Environmental Science and Engineering
Subseries: Environmental Engineering

Series Editors: Ulrich Förstner • Robert J. Murphy • W.H. Rulkens

Ioannis N. Athanasiadis · Pericles A. Mitkas ·
Andrea E. Rizzoli · Jorge Marx Gómez (Eds.)

Information Technologies in Environmental Engineering

Proceedings of the 4th International ICSC Symposium, Thessaloniki, Greece, May 28–29, 2009

Springer

DRAXIS

Editors
Dr. Ioannis N. Athanasiadis
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA)
Galleria 2
6928 Manno
Switzerland
ioannis@athanasiadis.info

Prof. Andrea E. Rizzoli
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA)
Galleria 2
6928 Manno
Switzerland
andrea@idsia.ch

Pericles A. Mitkas
Aristotle University of Thessaloniki
Greece

Prof. Dr.-Ing. Jorge Marx Gómez
Universität Oldenburg
Fak. Informatik
Abt. Wirtschaftsinformatik
Ammerländer Heerstr.
114-118
26129 Oldenburg
Germany
marx-gomez@informatik.uni-oldenburg.de

ISSN 1863-5520
ISBN 978-3-540-88350-0 e-ISBN 978-3-540-88351-7
DOI 10.1007/978-3-540-88351-7
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009926966

© Springer-Verlag Berlin Heidelberg 2009
This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.
The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: deblik, Berlin

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

ITEE-2009 Organizers

Chairs

Dr. Ioannis N. Athanasiadis
Dalle Molle Institute for Artificial Intelligence, Lugano, Switzerland

Prof. Pericles A. Mitkas
Aristotle University of Thessaloniki, Thessaloniki, Greece

Dr. Andrea-Emilio Rizzoli
Dalle Molle Institute for Artificial Intelligence, Lugano, Switzerland

Prof. Jorge C. Marx Gómez
Carl von Ossietzky University, Oldenburg, Germany

Program Committee

Abdullah Al-Zakwani
Aladdin Ayesh
Volker Barth
Bettina Bauer-Messmer
Antonio-Fernández Caballero
Quim Comas
Ioannis Dokas
Marcello Donatelli
Burkhardt Funk
Omar El-Gayar
Michael Georgiadis
Adrian Giurca
Vladimir Gorodetski
Renata S.S. Guizzardi
Geoffrey Holmes
Ralf Isenmann
Sander Janssen
Horst Junker
Vassilis Kaburlasos
Kostas Karatzas
Dimitrios Karras
Ayalew Kassahun
Dionisis Kehagias
Mikko Kolehmainen
Evaggelos Kosmidis
Konstantinos Kotis
Manolis Koubarakis
Adamantios Koumpis
Peter Krause
Helmut Lessing
Nikos Manouselis
Alexandros Mentes
Andreas Möller
Maurice Mulvenna
Yannis Mylopoulos
Peter Niemeyer
Carmel Polino
Nigel Quinn
Friedrich Recknagel
Miquel Sànchez-Marrè
Arno Scharl
Sahra Sedigh
Sabine Sedlacek
Athanasios Sfetsos
Michael Sonnenschein
Andreas Symeonidis
Panagiotis Symeonidis
Fani Tzima
Frank Teuteberg
Sotirios Tsaftaris
Monica Wachowicz
Gerd Wagner
Patrik Wallman
Jan-Erik Wien

Preface

This book is the result of the 4th International ICSC Symposium on Information Technologies in Environmental Engineering (ITEE-2009). Recent success stories in ecoinformatics, promising ideas and new challenges are discussed among computer scientists, environmental engineers, economists and social scientists, who showcase new computing paradigms for environmental problem solving and decision making.

The Symposium was held in Thessaloniki, Greece, in May 28-29, 2009. Local arrangements provided by the members of the Information Processing Laboratory of the Electrical and Computer Engineering Dept, at Aristotle University of Thessaloniki. Special thanks goes to Fani Tzima for her unreserved efforts towards the success of the Symposium.

Editors would like to express their gratitude to DRAXIS SA, that sponsored the publication of the present book, and personally Dr. Evangelos Kosmidis for his wholehearted support.

Finally, we would like to thank the personnel and the directors of the Macedonian Museum of Contemporary Art for their kind hospitality.

Ioannis N. Athanasiadis
Pericles A. Mitkas
Andrea E. Rizzoli
Jorge Marx Gómez

Thessaloniki, May 2009

Table of contents

Keynotes

Ecological Informatics: Current Scope and Future Directions 3
F. Recknagel

Semantically driven meta-modelling: automating model construction in an environmental decision support system for the assessment of ecosystem services flows 23
F. Villa

Regular papers

Session I: Environmental information systems

Governance of a Service-Oriented Architecture for Environmental and Public Security 39
A. Papageorgiou, S. Schulte, D. Schuller, M. Niemann, N. Repp and R. Steinmetz

Integration of Environmental Management Information Systems and ERP systems using Integration Platforms 53
B. Funk, A. Möller and P. Niemeyer

State of the art and future research in Environmental Management Information Systems – a systematic literature review 64
F. Teuteberg and J. Straßenburg

AKX – An Exchange for Predicting Water Dam Levels in Australia 78
S. Stathel, S. Luckner, F. Teschner, C. Weinhardt, A. Reeson and S. Whitten

Environmental Engineering and Facility Management of Berlin’s Public Buildings 91
V. Stantchev and R. Meyer

Spatial disaggregated emissions inventory for Cyprus 105
D. Gkatzoflias, G. Mellios, C. Kouridis, M. Theodosopoulou, S. Kleanthous and Z. Samaras

Session II: Environmental decision support systems

A web-based information system rating regional projects by considering sustainable development 120
N. Giesen and J.M. Gómez

Development of a Spatial Decision Support Framework for IMPETUS project in West Africa 132
A. Enders and B. Diekkrüger

Development of a decision support system for the operation of thermal power plants in Western Macedonia 149
A. Sfetsos, A. Syrakos, M. Politis, G. Efthimiou, N. Gounaris, D. Vlachogiannis, N. Koziakis, J.G. Bartzis, G. Nikolaou and M. Voutsinas

Towards a Novel Early Warning Service for State Agencies: A Feasibility Study.. 162
I.M. Dokas, R.J. Wallace, R. Marinescu, S. Imran and F. Foping

Session III: Tools and measurement techniques

Application of Optimization Algorithms for the Improvement of Water Quality Monitoring Systems 176
M.G. Erechtchoukova, S.Y. Chen and P.A. Khaiter

SQCB - Sustainability Quick Check Tool for Biofuels 189
T. Ziep, V. Wohlgemuth, M.F. Emmenegger, J. Reinhard and R. Zah

Volatile Organic Compounds measurements at schools in Kozani, Greece........ 203
D.A. Missia, E.I. Tolis, D. Karanikolas, C. Kosmidis, A. Markou, S. Mavrovouniotou and J.G. Bartzis

NoiseTube: Measuring and mapping noise pollution with mobile phones.......... 215
N. Maisonneuve, M. Stevens, M.E. Niessen and L. Steels

Monitoring of environment by energy efficient usage of Wireless Sensor Networks 229
S. Gancev and D. Davcev

Plant Phenotyping with Low Cost Digital Cameras and Image Analytics 238
S.A. Tsaftaris and C. Noutsos

Session IV: Formal methods; data processing and data mining (part 1)

Reproducing human decisions in reservoir management: the case of lake Lugano 252
G. Corani, A.E. Rizzoli, A. Salvetti and M. Zaffalon

Predicting chemical parameters of the water from diatom abudance in lake Prespa and its tributaries 264
A. Naumoski, D. Kocev, N. Atanasova, K. Mitreski, S. Krsti and S. Džeroski

Analysing chromatographic data using data mining to monitor petroleum content in water 278
G. Holmes, D. Fletcher, P. Reutemann and E. Frank

An Experimental Evaluation of ZCS-DM for the Prediction of Urban Air Quality 291
F.A. Tzima, H. Niska, M. Kolehmainen, K.D. Karatzas and P.A. Mitkas

An investigation of the effectiveness of advanced modeling tools on the forecasting of daily PM10 values in the Greater Athens Area 305
A. Sfetsos and D. Vlachogiannis

Session V: Formal methods; data processing and data mining (part 2)

Traffic Related PM Predictor for Besiktas, Turkey 317
F. Karaca, I. Anil, O. Alagha and F. Camci

Development and Evaluation of Data Mining Models for Air Quality Prediction in Athens, Greece 331
M. Riga, F.A. Tzima, K. Karatzas and P.A. Mitkas

Artificial neural networks based modelling of carbon monoxide: Effects of spatial parameters 345
A. Kurt, A.B. Oktay, F. Karaca and O. Alagha

Using Preprocessing Techniques in Air Quality forecasting with Artificial Neural Networks 357
I. Kyriakidis, K.D. Karatzas and G. Papadourakis

Approach for Dynamic Problems in Clustering 373
A. Minke, K. Ambrosi and F. Hahne

Session VI: Tools for environmental policy and management

Stakeholder Interaction in Sustainability Reporting with Web 2.0 387
D. Süpke, J. Marx-Gómez and R. Isenmann

Meta-modelling for cross-sectoral land use policy impact analysis: A model requirements analysis of the SIAT 399
S. Sieber, K. Müller, K. Tscherning, K. Helming, D. Pohle, K. Fricke and P. Verweij

An interdisciplinary approach to resolving conflict in the water domain 411
M. Nastar and P. Wallman

Conceptual design and implementation of a model for the integrated simulation of large-scale land-use systems 425
R. Schaldach and J. Koch

Enterprise systems applications development plans to build and develop robust and environmentally sensitive green logistics systems management software 439
T. Yildiz and F. Yercan

Session VII: Modelling and simulation

An investigation of the parameters influencing the determination of the number of particulate matter sources and their contribution to the air quality of an indoor residential environment 453
D.E. Saraga, A. Sfetsos, S. Andronopoulos, A. Chronis, T. Maggos, D.Vlachogiannis and J.G. Bartzis

Evaluating long-term contamination in soils amended with sewage sludge 465
A. Passuello, A. Bojarski, M. Schuhmacher, L. Jiménez and M. Nadal

Evaluation by simulation of the social acceptability of agricultural policies for water quality 478
C. Vautier, P. Roggero, F. Adreit and C. Sibertin-Blanc

COPERT: A European Road Transport Emission Inventory Model 491
L. Ntziachristos, D. Gkatzoflias, C. Kouridis and Z. Samaras

Session VIII: Environmental knowledge, software agents and the semantic web

A survey of knowledge Organization Systems in Environmental Sciences 505
N. Palavitsinis and N. Manouselis

Semantic modeling of temporal information in a database for biotope inventories 518
B. Bauer-Messmer and R. Grütter

Environmental notions representation and description: Towards a redefinition of the relationships between information systems development and individual cognition 535
H. Ravononarimanga-Raherimandimby, D. Le Boulch, D. Bouyssou and M. Grundstein

Multi-agent-based System Technologies in Environmental Issues 549
M.V. Sokolova and A. Fernández-Caballero

Specimen's behavior management: A masterpiece for Biodiversity Information System's representation 563
D. Sébastien, R. Courdier, N. Conruyt, N. Sébastien and T. Tanzi

MACE – Multiagent Control for Energy Infrastructures 579
S. Beer and M. Tröschel

Concluding talk

Corporate Environmental Management Information Systems (CEMIS) for Sustainable Development 595
H. Junker

Keynotes

Ecological Informatics: Current Scope and Future Directions

Friedrich Recknagel

Ecology and Evolutionary Biology, University of Adelaide
Adelaide 5000, Australia

Abstract

Ecological informatics emerges as a new discipline that studies principles of information processing in ecosystems as well as data analysis and synthesis for hind- and forecasting of ecosystems. It also focuses on integration and sharing of ecological data from genomic to landscape levels at different spatial scales by web-based data warehousing, GIS and remote sensing. Ecological informatics takes advantage of steadily advancing computational technology in order to better cope with extreme complexity and distinct nonlinearity of ecological data. It utilises cellular automata, neural, evolutionary and immunological computing to unravel ecological complexity as well as explain and forecast ecosystem responses to habitat and climate change.

1 Introduction

Ecological informatics (ecoinformatics) is an interdisciplinary framework for the management, analysis and synthesis of ecological data by advanced computational technology (Recknagel 2006). Management of ecological data aims at facilitating data standardization, retrieval and sharing by means of metadata and object-oriented programming (e.g. Eleveld, Schrimpf and Siegert 2003; Michener 2006). Analysis and synthesis of ecological data aim at elucidating principles of information processing, structuring and functioning of ecosystems, and forecasting of ecosystem

I.N. Athanasiadis et al., *Information Technologies in Environmental Engineering*,
Environmental Science and Engineering, DOI 10.1007/978-3-540-88351-7_1,
© Springer-Verlag Berlin Heidelberg 2009

behaviours by means of bio-inspired computation and hybrid models (e.g. Fielding 1999; Recknagel 2006).

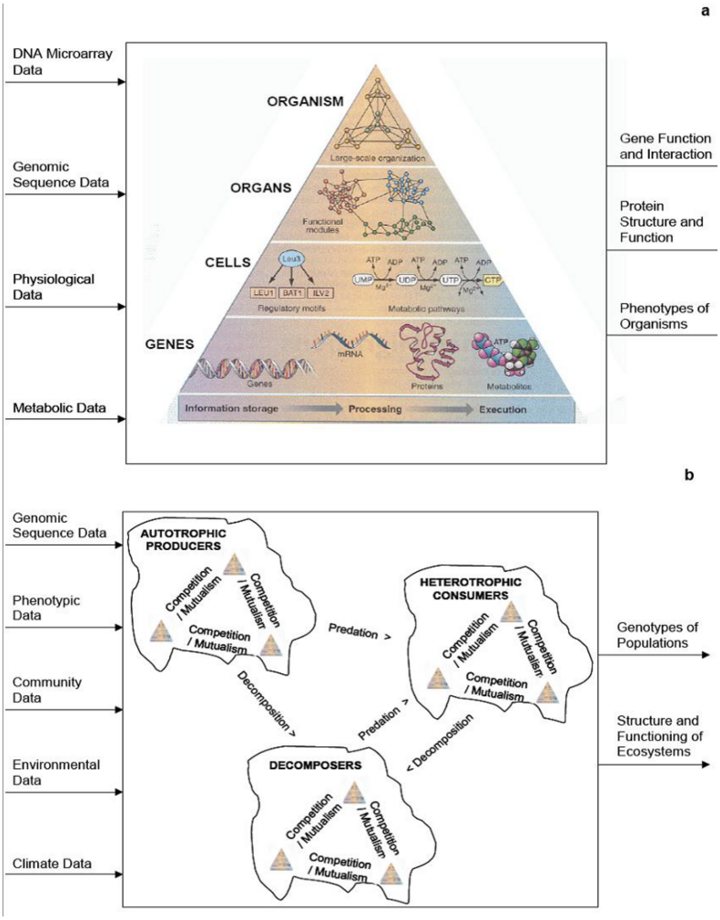

Fig. 1: Ecological informatics versus bioinformatics (from Recknagel 2006) , a) Scope of bioinformatics, b) Scope of ecoinformatics

Ecological informatics currently undergoes the process of consolidation as a discipline. It corresponds and partially overlaps with the well-established disciplines bioinformatics and ecological modeling but is taking its distinct shape and scope. In Fig. 1 a comparison is made between ecological informatics and bioinformatics. Even though both are based on the same computational technology their focus is different. Bioinformatics focuses very much on determining gene function and interaction (e.g. Wolf et al. 2001), protein structure and function (e.g. Henikoff *et al.* 1999) as well as phenotypes of organisms utilizing DNA microarray, genomic, physiological and metabolic data (e.g. Lockhardt and Winzeler 2000) (Fig. 1a). By contrast ecological informatics focuses to determine genotypes of populations by utilizing genomic, phenotypic and environmental data (e.g. Doney 2004) as well as structure and functioning of ecosystems by utilizing community, environmental and climate data (e.g. Lek *et al.* 2005) (Fig. 1b).

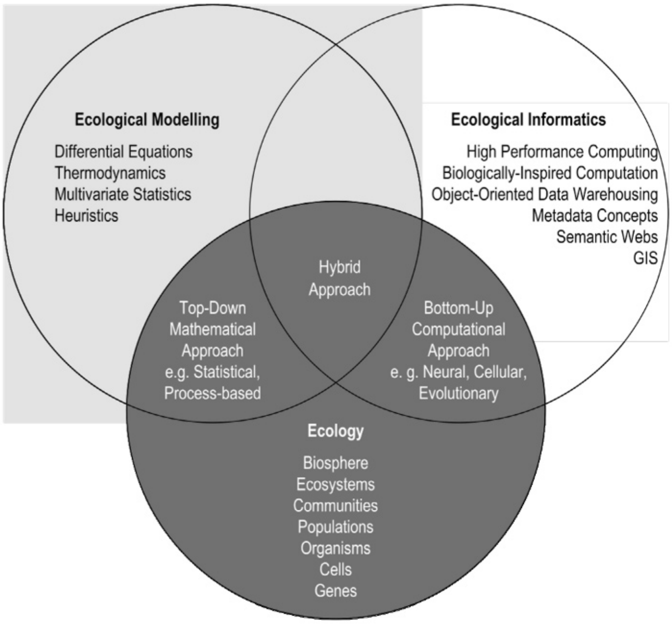

Fig. 2: Ecological informatics versus ecological modeling from Recknagel (2006)

A comparison is made between ecological modeling and ecological informatics in Fig. 2. Even though both rely on similar ecological data they adopt different approaches in utilizing the data. Whilst ecological modeling processes ecological data top down by ad hoc designed statistical or mathematical modeling methods (e.g. Jorgensen, Chon and Recknagel 2009), ecological informatics infers ecological processes from ecological data patterns bottom up by computational techniques. The cross-sectional area between ecological modeling and ecological informatics reflects a new generation of hybrid models that enable to predict emergent ecosystem structures and behaviours, and ecosystem evolution (e.g. Hraber and Milne 1997). Typically hybrid models integrate biologically-inspired computation and deterministic ecological models.

2 Feature Areas

Current research in ecological informatics focuses at four major feature areas: (1) understanding information processing and evolution in ecosystems, (2) computational management of ecological data,
(3) computational analysis and synthesis of ecological data, and (4) hybrid modelling of ecological data.

Great efforts are undertaken to address feature area (1) by studying both intraspecific population adaptations to changing climate and habitat conditions (e.g. Hairston et al. 1999) as well as interspecific population relationships controlled by info chemicals and allelopathy (e.g. Voss et al. 2006; van Donk 2007).

The feature area (2) aims at standardised archiving of highly complex and fragmented ecological data in order to allow ecological data sharing. The ecological metadata language EML (http://knb.ecoinformatics.org /software/eml/) is an example for developing computational tools based on meta data concepts (e.g. Michener 2006) that will facilitate ecological data warehousing at global scale.

The feature area (3) has been largely stimulated by both the availability of complex ecological data including genomic and phenotypic data, and the development of bio-inspired computational techniques. The study of population genomics in their natural habitats without the need for isolation and lab cultivation of individual species has led to the new research area of ecogenomics that promises to determine the impact of environmental and climate changes on biodiversity (e.g. Doney et al. 2004). Bio-inspired computational techniques prove to be superior in unravelling highly complex ecological data, coping with distinct nonlinearities and inducing predictive models by learning from temporal and spatial patterns. Section 2.1. illustrates applications of artificial neural networks and evolutionary algorithms for ecological informatics by.

Research on hybrid modelling in the feature area (4) promises ecosystem models with improved accuracy and generality. Cao and Recknagel (2009) provide a case study for multi-objective optimisation of process and parameter representations in process-based ecosystem models by the embodiment of evolutionary algorithms in ordinary differential equations for food web dynamics and nutrient cycles in lakes. Chen and Mynett (2006) integrated cellular automata and fuzzy logic to simulate spatio-temporal dynamics of algal blooms in coastal waters,

3 Ecological informatics by computational analysis and synthesis of ecological data

3.1 Artificial neural networks (ANN)

Artificial neural networks are computer programs designed for inducing problem solutions (models, knowledge) from complex data by means of principles of information processing similar to biological neurons in the human brain. A biological neuron consists of three major component: the cell body, dendrites and the axon (Fig. 3a). Connections between neurons are formed at synapses. Information is represented and transmitted by chemically generated electrical activity within the cell. Both excitatory and inhibitory inputs to the neuron enter through synaptic connections with other neurons. Input potentials are summed up within the cell body. If the total input potential is sufficient (e.g. meets a certain threshold value) then the neuron acts. Ultimately an action potential is generated and propagated down the axon towards the synaptic junctions with other nerve cells.

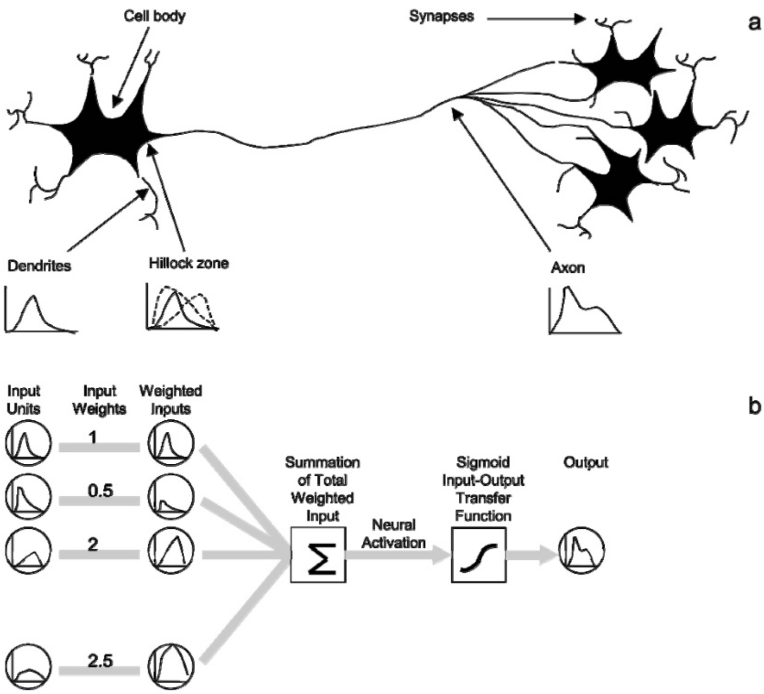

Fig. 3: Conceptual structures of biological and artificial neurons

The design of artificial neural networks (Fig. 3) has been inspired by the structure and functioning of biological neurons. The dendrites which are acting as input receptors were represented by input units. The cell body that acts as information accumulator was represented by activation units adjusting and summing up the weights of inputs, and the input-output transfer function. The axon that acts as the biological output channel was represented as the output.

ANN gain there adaptive capability by undergoing training similar to neural learning where two basic training modes are distinguished: supervised and non-supervised training. The supervised training aims at the optimal approximation of the calculated output Y_c to the observed (desired) output Y_o. An iterative adjustment of input weights takes place in order to minimise the error (Y_o - Y_c).

After training, the generalisation of the supervised ANN is as assessed by feeding it only with input values, not observed output values, and testing how close calculated outputs match observed outputs. The two most common methods for assessing generalisation are the *split-sample validation* and the *cross-validation*. The *split-sample validation* means that part of the data is reserved as a test set, which must not be used in any way during training. The test set must be representative for the problem to be modelled by the ANN. After training, the ANN is run on the test set, and the error on the test set provides an estimate of the generalization error usually expressed by the root mean square error (RMSE) or the correlation coefficient r^2. The disadvantage of split-sample validation is that it reduces the amount of data available for both training and validation (Weiss and Kulikowski 1991). By contrast *cross-validation* allows you to use all of the data for training. In *k*-fold cross-validation, the data is divided into k equal sized subsets. The net is trained *k* times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the generalisation error. If *k* equals the sample size, this is called "leave-one-out" cross-validation. The disadvantage of cross-validation is that the ANN need to be retrained many times.

Depending on using external inputs only or feedback inputs as well, supervised ANN are differentiated into feedforward or feedback ANN (see Fig. 4 a and b). By contrast non-supervised ANN process external inputs only without adjusting calculated outputs to known outputs (Fig. 4c).

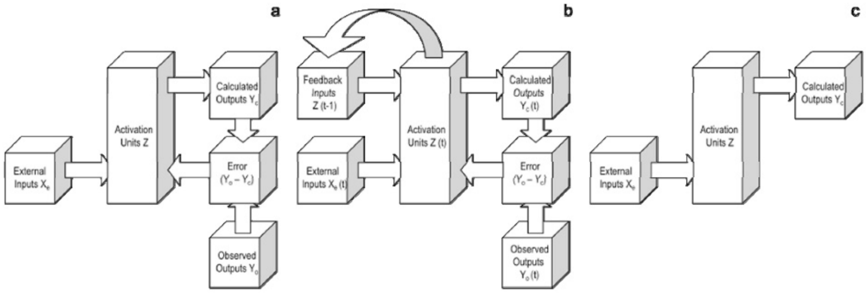

Fig. 4: Basic types of artificial neural networks (ANN): a) Supervised feedforward ANN; b) Supervised feedback ANN; c) Non-supervised ANN

3.1.1 Supervised feedforward ANN

The supervised feedforward ANN proves to be a universal approximator of multivariate nonlinear functions and is usually implemented as multi-layer perceptron with back-propagation training. The multi-layer perceptron (Minski and Pappert 1969) represents input units as input layer, adjusted and accumulated input weights as hidden layer(s) and outputs as output layer. The back-propagation algorithm (Rummelhardt *et al.* 1986) performs the iterative adjustment of input weights (activation units) in order to minimise the approximation error (Y_o - Y_c).

Supervised feedforward ANN are widely applied in ecology either using cross-sectional data to predict discrete ecosystem states or using time-series data to predict continuous ecosystem behaviour. Successful applications by means of cross-sectional data have been demonstrated for fish communities in streams (e.g. Lek et al. 1996), macroinvertebrate communities in streams (e.g. Walley and Fontana 1998), river salinity (e.g. Huang and Foo 2002), primary productivity in estuaries (e.g. Scardi 1996), chlorophyll *a* concentrations in lakes (e.g. Karul and Soyupak 2006), coastal vegetation (e.g. Foody 2000) and bird populations (e.g. Lusk et al. 2001).

Successful applications by means of time-series data have been demonstrated for marine fish and zooplankton communities (e.g. Aoki and Komatsu 1997; Reick, Gruenewald and Page 2003), river hydrology (e.g. Poff, Tokar and Johnson 1996), macroinvertebrate communities in streams (e.g. Schleiter et al. 2006), freshwater phyto- and zooplankton communities (e.g. Recknagel 1998).

The majority of the supervised feedforward ANN documented in the above mentioned papers achieved forecasting results that were superior to conventional modelling techniques such as multiple linear regression (e.g. Lek et al. 1996; Karul and Soyupak 2003). Even though supervised ANN don't provide explicit mathematical representations of the underlying ANN model, most of the authors have conducted sensitivity analyses in order to identify inputs as key driving forces of the predictive ANN. An example for revealing input-output relationships by both sensitivity and scenario analysis was documented in Recknagel and Wilson (2000).

3.1.2 Supervised feedback ANN

Supervised feedback or recurrent ANN (Pineda 1987) are designed to use not only external inputs for training but also activation levels of the previous training iteration which are constantly fed back (see Fig. 4b). Their functioning can be compared with ordinary differential equations that calculate the current system state Z (t) by taking into account current external inputs $X_e(t)$ and the system state Z(t-1) of the time step before:

$$dZ(t)/dt = f(X_e(t), Z(t-1), P)$$

where P are constant parameters.

Supervised feedback ANN prove to be very powerful for modelling time-series data where the fed back activation levels provide extra training information on the system state of the time step before.

The Fig. 5 shows an example for a supervised feedback ANN that has successfully been trained and tested by split-sample validation for the forecasting of the algal populations *Microcystis* and *Stephanodiscus* in the River Nakdong in South Korea (Jeong, Recknagel and Joo 2006).

The weekly measured limnological data of the river study site were interpolated to daily values. The interpolated data from 1995 to 1998 were used as training set, and the interpolated data of 1994 were used as testing set. In order to achieve a 4-days-ahead forecasting a four days time lag was imposed between the measured inputs and the measured outputs of the training data set. The design of the feedback ANN considered the following 18 external input variables: irradiance, precipitation, discharge, evaporation, water temperature, Secchi depth, turbidity, pH, DO, nitrate, ammonia, phosphate, silica, rotifera, cladocera, copepoda, 21 hidden activation units and the two output variables: *Microcystis aeruginosa* and *Stephanodiscus hantzschii.*

After 2,100 training iterations a root mean square error (RMSE) of 0.0017 was achieved and the generalization of the trained ANN was tested based on testing data of 1994. The Fig. 6 shows the visual comparison between the observed and the 4-days-ahead predicted data for *Microcystis aeruginosa* (r^2= 0.68) and *Stephanodiscus hantzschii* (r^2= 0.73). The results indicate a high degree of accuracy in the forecasting regarding both the timing and the magnitudes of populations dynamics of the two algal species, which have their distinctive seasonal patterns.

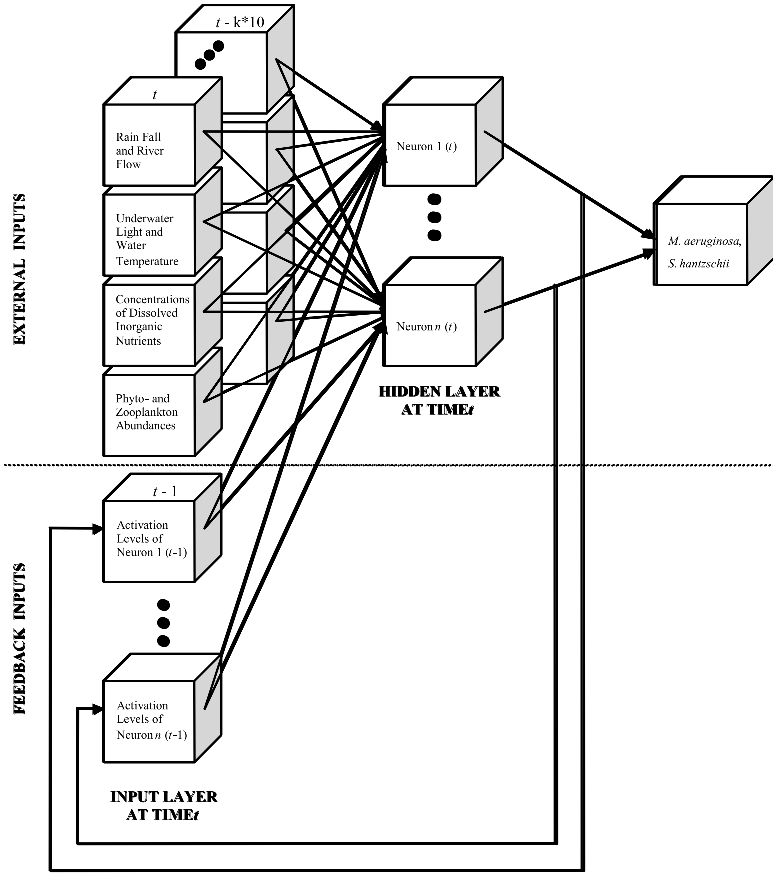

Fig. 5: Supervised feedback ANN for four-days-ahead forecasting of population densities of Microcystis aeruginosa and Stephanodiscus hantzschii in the River Nakdong (South Korea) from Jeong, Recknagel and Joo (2006)

Microcystis under warm and calm conditions in mid and late summer as observed in the River Nakdong in 1994 were well reflected by the predicted data in Fig. 6a. By contrast diatoms tend to be abundant at moderate temperatures and turbulent conditions. Both observed and predicted data for *Stephanodiscus hantzschii* in the River Nakdong correspond well by showing highest population densities in spring and autumn (Fig. 6b).

Successful applications have been demonstrated for time-series modelling of macroinvertebrate communities in streams by e.g. Chon et al. (2006) and of phytoplankton communities in freshwater lakes and rivers by e.g. Recknagel et al. (2006).

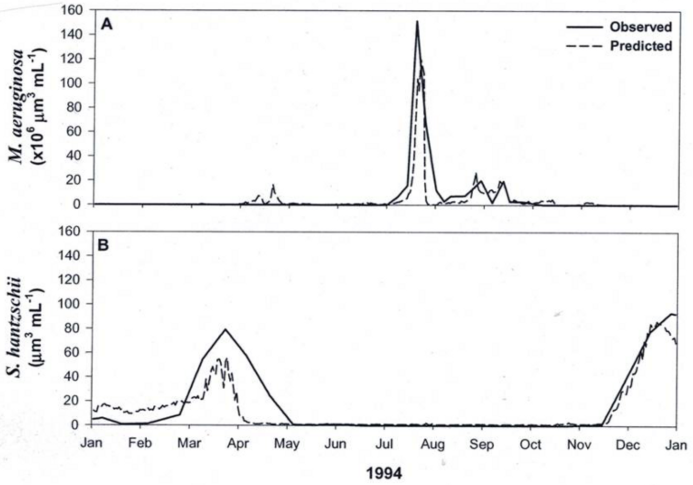

Fig. 6: Four-days-ahead forecasting of population densities of *Microcystis aeruginosa* and *Stephanodiscus hantzschii* in the River Nakdong (South Korea) by means of a supervised feedback ANN from Jeong, Recknagel and Joo (2006)

3.1.3 Non-supervised ANN

Non-supervised ANN are designed to identify unknown input patterns based on similarities between inputs. So-called self-organising maps developed by Kohonen (1989) are the most popular non-supervised ANN, which can be applied to ordination, clustering and mapping of complex non-linear data.

The principal approach of non-supervised ANN according to Kohonen (1989) is represented in a simplified manner in Fig. 7. It shows that the neurons of the non-supervised ANN learn to distinguish between similar and dissimilar features of the normalised input data, which are mapped as clustered inputs. The term non-supervised in this context means that the learning algorithm is not guided by known output patterns but learns the patterns from features of the inputs. Those features can be expressed by Euclidean distances, which are calculated between the inputs and weights. Similarities between inputs in terms of Euclidean distances can be visualised and partitioned by the unified distance matrix (U-matrix) and the K-means map.

In order to illustrate opportunities of applications of non-supervised ANN to ecological time-series data, Figs. 8 to 10 show results of a case study carried out for limnological data of Lake Kasumigaura in Japan (Recknagel *et al.* 2006). The Fig. 8 represents seasonal clusters for Lake Kasumigaura as mapped by the U-matrix and K-means partitioning using the SOM Toolbox of MATLAB 5.3 (Vesanto *et al.* 2000). The U-matrix map in Fig. 8a visualises the relative distances between neighbouring data of the input data space as shades of grey. The light areas in the U-matrix visualise neighbouring data with smallest distances belonging to a region or cluster. The black colours represent the biggest distances between neighbouring data and denote borders between clusters. The K-means algorithm partitions the input data space into a specified number of clusters based on the U-matrix. Fig. 8b represents the corresponding partitioned map for five seasons.

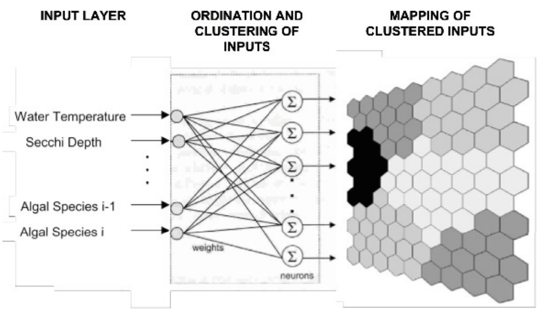

Fig. 7: Conceptual diagram of the structure and functioning of non-supervised ANN

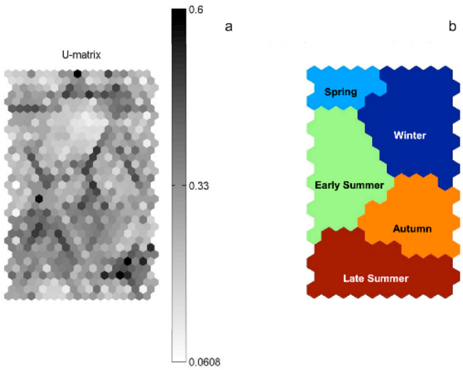

Fig. 8: Ordination and clustering of seasons of Lake Kasumigaura by means of non-supervised ANN visualised as unified distance matrix map (U-matrix) (a), and as partitioned map (K-means) (b); the seasons were defined as follows: winter from December 1st, spring from March 15th, early summer from June 1st, late summer from August 1st, autumn from October 1st

The Fig. 9 visualises seasonal distributions of abundances of the blue-green algae *Microcystis* and *Oscillatoria* in Lake Kasumigaura based on data of the years 1984 to 86 (left column) and 1987 to 89 (right column). The Fig. 10 represents the seasonal distributions of concentrations of NO_3-N and PO_4-P in Lake Kasumigaura in correspondence with the time periods differentiated in Fig. 8. Fig. 9 highlights that whilst Microcystis declines in cell numbers by more than 50% between 1984 to 86 and 1987 to 89, Oscillatoria doubles in cell numbers. It also shows that seasonal dominance of two algal populations for the early and the late 1980s shifted for Microcystis from late summer to autumn, and for Oscillatoria from early summer to late summer. Takamura et al. (1992) pointed at changes of NO_3-N/PO_4-P ratios as possible explanations for the succession of the two blue-green algal populations during the 1980s in Lake Kasumigaura, that are indicated by the component planes in Fig. 9. From the early 1980s to the late 1980s the NO_3-N concentrations increase by 50% whilst PO_4-P concentrations dropped to 50% causing a significant change of the NO_3-N/PO_4-P ratios from 8.5 to 32.

Successful applications of non-supervised ANN have been demonstrated for cross-sectional data of macroinvertebrate communities in streams (e.g. Chon et al. 1996) and vegetation types (e.g. Foody 2000).

Successful applications of non-supervised ANN have been demonstrated for time-series data of plankton communities in lakes and rivers (e.g. Recknagel, Talib and van der Molen 2006).

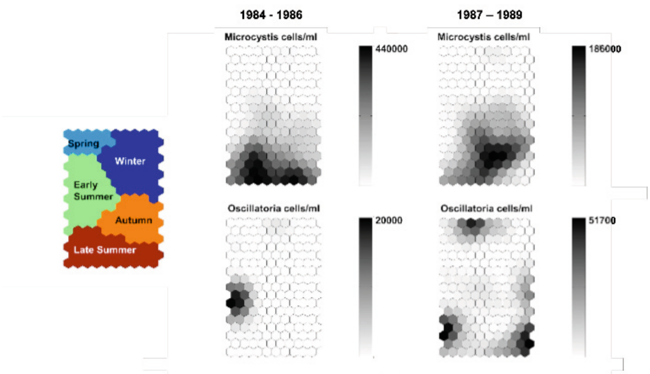

Fig. 9: Component planes for seasonal abundances of *Microcystis* and *Oscillatoria* populations in Lake Kasumigaura for the years 1984 to 86 (left column) and 1987 to 89 (right column)

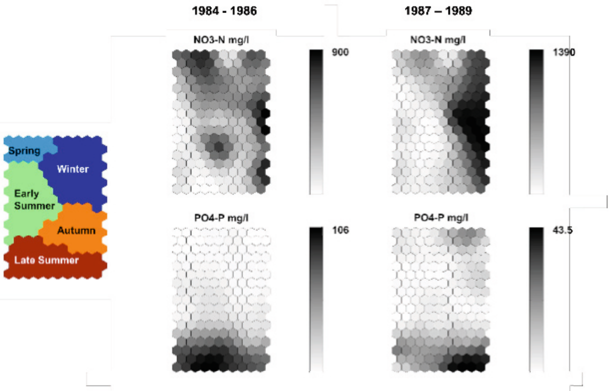

Fig. 10: Component planes for seasonal concentrations of PO_4-P and NO_3-N in Lake Kasumigaura for the years 1984 to 86 (left column) and 1987 to 89 (right column)

3.2 Evolutionary algorithms

Evolutionary algorithms (EA) are adaptive methods for finding problem solutions (models, knowledge) based on principles of biological evolution by natural selection, genetic variation and "survival of the fittest". Holland (1975) provided the theoretical framework for the development of genetic and evolutionary algorithms that are being widely used for pattern recognition, forecasting, knowledge discovery, optimum control and parallel processing. Useful guides for history, current developments and applications of genetic and evolutionary algorithms are provided by Goldberg (1989).

Successful implementations of EA as tools for solving complex economic and engineering problems have stimulated their application to solving ecological problems, which exhibit highest complexity. They allow to induce predictive models from ecological data sets similar to supervised ANN but rather than lacking an explicit model representation as typical for ANN, EA are distinctively designed for assembling the explicit model represented as multivariate functions or rule sets. Therefore EA serve as powerful tools for knowledge discovery as well.

The hybrid evolutionary algorithms (HEA) (Cao et al. 2006) has been *ad hoc* designed as flexible tool for inducing predictive multivariate functions and rule-sets from ecological time-series data. The detailed algorithm for the rule discovery and parameter optimization by HEA is shown in Fig. 11. HEA uses genetic programming (GP) to generate and optimize the structure of rule sets and a genetic algorithm (GA) (e.g. Mitchell 1996) to optimize the parameters of a rule set. GP (e.g. Banzhaf *et al.* 1997) is an extension of GA in which the genetic population consists of computer programs of varying sizes and shapes. In standard GP, computer programs can be represented as parse trees, where a branch node represents an element from a function set (arithmetic operators, logic operators, elementary functions of at least one argument), and a leaf node represents an element from a terminal set (variables, constants and functions of no arguments). These symbolic programs are subsequently evaluated by means of "fitness cases". Fitter programs are selected for recombination to create the next generation by using genetic operators, such as crossover and mutation. This step is iterated for consecutive generations until the termination criterion of the run has been satisfied. A general genetic algorithm (GA) is used to optimize the random parameters in the rule set. More details on the design and functioning of HEA including a demo version can be found in Cao *et al.* (2006).

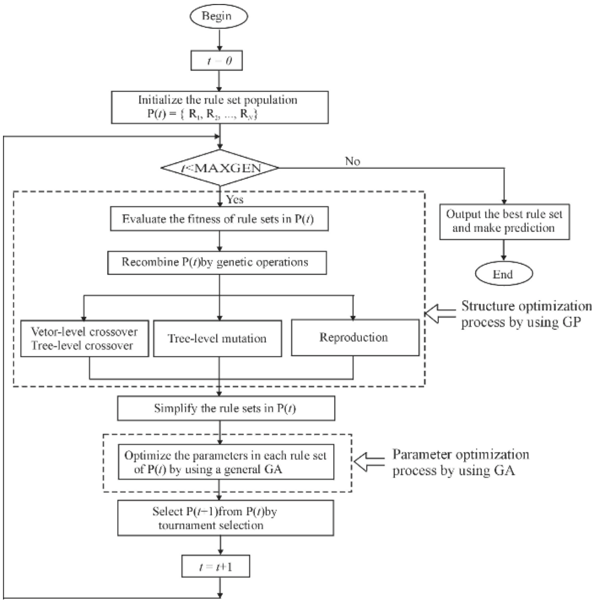

Fig. 11: Flowchart of the hybrid evolutionary algorithm HEA for rule discovery from Cao et al. (2006)

The Figs. 12 and 13 illustrate the structure, input sensitivity and *k*-fold cross-validation of a rule-based agent for 7-days-ahead forecasting of *Microcystis* biomass developed by HEA (Recknagel et.al. 2008).

The rule in Fig. 12a is the result of using 42 years of merged limnological data of the South African lakes Hartbeespoort, Roodeplaat and Rietvlei for the training of HEA. The sensitivity analysis in Fig. 12b indicates that both water temperature and Secchi depth are key driving variables for low biovolumes of *Microcystis* of up to 14 cm^3/m^3 reflected by the THEN branch of the rule as well as for high biovolumes of up to 350 cm^3/m^3 reflected by the ELSE branch of the rule. As a result of *k*-fold cross-validation the parameters p_1 and p_2 have been evolved to water temperature functions which provide the agent an extra mechanism for adaptation to lake specific seasonal conditions.

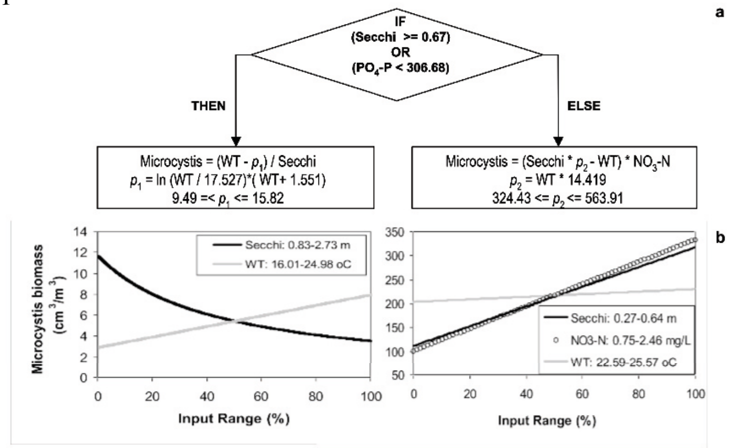

Fig. 12: Structure and input sensitivity analysis of a rule-based agent for 7-days-ahead forecasting of *Microcystis* biomass discovered in merged time-series data of the South African lakes Hartbeespoort, Roodeplaat and Rietvlei by HEA (from Recknagel et.al. 2008)

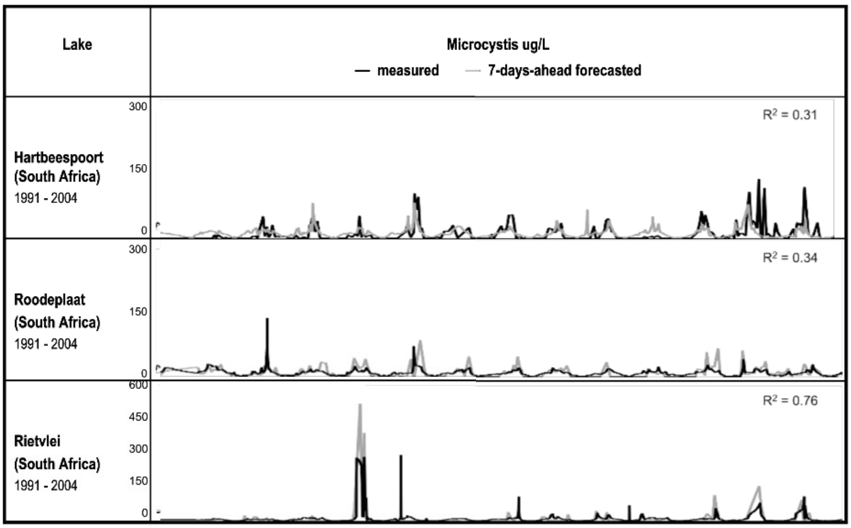

Fig. 13: *k*-fold cross-validation of a rule-based agent for 7-days-ahead forecasting of *Microcystis* biomass by means of merged time-series data of the South African lakes Hartbeespoort, Roodeplaat and Rietvlei (from Recknagel et.al. 2008)

The *k*-fold cross-validation of the rule-based agent for *Microcystis* achieved r2-values of 0.31 for Lake Hartbeespoort, 0.34 for Lake Roodeplaat and 0.75 for Lake Rietvlei (Fig. 13).

Successful applications of EA have been demonstrated for cross-sectional data of fish populations (e.g. D'Angelo et al. 1995) as well as macroinvertebrate communities in streams (e.g. Horrigan et al. 2005), and for time-series data of plankton communities in lakes and rivers (e.g. Cao et al. 2006; Chan et al. 2007), and biological waste water treatment (Hong and Bhamidimarri 2003).

4 Future directions

Making informed decisions on conservation of biodiversity and sustainable environmental management in spite of ongoing pollution, eutrophication and climate change is of vital importance for the habitat earth in the 21st century. Ecological informatics is challenged to improve ecological understanding and provide tools for integrating, analysing and synthesising the wealth of ecological knowledge and data for informed decision making at local, regional and global scale.

It is anticipated that at the next stage ecological informatics will focus in particular on: (1) integrated analysis of genomic, phenotypic and ecological data in order to better understand biodiversity and ecosystem behaviour in response to environmental and climate changes; (2) facilitating data sharing by www-based generic data warehousing tailored for ecosystem categories at global scale, and (3) implementing hybrid model libraries generic for ecosystem categories at global scale by object-oriented programming and interactive www-access.

References

Aoki, I. and T. Komatsu, 1997. Analysis and prediction of the fluctuation of sardine abundance using a neural network. *Oceanologica Acta* 20, 1, 81-88.

Banzhaf, W., Nordin, P., Keller, R.E., Francone, F.D., 1997. Genetic Programming: An Introduction on the Automatic Evolution of Computer Programs and its Applications. Morgan Kaufmann.

Cao, H., Recknagel, F, Welk, A. , Kim, B. and N. Takamura, 2006. Hybrid Evolutionary Algorithm for Rule Set Discovery in Time-Series Data to Forecast and Explain Algal Population Dynamics in Two Lakes Different in Morphome try and Eutrophication. In: Recknagel, F. (ed.), 2006. *Ecological Informatics. 2ndEdition*. Springer-Verlag Berlin, Heidelberg, New York, 330-342.

Cao, H. and F. Recknagel, 2009. Hybridisation of process-based ecosystem models with evolutionary algorithms: Multi-objective optimisation of process representations and parameters of the lake simulation library SALMO-OO. In:Jorgensen, S.E., Chon, T.S. and F. Recknagel, (editors), 2009. *Handbook of Ecological Modelling and Informatics*, WIT Press, Southampton, Chapter 10, 169-185.

Capcarrere, M. Tettamanzi, A. Tomassini, M. and M. Sipper, 1998. Studying parallel evolutionary algorithms: The cellular Programming Case. In: Eiben et al. (editors), 1998. *Parallel Problem Solving from Nature V.* Springer-Verlag, New York, 573-582.

Chan, W. S., Recknagel, F., Cao, H. and H.D. Park, 2007. Elucidation and short-term forecasting of microcystin concentrations in Lake Suwa (Japan) by means of artificial neural networks and evolutionary algorithms. *Water Research* 41, 2247- 2255.

Chen, Q. and A.F. Mynett, 2006, Modelling Algal Blooms in the Dutch Coast Waters by Integrated Numerical and Fuzzy Cellular Automata Approaches, *Ecological Modelling*, 199(1): 73-81

Chon, T.S., Park, Y.S., Moon, K.H. and E.Y. Cha, 1996. Patternizing communities by using an artificial neural network. *Ecological Modelling*, 90, 69-78.

Chon, T.S., Park, Y.S., Kwak, I.-S.. and E.Y. Cha, 2006.Non-linear approach to grouping, dynamics and organizational informatics by benthic macroinvertbrate communities in streams by artificial neural networks. In: Recknagel, F. (ed.), 2006. *Ecological Informatics. 2nd Edition*. Springer-Verlag Berlin, Heidelberg, New York, 187-238.

D'Angelo, D.J., Howard, L.M., Meyer, J.L., Gregory, S.V. and L.R. Ashkenas, 1995. Ecological uses of genetic algorithms: predicting fish distributions in complex physical habitats. *Can.J.Fish.Aquat.Sci.* 52, 1893-1908.

Doney, S.C., Abbott, M.R., Cullen, J.J., Karl, D.M. and L. Rothstein, 2004. From genes to ecosystems: the ocean's new frontier. *Front. Ecol. Environ.*, 2, 9, 457 466.

Eleveld, M.A., Schrimpf, W.B.H. and A.G. Siegert, 2003. User requirements and information definition for the virtual coastal and marine data warehouse. *Ocean & Coastal Management* 46, 487-505.

Fielding, A., 1999. Machine Learning Methods for Ecological Applications. Kluwer, 1-262.

Foody, G., 2000. Soft mapping of coastal vegetation from remotely sensed imagery with a feed-forward neural network. In: Lek, S. and Guegan, J.F. (Eds.), Artificial Neuronal Networks: Application to Ecology and Evolution. Springer-Verlag, Berlin, 45-56.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley, Reading, MA.

Gyllström, M., Hansson, L-A., Jeppesen, E., Garcia-Criado, F., Gross, E., Irvine, K., Kairesalo, R., Kornijow, M.R., Miracle, M., Nykänen, T., Nõges, T., Romo, S., Stephen, D., Moss, B. and E. van Donk, 2005. The role of climate in shaping zooplankton communities of shallow lakes. *Limnology and Oceanography* 50, 2008-2021.

Hairston, N.G., Lampert, W., Caceres, C.E., Holtmeier, C.L., Weider, L.J., Gaedke, U., Fischer, J.M., Fox, J.A. and D.M. Post, 1999. Rapid evolution revealed by dormant egg. *Nature* 401, 446.

Henikoff, S., Henikoff, J.G. and S. Pietrovski, 1999. Blocks+: a non-redundant database of protein alignment blocks derived from multiple compilations. Bioinformatics 15, 471-479.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbour, MI.

Hornik, K., Stinchcombe, M. and A. White, 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2 , 5, 359 – 366.

Hong, Y.-S. and R. Bhamidimarri, 2003. Evolutionary self-organising modeling of a municipal wastewater treatment plant. *Water Research* 37, 1199-1212.

Horrigan, N., Bobbin, J., Recknagel, F. and L. Metzling, 2005. Patterning, prediction and explanation of stream macroinvertebrate assemblages in Victoria (Australia) by means of artificial neural networks and genetic algorithms. In: Lek, S., Scardi, M., Verdonschot, P.F.M., Descy, J.-P. and Y.-S. Park (Eds.) (eds.), 2005. *Modelling Community Structures in Freshwater Ecosystems.* Springer-Verlag, Berlin, Heidelberg, New York, 252-260.

Hraber, P. and B.T. Milne, 1997. Community assembly in a model ecosystem. *Ecological Modelling* 103, 267-285.

Huang, W. and S. Foo, 2002. Neural network modelling of salinity variation in Apalachicola River. *Water Research* 36, 356-362.

Huong, H. Recknagel, F., Marshall, J. and S. Choy, 2001. Predictive Modelling of Macroinvertebrate Assemblages for Stream Habitat Assessments in Queensland (Australia). *Ecological Modelling* 146, 1-3, 195-206.

Jeong, K.-S., Recknagel, F. and G.-J. Joo, 2006. Prediction and elucidation of population dynamics of the blue-green algae *Microcystis aeruginosa* and the diatom *Stephanodiscus hantzschii* in the Nakdong River-Reservoir System (South Korea) by a recurrent artificial neural network. In: Recknagel, F. (ed.), 2006. Ecological Informatics. Scope, Techniques and Applications. 2nd Edition. Springer-Verlag, Berlin, Heidelberg, New York, 255-273.

Jorgensen, S.E., Chon, T.S. and F. Recknagel, 2009. *Handbook of Ecological Modelling and Informatics*. WIT Press, Southampton, UK , 1 – 431.

Karul, C. and S. Soyupak, 2003. A Comparison between Neural Network Based and Multiple Regression Models for Chlorophyll-*a* Estimation. In: Recknagel, F. (ed.), 2003. *Ecological Informatics. Understanding Ecology by Biologically-Inspired Computation*. Springer Verlag-Berlin, Heidelberg, New York, 249-264.

Kohonen T, 1989. Self-Organization and Associative Memory. Springer-Verlag, Berlin.

Lek, S., Delacosta, M., Baran, P., Dimopoulos, I., Lauga, J. and S. Aulagnier, 1996. Application of neural networks to modeling nonlinear relationships in ecology. *Ecological Modelling* 90, 39-52.

Lek, S., Scardi, M., Verdonschot, P.F.M., Descy, J.-P. and Y.-S. Park (eds.), 2005**.** Modelling community structure in freshwater ecosystems. Springer, Berlin, Heidelberg, New York.

Lockhardt, D. and E. Winzeler, 2000. Genomics, gene expression and DNA arrays. *Nature* 405, 827-836.

Lupas, A., Van Dyke, M. and J. Stock, 1991. Predicting coiled coils from protein sequences. *Science* 252, 1162-1164.

Lusk, J.J., Guthery, F.S. and S.J. DeMaso, 2001. Northern bobwhite (*Colinus irginianus*) abundance in relation to yearly weather and long-term climate patterns. *Ecological Modelling* 146, 3-15.

Michener, W.K., 2006. Meta-information concepts for ecological data management. *Ecological Informatics* 1, 3-7.

Minski, M.L. and S. Pappert, 1969. Perceptrons. MIT Cambridge.

Mitchell, M., 1996. An Introduction to Genetic Algorithms. MIT Press, Cambridge, MA.

Mulderij, G. Smolders, A.J.P. and E. van Donk, 2006. Allelopathic effect of the aquatic macrophyte, Stratiotes aloides, on natural phytoplankton. *Freshwater Biology*, 51, 554-561.

Overbeck , R., Fonstein, M., D'Souza, M., Pusch, G.D. and N. Maltsev, 1999. The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. USA 96, 2896-2901.

Pineda, F.J., 1987. Generalisation of back-propagation to recurrent neural networks. *Physical Review Letters*, 59, 19, 2229–2232.

Poff, N.L., Tokar, S. and P. Johnson, 1996. Stream hydrological and ecological response to climate change assessed with an artificial neural network. *Limnology and Oceanography* 41, 5, 857-863.

Recknagel, F. (ed.), 2006. Ecological Informatics. Scope, Techniques and Applications. 2nd Edition. Springer-Verlag, Berlin, Heidelberg, New York.

Recknagel, F., van Ginkel, C., Cao, H., Cetin, L. and B. Zhang, 2008. Generic limnological models on the touchstone: Testing the lake simulation library SALMO-OO and the rule-based *Microcystis* agent for warm-monomictic hypertrophic lakes in South Africa. *Ecological Modelling* 215, 144-158.

Recknagel, F., French, M., Harkonen, P. and K. Yabunaka, 1997. Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling* 96, 1-3, 11-28.

Recknagel, F. and H. Wilson, 2000. Elucidation and prediction of aquatic ecosystems by artificial neural networks. In: Lek, S. and J.F. Guegan (eds.), *Artificial Neural Networks in Ecology and Evolution. Springer-Verlag, New York.* 143-155.

Recknagel, F, Kim, B., Takamura, N. and A. Welk, 2006. Unravelling and Forecasting Algal Population Dynamics in Two Lakes Different in Morphometry and Eutrophication by Neural and Evolutionary Computation. *Ecological Informatics* 1, 2, 133-151.

Recknagel, F. , Talib, A. and D. van der Molen, 2006. Phytoplankton community dynamics of two adjacent Dutch lakes in response to seasons and eutrophication control unravelled by non-supervised artificial neural networks. *Ecological Informatics* 1, 3, 277-286.

Reick, C.H., Grünewald, A. and B. Page, 2003. Multivariate time series prediction of marine zooplankton by artificial neural networks. In: Recknagel, F.