensat.fr
[2] CESAC, UMR 5576, CNRS-Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 4, France. E-mail: lek@cict.fr

Ecosystem health is seen as a key feature in water quality management. Therefore, we need to understand aquatic community ecology for giving sound advice to decision makers. The Garonne basin represents an interesting example for fish community ecology because it has different features than other Western European rivers.
The objectives of this study are: a) to look into the structure of fish assemblages in the Garonne basin, b) to test the contribution of environmental variables for explaining fish community structure, and c) To assess the environmental quality of human-influenced rivers according to the European Water Framework.
Species richness data were collected between at 108 sampling sites spread over the Garonne basin, from 1986 to 1996, by electrofishing during low-flow periods; wading in streams and small rivers, and seining in larger rivers. Five environmental descriptors were selected for explaining local fish assemblages: altitude (m), distance from the river source (km), surface of catchment area (km^2), average annual water temperature (°C), and average annual water flow (m^3/s). Modelling was carried out by the Kohonen's self organising map (SOM) and back-propagation network (BPN).
The main result from the SOM is an upstream-transitional-downstream longitudinal pattern in both tributaries and the main channel of the Garonne basin. With respect to the BPN model, surface of catchment area and annual average water flow were the most important environmental descriptors of fish guilds composition. Both variables imply strong human influence (i.e. land-use and flow regulation) on certain species which are of interest to environmental managers.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Cyanobacterial dynamics in the lower Nakdong River (S. Korea): pattern recognition of genus shift using an unsupervised artificial neural network

Kwang-Seuk Jeong [*], Gea-Jae Joo, Dong-Kyun Kim, and Kyong Ha

Dept. of Biology, Pusan National Univ., Jang-Jeon Dong, Gum-Jeong Gu, Busan 609-735, S. Korea. E-mail: pow5150@hananet.net

An unsupervised artificial neural network, Self-Organizing Map (SOM), was applied to the abundance data of bluegreens from a river-reservoir system. From 1994 to 1998, the lower Nakdong River presented eutrophic status, and dynamic inter-annual variability was observed. Five years' genus data of dominant cyanobacteria (*Microcystis*, *Anabaena*, and *Oscillatoria*) were fed to the SOM. To reduce the unexpected error of data normalization, all values were transformed into logarithm scale. The algorithm clearly clustered the dataset into seven groups (quantization error, 5.443; topographic error, 0.005) which mainly consisted of the presence and absence of each genus. From the results, the changes of cyanobacteria seemed to be caused by water temperature and N/P ratio. This could emphasize the importance of heat energy of water body and nutrient dynamics by means of a non-linear method. The result could reflect the applicability of SOM to a river-reservoir hybrid system, and could be challenged for the purpose of pattern-recognizing plankton dynamics.

Evolving neural network algorithm to freshwater ecological modelling: predicting phytoplankton blooms in the lower Nakdong River (S. Korea)

Kwang-Seuk Jeong[1], Gea-Jae Joo, Hyun-Woo Kim[2], and Ga-Ik Cho[1]

[1]Dept. of Biology, Pusan National Univ., Jang-Jeon Dong, Gum-Jeong Gu, Busan 609-735, S. Korea. E-mail: pow5150@hananet.net
[2]Dept. of Environmental Education, Sunchon National Univ., Maegok-Dong, Sunchon 540-742, S. Korea.

In this study, a hybrid computational algorithm was applied to freshwater ecological modelling. Evolving neural network algorithm was adopted to develop a time-series predictive model in a river-reservoir hybrid system. Five years' (1994-1998) limnological study revealed that the lower Nakdong River resembles reservoir-like ecosystem due to intensive flow regulation, and eutrophic state is persisted (five year average chl. *a*, 50.2 □ l^{-1}) Even though there was inter-annual variability, severe algal bloom events occurred in summer and winter. A time-series Artificial Neural Network (ANN) was constructed to predict changes of two important bloom-forming algal species: *Microcystis aeruginosa* and *Stephanodiscus hantzschii*. Different from traditional challenges of ANN modelling, this model was evolved by means of Genetic Algorithm (GA) to find the suitable topology of network. The evolved model thereafter was trained with Backpropagation (BP) algorithm. Input-output dataset was selected from the data of 1995 to 1998 for training and tested on the independent year 1994 which showed severe algal proliferation. The network was well trained (RMSE lower than 0.001), and the predictability was as accurate as the ANN model previously developed. The result of sensitivity analysis showed that an evolving neural network is a good tool for developing algal dynamics model. By reducing time cost as well as efforts on performing network construction, this algorithm can be a feasible alternative to traditional modelling techniques.

How many Eco-sub-disciplines do we need?

S. E. Jørgensen

DFH, University park 2, Copenhagen Ø, Denmark. E-mail sej@dfh.dk

A new eco-sub-discipline was born at the meeting in Adelaide November 2000: Ecological Informatics or shorter Eco-informatics. It could be defined as the science about how to provide ecological information. It is naturally to discuss what this definition actually covers and where this new sub-discipline has overlap with other ecological sub-disciplines. During the last 30-40 years several ecological sub-disciplines have been developed because there was a need, in most cases due to the growing interest for environmental problems. These sub-disciplines are defined below and it is indicated approximately when the sub-discipline has emerged and been accepted as sub-discipline.
Ecological informatics covers the following activities according to the definition given above:

- ecological information on internet;
- ecological data bases and how to erect and develop them;
- ecological statistics;
- how to use models (which types?) to provide ecological information;
- parameter estimation methods;
- which computational techniques to apply in ecology?
- uncertainty and ecological data;
- ???

The overlap to the other sub-disciplines is obvious and can be derived from the above mentioned topics. It will be illustrated by use of figures and the border line between two sub-disciplines will be discussed.

Eco-Sub_Disciplines

Ecological Sub-discipline	Definition	Was accepted as sub-discipline
System Ecology:	the ecology of ecosystems	1960-1965
Evolutionary Ecology:	the ecology of the evolution	1966-1970
Ecological Modelling:	the science on how to model ecosystems	1968-1970
Ecological Engineering:	engineering of ecosystems	1975-1978
Ecological Economics:	the science integrating ecology and economy	1982-1985
Ecosystem Health:	the science on how to assess the health of Ecosystems	1988
Ecological Informatics:	the science of ecological information	2000

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Neural network modelling of freshwater fish and macro-crustacean assemblages for biological assessment in New Zealand

Michael K. Joy and Russell G. Death

Institute of Natural Resources-Ecology, Massey University, Private Bag 11 222, Palmerston North, New Zealand, E-mail: mikejoy@clear.net.nz

The comparison between the number of taxa observed and that expected in the absence of human impact is an easily understood and ecologically meaningful measure of biological quality. This comparison has been successfully applied to macroinvertebrates to assess the biological quality of flowing water sites using with the River Invertebrate and Classification System (RIVPACS) and its derivatives. We developed a methodology based on the comparison between observed and expected freshwater fish and macro-crustacean assemblages to assess the biological quality of stream sites in the Auckland region, New Zealand. Freshwater fish and macro-crustacean assemblages were sampled at 118 least impacted (reference) sites and individual artificial neural network (ANN) models developed using environmental measures taken at those sites to predict the presence or absence of the 12 most common species. The ANN models used only the environmental variables least likely to be influenced by human impacts and outputs from models for all species were combined so that the predictions were for the assemblage expected in the absence of human impacts. The output from the model is the ratio of observed to expected species number so that the biological quality of a site can be assessed under the assumption that human impacts reduce species richness.

Learning Metrics

Samuel Kaski

Neural Networks Research Centre
Helsinki University of Technology
P.O. Box 9800, FIN-02015 HUT, Finland
E-mail: Samuel.Kaski@hut.fi

Visualization and clustering of multivariate data is usually based on unsupervised learning that suffers from the garbage-in, garbage-out problem: The results are not useful unless the data variables and the metric of the data space are selected carefully. This implies that any successful application of unsupervised learning must have been implicitly supervised by proper selection of the model, the variables, and the metric.
The goal of learning metrics research is to automate part of this implicit manual supervision by learning the metric from data. It is assumed that variation of the data is important only to the extent it causes variation in auxiliary data which is available paired to the primary data. An example of suitable auxiliary data, in analysis of the financial state of companies, is indication of whether the company has gone bankrupt or not. A metric measuring the important differences in primary data is defined in terms of the Fisher information matrix
and learned from paired samples of primary and auxiliary data. Learning approximations to the metric and using them in data analysis has been coined the learning metrics principle. The principle is applicable whenever suitable auxiliary data is available during learning; so far it has been applied to clustering and self-organizing maps of text documents, gene expression, and bankruptcy data.

Implementation of Wavelets and Artificial Neural Networks to Pattern Recognition of Response Behaviors of Chironomids (Chironomidae: Diptera) for Water Quality Monitoring

Cheol Ki Kim[2], Inn-Sil Kwak[1], Eui-Young Cha[2] and Tae-Soo Chon[1]

[1]Division of Biological Sciences, Pusan National University, 30 Changjeon-dong, Keumjeong-ku, 609-735, Pusan, Korea
E-mail: kck@harmony.cs.pusan.ac.kr
[2]Division of Computer Science, Pusan National University, 30 Changjeon-dong, Keumjeong-ku, 609-735, Pusan, Korea
E-mail: tschon@pusan.ac.kr

Using an automatic tracking system, behavior of *Chironomus flaviplumus* was observed in semi-natural conditions in response to sub-lethal treatments of carbofuran, an anticholinesterase insecticide. The fourth instar larvae were placed in an observation cage (6cm X 7cm X 2.5cm) under the light condition of 10LL: 14DD. Sublethal concentration of carbofuran (0.1*mg/l*) was exposed to test specimens individually in observation cages, and a tracking system was devised to record the movement tracks of specimens continuously for four days (2 days; before treatment, 2 days; after treatment). After treatments of the insecticide, activities of test specimens were depressed in general, and the movement pattern appearing as the shape of "compressed zig-zag" was more frequently observed, along with other irregular behaviors. In order to characterize behavioral patterns of specimens, wavelet analysis was conducted on the spatial-temporal data of the locomotive tracks. Basis functions including Daubechie's series were implemented to the movement tracks and Discrete Wavelet Transforms (DWT) mappings were produced to extract parameters of the movement data to characterize different patterns of response behaviors. The extracted parameters were subsequently provided to artificial neural networks (e.g., multiplayer perceptron) as input, and were trained to represent different patterns of the movement tracks before and after treatments of the insecticide. This combined model of wavelets and artificial neural networks was able to point out the time occurrence of different patterns treated with the insecticide in the input data, and could be an alternative tool for automatically detecting presences of toxic chemicals for water quality monitoring on the real time basis.

Application of information theory for ecological interpretation of biological data

Markus Knoflacher

Systems Research, Austrian Research Center, A-2444 Seibersdorf
E-mail: markus.knoflacher@arcs.ac.at

Species distribution within ecosystems is determined by the individual eco-physiological potential. Based on the knowledge of the specific eco-physiological characteristics of species it is theoretically possible to identify abiotic conditions at a selected sample site. However, the occurrence of a particular species on a site is restricted by interacting abiotic and biotic factors. Consequently, the information about abiotic conditions given by occurring species is biased by interspecific relationships. Additionally the spatial and temporal variability of abiotic factors at different scales have to be considered, which influences the probability of species occurrence on a specific place and at the specific date. Information about abiotic condition, gathered by species analysis is therefore always incomplete and biased. This raises the question how the uncertainties of information can be reduced. If each species found on a site is interpreted as a partial sequence of information, the task will be to identify the proper number of partial sequences and to bring it to an structurised order. Approaches for solutions of the problem in macroinvertebrates will be presented in the paper by application of information theory.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Forecast Estimation In A Soils

Tatiana Koroleva

Institute of Ecology of TAS, Russia
Daurskaya st., 28, Kazan, 420089, Russia
E-mail: Tata.Koroleva@ksu.ru

Recently problem of accumulation of heavy metals in an atmosphere, soils both waters in the natural and antropogently-modified landscapes is characteristic for many countries of the world. The soil can be considered as the integrated indicator of long-term process of pollution of an environment giving representation about quality of environments connected with soil - air and waters. The various analytical methods are applied to research of the spatial concentration of heavy metals in soils. At the same time it is necessary to carry out a polyelement estimation of soil and to take into account variety of the factors at absence of the information a priory on probable distribution of heavy metals in a soils, which, is rather complex systems. Such necessity dictated development of a method of a grouping soil given for a forecast estimation of the concentration of heavy metals in soils on a basis of the cluster analysis. As model object were used of soils of Predvolgie of Tatarstan Republic.

The forecast estimation of the concentration of heavy metals in soils is carried out by application of a method of a grouping of soil date by concentration of heavy metals, on the basis of linguistic clusterisation, consecutive estimation procedure and fuzzy sets.

With the purpose of reception of estimated concentration of heavy metals in soils of Predvolgie the data on the basic physical and chemical parameters of soils, concentration of six heavy metals were used: lead, copper, nickel, zinc, chromium, manganese, and also rank estimations of an ecological condition of territory and total emissions of polluting substances in an atmosphere on a card of an ecological situation of Tatarstan Republic. The sharing of quantitative and qualitative parameters is characteristic for the analysis of a condition of soils for the complex approach to an estimation of territory.

For a grouping of soil data by content of heavy metals in soils the program complex on the programming language Visual FoxPro 5.0 for Windows is developed. Result of work of a program complex are the tables, which contain the items of information on functional value of quality of splitting of an initial matrix for different groupings of parameters, quantity of clusters and groups of parameters, average and standard deviation for each parameter of cluster.

With the help of the offered method the connection of the concentration of the total and mobile forms of the investigated heavy metals with clay and silt fraction in soil: concentration them in humus and absorbed bases are revealed; are received intervals of forecast of the concentration of heavy metals in soils, which are checked up on soils of two region of Tatarstan Republic. The high concurrence of experimental and settlement value is achieved.

The researchers in Tatarstan Republic have not enough the data of experimental and field researches necessary for construction of soil cards and an estimation of an ecological condition for territory. The offered method allow to solve the above mentioned problem.

Patterning on Community Dynamics of Benthic Macroinvertebrates in Streams by Using the Self-Organizing Mapping

Inn-Sil Kwak, Mi-Young Song, Young-Seuk Park, Hyun-Duk Cho, Eui Young Cha[1] and Tae-Soo Chon

Division of Biological Sciences, Pusan National University, 30 Changjeon-dong, Keumjeong-ku, 609-735, Pusan Korea
[1]Division of Computer Science, Pusan National University, 30 Changjeon-dong, Keumjeong-ku, 609-735, Pusan Korea
E-mail: tschon@pusan.ac.kr

Benthic macroinvertebrates have been good indicator groups for monitoring water quality in streams. Since community develops in a sequence on time domain, either in stressful or in favorable conditions, it is important to pattern community changes in response to environmental impacts. Pattering on community changes on time series is necessary for having comprehensive understanding on the target stream ecosystem, where communities are easily affected by disturbances caused by various natural and anthropogenic agents. However, it has been not an easy task to classify changes in communities, which are complex in spatial and time domains. An enhanced type of the Kohonen network was implemented to organize community changes in time series in this study. The field data used as input for training were monthly changes in density in the selected taxa of benthic macroinvertebrates collected in the Yangjae Stream in Korea from April 1996 to March 2000. The organized map by the networks was in accordance with ecological consequences in general, and the trained network was able to extract dynamic aspects of community changes. Features of ecological impacts on community changes on time series were revealed effectively through the self-organizing mapping, and further explanation was possible based on the patterning by the artificial networks.

Collective phenomena in ecological time series

Holger Lange

BITOEK, University of Bayreuth
Dr.-Hans-Frisch-Str. 1-3 D-95440 Bayreuth DE
E-mail: holger.lange@bitoek.uni-bayreuth.de

Using Singular System Analysis, Wavelets and a variety of other non-linear time series analysis methods, we demonstrate that a number of collective or synchronous dynamical structures is present in long-term time series from terrestrial ecosystems. Specifically for runoff from rivers and streams, where data with daily resolution exist for up to a century, interannual and decadal regular patterns are simultaneously present within a larger geographical region, indicating that long-range forces have significant influence. This has implications for global change and climate research as well as flood prediction. We elucidate the possible role of several prominent indicators (like sunspot numbers, North Atlantic Oscillation etc.) in this context.

Extracting information from noisy survey data on temporal change in vegetation following disturbance

M.G. Le Duc[1], R.J. Pakeman[2], R.H. Marrs[1]

1 Applied Vegetation Dynamics Laboratory, School of Biological Sciences, University of Liverpool, PO Box 147, Liverpool, L69 3GS, UK, E-mail: mled@liverpool.ac.uk
2 Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK

An ecological survey of upland and marginal vegetation in Great Britain, carried out to examine the effect of herbicide spraying for control of *Pteridium aquilinum*, produced a large and noisy dataset. This has been examined for the temporal impact on certain classifications of vegetation using a number of techniques, including multivariate analysis and a phylogenetic clustering technique. By means of calculated standard-deviational ellipse parameters, it was shown that at least three of the vegetation classes present showed different responses in time.

Tool for Predicting Aquatic Ecosystem Quality using Artificial Neural Networks (EU PAEQANN project)

Lek S.[1], Bretin L.P. [1], Coste M.[2], Descy J.P.[3], Ector L.[4], Gevrey M. [1], Giraudel J.L. [1], Knoflacher M.[5], Jørgensen S.E.[6], Park Y.S. [1], Scardi M.[7], Verdonschot P.[8]

[1]CESAC, University Paul Sabatier, Toulouse, France.
[2]CEMAGREF, Bordeaux, France.
[3]Laboratory of Freshwater Ecology (LFE), FUNDP, Namur, Belgium.
[4]CRP – Gabriel Lippmann, CREBS, Luxembourg.
[5]Austrian Research Center Seibersdorf, Austria.
[6]DFH, Environmental Chemistry, Univ. Copenhagen, Denmark.
[7]Dep. Zoology, Univ. Rome "Tor Vergata", Italy.
[8]Alterra, Green World Research, team of Freshwater Ecology, Wageningen, The Netherlands.

The aim of the PAEQANN project is to develop methodologies which allow: i) to provide predictive tools that can be easily applied to define the most effective policies and institutional arrangements for resource management; ii) to apply the most effective and innovative techniques (mainly artificial neural networks) to identify problems in ecosystem functioning, resulting in ecosystem degradation from human impacts, and to model relevant biological resources; iii) to fully exploit existing information, reducing the amount of field work (that is both expensive and time consuming) needed in order to assess freshwater ecosystem health; iv) to explore specific actions to be taken for restoration of ecosystem integrity; and v) to promote collaboration among scientists of different interested countries and research fields, encouraging collaboration and dissemination of results and techniques.
The data-processing user-friendly interface and the associated algorithms used in this project will be presented. This tool allows firstly to pattern the community structure of diatoms, macro-invertebrates and fishes in several Europe countries, and secondly to predict these communities using a set of environmental variables.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

PAEQANN project: Predicting Aquatic Ecosystem Quality using Artificial Neural Networks. Impact of Environmental characteristics on the Structure of Aquatic Communities (Algae, Benthic and Fish Fauna).

Lek S.[1], Coste M.[2], Descy J.P.[3], Ector L.[4], Knoflacher M.[5], Jorgensen S.E.[6], Scardi M.[7], Verdonschot P.[8]

[1]CESAC, University Paul Sabatier, Toulouse, France. Email: lek@cict.fr
[2]CEMAGREF, Bordeaux, France
[3]Laboratory of Freshwater Ecology (LFE), FUNDP, Namur, Belgium
[4]CRP – Gabriel Lippmann, CREBS, Luxembourg
[5]Austrian Research Center Seibersdorf, Austria
[6]DFH, Environmental Chemistry, Univ. Copenhagen, Denmark
[7]Dep. Zoology, Univ. Bari, Italy
[8]Alterra, Green World Research, team of Freshwater Ecology, Wageningen, The Netherlands.

The aim of this project is to develop methodologies which allow: **i**) to provide predictive tools that can be easily applied to define the most effective policies and institutional arrangements for resource management; **ii**) to apply the most effective and innovative techniques (mainly goal function and artificial neural networks) to identify problems in ecosystem functioning, resulting in ecosystem degradation from human impact, and to model relevant biological resources; **iii**) to fully exploit existing information, reducing the amount of field work (that is both expensive and time consuming) that is needed in order to assess freshwater ecosystems health; **iv**) to explore specific actions to be taken for restoration of ecosystem integrity; and **v**) to promote collaboration among scientists of different interested countries and research fields, encouraging collaboration and dissemination of results and techniques.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Comparing of classical and modern modelling techniques to predict macroinvertebrate community in the province of Overijssel (The Netherlands)

Lek S.[1], Gevrey M. [1], Giraudel J.L. [1], Park Y.S. [1], Scardi M.[2], Verdonschot P.[3]

[1]CESAC, UMR 5576, CNRS-Univ. Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex, France. E-mail: lek@cict.fr
[2]Dept. of Biology, Univ. of Rome "Tor Vergata", Via della Ricerca Scientifica, 00133 Rome, Italy, E-mail: mscardi@mclink.it
[3]ALTERRA-FE, Green World Research, team of Freshwater Ecology, Droevendaal, sesteeg 3a, 6700AA, The Netherlands. E-mail: pfmverdonschot@hetnet.nl

The complexity of the ecological systems often results in complex relations between the variables, justifying the uses of multiple modelling techniques. This paper aims to apply three traditional statistical techniques: GLM (Generalised Linear Models), GAM (Generalised Additive Models) and CART (Classification And Regression Tree), and two modern methods of artificial neural networks (Self-Organizing Map, SOM and backpropagation, BP) which were applied to model the EKO database as a part of the PAEQANN EU-project.
The EKO database consists of samples collected from 664 sites situated in the province of Overijssel (The Netherlands). The objective was to capture the majority of the macro-invertebrate species to study their relative abundances present at a given site according to the characteristics of the environment. At each site, major habitats were selected over a 10 to 30 m long stretch of the water body and were sampled with the same sampling effort. The sampling effort was thus standardised for each site. A total of 23 environmental variables were used to explain 40 cenotypes of macro-invertebrate species collected in all 664 study sites.
The modelling processes included 3 steps:
Firstly, the SOM was used to pattern the macro-invertebrate data to define the community structure among 40 cenotypes in the original database;
Secondly, we used 23 environmental variables to explain the variability of several groups and sub-groups of macroinvertebrates. The predictive models (GLM, GAM, CART and BP) were used to predict the macroinvertebrate community structure. The predictive powers, i.e. the percentages of explanatory variance of several macroinvertebrate cenotypes varied from 50% to 90%.
Thirdly, the roles of explanatory variables were discussed, by the way of the sensitivity analysis. The procedure allows to select relevant environmental variables which were used to build the final predictive and illustrative models in terms of macroinvertebrate community structure.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Importance of the use of multivariate analyses (AFC and ACPN) in structure studies of macroinvertebrates in Zegzel-Cherraa river, Eastern Morocco

A. Maamri

Laboratory of Hydrobiologie & General Ecology,
Faculty of Science, University Mohamed I, B.P. 524, OUJDA-MAROC.
E-mail: maamri@sciences.univ-oujda.ac.ma

The longitudinal distribution of benthic fauna was undertaken in two wadis (rivers) located in the prolongation one of the other: one permanent (Zegzel), the other temporary one (Cherraa). The study of these two rivers was based on faunistic taking away and physicochemical analyses carried out in 5 stations chosen to characterize the longitudinal zonation and fauna diversity of the temporary rivers compared to that of the permanent rivers. The data file was subjected to various multivariate analyses (AFC and ACPN) to facilitate the exploitation of all data analyses.

The longitudinal structure of the trophic groups was closely related to the bordering vegetation (good correlation). Indeed, in Zegzel, it varied since the dominant shredders in the upstream (abundance of the bordering vegetation) to the collectors of fine particles in the lower course (poverty or absence of vegetable cover) thus answering the theory of River Continuum Concept (RCC). The temporary river, on the contrary, does not obey this concept, partly because of the installation of terrestrial plants in the bed him even of the river. The longitudinal distribution of the trophic groups in this system corresponded better to the theory of " River Mosaïc " which rather cuts out the system in series of discrete parts having distinct limits than in a progressive gradient or continuum.

Spatial Subgroup Discovery Applied to the Analysis of Vegetation Data

Michael May and Lemonia Ragia

Fraunhofer Institut für Autonome Intelligente Systeme, Knowledge Discovery Team
E-mail: michael.may@ais.fhg.de
E-mail: lemonia.ragia@ais.fhg.de

This paper explores the application of Spatial Data Mining, the partially automated search for hidden patterns in georeferenced databases, to the analysis of vegetation data. *Phytosociological relevès* (vegetation records) contain information about which plants occur together at a certain site. They are the basis for classifying plant communities and are used for determining ecological conditions favourable for the existence of a community. They are of great practical importance in various fields, e.g. environmental impact studies, landscape-, agricultural- and forestry-planning. Monitoring activities use them to detect site changes due to immission, drainage, or climatic change.
The vegetation records are georeferenced and associated with a set of environmental parameters, e.g. soil conditions. Yet plant growth is influenced not only by local conditions, but also by larger spatial configurations such as being close to a river, being far from the sea, being in a mountainous region, and so on. Knowledge about spatial relations between objects is traditionally encoded in geographical maps. In the last decades, this information is increasingly stored in digital format, using Geographic Information Systems (GIS).
We describe a method that automatically incorporates spatial information stored in a GIS into the hypothesis space of a data mining search, and show that this method can be usefully applied to the analysis of vegetation data. The challenge arises from the fact that standard data mining algorithms represent data in a matrix (single table) containing only atomic values. Yet geographic information as represented in modern GIS is typically stored in multiple relations, allowing for complex values (e.g. points, lines and polygons to represent spatial objects). The key to spatial data mining is to exploit spatial information inherent in the data by extending the representational capabilities of data mining algorithms. While traditional attribute-value based learning methods have difficulties in expressing topological features such as being_inside, adjacent_to etc. in a natural and general way, they can be easily expressed in first-order-logic. This makes *inductive logic programming* (ILP), which uses a first-order representation, a natural approach to spatial data mining.
Potential disadvantages of using ILP in a practical setting are concerns about scalability and embedding into existing IT environments, especially interfacing with data stored in a GIS. To avoid these problems we have translated an ILP-algorithm for multirelational subgroup discovery (Klösgen 1996, Wrobel 1997) to extended relational algebra and have adapted it to the spatial domain (Klösgen, May, submitted). Subgroup mining searches for *deviation patterns*. It is practically implemented in SQL, using a sufficient statistics approach for better scalability. This allows to carry out a search directly in a object-relational database such as Oracle. Geographical knowledge is stored as a set of tables, where each table is either explicitly or implicitly georeferenced and corresponds to a geographical *layer* (a set of geographical objects described by the same attributes). Operations defined on these objects allow to compute topological (overlap, adjacency, covering etc.; cf. Egenhofer 1991) and proximity queries. Object-relational databases are soon becoming the method of choice for storing geographical information, giving our approach great practical relevance. Another

benefit is that Data mining results can be visualized on a geographic map using a GIS, producing highly understandable results, which a central aim of data mining.
The paper discusses the general issues of applying data mining to biodiversity data, describes in the spatial subgroup mining algorithm and then focuses on the application to a multi-relational data set combining vegetation records, data on climate and soil conditions, and location of spatial objects like rivers, streets and cities.

Fisher Information and Dynamic Regime Changes in Ecological Systems

Audrey L. Mayer, Christopher W. Pawlowski, and Heriberto Cabezas

U.S. Environmental Protection Agency
Sustainable Environments Branch, Sustainable Technology Division
National Risk Management Research Laboratory
26 West Martin Luther King Drive Cincinnati, Ohio 45268 USA
E-mail: mayer.audrey@epamail.epa.gov
E-mail: cabezas.heriberto@epa.gov

The sustainable nature of particular dynamic regimes of ecosystems is an increasingly integral aspect of many ecological, economic, and social decisions. Sustainability usually refers to a human preference for one particular regime versus another, and whether that regime is relatively stable with regard to the human and natural perturbations exacted on the system. As ecosystems experience perturbations of varying regularity and intensity, they may either remain within the state space neighborhood of the current regime, or "flip" into the neighborhood of a regime with different characteristics. Previous research has identified the presence of stable states and transitions between them in several time series. We used this data to test the ability of an Information Theory-based index to differentiate between dynamic regimes and transitions between them.

Information theory has significantly advanced our ability to quantify the organizational complexity inherent in systems in spite of imperfect observations or 'signals' from the source system. Fisher Information (FI) is one of several metrics developed under the rubric of estimation theory. FI can be described in three ways: as a measure of the degree to which a parameter (or state of a system) can be estimated; as a measure of the relative amount of information that exists between different states of a system; and as a measure of the disorder or chaos of a system. Highly disordered, chaotic systems have a low probability of being observed in any one particular state, and therefore have low information. Conversely, systems that are more ordered and follow a regular or repeating trajectory have higher information. FI may be a very useful measure to apply to the state of the system in order to identify the degree to which a system is at risk of "flipping" into a different steady state.

We have developed an FI index for dynamic systems and applied it to a simple, two species Lotka-Volterra predator-prey model. As we increased the carrying capacity (size) of the system, FI decreased when the system entered a transient phase between stable equilibria. This result indicated that the index was sensitive to transitions or "flips" from one state to another. We also examined data collected from three ecological systems (of increasingly large spatial and temporal scale) that have demonstrated regime changes. In the Bering Strait/Pacific Ocean food web, regime flips in 1977 and 1989 occurred in many of the environmental and biological variables collected over a thirty-year time period. Two flips in western Africa, from an arid system to a humid system and back again, have been recorded in western Africa (in ocean sediment cores) in the past 25,000 years. Similar flips have occurred in Florida ecosystems (in pollen records) across the Atlantic. Both systems are influenced by climate conditions modulated by the oceanic thermohaline conveyor of the Northern Atlantic. Finally, over the past 160,000 years, the Earth's climate has fluctuated between warm and cold (interglacial and glacial) periods (as recorded in ice cores), which have demonstrated differing stability and persistence. All of these datasets are noisy, and reflect several to many cycles that are out of phase and operate over a range of timescales.

Our paper will investigate the degree to which FI is useful in not only distinguishing a system's transitions between regimes in past data, but whether it also can indicate when systems currently in a stable dynamic regime are entering a transition phase. Humans may be able to reverse behavior or inputs into the system to prevent the system's flip into a less-desirable steady state (or continue the behavior if the resultant steady state is desirable, such as in ecosystem restoration efforts), if systems entering these transition phases could be detected early enough.

A comparison of various fitting techniques for predicting yield for the Ubolratana reservoir (Thailand) from a time series data on catch and hydrological features

Jacques Moreau[1], Sovan Lek [2], Wattana Leelaprata[3], Boonsong Sricharoendham[3] and Maria Concepcion Villanueva[1]

[1]INP/ENSAT BP 107 Auzeville Tolosane 31326 Castanet Tolosan, France
E-mail: moreau@ensat.fr
[2]CESAC, UMR 5576, CNRS-UPS, 118 route de Narbonne 31062 Toulouse cedex 4, France.
E-mail: lek@cict.fr
[3]The Department of Fisheries of Thailand Kajutjak 10120 Bangkok Thailand
E-mail: boonsosr@fisheries.go.th

Actual catch (commercial catch and local consumption) have been made available by the Royal Thai Department of Fisheries for the Ubolratana reservoir, North East Thailand, since the impoundment of the lake in 1965. The data have been used to assess the possible relationship between the actual catch and morphometric parameters of the lake (maximum area, depth, shore line, difference between maximum and minimum area) and actual catch the year before. A comparison of the predictive power is carried out between multiple linear regression analysis, a generalised additive model, a regression tree analysis, an autoregression analysis and an artificial neural network. Results show the poorly predictive power in linear system (around 40% of explained variances), compared to the non-parametric and non-linear systems, essentially the artificial neural network (more than 85% of explained variances). The morphometric parameters which displays the maximum contribution are the difference between the maximum and minimum area the year before for both pelagic Clupeid catches and littoral catch targeting other species. In addition, the catch during the previous year and the maximum area also influence the clupeid catch and the littoral catch respectively. The ecological significance of these results in terms of spatial distribution of fish populations and fisheries management is discussed.

Lake Ladoga thermal database: design, opportunities and results

Naumenko M. A., Karetnikov S.G.

Limnology Institute, Russian Academy of Sciences,
Sevastyanov st.,9, St.-Petersburg, 196105, Russia
E-mail: naumenko@limno.org.ru

At the present stage of monitoring and research of limnological processes can not be effective without the carefully organized system of accumulation, storage and fast call of a necessary information and submission it in visual and accessible to the form. The harmonisation of Lake Ladoga ecological parameters is a very important procedure for the solution of a many limnological tasks, connected with the environment problems and its ecological modelling.
The best realisation of these circumstances can achieve using thematic databases.
Many of Lake Ladoga researchers (Homen, Molchanov, Tikhomirov) tried to describe a main thermal features of the largest European freshwater basin (volume 838 km^3, mean depth 47 m.) by thermal surveys conducted .
Main efforts of authors were directed on the collection of archive information of Lake Ladoga hydrophysical water parameters and making a thermal database. At present time the database is firstly created at Institute for Lake Research Russian Academy of Sciences. It includes the field measurements of water thermal characteristic and meteorological parameters for period from 1898 to 2001 years. This database consists of more then 150 thousand strings of water temperature with accompanying parameters, covered 75 years . Total data density equals about 175 measurements per cu. km. It is the very large database among word lakes. Morphometrical model of Lake Ladoga was produced earlier.
Various statistics of water temperatures were calculated for different seasons, regions and horizons of the lake. It has shown there are a considerable temperature variations up to 50 m depth from June to September.
The portion of nonseasonal temperature variability in total dispersion has been estimated by the compositional approximation.
Initially the developed information-diagnostic system (IDS) has allowed to receive the daily spatial distributions of Lake Ladoga water surface and air temperature for open water period, which can serve as mean climatological distributions at synoptic variability analysis and interannual variations.
The scheme of spring frontal zone (thermobar) evolution was built.
The scheme of "biological summer" duration (period which the water has a temperature above 10 ºC) for water surface was built.
The schemes of values and dates of water temperatures maximum coming are presented.
Lake Ladoga information-diagnostic system was used for investigation of water surface layer and atmosphere thermal interaction and Volkhov bay ecological conditions.
Created information-diagnostic system gives new possibilities at a solution of the various tasks on thermodynamic of Lake Ladoga and its simulation.

Production of rules for adaptive behavior by a simple dynamic system

V.A. Nepomnyashchikh, K.A. Podgornyj

Institute for Biology of Inland Waters, Borok, Yaroslavl region,
152-742 Borok, Russia
E-mail: nepom@ibiw.yaroslavl.ru

Individual "organisms" created for an Artificial Life research are typically provided with explicitly defined rules which control, for example, adaptive searching behavior. This approach allows for a study of holistic behavior emerging in individuals from interactions among rules, as well as behavior emerging in ecosystems from interactions among individuals. However, Alife also provides an opportunity to raise an another fundamental problem: what type of control mechanism could produce adaptive behavioral rules in real organisms?

One answer to the problem is that the rules are developing gradually, starting from a scratch, in a course of natural selection. An alternative answer is that basic behavioral rules shared by many living beings could emerge as a whole from basic principles of non-linear dynamics which underlies animal's behavior. In order to support the last view, we have developed a virtual agent to search for sources of "odor", as well as for patchily distributed food. The agent has three "neurons" and single odor sensor. Neurons are spontaneously active, while external stimuli modify their activity via the sensor and food consumption. The agent reveals searching rules which had not been pre-programmed explicitly, but are typical of many real organisms:

If deprived of guiding ambient stimuli, the agent performs a non-Brownian walk, which consists of spontaneous switches from relatively straight long runs to tight loops and vice versa. This walk results in an anomalous diffusion observed in population of agents and typical for a variety of organisms.

It switches to the oriented movement when odor gradient is found. When moving to the source of odor, it does not follows changes in stimulation continuously, but leaps roughly toward a source and corrects the chosen direction only rarely. This tactics had been documented in as different organisms as bacteria, nematodes and crustaceans.

If there is a gap in gradient, the agent persists with a chosen direction for a while, which help to get through narrow gaps and don't lost a way to source. If gap is wide, the agent eventually resumes non-Brownian walk which still help to find a way out of gap more efficiently as compared to ordinary Brownian walk.

Upon finding one or more food items in some area, the agent initiates a thorough search within restricted area and keeps doing so for a while, even if there are no more food to stimulate the search. This "area-restricted search" (ARS) upon prey finding is also a basic property of searching behavior in living beings. As opposite to ARS, the agent leaves an area if captures a repulsive prey which suppresses feeding.

The agent reveals a "sampling" behavior known in foraging animals. Upon reaching an odor source, it makes "excursions" to various directions, taking a chance to find other possible sources. When within a food patch, the agent may leave it and "sample" elsewhere, which provides an opportunity to find more rich patches.

These results do suggest hypothesis that basic behavioral rules shared by a variety of organisms may be rooted in basic properties of non-linear systems, even as simple as we have studied. A natural selection may modify the systems and, therefore, rules in different

environments, but it need not to create rules anew. This hypothesis could be verified in Alife by studying evolution of the same basic non-linear system in different environments. Also, such a study may answer the question: how simple could be a minimal system capable of producing rules which make individuals and populations viable? Finally, we suggest that dynamic agents similar to one described here could be used for individual-based ecological modeling of real populations: while physiological mechanisms differs in different organisms, the same dynamic system may serve as a common phenomenological description of these mechanisms.

LIMPACT: An expert system to estimate the pesticide contamination of small streams using benthic macroinvertebrates as bioindicators

Michael Neumann[1] and Joachim Baumeister[2]

1Institute of Ecology, Department of Limnology; Friedrich-Schiller-University of Jena, Winzerlaerstr. 10, D-07745 Jena, Germany
2Department of Artificial Intelligence and Applied Computer Science; University of Wuerzburg, Am Hubland, D-97074 Wuerzburg, Germany

Small streams form the beginning of the water circuit and they represent an important habitat for the aquatic fauna on the landscape level. In agricultural used catchments these habitats are influenced by short-term impact from non-point sources, involving factors such as hydraulic stress and the input of nutrients and pesticides. Usually, no regular monitoring systems are established for these agricultural non-point sources of pollution. The main advantage of bioindicator systems is their easy and cost-efficient application. When they are used to monitor toxic contamination, they additionally indicate the ecotoxicological effect of the contaminant.

The development and the evaluation of a biological indicator system for pesticide pollution in streams are presented. For small headwater streams with an agricultural catchment area, the expert system LIMPACT estimates the pesticide contamination according to the four classes Not Detected, Low, Moderate and High contamination without any specification of the chemical agents. The input parameters are the abundance data of benthic macroinvertebrate taxa within four time frames in a year (March/April; May/June; July/August; September/October) and 9 basic water-quality and morphological parameters. The heuristic knowledge base was developed with the shell-kit D3 and contains 921 diagnostic rules with scores either to establish or to de–establish a diagnosis. We differentiate between positive indicator (PI) taxa, which indicate contamination by high abundance values and positive abundance dynamics, and negative indicator (NI) taxa, a high abundance of which rules out contamination and indicates an uncontaminated site. We analysed 39 taxa and found 13 positive and 24 negative indicators. The database comprised 157 investigations per stream and year with rainfall event-controlled pesticide sampling and repeated benthic sampling.

For the evaluation of LIMPACT, we used the same cases. The correct diagnosis for the 157 investigations per stream and year is established by LIMPACT in 66.7 to 85.5% of the cases, with better results for uncontaminated sites. If each stream is considered only once in the system (n = 104), the correct diagnosis is established by LIMPACT in 51.9 to 88.6% of the cases. In most of the remaining cases no diagnosis is established instead of a wrong one.

The potential application of LIMPACT could be a yearly monitoring of streams and would reduce chemical analysis to the mandatory cases. Furthermore, it could be used to evaluate the success of risk mitigation strategies in the catchment to reduce the impact of pesticides.

Developing modelling techniques for predicting naturalness of Dutch streams

R.C. Nijboer[1], Y.S. Park[2], S. Lek[2] & P.F.M. Verdonschot[1]

[1] Alterra, Green World Research. P.O. Box 47, 6700 AA Wageningen, The Netherlands. Email: r.c.nijboer@alterra.wag-ur.nl
[2] CESAC, UMR 5576, CNRS-University Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex, France. Email: lek@cict.fr

Predicting the effects of stream restoration or anthropogenic disturbance on stream communities is important in Dutch water management. However, suitable predictive tools are still lacking. In particular, techniques for predicting the naturalness of streams (the distance to the reference condition) have to be developed for implementing the European Water Framework Directive. The aim of this study was to model macroinvertebrate communities in Dutch streams with help of environmental variables.
A data set including 614 streams sampled by water district managers, 858 macroinvertebrate taxa and 21 environmental variables was analysed using artificial neural networks (ANNs). A non supervised (selforganising map) and supervised (multilayer perceptron) algorithms were used to model macroinvertebrate communities relating to the environmental variables. Because a high number of species is problematic for ANN modelling, some parameters that are indicative for the macroinvertebrate community structure are modelled separately: number and fraction of rare species, dominant species, indicator species, species richness and Shannon diversity index. The predictive power of modelling the different community parameters was compared. The most explaining environmental variables were compared and discussed for each of the models, focussing on the variables indicating natural conditions or disturbance.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

A framework for computer-based pattern recognition and visualisation for the interpretation of ecological data

Mark A. O'Connor & William J. Walley

School of Computing, Staffordshire University
The Octagon, Beaconside ST18 0DG Stafford GB
E-mail: mo3@staffs.ac.uk

In a wide variety of situations, masses of data must be analysed and condensed in order to extract meaningful information. Computer-based methods of data analysis and visualisation offer a means by which such data can be used to its full potential. Research in the domain of river water quality monitoring led the authors to develop a software system to provide data analysis and decision support for biologists and others working with data from the UK Environment Agency's river biological monitoring programme. The software comprises a 'training' element – Mutual Information and Regression Maximisation (MIR-max) - and a user-friendly interface for presentation and analysis of results – River Pollution Diagnosis System (RPDS). Although the systems were developed for the domain of biological river water quality monitoring, they are almost entirely data-driven rather than relying on pre-determined models or 'knowledge' of this specific application area. Thus, the underlying processes of the system are applicable to a wide range of scenarios, not just river biology – in fact, to almost any situation in which a large amount of multi-dimensional data needs to be analysed and understood. It is clear that the same methods could be readily adapted to many other ecological applications. This paper describes a framework for decision support systems based on data analysis and visualisation by pattern recognition, essentially defining a generic version of RPDS. Pattern recognition is introduced as a key component of expertise, and the need for advanced visualisation methods to aid data interpretation is outlined. Pattern recognition by generalised 'clustering' methods is described, together with the main data visualisation techniques that underpin the framework. Throughout, RPDS is used as a concrete example to demonstrate how the framework can be used in practice. It is hoped to produce general-purpose data analysis software based around this framework in the near future; the authors have already produced a system based on the MIR-max technique (which is introduced, together with examples) but it is possible that for some application areas different clustering or visualisation methods may be more appropriate. It may also be possible to combine these pattern recognition techniques with methods based on another core aspect of human expertise, plausible reasoning (modelled, for example, by Bayesian belief networks), in order to produce a 'true' expert system. In conclusion, the systems and framework presented in the paper demonstrate the immense potential of pattern recognition systems for the interpretation of ecological data and provision of decision support.

Methodological issues in building, training, and testing artificial neural networks

S. L. Ozesmi[1], C. O. Tan[2], U. Ozesmi[1]

[1]Environmental Science Branch, Department of Environmental Engineering, Erciyes University, 38039 Kayseri, Turkey.
E-mail: stacy@erciyes.edu.tr
E-mail: uozesmi@erciyes.edu.tr
[2]Middle East Technical University, Department of Biology, 06531 Ankara, Turkey

We review the use of artificial neural networks, particularly the feedforward multilayer perceptron with back-propagation for training, in ecological modelling and what is practiced. Based on our experience we discuss how to build an artificial neural network model and how to optimize the parameters and architecture. Varying the weight range and learning rate did not result in large changes in model performance. Including a momentum term did not change the model's performance. Standardizing input data resulted in better model performance. We varied the number of hidden nodes and number of hidden layers to optimize the architecture. For our data two hidden layers did not result in higher accuracy than one hidden layer. The accuracy leveled off after a certain number of hidden nodes had been reached. The variation in model performance based on different random starts is as much as the variation from using different parameter values and different architectures. Based on our review of the use of artificial neural networks in ecology we make recommendations for best practices. Problems we saw in reporting on the use of neural networks included unclear explanation of the modelling process, for example why certain variables were chosen for a final model. However, the major problem was overtraining on data or giving vague references to how it was avoided. Unfortunately it seems that often studies do not make sufficient effort to avoid overtraining. Various methods can be used to determine when to stop training artificial neural networks: 1) early stopping based on cross-validation, 2) stopping after a analyst defined error is reached or after the error levels off, 3) use of a test data set. We do not recommend the third method as the test data set is then not independent of model development. Many studies used the testing data to optimize the model and the amount of training. Although this method may give the best model for that set of data it does not give generalizability or improve understanding of the study system. The importance of an independent data set cannot be overemphasized as we found dramatic differences in model accuracy assessed with prediction accuracy on the training data set, as estimated with bootstrapping, and from use of an independent data set. We also recommend the comparison of the artificial neural network with a general linear model as a standard procedure because a general linear model may perform as well or better than the artificial neural network. If the artificial neural network model does not predict better than a simple general linear model, then there are no interactions or nonlinear terms which need to be modeled and it will save time to use the general linear model. Artificial neural network models should not be treated as black box models but instead techniques such as sensitivity analyses, input variable relevances and neural interpretation diagrams should be used to make the model more transparent. This will further our ecological understanding which is an important goal of the modelling process.

The generalizability of artificial neural network models: the relationship between breeding success and occurrence

U. Ozesmi[1], S. L. Ozesmi[1], C. O. Tan[2] and R. J. Robertson[3]

[1]Environmental Science Branch, Department of Environmental Engineering, Erciyes University, 38039 Kayseri, Turkey.
E-mail: uozesmi@erciyes.edu.tr
E-mail: stacy@erciyes.edu.tr
[2]Middle East Technical University, Department of Biology, 06531 Ankara, Turkey
[3]Department of Biology, Queen's University, Kingston, Ontario, Canada

Two different models for a marsh-nesting bird, the red-winged blackbird Agelaius phoeniceus, were developed in two geographical regions and years apart. The first model was developed to predict occurrence of nests in Sandusky Bay Wetlands, Stubble Patch and Darr, in Lake Erie, Ohio in 1995 and 1996. The independent variables were vegetation durability, stem density, stem height, distance to open water, distance to edge, and water depth. The second model was developed to predict breeding success (fledged nestlings) in two marshes in Connecticut, USA, Clarkes Pond and All Saints Marsh, in 1969 and 1970. The independent variables were vegetation durability, nest height, distance to open water, distance to edge, and water depth. The nest occurrence model performance was at an average cross entropy or concordance index (c-index) of 0.730. The within geographical region testing resulted in a c-index of 0.6685 and 0.5463 in two different wetlands. The breeding success model performance was at a c-index of 0.7458. The within region testing resulted in c-indices of 0.4741 and 0.5321. When we tested the nest occurrence model on fledged nestling data we obtained c-indices of 0.6896 and 0.4704 in Clarkes Pond in 1969 and 1970 respectively, and 0.4303 and 0.5203 in All Saints Marsh in 1969 and 1970 respectively. When we tested the fledged nestling model on the nest occurrence data, we obtained c-indices of 0.7015 and 0.4131 in Stubble Patch in 1995 and 1996 respectively, and 0.5440 and 0.5482 in Darr in 1995 and 1996 respectively. With the input variable relevances, sensitivity analyses and neural interpretation diagrams we were able to understand how the different models predicted nest occurrence and breeding success. The generalizability of the models was poor except for when the marshes had similar values of important variables in the model, for example water depth. These results indicate that the artificial neural network models developed are not generalizable to marshes with different sizes and structures. These results also indicate that nest occurrence is a poor predictor of breeding success as red-wing blackbirds may choose nests not in the best location but choose good mates.

Development of methods for understanding ecological data using self-organising map

Park Y.S.[1],Chon T.S.[2], Lek S.[1]

[1] CESAC, UMR 5576, CNRS-Univ. Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 4, France.
E-mail: park@cict.fr
E-mail: lek@cict.fr
[2] Division of biological sciences, Pusan National University, Pusan 609-735, KOREA. E-mail: tschon@cict.fr

Understanding the relationships between biological attributes and their environmental variables is the fundamental basis in ecosystem management, because natural distributions of organisms are primarily determined by their environments. Recently a self-organising map (SOM) has become a popular tool for patterning ecological communities and for exploration of ecological data as an ordination method. The SOM is an unsupervised neural network algorithm and a method for clustering, visualization, and abstraction, the idea of which is to show the data set in another, more usable, representation form. After learning process, the output of the SOM results in a smoothing effect on the reference vectors of neurons. These reference vectors tend to approximate the probability density function of the input vector. Therefore, the visualization of these vectors at different input variables is convenient to understand the contribution of each input variable to the clusters on the trained SOM. In previous study the SOM effectively visualized the contribution of each environmental variable to biological attributes and their classification. However it was not easy to compare the effects of environmental variables to biological attributes because there was no quantity to relate each other. They were compared only qualitatively based on visualization patterns. Therefore, in this study we propose to quantify the relationships between ecological communities and their environmental variables after visualization of the SOM results. To quantify the values of relations between variables, correlation coefficient, species and site scores, and coefficients of a trend surface analysis were used. These quantities effectively represented the relationships between ecological communities and environmental variables. Therefore, the results supported that the SOM can be powerful tool for exploration of ecological data.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Patterning and predicting fish assemblages in large scale using artificial neural networks

Park Y.-S. [1], Lek S.[1], Oberdorff T. [2]

[1]CESAC, UMR 5576, CNRS-Univ. Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 4, France.
E-mail: park@cict.fr
E-mail: lek@cict.fr
[2]Laboratoire d'Ichtyologie, Muséum national d'histoire naturelle, 43, rue Cuvier, 75005 Paris, France. E-mail: oberdorf@mnhn.fr

Two artificial neural networks (ANNs) were used to develop models of local fish species richness and composition using a fish dataset of 650 reference sites fairly evenly distributed across French rivers and defined by some easily measured regional and local characteristics (i.e. distance from source, catchment area, elevation, mean width, river slope, mean depth, minimum and maximum air temperatures, hydrological units). The modelling was proceeded in two steps. First a self-organizing map (SOM) was used to classify fish communities and to assess the relationships between fish communities and corresponding environmental variables. Then a backpropagation network (BP) was applied to predict the local fish species richness and the local fish composition in the whole France area using a set of environmental variables.

The SOM have classified fish communities according to the gradient of species composition and their densities, and the classification was explained with the gradient of environmental variables such as drainage basin area, distance from source, width, slope, depth, and altitude. Each species showed different preferences of their living environmental variables. The relationships between species as well as between species and their environmental variables were analysed. After exploration of datasets the BP modelling was carried out to predict fish assemblages using the 650 datasets which were divided into two independent data sets to build and test the model. The predictabilities of the model both in the training and testing sets were significantly high ($r > 0.9$, $p < 0.001$ in both data sets). The ANN model obtained using only eight environmental variables succeeded in explaining more than 80 per cent of the total variation in local fish species richness and assemblage composition. The results of the models were well agreed with current ecological knowledge and it has a capability of synthesis of the natural phenomena hitherto difficult to apprehend by conventional methods. Therefore, ANNs are powerful for exploration of ecological data and prediction of considering variables, and can be used as a powerful tool for ecological researches and ecosystem management, i.e. tool to predict fish community structure.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Patterning exergy of benthic macroinvertebrate communities using artificial neural networks

Park Y.S.[1], Lek S. [1], Scardi M. [2], Verdonschot P.F.M. [3] and Jørgensen S.E. [4]

[1] CESAC, UMR 5576, CNRS-Univ. Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 4, France. Email: {park, lek}@cict.fr
[2] Department of Biology, University of Roma, Via della Ricerca Scientifica, 00133 Roma, Italy. Email: mscardi@mclink.it
[3] Alterra, Green World Research, Department of Freshwater Ecosystems, P.O. Box 47, 6700 AA Wageningen, The Netherlands. Fax: 317 424988. E-mail: p.f.m.verdonschot@alterra.wag-ur.nl
[4] Department of Environmental Chemistry, University of Copenhagen, Copenhagen, Denmark. Email: sej@dfh.dk

Exergy is a measure of the free energy of a system with contributions from all components including the energy of organisms, and it is used as an ecological indicator. In this study, we implemented a self-organizing map (SOM) for patterning exergy of benthic macroinvertebrate communities. The datasets were extracted from the database EKOO consisting of 650 sampling sites in the Netherlands. Out of 854 species 201 species have been selected on a preliminary neural network modelling, and the eighteen environmental variables have been also selected out of 67 variables on the basis of a sensitivity analysis performed on a modelling. Using these datasets, exergy of five trophic levels (carnivores, detritivores, detritivores-herbivores, herbivores and omnivores) were calculated for each sampling site on the basis of the biomass data. Exergy of each trophic level was used as input data of the SOM. By training the SOM the sampling sites were classified into four clusters and the classification was mainly related to pollution status and habitat type of the sampling sites. Exergy of different trophic levels responded differently to water types displaying characteristics of target ecosystems. By comparing exergy and environmental variables on the SOM map, the relationships between variables were evaluated. Finally, the results show that exergy is an effective ecological indicator and patterning changes of exergy is an effective way to evaluate target ecosystems.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Use of interactive forest growth simulation to characterise stand structure

Lael Parrott[1] and Holger Lange[2]

[1]Département de géographie, Université de Montréal, Montréal, Canada
E-mail: lael.parrott@UMontreal.CA
[2]BITOEK, Univ. of Bayreuth, Bayreuth, Germany
E-mail: holger.lange@bitoek.uni-bayreuth.de

The development of forest stands is a field in which long-term experiential knowledge of practitioners as well as scientific knowledge exist but are not easily reconciled. By serving as both a decision support system and a visualisation tool, an interactive growth simulator might bridge this communicative gap and facilitate the transfer of knowledge in both directions. We introduce the forest stand growth simulator TRAGIC++, which places an emphasis upon visualisation techniques while at the same time providing detailed information on tree physiology and related parameters.
This approach allows for the validation of model runs to be performed from two directions: visual inspection (exploiting practitioner's experiential knowledge) and numerical analysis (as scientists are used to). These two methods are used in an attempt to characterise forest stand structure and to subsequently reproduce this structure in TRAGIC++ simulations. For the latter method, an application of the theory of marked point processes, which has a high discriminating power among models or different parameterizations of the same model, is used to describe the spatial structure (i.e., the locations and diameters of trees) of forest stands. By applying this analysis to several data sets, we demonst rate an ability to discriminate between forests having different management histories and species compositions. We then show that TRAGIC++ is capable of generating similar stand structure for several different model calibrations. Lastly, we explore the complementarity between this analytical method of characterisation and the intuitively based, visual inspections of practitioners.

Evaluating parameters of simulation models for aquatic ecosystems

K.A. Podgornyj, V.A. Nepomnyashchikh

Institute for Biology of Inland Waters, Borok, Yaroslavl region,
152-742 Borok, Russia
E-mail: kap@ibiw.yaroslavl.ru

Parameters of an aquatic ecosystem simulation model should be adjusted to achieve an extreme of some objective function. Herein, we describe basic properties of new algorithm to find the parameters, as well as results of it's application to the zero dimention imitation model for Neva Bay (Gulf of Finland). The model represents a system of ordinary differential equations which describe nitrogen and phosphorus transformation processes, as well as dynamics of dissolved oxygen and biotic components of the ecosystem (phyto- and zooplankton, protozoans and bacteria). The normalized Theil criterion, which depends on numerous parameters of a model, is used to evaluate a correspondence of model.

Undoubtedly, first and second power methods are more preferable to solve optimization tasks. These methods are most developed, their convergence had been proven, and convergence rates had been evaluated. However, one need know gradient values of objective function in order to use iterative procedures for these methods. On the other hand, zero power methods are less developed, they converge more slowly as compared to first and second power methods, and many of them are heuristic. Nevertheless, it is impossible to find explicit formulas for partial derivatives used in first and second power methods if unknown parameters are numerous, while an evaluation of derivatives using finite-difference approximations leads, as a rule, to numerical errors. Apparently, zero power methods provide a sole opportunity to solve practical multiextremal minimization tasks of simulation model parameters evaluation, if the models have simple constraints such as unequalities.

We propose the two-stage direct search algorithm to solve optimization tasks. At the first stage, a global random search is carried out. The task of this preliminary stage is to find a possible region of global maximum in parameter space. At the second stage a local search is carried out using the modified simplex Nelder-Mead method. The local search locates a maximum position more accurately. A deformable simplex is used to this end, which is able to adapt easily to a topology of objective function by stretching along long sloping plains, changing movement direction in curved troughs and reducing in a vicinity of objective function local minimum.

Numerical experiments have shown that:

The proposed algorithm solve a parameter evaluation task with a calculation accuracy which is consistent with an accuracy of row observation data.

The method's efficiency does not depend on how well a trial point was selected.

Optimal parameter sets obtained at different numerical experiments do not differ significantly, which leads to the conclusion on an (statistically) unambiguous solution of optimization task.

The local search procedure may fail in some cases, for example, if an initial simplex is degraded to lower dimension one in a vicinity of active constraints, or at plateaus, that is, at sufficiently small (zero) gradients of objective function.

Ecological Informatics: understanding ecology by biologically-inspired computation

Friedrich Recknagel

University of Adelaide
School of Earth and Environmental Science
Glen Osmond 5064, Australia
Email: friedrich.recknagel@adelaide.edu.au

Improved understanding and prediction of ecological systems regarding global climate change, biodiversity and sustainability is a permanent matter of concern for decision makers in natural resource and environmental management. However the distinct nonlinear and complex nature of ecological systems allows only gradual progress in adequate data analysis, synthesis and forecasting. Such progress relies always on existing ecological knowledge, empirical data, mathematical models and computational technology.

Fast advancing computational technologies such as biologically-inspired computation in conjunction with high performance computing and internet have currently a profound impact on ecological research. They are able to revolutionize both, ecological data management and modelling. Ecological data management based on object-oriented and associative representations significantly facilitate data standardization, access, sharing and analysis. Ecological data modelling based on cellular automata, artificial neural networks, evolutionary algorithms and adaptive agents facilitate the inference from data patterns to processes, agent-based adaptive modelling, and the integration of deductive and inductive modelling techniques.

The mission of *ecological informatics* is to promote interdisciplinary research between ecologists and computer scientists in ecological data management, analysis, synthesis and forecasting by means of biologically-inspired computation.

The paper will provide a scope of *ecological informatics* and case studies exemplary for the current and future trends in *ecological informatics*.

Use of Artificial intelligence (Mir-max) for Reference diatom communities definition in Rhone basin and Mediterranean region (France)

Frédéric Rimet[1], Valérie Peeters[2], Henri Vidal[3] and Luc Ector[1]

[1]: Centre de Recherche Public - Gabriel Lippmann, CREBS, 162a Avenue de la Faïencerie, L-1511 Luxembourg, Email: rimet@crpgl.lu
[2]: DIREN Bourgogne, 10 bd Carnot, F-21000 Dijon, France
[3]: Conseil Général des Alpes-Maritimes, BP 3007, F-06201 Nice Cedex, France

The southeast part of France is characterised by a wide variety of landscapes and climates, from alpine, continental to Mediterranean. The Rhone basin and the Mediterranean region in France cover an area of 130700 km^2, including Corsica. Since 1995, more than two hundred benthic diatom samples were taken in the rivers of these regions for bioindication. In the framework of the Water Directive, the European Union requires to assess the water quality with an Ecological Quality Ratio (EQR), comparing the observed status to its corresponding reference. The aim of this study is to characterise the most important structuring parameters for benthic diatom communities in the Rhone-Mediterranean–Corsica region and to define reference diatom communities for different river types.

To explore the benthic diatom assemblages of this complex region, it is necessary to use advanced techniques based on artificial intelligence techniques. Groups of samples with homogeneous diatom communities are defined using the Mir-max software (www.soc.staffs.ac.uk). Samples are clustered with a pattern recognition process, and are represented on a two dimensional map.

Using the benthic diatom inventories, the analysis clearly shows that the bottom part of the Mir-max map corresponds to rivers of high altitude whereas the upper part of the map corresponds to lowland rivers. In fact the main physico-chemical structuring parameters shown on the Mir-max map are altitude and distance from source. On a canonical correspondence analysis, those two parameters have also important contributions on the first two axes. Clear gradients of NO_3^-, biological oxygen demand and NH_4^+ are also shown on the Mir-max map, but don't have the same gradient directions.

To select the less polluted groups on the Mir-max map, we used an integrative index of organic pollution (IPO). 18 groups were considered as reference conditions with IPO values above 4.5/5 as criteria.

Among these 18 different reference conditions, there is for example a group of lowland rivers (75-100m) in the west part of the basin (Massif Central) characterised by common species as *Achnanthidium minutissimum, Cymbella excisa, Nitzschia fonticola* and also interesting taxa for Europe like *Eolimna comperei* and *Achnanthidium latecephalum* only known in Japan. Another group of reference conditions corresponds to mid altitude (230-780m) alpine rivers characterised by a catchment dominated by limestone, *Gomphonema tergestinum* is one of its most abundant taxa. Another reference group is composed of alpine rivers of high altitude (770-1300m) and high current velocity, small species with robust anchoring systems dominate the assemblages: *Achnanthidium biasolettianum, A. minutissimum, Gomphonema pumilum, Reimeria sinuata.*

This study is a first step in the establishment of a diatom typology for French rivers; a next step could be the validation with the epilithic diatoms of ecoregions generally based on the geology, the climate and the relief.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Modelling population and community dynamics with Qualitative Reasoning

Paulo Salles[1], Bert Bredeweg[2*]

[1]Universidade de Brasilia, Instituto de Cincias Biolgicas, Campus da Asa Norte, Brasilia - DF, 70.910-900, Brasil. E-mail: paulo.bretas@uol.com.br
[2]University of Amsterdam, Department of Social Science Informatics, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. E-mail: bert@swi.psy.uva.nl

Ecological knowledge has been characterised as incomplete, fuzzy, uncertain, sparse, empirical, and non-formalised. It is often expressed in qualitative terms, verbally or diagrammatically. There is a need for new and efficient computer-based tools for making this knowledge explicit, well organised, processable, and integrated with quantitative knowledge. Qualitative Reasoning (QR) is an area of Artificial Intelligence that creates representations for continuous aspects of the world to support reasoning with little information. Using QR techniques, many questions of interest in ecology can be answered in qualitative terms, and scientifically valid predictions can be made when numerical models are not available. Particularly relevant for our work are qualitative representations of differential equations, algebraic relations and the explicit representation of causal relations between quantities.
We present a qualitative simulation model about population and community dynamics in the Brazilian Cerrado vegetation designed to be used in a learning environment, representing a widely accepted hypothesis about succession, the Cerrado Succession Hypothesis (CSH). CSH model is divided into clusters of smaller predictive simulation models. The first cluster implements a general theory of population dynamics, with the explicit representation of processes such as natality, mortality, immigration, emigration, colonization and population growth. This knowledge is the basis for all the simulations in CSH model. The second cluster represents interactions between two populations, such as symbiosis, competition, predator-prey, comensalism and amensalism. The third cluster represents the CSH. It is assumed that fire frequency influences a number of environmental factors such as litter, cover, moisture, light, nutrients, which in turn influence natality and mortality of trees, shrubs and grass. The models show that when fire frequency increases, succession leads to open grasslands and, when fire frequency decreases, the vegetation becomes woody and denser. Our contribution will show the potential QR has for ecological modelling.

* Corresponding Author.

A fuzzy approach to land suitability analysis

A. Salski[1], P. Kandzia[1] and F. Bartels[2]

[1] Institute of Computer Science and Applied Mathematics, University of Kiel, Olshausenstr. 40, D-24098 Kiel, Germany. Email: asa@email.uni-kiel.de
[2] Knorrstr. 24, D-24105 Kiel, Germany

Land suitability analysis is an approach in land evaluation that concerns the assessment of land performance for specified land utilization purposes. Conventionally, land evaluation proceeds by identifying a land suitability class applied to a land unit. This land suitability classification is based on land characteristics and qualities, i.e. measurable or estimable attributes related to a specified land utilization type. Conventional land classification methods (e.g. available in most geographical information systems) ignore an internal inhomogeneity of the land unit, short-range spatial variation, measurement errors and the continuous nature of land characteristics. The classification results are related to homogeneous land units separated by sharp boundaries.

A fuzzy approach to land suitability analysis provides an useful alternative to conventional methods (Burrough 1989, Huajun 1991, Enea et al. 2001). In this case the land evaluation bases no more on Boolean algebra operations on land characteristics related to specified utilization, but on the joint degree of land suitability for defined uses. The evaluation criteria are defined as appropriate fuzzy sets (e.g. “low hydrologic conductivity”) and their membership functions can be used for the transformation of land characteristic data to a common scale (the interval [0,1]). After this transformation several fuzzy operations can be used as combination operators for the calculation of the joint degree of land suitability. The final results can be presented as isolines of the joint suitability degree using the fuzzy kriging procedure as the regionalization tool.

The values of the joint suitability degree can be calculated, interpolated and presented in the form of isolines using the Fuzzy Evaluation and Kriging System FUZZEKS which has been developed at the University of Kiel (Bartels 1997). FUZZEKS utilizes crisp measurement data as well as additional imprecise data subjectively estimated by an expert and defined as fuzzy numbers. This is particularly important in the case of an insufficient amount of data (Piotrowski et al. 1996). The analysis of the suitability of a specified land unit as a waste disposal site is presented in this paper as the simplified application example. Four land characteristics, namely water table depth, hydrologic conductivity, clay content and Cl concentration, are taken into account in this example.

Predicting fish assemblages in rivers: a neural network case study

Michele Scardi[1], Stefano Cataudella[1], Paola Di Dato[1], Giuseppe Maio[2], Enrico Marconato[2], Stefano Salviati[2], Lorenzo Tancioni[1], Paolo Turin[3] and Marco Zanetti[3]

[1] *Dept. of Biology, University of Rome "Tor Vergata", Via della Ricerca Scientifica, 00133 Roma, Italy, Email: mscardi@mclink.it*
[2] *Aquaprogram s.r.l., Via Borella 53, 36100 Vicenza, Italy*
[3] *Bioprogramm s.c.r.l., Via Tre Garofani 36, 35124 Padova, Italy*

Predicting the structure of fish assemblages in rivers is a very interesting goal in ecological research, both from a purely theoretical point of view and from an applied one, for instance when river management strategies are to be defined.
Estimates of the probability of presence/absence of fish species have been obtained using different approaches. Although conventional statistical tools (e.g. logistic regression) provided interesting results, the application of artificial neural networks (ANNs) has recently outperformed those techniques. ANNs are especially effective in reproducing the complex, non-linear relationships that link environmental variables to fish species presence and/or abundance.
Even though ANNs are usually regarded as "black-box" models, they allow to perform sensitivity analyses as well as to obtain useful insights into the ecological processes that determine the fish assemblage structure. In particular, sensitivity analyses point out the role that each environmental variable plays and define its predictive value with respect to the biotic component of the ecosystem.
Recent developments of the ANN training procedures, specifically aimed at solving ecological problems, allowed to optimize the prediction of species assemblages. The improvement in prediction involves not only the accuracy of the models, but also their ecological consistency.
A case history about fish assemblages in the rivers of the Veneto region (Northern Italy) will be presented not only to show the results obtained by using ANN models, but also to outline the particular strategies that have been used to adapt those ANN models to ecological problems.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Improving neural network models by means of theoretical ecological knowledge

Michele Scardi[1], Sovan Lek[2], Young-Seuk Park[2], Piet F.M. Verdonschot[3] and Sven E. Jørgensen[4]

[1] Dept. of Biology, University of Rome "Tor Vergata", Via della Ricerca Scientifica, 00133 Roma, Italy
E-mail: mscardi@mclink.it
[2] CESAC, UMR 5576, CNRS-Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 4, France
[3] Alterra, Green World Research, team of Freshwater Ecology, Wageningen, The Netherlands.
[4] DFH, Environmental Chemistry, Univ. of Copenhagen, Copenhagen, Denmark.

Sometimes the ecological relationships that we want to model are just too complex to be reproduced. In this case, we usually try to collect more data and/or to define a new modeling goal that is more simple and more feasible. Neural network models, even though they proved to be very effective in many cases, are not an exception to this rule.
However, the heuristic nature of neural network modeling allows developing strategies that can help to improve the accuracy of the results. This goal can be attained, for instance, by modifying the learning algorithms in order to improve their performance by adapting them to the specific problems that have to be solved.
Using a data set about benthic macroinvertebrates, we will provide an example of a complex ecological problem that has been approached by enhancing the neural network training algorithm. In order to attain this goal, we modified a conventional error back-propagation algorithm in a way that allows to use not only data, but also theoretical ecological knowledge. In particular, the latter source of information has been exploited by embedding a set of simple ecological "rules" into the computation of the mean square error during the validation phase of the learning algorithm.
This enhanced modeling strategy will be presented in comparison with a conventional approach and its potential developments will be thoroughly discussed.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Data mining and visualisation in biological and environmental processes

Subana Shanmuganathan, Philip Sallis and John Buckeridge

Auckland University of Technology, New Zealand.
Email: {subana.shanmuganathan,philip.sallis,john.buckeridge}@aut.ac.nz

Biomonitoring, the area in which biological indicators are used to measure the changes in the environment is not a new field. The emphasis in biomonitoring has been on finding the biological organism/s that show some form of response to environmental changes. Finding such organisms may not be easy. However, recent technological advances have given researchers the opportunity to observe, collect and store data in new territories with new devices. Analysing such ecological data sets and correlating these with the environmental changes still pose many constraints with the standard statistical methods. In this paper, we explore some data mining techniques using Kohonen's self-organising map (SOM) methods, an artificial neural network based on unsupervised training algorithm, for data analysis/ visualisation in biological and environmental systems.

Over the last two decades, the availability of increased computing power, improved network capabilities and data storage devices at constantly tumbling prices have increased the capture and storage of data significantly. Further, the use of recent technological advancements in embedded systems has improved the quality of data as well. This has led the exploration for new methods to analyse these data sets to their potential.

In recent years, the causes of global environmental degradation are increasingly seen to be anthropological. Many national and international institutions have introduced more programmes to address this issue i.e. Millennium Ecosystem Assessment, Ecological Assessment and Indicators Research Programme by the Environment Production Agency. In light of the complexity involved in ecosystem processes, traditional methods alone are becoming more and more unsuitable for analysing such large amounts data sets collected through environmental monitoring programmes.

Methodology: The use of SOM methods in ecological modelling is being proven to be very successful in the recent past. It has provided a means to reveal the relationships in highly complex ecosystem processes using measurable data alone without any physical models. We applied SOM techniques (i.e. cluster analysis, component dependency analysis and time series analysis) to model the poorly understood environmental and biological system processes in terms of ecological response caused due to human activities, with the data collected through some biodiversity monitoring programmes conducted by the Auckland City Council. We intend to show here how the use of SOMs provides a refined approach towards the research results in modelling complex ecosystem processes.

The example chosen here also illustrates how siltation, caused by rapid urbanisation along the east coast beaches of Auckland in northern New Zealand has been sufficient enough to cause observable degradation in the coastal biodiversity. The pressure imposed by the concerned public and environmentalists of the area paved ways for establishment of the Okure Marine Reserve (1996) in Long Bay, which was the first of its kind to be set up in an urban area in New Zealand.

The Auckland and North Shore City Councils have been carrying out monitoring programmes in an attempt to study the effects of urbanisation along these beaches. One of them is the project assigned to the Auckland University's UniServices. In this paper we look into the aspects of using SOM methods to explore this data to establish any relationships found between the environmental parameter and the community structure response in subtidal organisms. We also compare these results with that of the conventional statistical methods.

Neural network patterning and molecular biological analysis of fish behavior as a bio-monitoring system for detecting toxic chemicals in environment

Sung-Woo Shin[1], Tae-Soo Chon[2], Hyun-Duk Cho[2], Chang-Woo Ji[2], Wan-Sung Choi[3], Jae-Yoon Han[3], Jong-Sang Kim[4], Sung-Kyu Lee[5], and Sung-Cheol Koh[1]

[1]Division of Civil and Environmental Engineering, Korea Maritime University, Bussan 606-791, Korea. E-mail: skoh@hanara.kmaritime.ac.kr
[2]Divison of Biological Sciences, Pusan National University, Pusan 609-735, Korea. E-mail: tschon@pusan.ac.kr
[3]Dept. of Anatomy, College of Medicine, Gyeongsang National University, Chinju 660-751, Korea.
[4]Dept. of Science and Biotechnology, Kyungpook National University, Taegu 702-701, Korea.
[5]Toxicology Division, Korea Res. Inst. of Chemical Technology, Taejon 305-600, Korea.

The objective of this study is development of a biomarker used to monitor abnormal behaviors of Japanese medaka as a model organism caused by toxic and persistent chemicals in the ecosystem. Each fish was subjected to toxic chemicals such as fluoranthene, copper and diazinon after starvation for 24 hr. The fish was observed for their abnormal behaviors such as an enhanced surfacing activity, opercular movements, erratic locomotion, tremors and convulsions. These behaviors were concurrently observed on a real time basis using image processing and automatic data acquisition systems. The untreated individuals showed common behavioral characteristics (e.g., smooth and linear movements). The characteristic movements such as repeated linear back-and-forth movement and other irregular locomotion were chosen for patterns of response behaviors of fish treated with the toxic chemicals. Subsequently the selected parameters to characterize the movement patterns such as speed, meander and stop duration in a short time (e.g., 1 minute) sequence were sequentially provided as input for training with the artificial neural network with the backpropagation algorithm. After training with the network, new sets of data were given to the network for evaluation. The trained network was able to detect the characteristic patterns of response behavior of fish treated with different chemicals. The different response patterns were further investigated with molecular biological analyses through *in situ* immunohistochemistry and gene expression. The expressions of a catecholamine-producing enzyme, tyrosine hydroxylase (TH) at olfactory bulb and some brain regions of the fish were comparatively analyzed. The toxic effects on test specimens were also compared on the molecular genetic basis among different toxic chemicals. This study provides molecular and neurobehavioral bases of a bio-monitoring system of behavioral informatics projected from test specimens treated with toxic chemicals.

Reduction of a complex biogeochemical model with data mining techniques

Tobias A. Sperr[1] and Kai W. Wirtz

University of Oldenburg, Institute for Chemistry and Biology of the Marine Environment
P.O. Box 2503, 26111 Oldenburg, Germany
E-mail: tosp@uni-oldenburg.de

Recently build biogeochemical models of marine sediments are able to calculate the efflux of nitrous oxide, which is an important agent in the global climatic system. Due to the complex nature of the model, parameterization uncertainty is high, in particular when being applied to various sites. In order to reduce complexity without losing features relevant for nitrous oxide dynamics we first generated a large number of time series for important variables employing realistic Wadden Sea boundary conditions and random parameter sets. Secondly, we used Kohonen-SOMs to map the data onto a set of prototype vectors which are further reduced by means of different clustering algorithms. From the resulting set of basic states a Bayesian-type state-transition network was constructed to investigate the temporal behavior of the system. We found only a few constellations of the model with high gas efflux and also only a few number of antecedent states leading to them. The results confirm the assumed large contribution of coastal waters to the global nitrous oxide budget. In addition it is shown that there exists only a narrow domain in which those high emissions occur. The dynamics of these emissions can be formalized in a reduced model frame without losing specific aspects of interest.
Finally we conclude that with our approach it is possible to visualize complex model data, take into account model uncertainty and to deduce new knowledge.

Cellular automata models applied to landslides simulation on high performance computers

Giandomenico Spezzano

ICAR-CNR, c/o DEIS, Università della Calabria, Via Pietro Bucci cubo 41C
87036 Rende (CS), Italy, E-mail: spezzano@si.deis.unical.it

Landslides are natural disasters which can cause serious damages in terms of lives lost, homes destroyed, economies disrupted. By understanding how and where these natural events occur we can respond effectively when disasters strike. The simulation of landslide hazards is particularly relevant for the prevention of natural disasters, since it enables to compute risk maps and helps to design protection works. Landslides are complex natural phenomena that are hard to model and simulate. Predicting hazardous events such as landslides is particularly difficult because no laboratory exists that can preliminary measure the necessary variables, refine the techniques, and apply the results. Moreover, landslide simulations need to be accurate and often require massive amounts of computations.

Cellular automata (CA) theory is a significant approach to incorporating computer modeling in the study of natural phenomena because CA can capture the essential features of systems in which global behavior arises from the collective effect of numerous simple, locally interacting components. Moreover, CA are an inherently parallel computing abstract model. CA consist of a grid of cells, where each cell has one of a number of finite states. Each cell's state is updated in discrete time steps according to a local rule (*transition function*) or set of rules. These rules might depend on the state of the cell in its previous time step or the state of its nearest neighbors.

In my opinion, CA models, visualization methods, high performance systems, together with geographical data retrieved by the Worl Wide Web are the essential elements by which simulations that can help researchers to better understand the physics of these natural phenomena. To meet this goal, we have developed a high performance problem solving environment, called CAMELOT (**C**ellular **A**utomata environ**M**ent for syst**E**ms mode**L**ing **O**pen Technology), that allows interactive simulation and steering of cellular computations. CAMELOT is a multi-platform system that uses the cellular automata computational model both a tool to model and simulate dynamic complex phenomena and as a computational model for parallel processing. It combines simulation, visualization, control and parallel processing into one tool that allows to interactively explore a simulation, visualize the state of the computation as it progresses and change parameters, resolution or representation on the fly. Moreover, CAMELOT supports exploratory, algorithmic and performance steering of cellular computations.

Currently, we are applying CAMELOT within a research project that handles the landslide events that have interested the Campania Region in May 1998 in the Sarno area. The aim of the work is to improve the overall ability to predict landslides and mud flows and to develop countermeasures to limit its disastrous consequences. Our simulation is based on a cellular model of debris/mud flows defined in [1]. This model defines the ground as a two-dimensional plane partitioned into square cells of uniform size. Each cell represents a portion of land, whose altitude and physical characters of the debris column land on it are described by the cell states. The state evolution depends on a transition function, which simulates the physical processes of the debris flow. We have developed and tested the model of the landslide simulation on a Linux Beowulf cluster machine. We have selected the Chiappe of Sarno-Curti debris flow as a first case of study. For our simulation we have used a map of the

landslide at 1:5000 scale. The area of interest has been mapped on a cellular automata of 772 cells long and 880 cells wide, which each square cell being 2.5 meters on each side. The experiments required about 6000 iterations (update steps) for obtaining the shape of the landslide. First results show a good accordance between the real phenomenon and simulation result. A critical aspect of our model concerns the identification of accurate and accessible characterizations of both soils attributes and water content for the land units in our study area. We are evaluating the possibility to extract this information from hyperspectral data analysis acquired by MIVIS airbone platform by the LARA-CNR project.

Application of genetic algorithms and Internet computing to biodiversity science

David R.B. Stockwell[1], James H. Beach[2], Aimee Stewart[2], Gregory Vorontsov[2], David Vieglais[2], Ricardo Scachetti Pereira[3]

[1] San Diego Supercomputer Center, 9500 Gilman Dr, La Jolla, CA 92037, USA, Email: davids@sdsc.edu.
[2] Biodiversity Research Center, 1345 Jayhawk Boulevard, University of Kansas, Lawrence, Kansas 66045, USA
[3] Centro de Referencia em Informacao Ambiental, Av. Dr. Romeu Tortima, 388, Campinas, Sao Paolo, Brazil 13084520

This paper describes the selection and implementation of a genetic algorithm for mapping species distributions into the Lifemapper project (http://www.lifemapper.org). Using a screensaver version of GARP (Genetic Algorithm for Rule-set Production) for modeling species distributions, lifemapper harnesses vast computing resources through 'volunteers' PCs similar to SETI@home. Lifemapper's primary goal is to provide an up to date and comprehensive database of species maps and prediction models (e.g. geographic distribution of the bobcat or Lynx rufus). The models are developed using specimen data from distributed museum collections and an archive of geospatial environmental correlates. A central server maintains a dynamic archive of species maps and models for research, outreach to the general community, and feedback to museum data providers.

An independent test of an artificial neural network model for predicting breeding success

C. O. Tan[1], S. L. Ozesmi[2*], U. Ozesmi[2] and R. J. Robertson[3]

[1]Middle East Technical University, Department of Biology, 06531 Ankara, Turkey
[2]Environmental Science Branch, Department of Environmental Engineering, Erciyes University, 38039 Kayseri, Turkey, Email: : stacy@erciyes.edu.tr
[3]Department of Biology, Queen's University, Kingston, Ontario, Canada

We developed an artificial neural network model for breeding success of a marsh-nesting bird, the red-winged blackbird Agelaius phoeniceus, using data from two marshes in Connecticut, USA in the years 1969-1970. The independent variables were vegetation durability, nest height, distance to open water, distance to edge and water depth. The dependent variable was a binary index of whether or not any nestlings fledged. We developed the model using data from one of the marshes, Clarkes Pond in 1969-1970. We used bootstrapping to determine the maximum number of epochs to train the model. We stopped the training when corrected average cross entropy (c-index), which is approximately the area under the ROC curve, leveled off. The corrected c-index, calculated using 150 bootstraps, leveled off at 0.663 after 70 epochs. This corrected c-index was lower than the c-index of 0.746 obtained on the training data. According to the ANN model, the most important variables for predicting breeding success were nest height and water depth. Distance to edge was the least important variable. Sensitivity analyses indicated that breeding success increased with increasing water depth and increasing distance to edge but decreased with increasing nest height. Breeding success was high with lower vegetation durability and high vegetation durability while vegetation with mid-range durability had low breeding success. Breeding success generally decreased with increasing distance to open water. The model was tested independently on All Saints Marsh. The two wetlands were quite different from each other. Clarkes Pond was an open water pond surrounded by Typha vegetation where the red-winged blackbirds nested. All Saints Marsh was a shallow emergent marsh dominated by Typha with scattered shrubs throughout. It did not have any large areas of open water in either 1969 or 1970. The resulting c-index of the test for both years of All Saints Marsh was 0.511. When we tested separately on the two years, the c-indices for All Saints marsh in 1969 and 1970 were 0.474 and 0.532 respectively. The artificial neural network model did not perform better than a general linear model. The GLM model had a c-index of 0.630 on the training data and a c-index of 0.536 for the test for both years of All Saints Marsh. The GLM c-indices for All Saints Marsh in 1969 and 1970 were 0.533 and 0.528 respectively. In addition to differences in vegetation and open water areas, an important reason for the failure of the model to generalize to the other wetland was different predation pressure in the wetlands. In these wetlands the most important factor influencing breeding success was predation. Predators are influenced by environmental variables such as water depth and distance to edge, which were included in the model. For example, predation decreased with increasing water depth under the nest. However, predators are also influenced by factors such as prey density, which was not included in this model. High nesting density reduced the percentage of nests that were destroyed by predators. Another reason for the poor generalizability of the artificial neural network model may be the limited number of data points used for training. In this study artificial neural networks did not provide any advantages over general linear data analysis techniques.

* Corresponding Author

Environmental molding of human life history evolution: modelling and data analysis

A.T. Teriokhin[1,2], F. Thomas[1], F. Renaud[1], E.V.Budilova[1,2] and J.F. Guégan[1]

[1] *Centre d'Etudes sur le Polymorphisme des Micro-Organismes, IRD, 911 Avenue Agropolis, 34032 Montpellier, France, Email: terekhine@mpl.ird.fr*

[2] *Section of General Ecology, Dept. of Biology, Moscow Lomonosov State University, Moscow 119899, Russia .*

One more or less current approach to modelling life history evolution is optimization modelling. Though this approach is not universal, there is a wide class of situations (at least on the level of modelling, but we hope in the reality too) where the natural selection can be interpreted as an optimization of some criterions of fitness. Usually these criterions are the intrinsic rate of population increase, r or the average life time reproductive success, R0. They are easily calculated from the Leslie model on the basis of age-specific fertilities bt and survival probabilities pt, either directly (as R0) or using the Euler-Lotka equation (as r). The characteristics bt and pt, in turn, depend on age-dependent variables describing the state of organism (such as its size wt and vulnerability qt), and on (possibly time- or season-dependent) variables describing the state of environment (such as availability of food aT and degree of aggressiveness AT). In the present model we do not take into account the dependence of environmental characteristics on time or season and simply associate them with the parameters a and A in the equations of Kleiber and Gompertz-Makeham, correspondingly. The first of these equations, et =a·(wt)b, with b about 0.75, expresses the rate of energy production for an organism of size wt as a function of this size, whereas the second equation, pt=exp(-A-qt), expresses the total rate of mortality as the sum of an age-independent environmental component A and an age-dependent individual component qt. The age dynamics of bt and pt (as well as the age dynamics of the state variables) depends also on the strategy of distributing the energy between different needs of the organism: growth, reproduction, survival, etc. This strategy can be presented as a vector function Ut of state variables Ut=Ut (wt, qt, ·), the components of Ut being the fractions of energy allocated to growth, reproduction, survival and other needs of the organism. Finally, the goal of the optimization modelling of life history evolution is to find a strategy providing the optimum (maximum) of the fitness. One way of solving this problem is to represent the function Ut as a neural network having the state variables as its inputs and the components of Ut as its outputs and to find the optimal parameters of this network (weights of links and thresholds of neurons) using a genetic algorithm. In the presented study we use the described model, under different underlying hypotheses, to predict how the changes in environmental characteristics such as food availability, environmental conditions and survival hazards might influence the basic life history parameters bt and pt (and thereby derivative parameters such as size at birth, size at maturity, age of maturity, fertility, life expectancy, etc.). The model predictions are compared with real tendencies drawn using the statistical analysis of global data on human demography and life conditions for about 170 countries.

Application of the Self Organizing Map algorithm combined with the Structuring Index to study diatom assemblages

J. Tison[1], J.L. Giraudel[2], M. Coste[1] and F. Delmas[2]

[1] Cemagref Bordeaux, U.R. Qualite des Eaux 50 av. de Verdun 33610 Cestas, France, Email: juliette.tison@bordeaux.cemagref.fr
[2] Centre d'Ecologie des Systèmes Aquatiques Continentaux, UMR 5576, 118 route de Narbonne, 31062 Toulouse cedex 4, France.

The aim of the present research was to demonstrate on a pilot data-set the efficiency of using the association of the SOM algorithm and the Structuring Index to classify different ecological conditions and to determine the most relevant species (called "structuring species") or assemblages characterising eco-regions.
The pilot reference data-set comes from water quality surveys led by the Cemagref in 1994 and 1998 for the Adour-Garonne Water Agency. All the information have been gathered in a database, each site being described with environmental and taxonomic descriptors. Used in the frame of the PAEQANN project, this database revealed weaknesses concerning the stations distribution over the basin, and then has been completed in this way in summer 2001.
In a preliminary step of analysis, a set of parameters as altitude and conductivity computed by PCA showed their ability to settle an eco-regional framework within this basin, as they cluster stations in a geographical way.
Keeping in mind this first level of aquatic ecosystem classification, the taxonomic information have then been analysed with the Kohonen Self Organizing Map algorithm (SOM) (Kohonen, 1995), combined with the Structuring Index (SI) (Giraudel, submitted).
SOM has already been successfully used in ecology for communities patternizing (Giraudel & Lek, 2001), because of non-linear relationships between abiotic and biotic parameters. This unsupervised neural network enables visualisation of the species assemblages in a two-dimensional space, preserving the topology of the input data. For each species environmental distribution can be observed directly, so this method appears to be very powerful and could be applied to strengthen the results already obtained with other techniques (for example ecological profiles by Daget & Godron 1982).
Combined with the SI, this method allowed us:

- to exclude structuring and non-structuring taxa, that is to say taxa which presence brings strong ecological information or not ;
- and then to determine, in the reference stations database, the species assemblages characterising the main Adour-Garonne eco-regions.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

A river pollution Bayesian Belief Network (RPBBN) for the diagnosis and prognosis of river health

David J. Trigg and William J. Walley

Staffordshire University, Beaconside, Stafford, Staffordshire. ST18 0AD
E-mail: D.J.Trigg@staffs.ac.uk

The paper describes a River Pollution Bayesian Belief Network (RPBBN) that was developed for the Environment Agency (England and Wales) as part of a R&D Project that used plausible reasoning and pattern recognition to diagnose and predict river health from field data. The background to the study is presented, including the findings of two preliminary studies that highlight some problems of network design, creation and testing. A brief description is also given of the data set and how it was validated and manipulated to form the project database. The final database contained the abundance levels of 76 macroinvertebrates, 12 site characteristics and up to 34 chemical variables recorded at 3556 sites in the spring and autumn of 1995.

Networks having nodes with multiple parents may produce excessively large conditional probability matrices. The problems that this creates are explained, and details of two tools that were developed to help overcome them are given. The first tool used multiple linear regression and information theory to optimise the causal links between the variables, whereas the second performed *n*-dimensional smoothing of the raw conditional probability matrices.

The paper includes details of the development process and final structure of RPBBN, and a description of a graphical user interface designed to interface with RPBBN. The interface allows users to input data from file and to perform diagnoses or prognoses (prediction) as required. The state of health of a river can be diagnosed, in terms of probabilistic predictions of five key chemical variables, using macro-invertebrate data, site characteristics and the season. Alternatively, the community composition of a river can be predicted from its chemistry and site characteristics. This flexibility allows users to model hypothetical situations or assess the impact of changes to an existing situation. For example, one might explore how a change in river chemistry would affect community composition. These aspects of RPBBN are demonstrated in the paper through typical examples. The strengths and weaknesses of the BBN approach to ecological modelling are discussed and some examples of its use in other environmental fields are given. The paper concludes that BBNs provide a very powerful means of interpreting ecological data and have considerable potential for use in all environmental fields.

BASIS, a case-based reasoning system for lake management

Egbert H. van Nes & Marten Scheffer

Wageningen University
6700 DD Wageningen NL
E-mail: Egbert.vanNes@aqec.wkao.wau.nl

A central theme in water management research is the problem how to select the optimal measure for restoration. Solving this problem requires prognosis of the expected response of the ecosystem. Up till now, the dominant approach has been to combine all available quantitative information on important processes in computer models for prediction. Such a compilation of information, however, is accompanied by a compilation of errors and uncertainties, which is the main reason that this approach is losing some popularity.
Most experts do better than most models. We think that this is because experts tend to reach their conclusion in a safer way than models do. It has been indicated a predominant problem solving method of experts is the use of past analogous cases. In lake management for example, the response of a specific lake to removal of fish is expected to be similar to that of comparable lakes, where this has been done before. Obviously all lakes are different, therefore some nuance has to be made as well.
The search for analogous cases has been formalized in the technique of case-based reasoning (CBR). This technique is rooted in the 1980's in the rather specific research area of artificial intelligence. In its essence, the computer selects the most similar cases (lakes) from some relevant variables. The characteristics of the selected cases are visualized in such way that differences and similarities are easily recognized.
The computer information system that we developed is called *Best Analogous Situations Information System (BASIS)* and is an aid for ecologists to make predictions of the response of aquatic systems to measures. Besides the CBR technique, the program includes an option for exploratory analysis of the database. Two and three-dimensional plots can be made and outliers can easily be examined and selections can be highlighted. Up till now, we have collected data of 70 Dutch lakes with several relevant variables.

A macrofauna-environment based prediction model using multinomial logistic regression

P.F.M. Verdonschot, P. Goedhart, & R.C. Nijboer

Alterra, Green World Research, team of Freshwater Ecology, Wageningen, The Netherlands
E-mail:P.F.M.Verdonschot@alterra.wag-ur.nl

Water managers more and more ask for prediction of the measures they intend to take. Prediction of ecological effects is complex and often not quantified. This contribution takes communities of macrofauna taxa and theier environment as an important example. To use community-environment relationships, a data-analytical approach is followed that consists of three steps: (1) macrofauna samples are clustered into groups; (2) the groups are related to the environment data by ordination; and (3) a prediction model based on the environment of the groups by using multinomial logistic regression is constructed. Communities or groups are used rather than taxa because groups have the simplicity that helps to communicate the results to water managers. Groups, when well described in a typology, have meaning and become real as if they already exist.
Three prediction models are constructed: one for a large number of regional waters and two for streams and ditches, respectively for national purposes.
The models are tested on different water management cases in the Netherlands. The results validated the models and are promising for the end-users.

This work has been supported by the EU through the PAEQANN project (5th Framework Programme, contract EVK1-CT1999-00026)

Projection persuit and robust indices for the classification of ecological data

Heinrich Werner

University of Kassel, FG Neuronale Netzwerke,
Heinrich Plett Str. 40 D-34109 Kassel
E-mail: werner@neuro.informatik.uni-kassel.de

The projection persuit algorithm for the navigation through high dimensional data-spaces has been developed by Friedman & Tukey in 1974 (IEEE Trans. Comp). It is based on a random selection of 2-dimensional projections of the high dimensional data optimized with respect to a given index-function measuring the degree of data separation seen in the 2-dimensional projection. The advantage of this algorithm is the good visualization of the data, however the high calculation load did not allow ist application to very large datasets for some time.
Nowadays with large computing power even in desktop computers this problem has vanished and makes this tool available in many different areas.
In this paper we present an implementation of this tool together with some new index-functions which have proved to be very useful in various applications in quality assurance in industry, in sleep research in medicine both with data sets of some 104 vectors of dimension >500. The indices developed have a behaviour which is robust with respect to outliers in the data set. The originally pure random choice of two projection directions is made more effective by using annealing techniques and genetic algorithms.
The algorithm can be applied both supervised, with preassigned classes, as well as unsupervised, separating subclasses differing in their distribution from the expected overall pattern.
The software presented provides an easy to use GUI for the visual inspection of large data sets in 2-dimensional projections. Classifications can be found automatically or interactively by user selections.
We prove its usefulness for ecological data in applications to several data sets on water ecology in some small rivers in central Germany.

Applying case-based reasoning to explore freshwater phytoplankton dynamics

P.A. Whigham

Information Science Department, University of Otago, PO Box 56, Dunedin, New Zealand,
Email: pwhigham@infoscience.otago.ac.nz

The prediction and explanation of algal abundance and succession is of major interest to freshwater ecologists. The use of machine learning techniques, such as rule induction and neural networks, has been successfully applied to time series data of freshwater systems. Case-based reasoning (CBR) is a machine learning technique based on concept formation and similarity measures that allows the user to incorporate background knowledge and interrogate explanations of predicted values. Although CBR has been applied to many domains it has only been applied in a limited number of cases to freshwater systems, and to time-based modeling in particular. This paper describes the application of CBR to predicting chlorophyll-*a* concentration and *Microcystis* abundance for a freshwater lake environment, based on water quality variables, light, temperature and zooplankton populations. The ability for CBR to measure concepts of similarity allows an exploration of conditions that occur during different bloom periods and to support process understanding by exploring the variables that give significant improvement to prediction. This paper demonstrates the capability of CBR for time series ecological data, suggests several directions for research and concludes that CBR has a significant role to play in the domain of predictive freshwater ecology.

Exploring seasonal patterns using process modelling and evolutionary computation

P.A. Whigham[1], G. Dick[2] and F. Recknagel[3]

[1] *University of Otago, Department of Information Science, Dunedin, New Zealand, Email: pwhigham@infoscience.otago.ac.nz*
[2] *University of Otago, Department of Information Science, Dunedin, New Zealand*
[3] *University of Adelaide, Department of Soil and Water, Waite Campus, Glen Osmond, South Australia 5064*

Process models of chlorophyll-*a* concentration for freshwater systems, and in particular lake environments, have been developed over many years. Previous work has demonstrated that the optimisation of constants within these models has been able to significantly improve the quality of the resulting model on unseen data. This paper explores two properties of one particular process model: can the model predictions be improved by optimising the constants over different temporal scales; and can seasonal patterns be identified, based on monthly training scales, that allow further understanding of the response of the freshwater system to changing environmental conditions. The results show that there is some improvement on the prediction of unseen data when using constants of the process model optimised for individual months, versus constants trained over a yearly cycle. Additionally, by studying the patterns of the constants over various time scales some underlying seasonal patterns can be observed. These patterns can be further studied by exploring how the various elements of the process model vary with monthly versus yearly training constants. This work demonstrates some possible directions for understanding how the behaviour of freshwater systems at different time scales can be used to understand the properties of these complex, non-linear systems.