Faculteit

Bio-ingenieurswetenschappen

Academiejaar 2004-2005

DATA DRIVEN DEVELOPMENT OF PREDICTIVE ECOLOGICAL MODELS FOR BENTHIC MACROINVERTEBRATES IN RIVERS

GEGEVENSGEBASEERDE ONTWIKKELING VAN PREDICTIEVE ECOLOGISCHE MODELLEN VOOR BENTHISCHE MACRO-INVERTEBRATEN IN RIVIEREN

door

ir. PETER L.M. GOETHALS

Thesis submitted in fulfilment of the requirements for the degree of Doctor (PhD) in Applied Biological Sciences

Proefschrift voorgedragen tot het bekomen van de graad van Doctor in de Toegepaste Biologische Wetenschappen

op gezag van

Rector: **Prof. dr. A. DE LEENHEER**

Decaan:

Prof. dr. ir. H. VAN LANGENHOVE

Promotoren:

Prof. dr. N. DE PAUW

Prof. dr. ir. P. VANROLLEGHEM

Faculteit

Bio-ingenieurswetenschappen

Academiejaar 2004-2005

DATA DRIVEN DEVELOPMENT OF PREDICTIVE ECOLOGICAL MODELS FOR BENTHIC MACROINVERTEBRATES IN RIVERS

GEGEVENSGEBASEERDE ONTWIKKELING VAN PREDICTIEVE ECOLOGISCHE MODELLEN VOOR BENTHISCHE MACRO-INVERTEBRATEN IN RIVIEREN

door

ir. PETER L.M. GOETHALS

Thesis submitted in fulfilment of the requirements for the degree of Doctor (PhD) in Applied Biological Sciences

Proefschrift voorgedragen tot het bekomen van de graad van Doctor in de Toegepaste Biologische Wetenschappen

op gezag van

Rector: **Prof. dr. A. DE LEENHEER**

Decaan:

Prof. dr. ir. H. VAN LANGENHOVE

Promotoren:

Prof. dr. N. DE PAUW

Prof. dr. ir. P. VANROLLEGHEM

This research was performed at:

Laboratory for Environmental Toxicology and Aquatic Ecology, Department Applied Ecology and Environmental Biology, Faculty of Bio-engineering Sciences, Ghent University
J. Plateaustraat 22, B-9000 Gent (Belgium)
Tel. 0032 (0)9 264 37 65, Fax. 0032 (0)9 264 41 99

Department Applied Mathematics, Biometrics and Process Control, Faculty of Bio-engineering Sciences, Ghent University
Coupure Links 653, B-9000 Gent (Belgium)
Tel. 0032 (0)9 264 59 32, Fax. 0032 (0)9 264 62 20

ISBN 90-5989-046-9

Acknowledgements

First of all I would like to thank my promoters, Prof. Niels De Pauw and Prof. Peter Vanrolleghem, for their support during this scientific survival trip. I am particularly grateful for the high degree of freedom I received and the possibility I got to start this new line of research related to ecological informatics applications in water management. I also enjoyed and learned a lot from the foreign research activities in Greece, Russia, Vietnam, Ecuador, ... and am aware that I was very fortunate to get these opportunities. I would like to thank Prof. De Pauw and his spouse Mrs. Françine Moonen for the huge pile of extra-ecological knowledge, in particular related to culture, gastronomy, religion, psychology, gardening, … Thank you very much for your hospitality and the many good discussions we had over delicious meals at your home!

I am very grateful for the technical luggage I received from many persons for this scientific mission. Two Mr. Q's I would like to mention in particular. I learned a lot on classification trees during my stay with Prof. Saso Dzeroski at the Jozef Stefan Institute in Ljubljana (Slovenia). This technique was an important element in the accomplishment of my PhD research. I also thank Prof. Sovan Lek of the Université Paul Sabatier in Toulouse (France) for the hospitality and great help on the use of artificial neural networks during the five times I was in his laboratory. You are not only an inspiring and hard working scientist, but above all a good and friendly man. I am very grateful for the ANN toolbox that was established during the European PAEQANN project. It was a privilege to be one of the first to apply this software.

Many thanks to my official (Prof. Hendrik Blockeel, Prof. Bernard De Baets, Prof. Luc Tirry, Prof. Niko Verhoest, Drs. Harm Duel, Drs. Peter Van Puijenbroek, Prof. Niels De Pauw, Prof. Peter Vanrolleghem and Prof. Georges Hofman) and private (Andy, Ans, Jurgen, Katrien, Tom, Wim, Wouter) reading committees for their valuable remarks! This work has dramatically improved due to your co-operation. Andy and Ans, thank you for your kind support during the last rush to the finish line!

Dear fun loving criminals of the lab, it was great to have so many devoted as well as joking creatures around me (the Christmas parties are legendary!). I would like to thank Andy, Anke, Ans, Argaw, Diederik, Lien, Lieven, Luis, Martine, Mattias, Minh, Philippe, Rahmat, Steven, Tom, Veronique, Wim, Wouter of the Aquatic Ecology group and An, Barbara, Bart B, Bart C, Bram, Brita, Dago, Emmy, Frederik, Gertie, Hilde, Jill, Karel, Katrien, Koen, Leen, Marc, Marjan, Marianne, Marleen, Michiel, Nele, Peggy, Tim, Tom, Vincent, Yblin and Prof. Janssen of the Ecotoxicology group, the unbelievable Euras creatures Marnix and Patrick, as well as Kris and Prof. Persoone of Creasel. A special thank you to Guido and Gisèle for the professional technical management… what would our lab be without you?! I also thank Prof. Van den Heede and his assistants Helga, Isabel and Veerle for their kind support at the CES and our department chairman Prof. Lemeur and his team of the Plant Ecology Laboratory.

Thank you all thesis students for your valuable work, making me proud as your tutor. I hope I succeeded in my task as mentor and it is always nice to see you back or hear from you. Wish you all the best and stay in touch!

Doing research is only possible when there are enough financial means. Therefore I am very grateful to the representatives of the VMM, AMINAL, VLIR, FWO, IWT, VLM, DGIS, BTC, EU, … for their support of our research on ecological informatics in water management. It is my sincere wish that this research is useful for society in the short or long run. I am also grateful to many people spread in research institutions, universities and private companies all over the world, I had the pleasure to work and discuss with. Thank you Alain (CRP-GL), Andreas (U. Stuttgart), Annick (VMM), Ari (GFRI), Atle (SINTEF), Bart (Aquafin), Chris (UIA), Claude (IBW), Diederik (RIKZ), Diego (U. Madrid), Dirk (KULeuven), Edwin (IHE), Egbert (WUR), Emmanuel (VUB), Franz (U. Stuttgart), Friedrich (U. Adelaide), Gaby (VMM), Greet (VMM), Harm (WL Delft Hydraulics), Helmut (BOKU), Henk (VMM), Huub (IHE), Ilse (IBW), Jan (DGIS), Jan (IBW), Jan (RIVM), Javier (U. Granada), Jean-Luc (U. Bordeaux), Klaus (U. Idaho), Lowie (RIVM), Luc (VUB), Ludwig (VUB), Maarten (IHE), Marcel (W+B), Martin (TUDelft), Matthias (SJR), Mike (CEH), Morten (NERI), Nico (W+B), Nikolai (NERI), Péter (SINTEF), Piet (ALTERRA), Rebi (ALTERRA), Régis (UPS), Rudy (VMM), Stefan (BOKU), Sylvie (U. Lyon), Yann (CEMAGREF), …!

I would like to thank Prof. Verstraete of the Laboratory of Microbial Ecology and Technology (Ghent University), as he taught me the basic steps in research. I also greatly enjoyed the 'exergetic' discussions with Prof. S. Jorgensen on structural dynamical modelling of ecosystems in Namur, Toulouse and Adelaide. Thank you for your supportive talks and openness to young scientists, it is great to have some examples in life to look up to!

I am very grateful to all musicians that were responsible for a pleasant blend of hormones in my veins, especially during the lonesome writing periods. Thank you AC/DC, Arid, Iron Maiden, Jamiroquai, Metallica, Moby, Muse, Novastar, Pearl Jam, Radiohead, Smashing Pumpkins, U2, … I enjoyed your cocktail of emotions, shaking beats and encouraging lyrics very much! I also would like to ask a round of applause for the many cartoonists, whose work brings life in an original perspective during the stressing moments.

Jurgen, thank you for the many hilarious and exhausting moments at the fitness, as well as the social-economical, political and philosophical debates in pubs! Veerle, I look forward to seeing you back after your journey to Canada and having some enlightening job discussions as partners in crime! I also enjoyed the indoor soccer games very much with Jo, Joost, Peter, Georges, Pascal, Dries, Lieven…! Thank you Jo, Bart en Mik for the entertaining weekend meetings and the sport activities… Katrien, thank you for the charming talks about life…

Of course, I would like to thank my wonderful parents! I was lucky to grow up among such great and friendly people, who always believed in and supported my choices in life. You both deserve a *Doctor Honoris Causa* title for your hard work, courage and altruism!

Alexander, you were a great brother and best friend!

Peter,
Gent, 15th January 2005.

Calvin and Hobbes, by Bill Watterson

List of abbreviations

Abbreviation	Explanation
ANN	Artificial neural network, in this study the abbreviation X-Y-Z ANN model is applied for a feed forward back-propagation ANN model with X input neurons, Y hidden neurons (in one hidden layer) and Z output neurons
AQEM	The Development and Testing of an Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates
ASPT	Average Score Per Taxon
AUSRIVAS	Australian River Assessment Scheme
AVS	Acid volatile sulfides
BBI	Belgian Biotic Index
BBN	Bayesian belief networks
BEAST	Benthic Assessment of Sediment
BGBI	Bulgarian Biotic Index
BMWP	Biological Monitoring Working Party
CCI	Correctly Classified Instances
COD	Chemical oxygen demand
CPD	Conditional probability distributions
CPT	Conditional probability table
CT	Classification tree
CVE	Cross-validation error
DSFI	Danish Stream Fauna Index
DO	Dissolved oxygen (mg oxygen per liter)
DSS	Decision support system
EC	Electrical conductivity (microSiemens per centimeter)
EQI	Environmental Quality Index
EQR	Ecological Quality Ratio
FBI	Family Biotic Index
FN	False negative
FP	False positive
FV	Flow velocity (meter per second)
H’	Diversity Index
IBE	Indice Biotico Esteso
IBGN	Indice Biotique Global Normalisé
IBI	Index of Biotic Integrity
ITC	Index of Trophic Completeness
K	Cohen’s Kappa
LMLS	Least mean log squares
NID	Neural interpretation diagram
NMI	Normalized mutual information statistic
NOT	Number of taxa
MAE	Mean absolute error
MSE	Mean squared error
OM	Organic material
OrthoP	Orthophosphate (mg orthophosphate per liter)

PAEQANN	Predicting Aquatic Ecosystem Quality using Artificial Neural Networks
PCF	Pruning confidence factor
PhD	Philosophy Doctor
r	Correlation coefficient
r^2	Determination coefficient
RAE	Root absolute error
RIVPACS	River Invertebrate Prediction and Classification System
RMSE	Root mean squared error
RRSE	Root relative squared error
RSE	Relative squared error
RT	Regression tree
S	(German) Saprobic Index
SASS	South African Score System
SCI	Sequential Comparison Index
SEM	Simultaneously extracted metals
SOM	Self organizing map or Kohonen network
SS	Suspended solids (mg per liter)
SSE	Sum of squared errors
SWOT	Strengths, weaknesses, opportunities and threats (analysis)
T	Temperature (°Celcius)
TN	True negative
TotalN	Total nitrogen (mg nitrogen per liter)
TotalP	Total phosphorus (mg phosphorus per liter)
TOXR	Toxicity class variable (true or false) based on sediment pore water 72h growth-inhibition test with the species *Raphidocelis subcapitata* (alga)
TOXT	Toxicity class variable based (true or false) on sediment pore water 72h growth-inhibition test with the species *Tamnocephalus platyurus* (crustacean)
TP	True positive
VMM	Flemish Environment Agency (Vlaamse Milieumaatschappij)
WAECO-DSS	Water Ecology decision support system
WFD	European Water Framework Directive (EU, 2000)

Table of contents

Chapter 1: Introduction……………………………………………………………….1
1.1 Problem definition
1.2 Scope and objectives

Chapter 2: A general concept of integrated monitoring, assessment, modelling and management of rivers………………………………………9
2.1 Introduction
2.2 Concepts of economic value of ecosystems
2.3 Contemporary aquatic (eco)system assessment in Flanders
2.3.1 Biological monitoring and assessment for decision support in water management
2.3.1.1 Advantages and disadvantages of monitoring and assessment methods based on macroinvertebrates
2.3.1.2 Elements of biological monitoring and assessment methods based on macroinvertebrates
2.3.1.3 Different assessment approaches based on macroinvertebrates
2.3.1.3.1 Introduction
2.3.1.3.2 Saprobic approach
2.3.1.3.3 Diversity approach
2.3.1.3.4 Biotic approach
2.3.1.3.5 Multimetric approach
2.3.1.3.6 Ecological Quality Ratio approach
2.3.1.3.7 Multivariate approach
2.3.2 Ecological indices used in Flanders and limitations for cost-benefit analyses
2.4 Discussion on the need of (ecological) models and related monitoring strategies for decision support in water management
2.5 Conclusions

Chapter 3: State of the art of ecological modelling techniques to predict macroinvertebrates in rivers……………………..……………………….31
3.1 Introduction
3.2 The need of soft computing methods to predict aquatic communities on the basis of abiotic water system characteristics
3.3 Model performance measures and validation methods
3.3.1 Model performance measures
3.3.2 Data set splitting for training and validation
3.4 Soft computing methods for prediction of macroinvertebrates on the basis of abiotic water system characteristics
3.4.1 Fuzzy logic
3.4.2 Bayesian belief networks (BBNs)
3.4.3 Artificial neural networks (ANNs)
3.4.3.1. General description of ANNs
3.4.3.2 Predictive ANN development
3.4.3.2.1 Data processing
3.4.3.2.2 Bandwidth
3.4.3.2.3 Learning method

3.4.3.2.4 Model architecture
3.4.3.2.5 Model interpretation
3.4.3.3 Predictive ANN development studies of aquatic macroinvertebrates
3.4.3.4. Research needs on predictive ANN development studies of aquatic macroinvertebrates
3.4.4 Classification and regression trees
3.4.4.1 Description of the model development method
3.4.4.2 Predictive classification and regression tree development studies of aquatic macroinvertebrates
3.4.4.3 Research needs on predictive classification and regression tree development studies of aquatic macroinvertebrates
3.5 Comparison of different soft computing methods
3.6 Discussion
3.7 Conclusions

Chapter 4: Data and information collection to develop and validate models predicting macroinvertebrates in rivers……………………………………..75
4.1 Data collection on river sediments in Flanders (1996-1998)
4.1.1 Introduction
4.1.2 The TRIAD approach for the assessment of river sediments in Flanders
4.1.3 Monitoring strategy and sites
4.1.4 Sampling methods
4.1.5 Abiotic and ecotoxicological river sediment characteristics
4.1.6 Macroinvertebrate community analysis
4.1.7 Database setup
4.2 Data collection in the Zwalm river basin (2000-2002)
4.2.1 General characteristics of the Zwalm river basin
4.2.2 Monitoring strategy and sampling sites
4.2.3 Abiotic river characteristics
4.2.4 Macroinvertebrate community monitoring and analysis
4.2.5 Database setup
4.3 Information collection on the habitat preferences of *Gammarus* and *Asellus*
4.3.1 Introduction
4.3.2 Knowledge base on *Gammarus*
4.3.3 Knowledge base on *Asellus*

Chapter 5: Data driven development of habitat suitability models based on classification trees and artificial neural networks to predict *Gammarus* and *Asellus* in rivers…………………………………………………………..99
5.1 Introduction
5.2 Data driven model development methods
5.2.1 Overview of data analysis, model development and validation procedures
5.2.2 Data analysis procedures
5.2.2.1 Bandwidth and distribution of input and output variables
5.2.2.2 Correlation between input variables
5.2.2.3 Visual relation analysis between the input and the output variables
5.2.3 Model development methods
5.2.3.1 Dataset construction for model training and validation
5.2.3.2 Classification trees

5.2.3.3 Artificial neural networks
5.2.3.3.1 Artificial neural network parameter settings and architecture
5.2.3.3.2 Input variables contribution methods
5.2.4 Model validation methods
5.2.5 Model development scheme
5.3 Data analysis results
5.3.1 Bandwidth and distribution of input and output variables
5.3.2 Correlation between input variables
5.3.3 Visual relation analysis between input and the output variables
5.3.4 Discussion
5.3.5 Conclusions
5.4 Development of predictive habitat suitability models based on classification tree methods for *Gammarus* and *Asellus* in river sediments in Flanders
5.4.1 Introduction
5.4.2 Evaluation based on performance indicators
5.4.3 Ranking of importance of input variables combined with ecological interpretation and discussion of practical use of classification tree models
5.4.4 Conclusions
5.5 Application of backpropagation artificial neural networks predicting *Gammarus* and *Asellus* in river sediments in Flanders
5.5.1 Introduction
5.5.2 Predictive performance of classification and regression ANN models
5.5.3 Ranking of importance of input variables combined with ecological interpretation and discussion of practical use of ANN models
5.5.4 Conclusions
5.6 Development of predictive habitat suitability models based on classification tree methods for *Gammarus* and *Asellus* in the Zwalm river basin
5.6.1 Introduction
5.6.2 Evaluation based on performance indicators
5.6.3 Ranking of importance of input variables, ecological interpretation and discussion of practical use of classification tree models
5.6.4 Conclusions
5.7 Application of backpropagation artificial neural networks predicting *Gammarus* and *Asellus* in the Zwalm river basin
5.7.1 Introduction
5.7.2 Predictive performance of classification and regression ANN models
5.7.3 Ranking of importance of input variables combined with ecological interpretation and discussion of practical use of ANN models
5.7.4 Conclusions
5.8 A comparative discussion of the obtained results
5.8.1 Performance of classification tree and ANN models based on numerical indicators
5.8.2 Importance of different input variables and ecological relevance of models
5.9 Conclusions

Chapter 6: Application of predictive macroinvertebrate habitat suitability models for information and decision support in river management: case studies in the Zwalm river basin…………………………………..……….173
6.1 Introduction
6.2 Prediction of the effect of restoration options
6.2.1 Site selection
6.2.2 Remeandering project of the Traveinsbeek
6.2.3 Removal of a weir for water quantity control in the Zwalmbeek
6.3 Selection of monitoring sites based on predictive errors of data driven models
6.4 Delivering insight in the habitat preferences of taxa
6.5 Discussion
6.6 Conclusions

Chapter 7: General discussion and further research……………………..193
7.1 Introduction
7.2 Data collection
7.3 Model development
7.4 Model applications for decision support in water management
7.5 Further research
7.5.1 Overview of potential research
7.5.2 Improvement and extension of model inputs via automated data collection procedures
7.5.3 Extension of the habitat preference models with migration models
7.5.4 Model evaluation and optimization methods
7.5.5 Linking ecological models to climate, land-use, discharges, river water quality and other physical models
7.5.6 Linking ecological to socio-economical models and stakeholder information needs
7.6 General conclusions

References………..……………………………………………………….217

Summary……………..…………………………………………………….247

Samenvatting ..…………………………………………………………….251

Curriculum vitae………………………………………………………….255

Appendices ………………………………………………………………..291

Chapter 1
Introduction

1.1 Problem definition

The European Water Framework Directive (WFD) 2000/60/EC (EU, 2000) aims at a good ecological status for all water bodies in the member states of the European Union by 2015. A major part of these water bodies can be classified as running waters or rivers. According to the WFD, rivers are to be assessed by comparing the actual status to a reference status. To this end, reference communities must be described that represent a very good ecological status. Additionally, for the development of a representative set of metrics for ecological river assessment, one needs to gain insight in the relation between the aquatic communities and the human activities affecting these water systems. Insights in these relations will also be valuable for detection of causes of particular river conditions (environmental impact assessment) as well as for decision-making in river restoration and protection management to meet and sustain the requirements set by the WFD.

So far, ecological models have been rarely used to support river management and water policy. Models have however several interesting applications in this context. A summary of the potential value of (ecological) models in river management is presented in Figure 1.1 (Goethals and De Pauw, 2001). First of all, through these models a better interpretation of the river status can be possible, the causes of the status of a river can be detected and assessment methods can be optimised. Secondly, these models can allow for calculating the effect of future river restoration actions on aquatic ecosystems and supporting the selection of the most sustainable options. Thirdly, these models can help to find the major gaps in our knowledge of river systems and help to setup cost effective monitoring programmes (Vanrolleghem et al., 1999).

Before such ecological models can be effectively applied in river management, however, several research challenges still remaining need to be tackled. Two major contemporary problems are here at order:

- the need for reliable datasets, which are necessary to develop, train and validate ecological models seems to be an everlasting key problem. During the last decades, a lot of ecological data are being collected, but these efforts are often too fragmented, resulting in databases that are not compatible, lacking essential variables, etc. Often these problems are (were?) related to the organizational structure of water management boards and their too specific goals to allow an integrated water management;

- numerous modelling and data mining techniques have been developed, but the particular strengths and weaknesses of these techniques remain unclear. This is partly because there is a lack of sound methodologies and criteria (indicators) to assess the models' qualities for practical use in decision support. A major source of criticism on the application of ecological models in water management originates from a lack of success stories in which models play a glamorous role. Indeed model studies often end after their development and theoretical validation. However model developers and water managers can both benefit from feedback studies in which the added value of models in the decision making is analysed once the effect of management decisions has taken place.

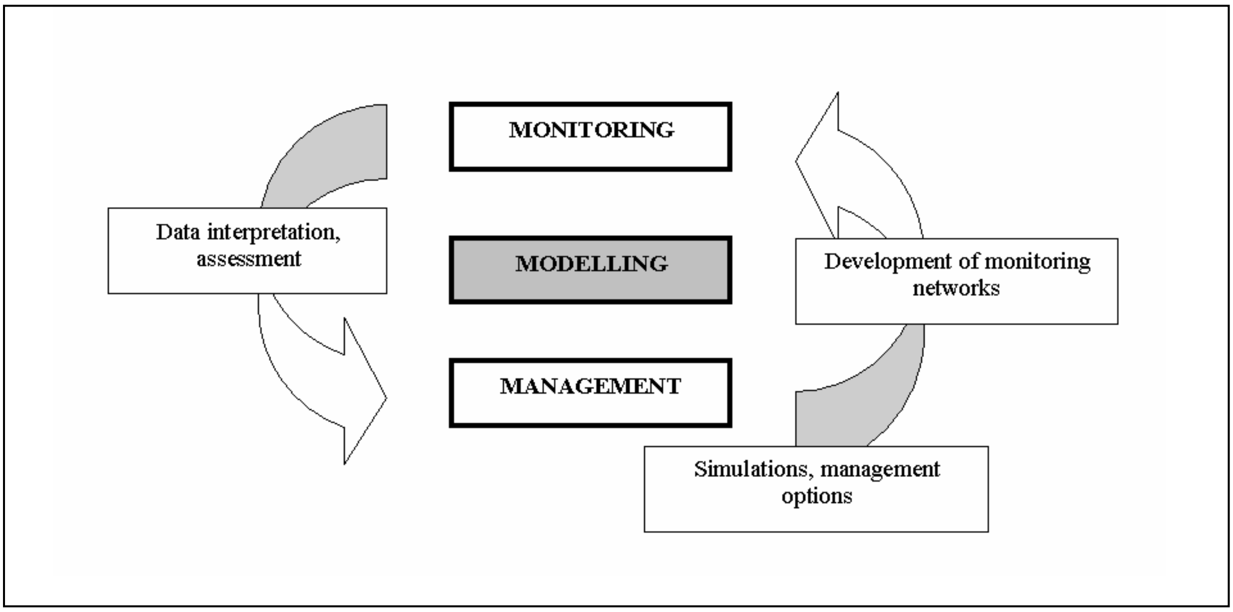

MONITORING
Data interpretation, assessment
MODELLING
Development of monitoring networks
MANAGEMENT
Simulations, management options

Figure 1.1 Potential applications of models for information and decision support in river management (Goethals and De Pauw, 2001).

The above challenges are the core research goals of this PhD study (Table 1.1). Evidently it is impossible to solve all problems related to the development and application of ecological models for decision support in water management in this study. However, several approaches presented in this manuscript can act as a basis to enhance the development of better monitoring methods, databases, ecological indices, models, validation techniques and decision support systems and encourage water managers to make use of ecological informatics techniques during their challenging task of satisfying all stakeholders in the water sector and this in a sustainable way.

Table 1.1 Major contemporary research challenges related to the development and application of ecological models for decision support in water management that have been taken into consideration in this thesis.

Monitoring methods and database set-up	**Model construction, assessment, comparison and application**
- selection and combination of environmental variables - type of variables - number of measurements (instances)	- model development techniques - model assessment methods - model comparison and application for decision support in river management

1.2 Scope and objectives

The present thesis aims at determining the appropriate variables and ecosystem processes by using different data mining and modelling techniques to predict biological communities present in rivers. This approach allows for deriving rules that contribute to a better understanding of river ecosystems and support of their management.

The research mainly focuses on macroinvertebrates in brooks and small rivers in Flanders (Belgium). The selected sampling sites are characterized by a gradient ranging from nearly natural situations to severely impacted (water pollution, physical habitat degradation) ones.

The applied modelling techniques in this research are all data driven approaches. In this manner, an *a priori* and often biased knowledge of ecological experts has not been used during the model development process. However, when discussing the results, the outcome of the data driven models has been compared to expert rules from literature.

The developed models have been put in practice to support decision making in water management. In this way, a crucial validation step, often lacking in many model development and assessment studies has been made and this can probably also help to pursue river managers of the added value of such ecological models. These models can as such become essential tools to convince stakeholders to make the necessary investments and/or activity changes as desired by society.

The thesis research is divided in four core parts dealing respectively with:

- State-of-the-art of ecological informatics, decision support in river management and habitat modelling techniques (with a focus on macroinvertebrates);
- Establishment of monitoring networks and ecological databases to develop models predicting aquatic macroinvertebrates in rivers;
- Development of predictive ecological models based on data driven methods (classification trees and artificial neural networks);
- Application of predictive ecological models for decision support in integrated river management.

The individual chapters grouped in these four parts are briefly described in the next paragraphs and further elucidate the specific goals of the research.

This first set of chapters (Chapter 2-3) reviews applications of ecological informatics in water management. In this first chapter (Chapter 2), the need of ecological models is illustrated. In particular insights are needed between river characteristics and biological communities for the optimization of ecological indices as well as for the prediction/allocation of the impacts of water uses as a prerequisite to make integrated cost-benefit analyses in river management. Both goals are very important for the implementation of the WFD. In the second chapter of this part (Chapter 3), different approaches to describe relations between river characteristics and biological communities are presented. Major focus is given on the development and use of data driven models for predicting macroinvertebrates. In an attempt to provide a straightforward methodology to compare the quality of the predictions on the basis of different techniques, the first part of Chapter 3 is dealing with the assessment of predictive models.

The second core part is dealing with the data and ecological information collection and only consists of one chapter (Chapter 4). It is merely a description of the monitoring networks and methods, the river site selection criteria, the database set-up and expert knowledge on *Gammarus* and *Asellus* habitat preferences found in literature. Two databases are presented in this chapter. The first database (constructed for the development of a river sediment assessment system) consists of measurements conducted in whole Flanders during the period 1996-1998. In total 360 sites were monitored. The macroinvertebrates were collected by means of a Van Veen grab sampler. Mainly river sediment variables were analyzed. The second database was developed during the period 2000-2002 and contains measurements conducted in 60 different sites in the Zwalm river basin, sampled each year. This database (and monitoring methods) was specifically constructed for the development of habitat suitability models. The macroinvertebrates were collected using nets and artificial substrates. In addition, water quality and physical habitat variables were used to describe the river characteristics of each site. A standardized methodology to describe and assess different structural and morphological characteristics is therefore presented. The last component of this chapter is devoted to an overview of expert knowledge related to *Gammarus* and *Asellus*, the two selected taxa for the data driven model development research.

The third part contains the core research of this PhD study and is divided in two chapters (Chapters 5 and 6). Based on the available datasets, two model development techniques were applied to the datasets: classification trees and artificial neural networks (ANNs). In addition and combination with the ANN models, several methods were used to analyse the contribution of environmental variables to predict macroinvertebrates in a reliable manner and to detect the major river characteristics to describe the habitat suitability for the different taxa. For the classification tree method, the major variables are automatically selected (and are for an important part influenced by the pruning settings of the method) and visualised. Both data driven model development methods can in this manner be more easily compared. In this manner, one also obtains insight in the effect of river conditions on the presence/absence (or abundance) of macroinvertebrates and the outcomes of the models can be compared with ecological expert knowledge from literature. These model development and habitat suitability studies focus on two taxa, the crustaceans *Gammarus* and *Asellus*.

In Chapter 5, seven transition components can be distinguished. The first component describes the methodologies behind the data analysis and preparation, development of predictive models based on data driven methods and the model validation methods. The second component presents the results of the application of data analysis and preparation methods, while the other five components present the following results:

- development of habitat suitability models based on classification trees to predict *Gammarus* and *Asellus* in river sediments in Flanders;
- application of backpropagation artificial neural networks predicting *Gammarus* and *Asellus* in river sediments in Flanders;

- development of habitat suitability models based on classification trees to predict *Gammarus* and *Asellus* in the Zwalm river basin;
- application of backpropagation artificial neural networks predicting *Gammarus* and *Asellus* in the Zwalm river basin;
- a comparative discussion of the obtained results.

By means of predictive ecological models the effects of specific management options can be evaluated in a more transparent and rational way. The aim of this second results chapter (Chapter 6) is to make a crucial validation step by analysing the feasibility of using this type of data driven models to solve practical management problems. This chapter tries in this manner to illustrate the added value of models in water management to select sustainable restoration options, but also to get insight in the basic relations between river characteristics and inhabiting communities. Also the selection of monitoring sites is aided by means of the models. In this manner, the feasibility of models to improve the processes as illustrated in Figure 1.1 is practically investigated.

The thesis ends with a general discussion (Chapter 7) about the scientific and practical meaning of the obtained results and the future prospects of continuing research.

Chapter 2
A general concept of integrated monitoring, assessment, modelling and management of rivers

<u>Redrafted after:</u>

Goethals, P.L.M., Bouma, J.J., François, D., D'heygere, T., Dedecker, A., Adriaenssens, V. and De Pauw, N. (2003). Coupling ecosystem valuation methods to the WAECO decision support system in the Zwalm catchment (Belgium). Vol. 3, p. 971-976. In: Post, D.A.(Ed.), Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ), Proceedings 'Integrative Modelling of Biophysical, Social and Economic Systems for Resource Management Solutions (MODSIM 2003)', 14-17 July 2003, Townsville, Australia. 2066 p.

De Pauw, N., Gabriels, W. and Goethals, P.L.M. (accepted). Monitoring and assessment of macroinvertebrates in freshwaters. In: Biomonitoring. Wiley, New York, USA.

2.1 Introduction

In Belgium, different water policies are being developed in the Flemish, Brussels and Walloon regions. Because parts of the major river basins (the Scheldt and the Meuse river basins, Figure 2.1) are situated in these three regions, the water policies are often conflicting between the different regions, resulting in ineffective and inefficient management of these water systems (e.g. many investments during the nineties, so far did not result in a clear improvement of the ecosystem quality in several river basins). Particular examples are water quality management, flood control and restoration of fish migration. These are issues that need an integrated approach over all regions, because one particular region is not able to restore or control these aspects within the borders of its territory and related responsibility. On top of this, contemporary river management is scattered in different manners in Belgium, often resulting in specific and conflicting targets for the responsible managers. The major divisions are based on river system sizes, system components (surface water, sediments, groundwater, aquatic ecosystems, etc.) and stakeholder (water uses) related issues.

The high need for a more integrated approach resulted in the very recent development of river basin committees (in total eleven in Flanders, Figure 2.1), in which delegated managers of the different administrations interact to obtain the best integrated management solution for that water system. The different stakeholders take part in these debates (water quantity managers, land use planners, wastewater collection and treatment managers, drinking water production companies, ecologists, etc.). However, gaining more insight in the water system and its social and economic functions is therefore needed. Information on the specific interests of certain stakeholders as well as the application of tools that allow a multi-criteria analysis can ease the discussion between all involved managers. This stresses the need of an interdisciplinary scientific approach to develop the required tools for this purpose (Scoccimarro et al., 1999; Pavlikakis and Tsihrintzis, 2003; Mustajoli et al., in press).

In particular scientific tools which can bridge the gap between the economic market and the natural market of (aquatic) ecosystems are of paramount importance to attain sustainability characterized by a growth of the economic development allowing ecosystem repair and regeneration (Costanza et al., 1997; Hansell et al., 2003). For this purpose, data collection and preparation should be based on insights in the river processes at different spatial and temporal scales (Lau et al., 1999), but also include the needs of the managers, allowing discussions among all involved participants and thus making decisions more transparent (Denzer et al., 2000). Cost-benefit analyses are good instruments for this goal and predictive models embedded in a decision support system (DSS) can be valuable tools to deliver the necessary data for the in depth analysis of the different restoration options (Jolma et al., 1997; Alkemade et al., 1998; Lam et al., 2004). These models allow the prediction of water quality (e.g. Jolma et al., 1997, Sigua and Tweedale, 2003) or biological communities (e.g. Larson and Sengupta, 2004) in river stretches for different restoration scenarios, being the necessary basis for cost-benefit analyses enabling the selection of the most feasible management plan. An example of a concept of such a DSS is presented in the final part of this PhD research.

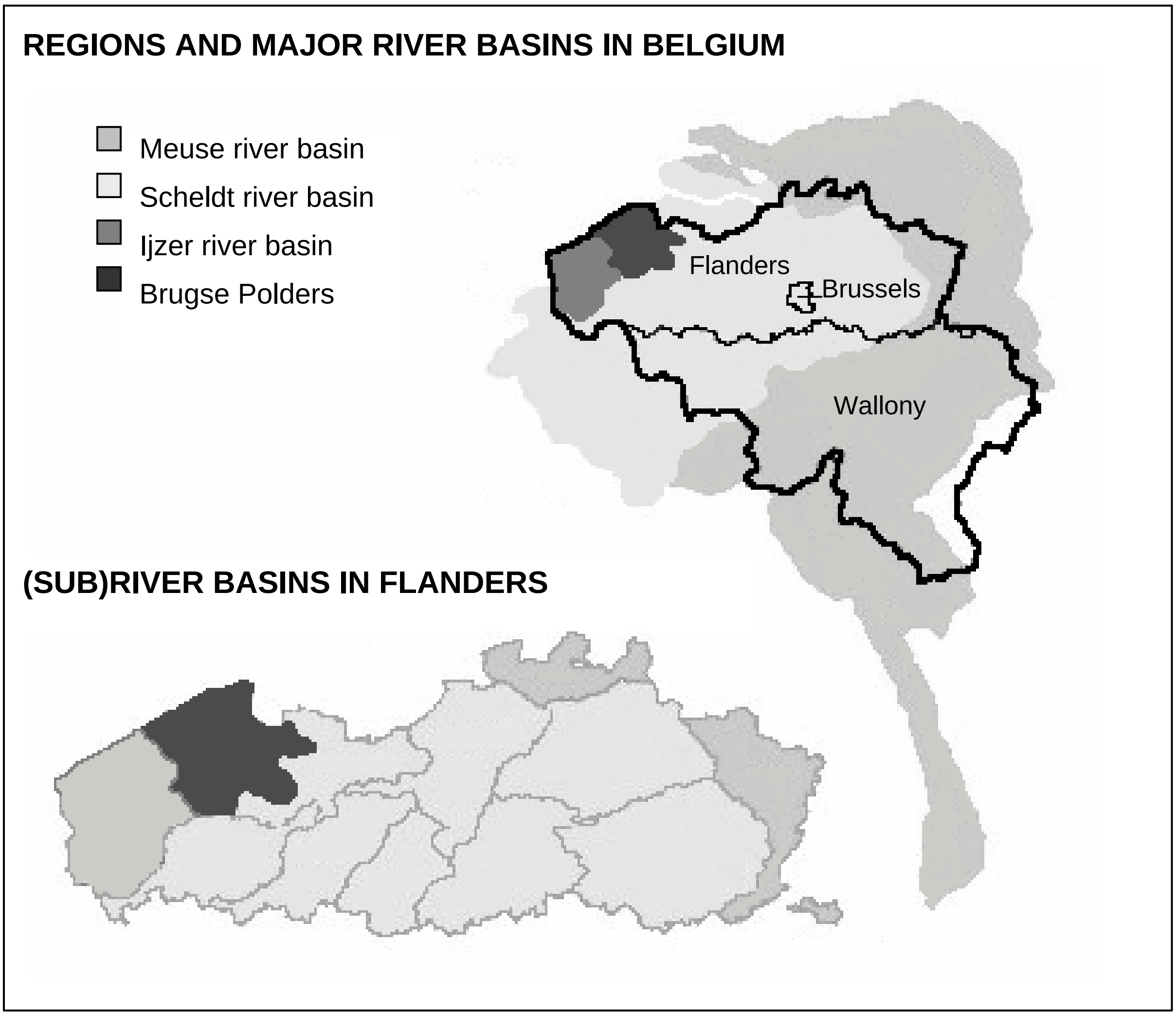


Figure 2.1 *The Scheldt and Meuse river basins in France, Belgium and The Netherlands. Major part of the Scheldt river basin is located in Flanders (northern part of Belgium). Recently Flanders is from a water management perspective divided in eleven river drainage basins.*

2.2 Concepts of economic value of ecosystems

The concept of economic value of ecosystems can be a useful guide when distinguishing and measuring where trade-offs between society and the rest of nature are possible and where they can be made to enhance human welfare in a sustainable manner (Farber et al., 2002). From the perspective of welfare economics a useful common terminology regarding economic valuation is provided. This perspective regards values as the assessment of human preferences for a range of natural or non-natural 'objects', services and attributes (Turner et al., 2001). The total economic value of a resource can be broken down into different categories (Turner et al., 2001):

- Use values involve some interaction (actual use) with the resource, either directly or indirectly. Indirect use value derives from services provided by the ecosystem (e.g. the prevention of downstream flooding). Direct use value involves interaction with the ecosystem itself rather than via the services it provides and can be consumptive or non-consumptive (for example recreational and educational activities);
- Non-use values are associated with benefits derived simply from the knowledge that a resource is maintained. They suggest non-instrumental values that are in the real nature of the thing but not associated with actual use, or even the option to use the thing (Turner et al., 1994). Existence values (derived from the satisfaction of knowing that some feature of the environment continues to exist), bequest values (associated with the knowledge that a resource will be passed on to descendants to maintain the opportunity for them to enjoy it in the future) and philanthropic values (associated with the satisfaction from ensuring resources are available to contemporaries of the current generation) are examples of non-use values. Two other categories of values can be mentioned, not related to the initial distinction between use and non-use values. Option value refers to the fact that an individual derives benefit from ensuring that a resource will be available for use in the future, it reflects the value people place on a future ability to use the resource. Quasi-option value is associated with the potential benefits of awaiting improved information before giving up the option to preserve a resource for future use.

Some of these values can be relatively easy monetised, others however are less tangible. Table 2.1 (Turner et al., 2001) gives a general overview of different valuation methods that have been developed to estimate the value of resources.

Table 2.1 Valuation methodologies relating to ecosystem functions (Turner et al., 2001).

Valuation Method	**Description**	**Direct use values**	**Indirect use values**	**Non-use values**
Market Analysis	use of market prices	x	x	
Public Pricing	public investment as a surrogate for market transactions	x	x	x
Hedonic Price Method	derive an implicit price for an environmental good from analysis of goods for which markets exist and which incorporate particular environmental characteristics	x	x	
Travel Cost Method	costs incurred in reaching a recreation site as a proxy for the value of recreation	x	x	
Contingent Valuation Method	construction of a hypothetical market by direct surveying of a sample of individuals and aggregation to encompass the relevant population	x	x	x
Damage Costs Avoided	costs that would be incurred if an ecosystem function were not present		x	
Defensive Expenditures	costs incurred in mitigating the effects of reduced environmental quality		x	
Relocation Costs	expenditures involved in relocation of affected agents or facilities		x	
Replacement Costs	potential expenditures incurred in replacing the function that is lost	x	x	x
Restoration Costs	costs of returning a degraded ecosystem to its original state	x	x	x

The application of assumptions behind valuation methods shows that not all the effects can be monetised by each method. Therefore the inclusion of some effects into an assessment puts its limits on the choice of freedom regarding the selection of techniques. Furthermore, one should proceed with caution when using the results of different valuation studies that are based on different methods. The integration of the outcomes of valuation studies can be questioned when different assumptions have been made. Therefore, as a policy science, ecological economics is context-sensitive and action oriented (Shi, 2004), nevertheless it can help to make water management more sustainable by integrating relevant interactions

(McCoy, 2003), tuning decision making to social needs (Melloul and Collin, 2003), allowing the calculation of ‘true’ amounts of compensation for losses resulting from environmental disasters (Dunford et al., 2004), etc. In the next paragraph, the potential of ecosystem valuation methods will be analysed in the perspective of (surface) water management in Flanders.

2.3 Contemporary aquatic (eco)system assessment in Flanders

2.3.1 Biological monitoring and assessment for decision support in water management

2.3.1.1 Advantages and disadvantages of monitoring and assessment methods based on macroinvertebrates

Indicators are now widely used in many counties and regions to steer sustainable development (Yuan et al., 2003). The biotic component of an aquatic ecosystem may indeed be considered as an ‘integrating-information-yielding unit’ for assessment of its quality. Biological communities also integrate the effects of mixed types of stress and in certain cases already respond before analytical detection allows for (De Pauw and Hawkes, 1993).

Among the biological communities, the macroinvertebrates are by far the most frequently used group for bioindication in standard water management (Woodiwiss, 1980; Helawell, 1986; De Pauw et al., 1992; Rosenberg and Resh, 1993; Metcalfe-Smith, 1994; Hering et al., 2004). The term ‘macroinvertebrates’ however, does not respond to a taxonomical concept but to an artificial delimitation of part of the groups of invertebrate animals. In general, in running waters, one considers macroinvertebrates as those organisms large enough to be caught with a net or retained on a sieve with a mesh size of 250 to 1000 µm, and thus can be seen with the naked eye. In fact most of them are larger than 1 mm (e.g. Cummins, 1975; Sladecek, 1973; De Pauw and Vanhooren, 1983; Rosenberg and Resh, 1993; Ghetti, 1997; Tachet et al., 2002).

The majority of aquatic macroinvertebrates has a benthic life and inhabits the bottom substrates (sediments, debris, logs, macrophytes, filamentous algae, etc.) and for this reason in the literature about biological water quality assessment methods one is often referring to them as benthic macroinvertebrates or macrozoobenthos (Rosenberg and Resh, 1993). Other representatives of the macroinvertebrates, however, also serving as bioindicators, are pelagic and freely swimming in the water column, or pleustonic and associated with the water surface (Tachet et al., 2002).

The reasons for macroinvertebrates being so popular in bioassessment are numerous (e.g. Hawkes, 1979; Sladecek, 1973; Helawell, 1986; Metcalfe, 1989; Norris and Georges, 1993; Hering et al., 2004). Macroinvertebrates are ubiquitous and abundant throughout the whole river system in the crenal, rhitral as well as the potamal part (Illies, 1961). They play an essential role in the functioning of the river continuum food web (e.g. Vannote et al., 1980; Giller and Malmqvist, 1998).

Since macroinvertebrates are a heterogeneous collection of evolutionary diverse taxa, this means that at least some will react to specific changes in the aquatic environment, natural as well as imposed. They are not merely affected by different types of physical-chemical pollution (e.g. organic enrichment, eutrophication, acidification), but as well by physical changes and anthropogenic manipulation of the aquatic habitat (e.g. canalisation, impoundment, river regulation) (cf. Figure 2.2). Macroinvertebrates can thus be used for the assessment of the water as well as the habitat quality (Armitage et al., 1983) and enable a holistic assessment of streams.

Macroinvertebrates have furthermore the advantage to be relatively easy to collect and identify, and to be confined for most part of their life to one locality on the river bed and are therefore indicative of the changing water qualities. As such, they act as continuous monitors of the water flowing over them as opposed to chemical samples of the water taken at one time. Having long life spans, macroinvertebrates integrate environmental conditions over longer periods (weeks, months, years) and thus sampling may be less frequent (De Pauw and Hawkes, 1993; Giller and Malmqvist, 1998; Tachet et al., 2002).

Using macroinvertebrates as monitors of river (water) quality however has also its limitations. Quantitative sampling for example is difficult because of their non-random distribution in the river bed. Because of the seasonality of the life cycles of some invertebrates, e.g. insects, they may not be found at some times of the year (e.g. Linke et al., 1999; D’heygere et al., 2002; Tachet et al., 2002). An appreciation of this seasonality enables this to be taken into account in interpreting the data. As shown in Figure 2.2, factors other than water quality are also important determinants of benthic communities. Of these the related factors of current velocity and nature of the substratum are overriding ones determining the nature of the community, especially in relation to invertebrates. Since these factors differ along the river in different zones, different communities become established at different sites with the same water quality (Giller and Malmqvist, 1998). Therefore, in practice where possible, sampling sites having similar benthic conditions are selected or a typology is developed consisting of distinct river types with adapted sampling and assessment systems (e.g. Hering et al., 2004). Some assessment systems, e.g. RIVPACS (Wright et al., 1993), even predict the reference communities on the basis of a set of local river features as a basis for the assessment.

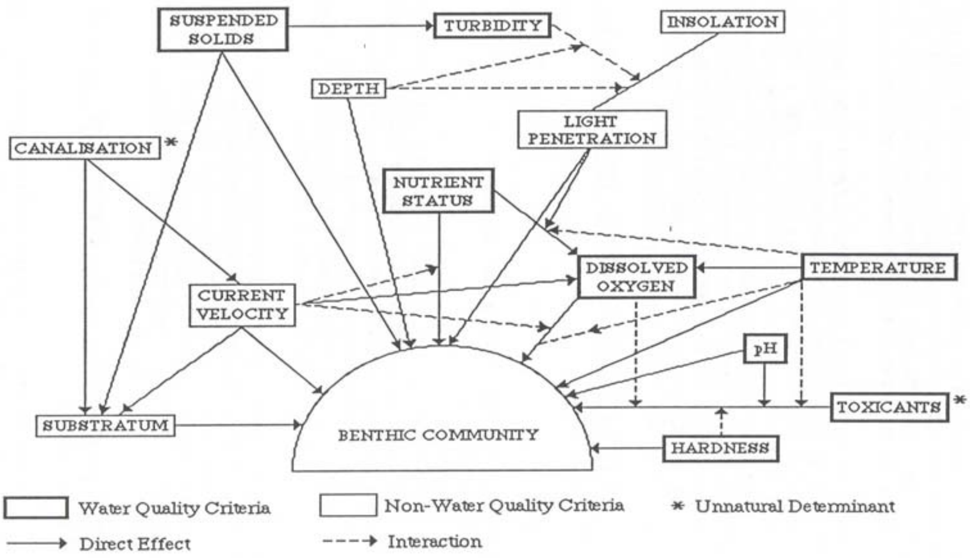


Figure 2.2 Water quality and non-water quality determinants of benthic communities in rivers (after Hawkes, 1979; De Pauw and Hawkes, 1993).

A last limitation of macroinvertebrates is their restricted geographic distribution, the incidence and frequency of occurrence of some species being different in rivers throughout the region. Furthermore, because of their geographic distribution, species at the edge of their natural distribution range are theoretically more sensitive to additional stress – pollution than those at the centre of their distribution. It would therefore not be possible to have a universal system of biological assessment based on the response of the same species/taxa (Sandin et al., 2000).

2.3.1.2 Elements of biological monitoring and assessment methods based on macroinvertebrates

The main elements of biological monitoring and assessment methods are summarized in Figure 2.3. Monitoring includes the sampling and sample analysis that is the collection of information, while assessment on the other hand is the interpretation of the data (Chapman, 1992). The assessment involves the numerical evaluation and index calculation, the classification of the indices into quality classes, the testing of compliance with standards, and finally the graphical presentation. Not all monitoring and assessment methods however apply all the elements presented.

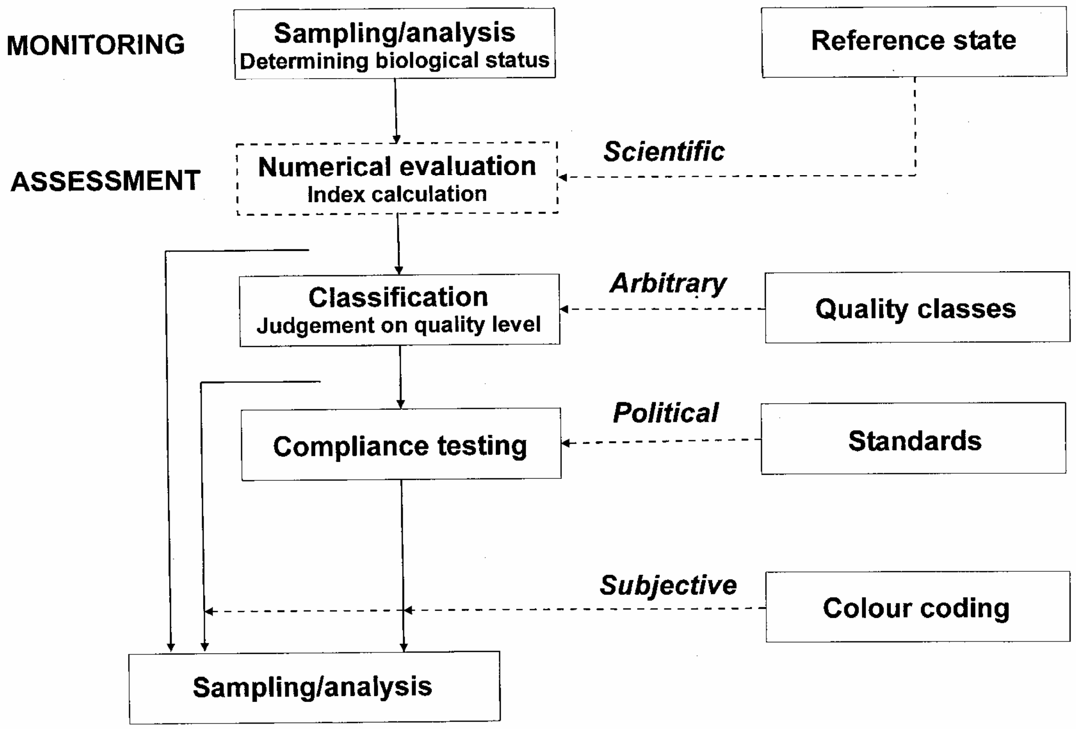


Figure 2.3 Elements of biological monitoring and assessment methods (after Knoben et al., 1995).

The history of bio-assessment of rivers is a good hundred years old taking a definite start in Europe in 1902 with the development of the saprobic system introduced by Kolkwitz and Marsson and in the US in 1913 with the development of a river water quality classification system by Forbes and Richardson (Richardson, 1928) (cf. reviews by Hynes, 1971; Sladecek, 1973; Persoone and De Pauw, 1979; Woodiwiss, 1980; Metcalfe, 1989; De Pauw et al., 1992; Roldan, 1992; Rosenberg and Resh, 1993; Sandin et al., 2000; Hering et al., 2004). Although the main focus in the beginning was on micro-organisms (a.o. plankton), macroinvertebrates as bio-indicators rapidly gained in importance (cf. Bartch and Ingram, 1966; Mackenthun, 1969; Sladecek, 1973; Rosenberg and Resh, 1993; Hering et al., 2004). The earlier systems were purely descriptive or qualitative and mainly based on the presence or absence of indicator species in the first place related to discharges of domestic sewage, i.e. organic pollution. Since early 1950 however biologists felt the need to convey their complex biological data in a numerical form such as indices or scores (e.g. Beck, 1954; Knöpp, 1954; Pantle and Buck, 1955). Today over 100 different biotic indices have been described (De Pauw et al., 1992; Ghetti and Ravera, 1994; Hering et al., 2004). Yet, many ecologists remain sceptical regarding the possibility and advisability of expressing complex biological communities in terms of a single numerical value. Nevertheless the pseudo-accuracy of a biotic index, is apparently more acceptable to the non-biologist and administrator than biological survey data expertly interpreted. To ensure biological information is made more comprehensible and therefore more acceptable in decision-making, the use of indices is probably justifiable although by using them information is inevitably lost. Having reduced the original data to a number there is a danger that it can then be more readily misused (Seegert, 2000).

2.3.1.3 Different assessment approaches based on macroinvertebrates

2.3.1.3.1 Introduction

Analysis of the macroinvertebrate communities in rivers can theoretically be structural, functional, taxonomical and non-taxonomical in approach (Matthews et al., 1982). Most of the actually used bio-assessment systems are however structural and taxonomical, what means relying for example on the presence or absence of particular taxa, the sensitivity of particular taxa, the taxa richness, taxa abundance, taxa diversity. All that information can be converted into numerical values, including indices and scores. Most assessment methods are based on the analysis of species assemblages or populations of particular taxonomic groups of benthic macroinvertebrates (e.g. oligochaetes, chironomids). Assessment methods based on organism-level indicators (biochemical, physiological, morphological deformities, behavioural responses and life-history responses) are not considered here.

Reviewing the common assessment methods in Europe based on structural-taxonomical analysis, Metcalfe (1989) distinguishes three major approaches to assess the response of macroinvertebrate communities to pollution: namely the saprobic, biotic and diversity approaches. In recent years however, also several new approaches were developed. Since the eighties for example, the use of multi-metric assessment systems, like the Index of Biotic Integrity, initiated in the US became more and more popular (Karr and Chu, 1999). Another approach was the introduction in the UK of RIVPACS which led to methods in which the comparison with reference conditions became central, a principle which was later adopted in the assessment proposal of the European Water Framework Directive (WFD) (EU, 2000). A last approach, although existing for some time already and for which a growing interest exists, is the use of multivariate analysis to distinguish among different river typologies and communities and which can be considered as a type of similarity indices. In the next paragraphs, a brief overview of these major approaches is presented.

2.3.1.3.2 Saprobic approach

The saprobic approach was the first river assessment system to be developed, already at the beginning of the 20th century by Kolkwitz and Marsson (1902), and later on expanded by a.o. Zelinka and Marvan (1961), Liebmann (1962) and Sladecek (1973). The objective is to provide a water quality classification based on the pollution tolerance of the indicator species present (= Response A in Table 2.2). Every species has a specific dependency of organic substances and thus of the dissolved oxygen content: this tolerance is expressed as a saprobic indicator value. The advantage is a quick classification of the investigated community by means of a saprobic index, which can be made on a universal scale (e.g. DEV, 1988-91). A major problem is the identification of the organisms up to species level. The saprobic index calculation also requires the assessment of abundances. The indicator system furthermore implies more knowledge than actually exists: pollution tolerances are highly subjective and based on ecological observations and are rarely confirmed by experimental studies.

2.3.1.3.3 Diversity approach

The diversity approach uses three components of community structure: richness (B), evenness (D) and abundance (C) (Washington, 1984). Diversity indices developed from information theory methods (Shannon and Weaver, 1949) have been used by Patten (1962), Wilhm and Dorris (1966), etc. The objective aims at evaluating the community structure with respect to occurrence of species. The diversity indices relate the number of observed species (richness) to the number of individuals (abundance). The principle is that disturbance of the water ecosystem or communities under stress leads to a reduction in diversity. The advantages of diversity indices lay in the fact that they are easy to use and calculate, applicable to all kinds of watercourses, have no geographical limitations and are best used for comparative purposes. Having no clear endpoint or reference level is however the main problem; the diversity in natural undisturbed waters can indeed vary considerably, moreover, all species have equal weight. This is probably the reason why not one country in Europe has adopted a diversity index as a national standard for biological water quality assessment (see De Pauw et al., 1992; Ghetti and Ravera, 1994; Nixon, 2003).

Table 2.2 Biocoenotic responses of indicator value induced by polluting discharges (after De Pauw and Hawkes, 1993).

Response class	**Species vs. community response**	**Response description**
A	species	the appearance or disappearance of individual species
B	community	a reduction in numbers of species/taxa present i.e. a reduction in diversity
C	community	a change in the population of individual species
D	community	a change in the proportional species composition of the community

2.3.1.3.4 Biotic approach

The biotic approach on the other hand incorporates desirable features of the saprobic and diversity approaches combining a quantitative measure of species diversity (B) with qualitative information on the ecological sensitivities of individual taxa (A) into a single numerical expression (cf. Table 2.2). Woodiwiss (1980) rightly distinguishes between biotic indices and biotic scores which although using the same responses A + B do so in quite different ways. In the biotic index approach the index is directly taken from a table in which the taxa richness is combined with the presence of the most sensitive taxon (e.g. the Trent Biotic Index, Woodiwiss, 1964). In the biotic score system on the other hand a score is allocated to each taxon. The score for the site is then derived by summing the individual scores. The biotic score may also include a measure of abundance of the organisms (e.g. Chandler, 1970). The objective of biotic indices or scores is to assess the biological water quality of running waters, in most cases based on macroinvertebrates, and to measure various types of environmental stress, organic waters, acid waters, etc. The principle is that macroinvertebrate groups disappear as pollution increases and that the number of taxonomic groups is reduced as pollution increases. Mackenthun (1969) identified the following stepwise disappearance of macroinvertebrates subsequent to increasing pollution: stoneflies (Plecoptera), mayflies (Ephemeroptera), caddisflies (Trichoptera), scuds (Amphipoda), aquatic sowbugs (Isopoda), midges (Diptera) and bristle worms (Oligochaeta).

The advantages are that only qualitative sampling is required and that identification is mostly at family or genus level and that there is no need to count abundances per taxon. The problems on the other hand are how to determine representative reference communities to which the investigated stations can be compared to. Also should an optimal biological assessment be achieved through regional adaptations.

2.3.1.3.5 Multimetric approach

The first true explicitly called multi-metric systems were developed in the US by Karr (1981) for assessments with fish. Recently similar systems are also being designed for benthic macroinvertebrate communities (US-EPA, 1996; Barbour et al., 1992; Karr and Chu, 1999; Hering et al., 2004). In multi-metric systems, several metrics representing different characteristics of the macroinvertebrate community are summed up into one index value or score (e.g. Barbour and Yoder, 2000) which is an expression of the overall quality. It is expected that working with more descriptors will result in an index being representative for a specific aquatic environment (e.g. the Acidification Index developed by Johnson, 1998). Multimetric systems may include structure metrics, community balance metrics, tolerance metrics, feeding group metrics and others (e.g. US-EPA, 1996). Within the context of the implementation of the EU Water Framework Directive (WFD), the European project AQEM (The Development and Testing of an Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates) has been proposing a strategy and methodology for the establishment of multi-metric assessment systems for different streams in Europe based on macroinvertebrates (Hering et al., 2004). Most of the multi-metric systems do not aim to separate the impact of different stressors (Lorenz et al., 2004). However, it has been recommended that the developed multi-metric systems should be stressor specific (e.g. for organic pollution, acidification, morphological degradation), to ease the cause allocation under conditions of deterioration. Examples of such stressor-specific systems can be found in Brabec et al. (2004) and Buffagni et al. (2004). For the assessments of the sediment quality in rivers, a TRIAD approach has been developed combining physical-chemical, ecotoxicological and biological information based on macroinvertebrates (De Cooman et al., 1999; De Pauw and Heylen, 2001).

2.3.1.3.6 Ecological Quality Ratio approach

The assessment value obtained with any index system can be compared with a reference status to be reached, by calculating the proportion between both values. This is called the Ecological Quality Ratio (EQR) according to the EU Water Framework Directive (EU, 2000). The reference can be based on real samplings, expert knowledge, historical data or predictive models, or a combination of these. An example of an EQR is the Environmental Quality Index (EQI) based on the 'River Invertebrate Prediction and Classification System' (RIVPACS) developed in the UK (Armitage et al., 1983; Wright et al., 1993; De Pauw, 2000; Wright et al, 2000). The principle of RIVPACS is that on the basis of the physical-chemical features of the river it is possible to predict which macroinvertebrate taxa should be present under these conditions. The predicted reference conditions can then be compared with the observed macroinvertebrate communities. The RIVPACS EQI can be calculated with different metrics or indices, for example the BMWP, the ASPT or the number of taxa (NOT) (Sweeting et al., 1992). Based on RIVPACS, other similar models have been developed in Australia (AUSRIVAS: 'Australian River Assessment Scheme', e.g. Davies, 2000; Smith et al., 1999) and Canada (BEAST: 'Benthic Assessment of Sediment', Reynoldson et al., 2000).

2.3.1.3.7 Multivariate approach

Several multivariate techniques have been applied in water quality assessment using macroinvertebrates (Norris and Georges, 1993). The basis for the multivariate approach is the similarity index (Sandin et al., 2000). The most commonly used similarity index is the Jaccard's index (Jaccard, 1908 and Washington, 1984). This index expresses the percentage of species shared between two sites. Other examples are the percentage similarity index (Whittaker, 1952), Bray-Curtis dissimilarity index (Bray and Curtis, 1957), Sorensen index (Sorensen, 1948) and Euclidean of ecological distance (Williams, 1971). All these indices give an indication how much a biological community at each sampled site is similar to the median of all reference communities and are not resulting in an assessment class as such.

Multivariate techniques are since the nineties also commonly applied for the development of multimetric systems. The selection of the metrics is based on how complementary or explanatory these are. The complementary of score systems is necessary to guarantee that correlated metrics do not dominate the overall assessment, while the explanatory aspects are interesting to get insight in the causes of deterioration. Since the new millennium, also a shift in use from multivariate statistical (classification, ordination, regression, clustering, etc. based on data distribution functions) to soft computing (based on heuristic search methods, e.g. artificial neural networks, inductive logic programming, etc.) techniques has started. Major examples of assessment systems using multivariate approaches are RIVPACS and AusRivAS (Davies, 2000). Examples of indices of the different assessment approaches based on macroinvertebrates are given in Table 2.3.

Presently, the most commonly applied indices in Europe are based on the saprobic and biotic approach. According to Nixon (2003) 11 countries (mainly Central and Eastern Europe) are assessing river water quality by means of the saprobic system, while another 11 are using one or another biotic index. The saprobic system would produce comparable results, whereas the biotic indices used by one country may not necessarily be comparable with that used in another. Recently, however, as an incentive of the European WFD also (stressor-specific) multi-metric systems, originating from the US, are now being developed and introduced. In contrast with the saprobic and biotic indices which are solely based on a community structure analysis, the multi-metric assessment systems may also include functional and non-taxonomic characteristics. Also, diversity indices which are nowhere used as a national standard are now being included as a separate metric in these multi-metric systems (e.g. Hering et al., 2004). For the assessments based on macroinvertebrates, also more and more use is being made of multivariate analysis which has the advantage to clearly link the biological communities to the river typology. Other characteristics which received attention during the last decade in assessments are the macroinvertebrate community structure related to the feeding strategy, migration or habitat use (e.g. Index of Trophic Completeness (ITC); Pavluk et al., 2000) and the use of key or target species (in how far does the species or taxa composition correspond with the expected composition of a particular type of surface water), (e.g. Lorenz et al., 2004).

2.3.2 Ecological indices used in Flanders and limitations for cost-benefit analyses

Also in Flanders, a number of ecological/biological indices are available that inform decision makers in a condensed way about the potential changes in the ecological quality as a result of their decisions. A brief overview of the two major contemporary river quality assessment methods used to steer the water system management at a regional level (Flanders in particular) is given underneath (Goethals and De Pauw, 2001).

The Belgian Biotic Index (BBI) was developed as a policy tool to get insight in the biological condition of watercourses in Flanders (De Pauw and Vanhooren, 1983). The methodology was standardized to allow a convenient application of the methodology in whole Flanders (Belgium). The BBI method uses macroinvertebrates as indicators for the level of pollution. The methodology is based on the theorem that increasing pollution will result in a loss of in-stream biodiversity and a progressive elimination of certain pollution sensitive groups. Besides the BBI, a Fish Index or Index of Biotic Integrity (IBI) (e.g. Belpaire et al., 2000) is still under development for the Flemish watercourses. The index is based on a set of fish community characteristics, often referred to as metrics and related to species composition, trophic composition and fish condition. Similar indices are under development for the other biological communities in the context of the implementation of the Water Framework Directive.

Next to several misuses (Seegert, 2000; Failing and Gregory, 2003), a major drawback to these indices is that they only allow gaining insight into the quality of a particular system from a rather limited point of view, namely the ecosystem status from an ecologist’s perspective. Also, they often do not allow for allocating the causes of the water system condition. For these reasons, it is difficult to make cost-benefit analyses on the basis of these ecological indices. Therefore, the use of models linking stakeholder activities to ecosystem status might be more useful to solve this type of questions, and also allows for proactive monitoring important to protect natural resources (Lawson et al., 2003).

Table 2.3 Examples of commonly applied biological assessment methods based on macroinvertebrates.

Approach/Method	Country	Reference
Saprobic approach		
Saprobic Index (S)	Austria	Moog, 1995
German Saprobic Index (S)	Germany	DEV, 1988-1991
Diversity approach		
Diversity index (H’)	Various	Shannon and Weaver, 1949
Biotic approach		
Belgian Biotic Index (BBI)	Belgium (Flanders)	De Pauw and Vanhooren, 1983; IBN, 1984
Bulgarian Biotic Index (BGBI)	Bulgaria	Uzunov et al., 1998
Indice Biotique Global Normalisé (IBGN)	France, Belgium (Wallonia)	AFNOR, 1992 Vanden Bossche and Josens, 2003
Danish Stream Fauna Index (DSFI)	Denmark	Skriver et al., 2001
Indice Biotico Esteso (IBE)	Italy	Ghetti, 1997
BMWP, ASPT	UK	Armitage et al., 1983
IBMWP	Spain	Alba-Tercedor and Sanchez-Ortega, 1988
Family Biotic Index (FBI)	USA	Hilsenhoff, 1988
IBPAMP	Argentina	Rodriguez et al., 2001
NEPBIOS	Nepal	Sharma and Moog, 2001
South African Score System (SASS)	South Africa	Chutter, 1972
Acid Class	Germany	Braukmann, 2001
Sequential Comparison Index (SCI)	USA	Cairns et al., 1968
Multimetric approach		
Index of Biotic Integrity (IBI)	USA	Barbour et al., 1992
Acidification Index	Sweden	Johnson, 1998
EBEOSWA	The Netherlands	STOWA, 1992
Environmental Quality Ratio (EQR) approach		
Environmental Quality Index (EQI)	UK	Sweeting et al., 1992
RIVPACS	UK	Wright et al., 2000
AUSRIVAS	Australia Indonesia	Smith et al., 1999 Sudaryanti et al., 2001
SWEPACS	Sweden	Sandin, 2001
Other approaches		
Index of Trophic Completeness (ITC)	Russia, The Netherlands	Pavluk et al., 2000
Gammarus/Asellus Index	UK	MacNeil et al., 2002

The development of models needed for water management already has a fairly long history, e.g. Young and Beck (1974). Applications of models in the field of water management show their practical relevancy to decision makers (Pullar and Springer, 2000; Lam et al., 2004). However, it also reveals some shortcomings. Often, a striking shortcoming is the strict use of a hydrological/chemical/biological/ecological dimension in presenting information to decision makers. Policy studies on water management in The Netherlands show that decision makers have difficulties in understanding these dimensions and hence their importance to water management (Bouma, 1998). Indicating the economic value of the ecological quality would greatly facilitate the assessment made by decision makers while evaluating the interventions in ecosystems. If this statement is accepted as a starting point, the related questions to overcome the shortcomings of the use of modelling (in particular ecological modelling) are related to sound data collection and model development strategies and methodologies for decision support in river management.

Although economic values for biological resources are increasingly being incorporated in cost-benefit evaluations of projects and policies, values for biodiversity tend not to be (Pearce, 2001). Much of the literature on the economic valuation of biodiversity considers the value of biological resources and is linked only tenuously to the value of diversity. This is especially true for the studies that use stated preference techniques - questionnaire approaches which ask directly for willingness to pay for the resource (contingent valuation), or which elicit a value indirectly (conjoint analysis) (Pearce, 2001). Ecologists also draw attention to a wider insurance value of diversity in terms of its value in ecosystem integrity and functioning (Dietz and Adger, 2003). The diversity of plants, animals and micro-organisms appears to have a role in helping ecosystems to organise themselves to cope with shocks and stresses. Put in another way, diversity would appear to be linked to resilience, the capacity of ecosystems to deal with externally imposed change (Pearce, 2001).

Up to now, contemporary river management in Flanders merely delivers ecological information on the basis of community indices, such as BBI and IBI (Goethals and De Pauw, 2001). In this manner, the field data are filtered from the perspective of ecologists, aiming at restoring the system towards its natural condition. Therefore, these indices are very difficult to use for non-ecologists, e.g. fishermen, preferring the optimisation of only particular characteristics of fish communities during a valuation exercise. The value that people attach to the characteristics of fish communities that are embedded in an index, such as biodiversity, evenness, biomass, amount of invasive species, etc., in the case of the IBI could be revealed by using, for example, contingent valuation or, for certain characteristics, market prices. In other words, what is the price for a certain number of fish (what can be specified on the basis of the species, size and weight). However a different monetary value will be obtained when one does not only look at the characteristics themselves, but also at the underlying factors that have led to those characteristics. The information from the indices is therefore often too scarce and not straightforward enough to allow the development of an efficient and effective river restoration policy due to the unknown specific local conditions and needs. With this information, it is merely possible to find out what the sites are that need specific management programmes to restore the systems (sanitation of very bad sites) or protect them (very good systems), but one is not able to select the most appropriate and optimal actions. Nevertheless, these indices were chosen as a major tool to assess water system conditions for the implementation of the WFD and will therefore play a major role in the water management in Europe during the coming years. However, when this surveillance monitoring will reveal river deterioration, additional monitoring will be necessary to allocate the impact to the human activities and set up a sanitation programme (D’heygere et al., 2002), as is the case in this study. For this purpose a combination of models and valuation methods could be considered. Therefore, models that can predict biological communities under different river conditions and related human activities are necessary.

2.4 Discussion about the need for (ecological) models and related monitoring strategies for decision support in water management

The use of cost-benefit analyses will be a good help for supporting the selection of convenient restoration actions, but for this it is also necessary that the data are collected in a convenient manner to set up the models and provide the appropriate information for the valuation process. The interfacing between monitoring, modelling and ecosystem valuation is therefore probably the major bottleneck to develop and use cost-benefit analyses in water management in Flanders and the rest of Belgium, because the data collection and model development strategies will have to be drastically changed for this purpose.

Therefore, one can state that the water management will probably have to be adapted from a regional towards a water system approach (river basin) to be able to deal with the particularities of each water system and the involved social and economic activities even beyond the borders of Belgium. For this, clear standards will be necessary for the biological monitoring approaches within a river basin. Particular attention needs to be paid to the standardisation of the selected river characteristics (variables), monitoring techniques, data base set up, identification level of the different communities, to ensure a convenient and necessary trans-boundary data exchange within the river basins. In July 2003, the SCALDIT project (http://www.scaldit.org) aiming at a standardisation and visualisation of the river data of the Scheldt was started. This project constitutes an important step towards a more sustainable water management approach for the major river basin in Flanders.

In addition, also predictive models will be necessary, to relate the different components within a system and get insight in the effect of changing one or more variables on the other ones. In particular models that can predict communities (habitat suitability models) have to be constructed. Therefore new approaches need to be developed that go further than the so far used water quantity and quality models.

2.5 Conclusions

A major conclusion for the water management in Flanders is that the contemporary monitoring and assessment of water systems based on ecological indices presently only allow for the allocation of major impacts, and is merely useful for surveillance monitoring as requested by the WFD. Within the policy area of water management economic valuation can play an important role to analyze the costs and benefits for river restoration options. This chapter shows that attaching a monetary value to ecological quality asks for linking ecological data to the use of economic valuation methods. The use of models that allow for a better allocation of the contribution of all stakeholders to the deterioration of the water system (water quality problems, floodings, ecosystem destruction, etc.) is therefore an important step forward. These models can for instance help to determine restoration costs, and in this manner (see Table 2.1), the value of ecosystems can be set. As such, these models can be important instruments to deliver the needed data for an integrated economic valuation of the water system and can help to obtain a more sustainable use of one of the most critical natural resources for mankind. In the following chapter will be described what type of (ecological) models are in particular needed for this purpose and how they can be constructed.

Chapter 3
State of the art of ecological modelling techniques to predict macroinvertebrates in rivers

Redrafted after:

Goethals, P., Dedecker, A., Gabriels, W. and De Pauw, N. (2002). Development and application of predictive river ecosystem models based on classification trees and artificial neural networks, p. 91-107. In: Recknagel, F. (Ed.), Ecological Informatics: Understanding ecology by biologically-inspired computation, Springer-Verlag, Berlin Heidelberg, Germany. 398 p.

Adriaenssens, V., De Baets, B., Goethals, P.L.M. and De Pauw, N. (2004). Fuzzy rule-based models for decision support in ecosystem management. The Science of the Total Environment 319: 1-12.

Adriaenssens, V., Goethals, P.L.M., Trigg, D. and De Pauw, N. (2004). Application of Bayesian Belief Networks for the prediction of macroinvertebrate taxa in rivers. Annales de Limnologie – International Journal of Limnology 40(3): 181-191.

3.1 Introduction

This chapter gives an overview of modelling approaches that can be used to get insight in aquatic ecosystems, what is necessary to improve decision making in water management. Although many ecological modelling methods already exist for several decennia, their practical application to support river management is rather limited, mostly because the direct benefits of the use of models for this purpose are not straightforward. In particular for river restoration management, there is a need of tools to guide the investments necessary to meet a good ecological status as set by the European Water Framework Directive. The major aim of this chapter is to review the use of some recently developed soft computing techniques such as artificial neural networks, classification trees, fuzzy logic and Bayesian belief networks for predicting aquatic communities in rivers on the basis of abiotic stream characteristics. Part of this chapter is devoted to model performance measures seen their importance for the validation of models. A good selection of these measures is in particular important for data driven methods.

The availability of proper datasets and modelling techniques allows the development of ecosystem models with high reliability (Recknagel, 2003). Several techniques such as models based on partial differential equations (Jorgensen, 1999) and multivariate statistics (Adriaenssens et al., 2002, 2004d) are already used for several decennia in this context. However, during recent years, artificial neural networks (Lek and Guégan, 1999), fuzzy logic (Adriaenssens et al., 2004a,b; Barros et al., 2000), classification and regression trees (Dzeroski et al., 1997; Blockeel et al., 1999a,b; Dzeroski and Drumm, 2003), Bayesian belief networks (Adriaenssens et al., 2004c), etc. proved to have a high potential in ecological modelling as well, as they combine reliable predictions with gaining insight in ecosystem interactions (Recknagel, 2001, 2003). So far, mainly data driven methods (e.g. artificial neural networks and classification trees) are preferred in this context, seen their time efficient development. However also knowledge based methods (e.g. fuzzy logic, Bayesian belief networks) can be of considerable importance, in particular when enough data of good quality are missing to develop data driven models (Goethals et al., 2004). All four methods will be described and reviewed in this chapter, with an emphasis on their use for the prediction of macroinvertebrates in rivers and lakes.

3.2 The need for soft computing methods to predict aquatic communities on the basis of abiotic water system characteristics

The impact of human activities has led to dramatic shifts in water systems all over the planet. More and more water uses got threatened, and since several decennia, protection and restoration actions were undertaken. However, the optimal balance between the different stakeholder activities needs a very deep insight in the integrated water system. In this context, models can show the limitations of the carrying capacity of certain regions and systems, based on sound science and/or experience. However, the practical need to simplify certain aspects of reality in these models, has led to disbelieve in the use of models. A major difficulty in using models exists in the fact that they are only convenient to use within a certain spatial, temporal, processes and application frame. When applied in a 'wrong' context, nonsense is simulated on the basis of the models, what has lead to quite some controversy between model developers and (ab)users. An in-depth study about the related credibility and acceptability of water management models was in this context made by van der Molen (1999). This author illustrated that the development of models has to be tuned to the needs of water managers, but that in the mean time the models need to be based on a convincing amount of scientific knowledge.

Several studies proved the practical benefits of models, in particular related to the management of ecosystems. In these cases models not only allowed to simulate new insights in ecosystem shifts as a result of human activities (e.g. Sheffer, 1990), but also showed practical relevance for use in environmental management. An example of such a system is the PCLake model (Figure 3.1). This model calculates water quality parameters of ponds such as nutrient concentration, chlorophyll-a, transparency, phytoplankton types and the biomass of submerged macrophytes. It also calculates the distribution and fluxes of the nutrients nitrogen and phosphorus in water and sediment. Inputs to the model are lake hydrology, nutrient loading, dimensions (mean depth and size), sediment characteristics and initial conditions. A main feature of the model is simulation of a possible shift between algae and vegetation dominance as a function of nutrient loading and other factors (Janse, 1997). In this context it was successfully applied to set nutrient standards to protect and restore lakes in The Netherlands by combining it with the LakeLoad model (van Puijenbroek and Knoop, 2002).

In this manner the effect of changes in land use and lake management could be calculated (van Puijenbroek et al., 2004).

However, a shortcoming to this type of models is the translation of the human activities to a species or taxon level, allowing to the calculation of the ecological quality on the basis of indices as required by the WFD for instance. Up to now, most 'integrated' ecological models are merely able to calculate water quality variables and some overall biomass figures about phytoplankton, macrophytes, fish communities, etc. without details about the composition of these communities. For this purpose, the use of habitat suitability models, also sometimes referred to as mini-models (e.g. Scheffer, 1990) can be useful for this more detailed type of calculations, where direct relations between a set of variables is calculated, without incorporating feedback loops. Such models can be seen as response models as described by Verdonschot et al. (1998), who developed a layered model (5-S-Model) based on a logical dominant hierarchy in river ecosystem relations.

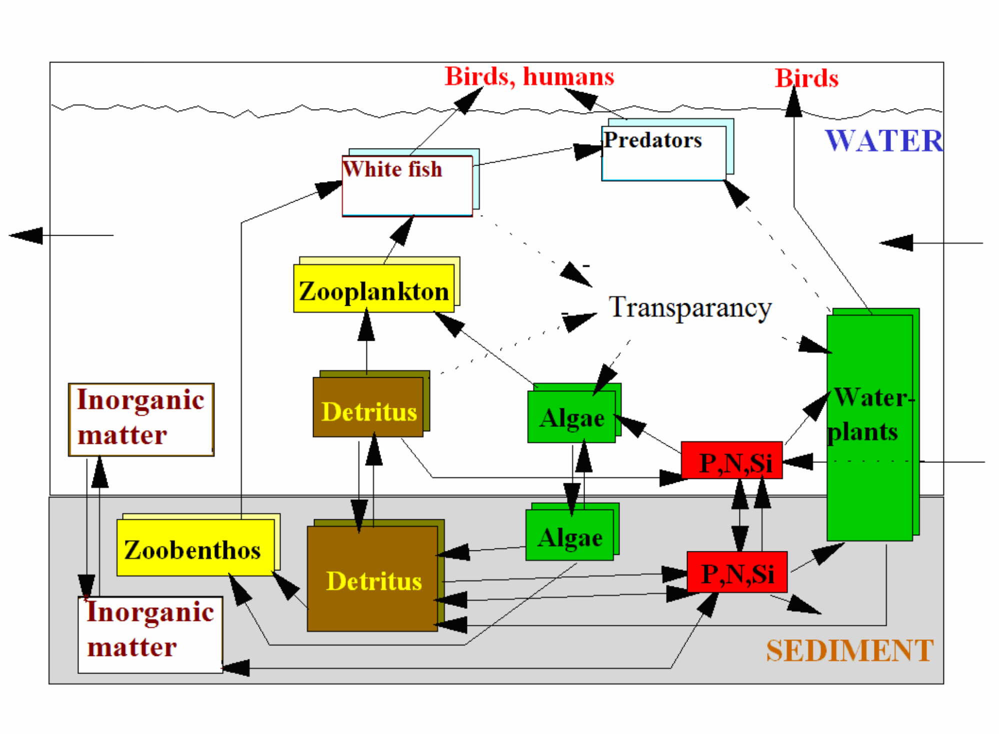


Figure 3.1 *PCLake model structure. Double blocks denote compartments modelled in both dry weight and nutrient units. Three functional groups of phytoplankton are distinguished: cyanobacteria, diatoms and other small edible algae. Solid arrows denote mass fluxes (e.g. food relationships), dotted arrows denote 'empirical' relationships (minus sign denotes negative influence, otherwise influence is positive). Egestion and mortality fluxes of animal groups and respiration fluxes are not shown. (after Jeuken et al., 1999).*

This 5-S-Model aims at enabling river managers to make proper choices based on a sound understanding of the functioning and interaction of the controlling factors. All the considerations on concepts, scales and hierarchies provide a conceptual basis for ‘catchment ecology’. The five main components of the model are (Verdonschot et al., 1998):

1. ‘System conditions’ comprises the structures and processes related to climate (temperature, rain-fall), geology and geomorphology (slope, soil composition) and set the boundary conditions for stream ecosystem functioning at a high hierarchical level in space (the catchment) as well as in time (±100 years). Generally, system conditions are not often changed by management.
2. ‘Stream hydrology’ comprises the hydrological processes of the catchment and the hydraulic processes of the stream and the habitat (Henry and Amoros, 1995). The two main directions of flow are one running from the boundary of the catchment towards the stream (lateral) and one running from source to mouth of the stream (longitudinal). Groundwater flow, precipitation and evaporation also play a role.
3. ‘Structures’ refer to the morphological features of the longitudinal and transversal stream bottom, banks and beds, as well as to the substrate patterns within. Cut of meanders, terrestrialization, sand deposits and other features of the stream valley are included here.
4. ‘Substances’ comprise the processes related to dissolved components like nutrients, organic matter, oxygen, major ions and contaminants. From catchment boundary towards the stream the amount of dissolved substances increases. This increase is also visible from source to mouth.
5. ‘Species’ are the response to the functioning of all above mentioned groups of controlling factors. ‘Species’ includes all taxonomic and non-taxonomic entities as well as biotic processes like production, respiration and so on. Species and their communities are the actual goal of ecological stream management and rehabilitation.

The five components mutually interact at different hierarchical levels and with different intensity. In general however, stream hydrology, structures and substances together compose the group of controlling factors that directly determine how the stream community functions. Nevertheless, numerous exceptions to this rule exist, e.g. species can adapt to stream hydrology and at the same time (e.g. trees) can impact stream hydrology and morphology. Thus, despite a dominant hierarchical effect, a feedback is always present. This feedback is in this research not taken into account and will be considered in future research. Also feedback within the same layer is not yet part of the modelled relations. Probably for presence/absence models it is of less importance, but for abundance models the interactions between species can become necessary to obtain good results. This will however also make part of future investigations. Therefore, in this research focus is laid on the interactions between the river characteristics and the inhabiting species by means of habitat suitability models.

A wide array of habitat suitability models has been developed to cover aspects as diverse as biogeography, conservation biology, climate change research and habitat or species management (Guisan and Zimmermann, 2000). The variety of model development techniques used is growing. The selection of an appropriate method should not depend solely on merely statistical performance considerations. Some models are better suited to reflect theoretical findings on the shape and nature of the species' response (or realized niche), while others are more convenient to support decision making in river management because of a good visual interface and ease to understand the model principles. A recent review of model design methods for such habitat models was provided by Verdonschot and Nijboer (2002) within the European PAEQANN-project (EVK1-CT1999-00026): 'Predicting Aquatic Ecosystem Quality using Artificial Neural Networks: Impact of Environmental Charateristics on the Structure of Aquatic Communities (Algae, Benthic and Fish Fauna)' (http://quercus.cemes.fr/paeqann). This overview illustrates that many techniques were not yet applied to relate river characteristics with biological communities. Mainly multivariate statistical techniques and ANNs are used so far. There is also a serious lack on studies comparing different techniques. Therefore, it is up to now very difficult to know what are the strengths and weaknesses of the different techniques, and under what conditions particular techniques can be applied and are best performing. Also the data preparation and model validation processes are often performed in different manners, making the comparison even more difficult between the available studies.

In Figure 3.2, a scheme for habitat suitability model development is presented. The first step consists of ecosystem component selection (input and output variables) to link habitat characteristics to community variables, followed by model training and model validation.

In the next parts of this chapter, validation methods and the major techniques for habitat suitability model development will be described and reviewed, with an emphasis on their use for the prediction of macroinvertebrates in rivers and lakes.

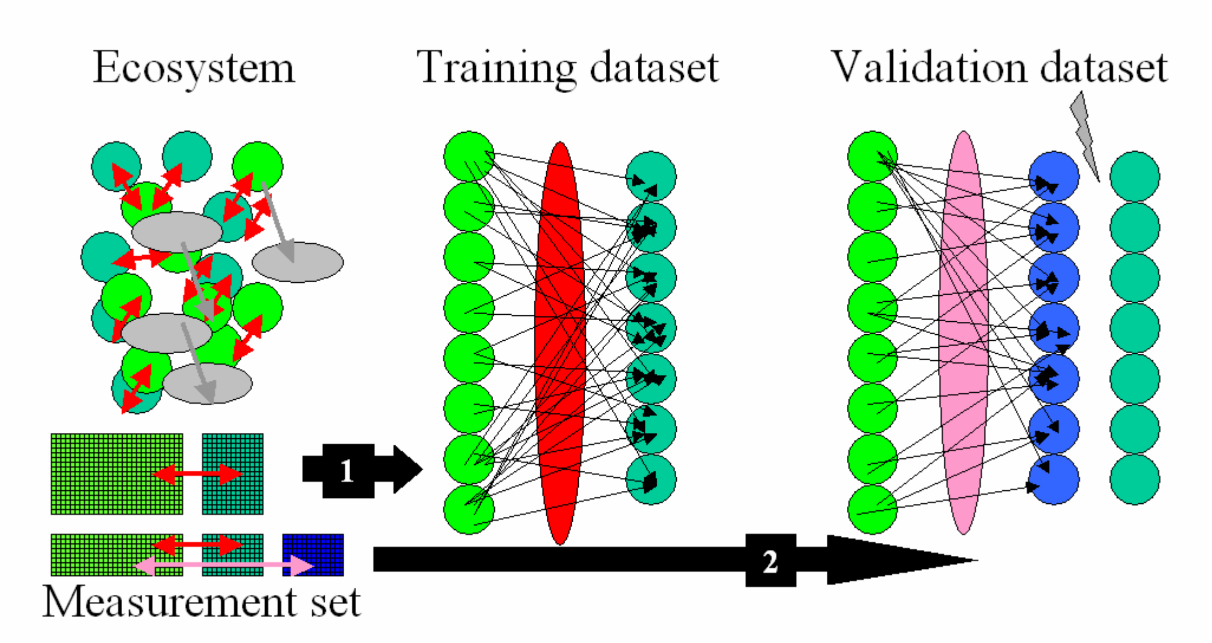


Figure 3.2 *Habitat suitability model development scheme: first step consisting of ecosystem component selection to link habitat characteristics to community variables (left), model training (middle) and model validation (right).*

3.3 Model performance measures and validation methods

3.3.1 Model performance measures

Given the importance to get insight in the model performance, the next paragraphs provide an overview of different indices that can be calculated to evaluate the quality of model predictions. This overview is an important start to get insight in what performance criteria can tell about the quality of models and which combinations need to be applied. The appropriate selection will be necessary to get more insight in which soft computing methods perform best for what type of problems.

Based on the output, different performance measures can be distinguished. When presence/absence of the macroinvertebrates are predicted, most of the papers apply the percentage of correctly classified instances (CCI) to assess model performance. There is however clear evidence that this CCI is affected by the frequency of occurrence of the test organism(s) being modelled (Fielding and Bell, 1997; Manel et al., 1999). This was practically illustrated for predictive models of macroinvertebrates by Goethals et al. (2002). Among the different measures, which are based on a confusion matrix (Table 3.2), proposed to assess the performance of presence/absence models (Table 3.3), Fielding and Bell (1997) and Manel et al. (1999) recommended the Cohen's kappa as a reliable performance measure. The effect of prevalence on the Cohen's kappa appeared indeed to be negligible (e.g. Dedecker et al., 2004a,c; D'heygere et al., 2004).

Table 3.2 The confusion matrix as a basis for the performance measures with true positive values (TP), false positives (FP), false negatives (FN) and true negative values (TN).

		Observed	
		+	-
Predicted	+	a (TP)	b (FP)
	-	c (FN)	d (TN)

Table 3.3 Measures based on the confusion matrix to assess the performance of presence/absence models (after Fielding and Bell 1997). NMI is the normalized mutual information statistic and N is the total number of instances.

Performance measure	**Calculation**
CCI	*(a+d)/N*
Misclassification rate	*(b+c)/N*
Sensitivity	*a/(a+c)*
Specificity	*d/(b+d)*
Positive predictive power	*a/(a+b)*
Negative predictive power	*d/(c+d)*
Odds-ratio	*(ab)/(cd)*
Cohen's kappa	$\frac{[(a+d)-(((a+c)(a+b)+(b+d)(c+d))/N)]}{[N-(((a+c)(a+b)+(b+d)(c+d))/N)]}$
NMI	$\frac{[-a.\ln(a)-b.\ln(b)-c.\ln(c)-d.\ln(d)+(a+b).\ln(a+b)+(c+d).\ln(c+d)]}{[N.\ln(N)-((a+c).\ln(a+c)+(b+d).\ln(b+d))]}$

When the output of a model consists of the species abundance, richness, diversity, density or a derived index, commonly used performance measures are the correlation (r) or determination (r^2) coefficient and the (root) mean squared error ((R)MSE) or a derivative between observed (O) and predicted (P) values (Table 3.4).

Table 3.4 Measures based on observed (O) and predicted (P) values to assess the performance of ANN models using abundance, richness, diversity, density or a derived index as model output. N is the total number of instances.

Performance measure	**Calculation**
Correlation coefficient (r)	$\dfrac{\sum(P \times O) - \dfrac{(\sum P \times \sum O)}{N}}{\sqrt{(\sum P^2 - \dfrac{(\sum P)^2}{N}) \times (\sum O^2 - \dfrac{(\sum O)^2)}{N})}}$
Determination coefficient (r^2)	$\left(\dfrac{\sum(P \times O) - \dfrac{(\sum P \times \sum O)}{N}}{\sqrt{(\sum P^2 - \dfrac{(\sum P)^2}{N}) \times (\sum O^2 - \dfrac{(\sum O)^2)}{N})}}\right)^2$
Root Mean Squared Error (RMSE)	$\sqrt{\frac{1}{N}\sum(P-O)^2}$
Mean Squared Error (MSE)	$\frac{1}{N}\sum(P-O)^2$

In the literature on the prediction of macroinvertebrates, r is most often used. Values of r larger than 0.4 can be accepted as an indication of good predictive performance of the concerning model.

Similar indices are developed for discrete variables characterized by more than two classes (e.g. Adriaenssens, 2004). Seen this type of models is not developed in this PhD thesis, they are not compared and discussed in this overview.

3.3.2 Data set splitting for training and validation

In most cases the amount of data for training and validation is limited. In particular for aquatic ecosystems, the data collection is very expensive (about 1,000 to 3,000 EUR for a combined measurement of biological communities with environmental characteristics at one site, mainly depending on the identification level and the type of environmental analyses), often resulting in relatively small datasets. Therefore the trade off between the size of the subsets for training and validation is of crucial importance and needs to be balanced to ensure that the training and validation are done in a ‘globally’ optimal manner.

The ‘holdout’ method reserves a certain amount of data for testing and uses the remainder for training (and even sets part of that aside for validation during the training process, if required by the model development algorithm). In practical terms, it is common to hold one-third of the data out for testing and use the remaining two-thirds for training (Witten and Frank, 2000). In general, to make the training and validation subsets as representative as possible, stratified training and validation subsets are created, what means that the classes of the predicted variable are evenly distributed over both subsets. In some cases also stratification of other variables can be useful.

However, to use all the data for training and validation, a subset swapping method is commonly applied. This technique is called ‘cross-validation’. In cross-validation, a fixed number of subsets, often called folds or partitions of the dataset are chosen. For instance, in threefold cross-validation, the dataset is split in three (approximately) equal partitions and each in turn is used for validation, while the remainder is used for training. This method is often applied for predicting the error rate of a learning algorithm. A standard way is based on stratified tenfold cross-validation (Witten and Frank, 2000). However, constraints due to data or time limitations can make fivefold or threefold cross-validation more convenient.

3.4 Soft computing methods for prediction of macroinvertebrates on the basis of abiotic water system characteristics

3.4.1 Fuzzy logic

Fuzzy logic, initiated by Zadeh (1965), can be seen as a soft computing technique making use of the fuzzy set theory. Fuzzy set theory enables to process imprecise information by means of an adaptable membership function, in contrast with binary crisp and limited functions (Zadeh, 1965; Zimmerman, 1990). The conventional membership of a crisp set takes only two values: one, when an element belongs to the set, and zero, when it does not. In the case of a fuzzy set an element can belong to the set with membership values ranging from zero to one. Real values can be transformed into linguistic values by an operation called fuzzification. Traditionally the symbol m has been used to represent the degree of fuzzy membership. If x represents the value of an environmental variable, then m(x) is the corresponding degree of membership in the set of acceptable conditions, and takes a value between zero and one. In the fuzzy logic inference system, the knowledge is represented by if-then linguistic rules. Fuzzy rules are evaluated for their degree of truth. Those that have some truth contribute to the final output state of the solution variable set (Meesters et al., 1997). The output can be a fuzzy or crisp value, depending on the inference method. In case of a fuzzy output of the inference engine, a final crisp value can be obtained by defuzzification. In most situations more than one environmental variable is important. Therefore the partial membership is defined to represent the acceptability of each environmental variable. The way in which partial memberships are combined depends on the application (Silvert, 2000).

By using fuzzy sets each measurement can be associated with more than one classification by specifying the partial membership in each set. For example in the field of water quality assessment, structural characteristics (degree of meandering, flow velocity, substrate type, etc.) which are often difficult to quantify or to classify as crisp sets can be valuable input variables in fuzzy models. Also some physical-chemical variables that are characterized by a high uncertainty and temporal variability, e.g. dissolved oxygen concentration, are often not appropriate to use as crisp values for prediction of biological communities and it can be more appropriate to use as a fuzzy set for this purpose. Because of the high non-linearity of ecological relations, fuzzy rules can be used to obtain a crisp or fuzzy output, depending on the applied inference system.

Fuzzy sets and systems have achieved success in business and engineering applications, mostly in the field of fuzzy control, in response to the need of flexible decisions in the face of rapid change, imperfect information, uncertainty, and ambiguous objectives (Terano et al., 1994). There are several applications in ecosystem management for which fuzzy models were designed. Most of the developed models are used for assessment of ecosystem integrity or sustainability. With regard to the prediction of organisms based on ecological variables and simulations of ecological interactions in the ecosystem based on fuzzy models, less research has been conducted so far. Kampichler et al. (2000) used fuzzy models for representing the soft knowledge of field-margin/spider-assemblage relationships. Mackinson (2000) used fuzzy logic in an expert system for predicting structure, dynamics and distribution of herring schools to capture and integrate scientific and local knowledge in the form of heuristic rules. Bock and Salski (1997) presented knowledge based modelling of the abundance of the yellow-necked mouse in a beech forest in northern Germany, based on a fuzzy logic model and a fuzzy knowledge based model has been constructed for the prediction of annual production of skylarks (Daunicht et al., 1996). Jorde et al. (2000) applied fuzzy rules, derived from experts' knowledge, to describe habitat preferences of fish species. A review is presented in Adriaenssens et al. (2004a).

Prototype models based on fuzzy logic were developed to predict macroinvertebrate taxa in the Zwalm river basin (Flanders, Belgium), based on expert knowledge and validated by ecological data of river basins in Flanders (Goethals et al., 2001; Adriaenssens et al., 2004b). Structural characteristics as well as physical-chemical variables were used as inputs to predict the presence/absence of different macroinvertebrate taxa. Predictive fuzzy models were constructed by means of a knowledge base that gave the basis for the fuzzification of the input variables and the construction of the fuzzy rule base. The knowledge base has been constructed using literature and an ecological data survey. Relevant and available input variables for prediction of these two taxa were selected based on multivariate analysis and fuzzificated into fuzzy sets. Each variable was divided into two trapezoidal fuzzy sets reflecting low and high values. Boundaries for the fuzzy sets were determined by the knowledge database. A fuzzy rule base system was constructed that connects the input variables to the output by means of if-then rules. These rules were implemented in a fuzzy inference system of the Sugeno type (Takagi and Sugeno, 1985), which produces a crisp output. 'And' was used as conjunction operator in the fuzzy rule base.

In the research of Adriaenssens et al. (2004b), four fuzzy models were constructed for predicting the macroinvertebrate taxa *Asellus* and *Gammarus*. The input variable selection has been based on the expert knowledge database and depended on the available variables in the ecological validation sets. Fuzzy predictive models for *Gammarus* were based on combinations of the input variables: conductivity, distance to source, habitat quality. Fuzzy predictive models for *Asellus* made use of the input variables conductivity, river width, dissolved oxygen, stream velocity and water level.

Until now, there are few fuzzy modelling tools that are directly useful in ecosystem management. There are two reasons: the exploration phase of fuzzy logic models development and the difficulty to convince managers to use these 'subjective' fuzzy models. There is also a need for easy-to-use tools and interfaces for fuzzy models in practice. This requires a good communication between ecologists and model developers.

The quality of the expert knowledge and ecological data as well as the methodology of fuzzy model construction has to be improved in the next research stages (Adriaenssens, 2004). This requires a good communication between field biologists and model developers. These future approaches in research together with the involvement of end-users in the model development process should enhance the reliability of fuzzy models with the final objective of their use in ecosystem management.

3.4.2 Bayesian belief networks (BBNs)

BBNs are models with a network structure that focus on the explicit representation of 'cause-and-effect' relationships between variables, representing in this case ecosystem components. The network architecture is linked to probability distributions that allow it to deal with variability and uncertainty in the models. This is particularly useful for the description of ecological systems (Regan, 2002). Despite some controversy (Dennis, 1996), Bayesian statistics have proven useful in ecology for evaluating and managing wildlife species and forests (Cohen, 1988; Haas et al., 1994; Crome et al., 1996; Lee and Riemann, 1997) and for other areas of environmental research and management (Dixon and Ellison, 1996; Ellison, 1996; Olson et al., 1990; Wolfson et al., 1996; Borsuk et al., 2002; Tattari et al., 2003; Borsuk et al., 2004). More recently, a computer based BBN system with potential for operational use in river management to diagnose river health has been developed in the United Kingdom, under the authority of the Environment Agency (Trigg et al., 2000; Walley et al., 2002).

Bayesian belief networks (Pearl, 1988) are probabilistic expert systems in which the knowledge base has two components: a network of causal relationships between variables; and a set of conditional probability matrices that relate each variable to its causal variables (Trigg et al., 2000). Bayes theorem lies at the heart of Bayesian inference. It is based on the use of probability to express knowledge and combine probabilities to characterize the advancement of knowledge. The simple, logical expression of Bayes theorem stipulates that, when combining information, the resultant (or posterior) probability is proportional to the product of the probability reflecting a priori knowledge (the prior probability) and the probability representing newly acquired knowledge (the sample information, or likelihood) (Reckhow, 2002). Expressed more formally, Bayes theorem states that the probability for y conditional on experimental outcome x (written p(y|x)) is proportional to the probability of y before the experiment (written p(y)) times the probabilistic outcome of the experiment (written p(x|y)).

The Bayesian analysis uses the knowledge gained from the previous analysis of data (= prior probability distribution, and when based on data = data driven prior) (Bernardo and Smith, 1994; Gelman et al., 1995). Conditional probability distributions (CPD) at each node need to be specified and if the variables are discrete, these can be represented as a table (CPT) that lists the probability that the child node takes on each of its different values for each combination of values of its parent. A new likelihood distribution is calculated from the new data. The new posterior distribution is intermediate to the prior likelihood and becomes zero where either the prior or likelihood becomes zero. The flows connecting the nodes, indicated by the arrows in a graphical model, represent causal relationships and represent conditional dependency (Reckhow, 2002). Conditional probability relationships can either be based on (1) experimental investigation, (2) collected field data, (3) process based models, or (4) elicited expert judgment. When appropriate and sufficient data do not exist, the elicited judgement of scientific experts may be required to quantify some probabilistic relationships (Borsuk et al., 2002).

One of the first applications of this technique on macrobenthos communities was made by Adriaenssens et al. (2004c). The paper describes a preliminary study evaluating the use of BBNs for prediction of two crustacean macroinvertebrate families, *Asellidae* and *Gammaridae* in rivers. Field data were used to represent the conditional probability relationships and expert judgement allowed the construction of the causal network. The authors compared the predictive success of one- and two-layered BBN networks.

In comparison with other predictive modelling techniques previously applied on data from the Zwalm river basin, such as ANN (Dedecker et al., 2004a) and fuzzy logic (Adriaenssens et al., 2004b), BBN networks showed a relative good predictive success based on only three input variables. However, a large inherent uncertainty was present in the predictions of all applied techniques, mainly because the applied database was rather small for training and validation. Other studies have found BBNs to perform well as predictive models (Walley and Dzeroski, 1995; Trigg et al., 2000; Fleishman et al., 2001), however in many of these, rigorous validation was not done or did not receive enough attention (Fleishmann et al., 2002).

The main highlights with regard to river management is that a BBN is (1) a medium that clearly displays the major influences on the wildlife population and their values and interactions, making them simple to explain (2) can combine information and different variables, (=multivariate approach) which can be both categorical or continuous variables, (3) combines empirical data with expert judgement, (4) expresses predicted outcomes as a likelihood, allowing analysis and risk management, and probability statements better represent the state of a population and the uncertainty involved in the prediction (5) can be used in a deductive way (Marcot et al., 2001). Based on the above-mentioned advantages, it is clear that the full potential of these techniques in river management can be directed to the support of policy decisions by coupling predictive bio-indicator models to mitigation, conservation and restoration actions (Marcot et al., 2001).

3.4.3 Artificial neural networks (ANNs)

3.4.3.1 General description of ANNs

Artificial Neural Networks (ANNs) are non-linear mapping structures that can be applied for predictive modeling and classification. Various types of neural networks exist, suitable to solve different kinds of problems. The choice of the type of network depends on the nature of the problem to be solved. The most popular ANNs are multilayer feedforward neural networks with the backpropagation algorithm, i.e. backpropagation networks (Rumelhart et al., 1986; Hagan et al., 1996) and Kohonen self-organizing maps, i.e. Kohonen networks (SOMs) (Kohonen, 1982). However, the latter are mainly interesting for clustering data and will not be further discussed.

The backpropagation network is based on the 'supervised' procedure. The network constructs a model based on examples of data with known outputs. It has to build up the model solely from the examples presented, which are together assumed to contain the information necessary to establish the relation. An example of a relation can be the abundances of a number of macroinvertebrate taxa (such as Gammaridae, Tubificidae, Chironomidae) which are being predicted based on a number of environmental variables such as flow velocity, percentages of clay, silt and sand in the sediment, river depth, dissolved oxygen, pH,… To make reliable predictions it is better to rescale the input variables, because they can have very different orders of magnitude. For example, the input variables can be rescaled to be included within the interval [-1, 1] by using the following equation:

$$V_n = 2 \times \frac{(V_0 - V_{\min})}{(V_{\max} - V_{\min})} - 1$$

in which V_0 and V_n are, respectively, the old and new value of the variable for a sampling point, V_{min} and V_{max} are the minimum and maximum values of that variable in the original dataset. The architecture of the backpropagation network is a layered feed-forward neural network in which the non-linear elements, the neurons, are arranged in successive layers, and the information flows from input layer to output layer, through the hidden layer(s) (Figure 3.3). As can be seen in Figure 3.4, nodes from one layer are connected to all nodes in the following layer, but no lateral connections within any layer, nor feedback connections are possible. In the example mentioned above, each input neuron would represent one environmental variable and each output neuron the abundance of one macroinvertebrate family. With the exception of the input neurons, which merely connect one input value with its associated weight values, all neurons can be visualised with their connections as in Figure 3.6. The inputs are indicated as x_1, x_2, … x_n, each associated with a quantity called weight or connection strength w_{j1}, w_{j2}, … w_{jn} for the input to the j-th neuron. The net input for each neuron is the sum of all input values, each multiplied by its weight, and z_j a bias term which may be considered as the weight from a supplementary input equalizing one:

$$a_j = \sum w_{ji}x_i + z_j$$

The output value, y_j, can be calculated by feeding the net input into the transfer function of the neuron:

$$y_j = f(a_j).$$

Many transfer functions may be used, e.g. a linear function or most often a sigmoid function. The number of input and output nodes depends on the number of the input and output objects.

For determining the values of weights and biases in a backpropagation network, all the weights and biases are initially set to small random numbers. Subsequently, a set of input/output ensembles is presented to the network. For example, the input can be a set of 15 environmental variables determined at a certain amount of sampling sites and the output the abundances of species or taxa (macroinvertbrates, fish, macrophytes, etc.) sampled at each of these sites. For each of the input sets, the output is calculated by the ANN, and an error term is calculated by comparing the calculated output with the desired output (the ‘target’). Using this error term, the weights and biases are updated in order to decrease the error, so future outputs are more likely to be correct. This procedure is repeated until the errors become small enough or a predefined maximum number of iterations is reached. This iterative process is termed ‘training’. After the training, the ANN can be tested using independent data.

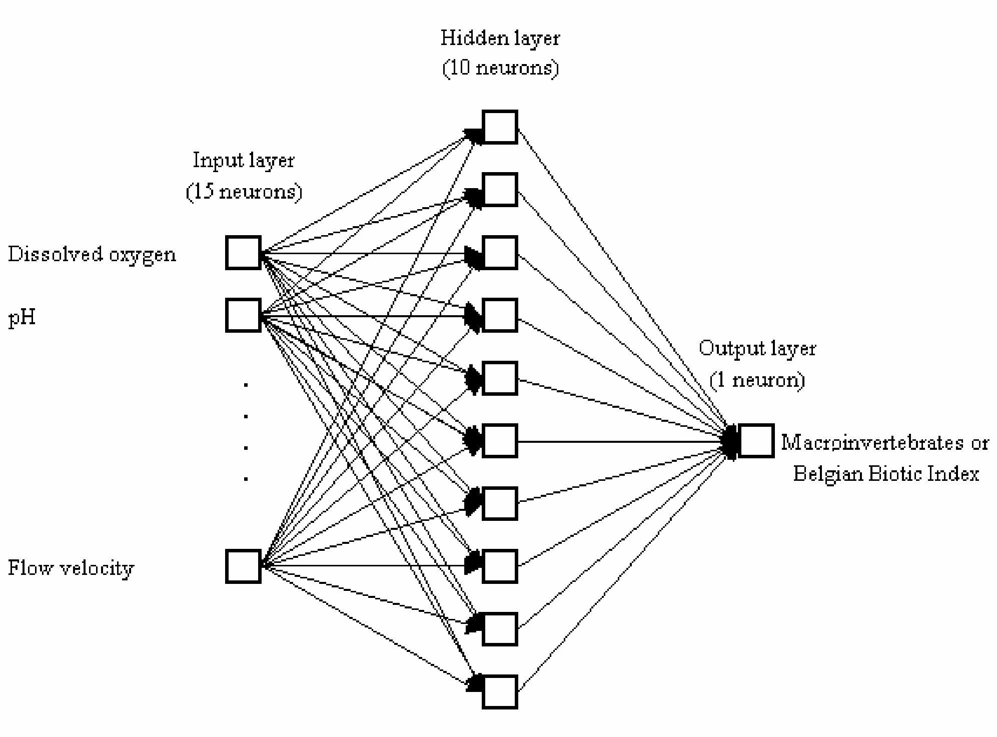


Figure 3.3 *Illustration of a three-layered (15-10-1) artificial neural network with input layer, one hidden layer with ten neurons and output layer (Dedecker et al., 2002).*

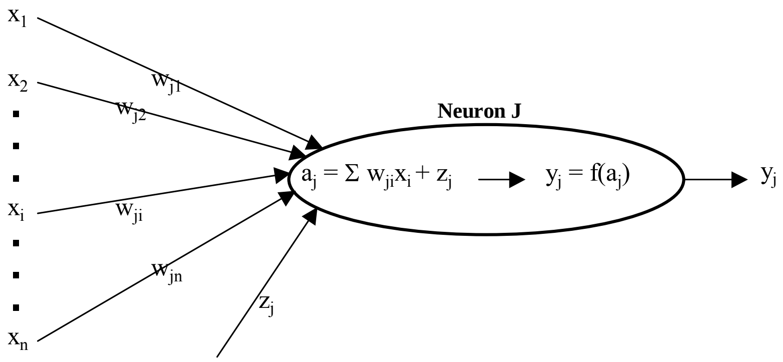


Figure 3.4 *Scheme of a neuron in a backpropagation network receiving input values from n neurons, each associated with a weight, as well as a bias z_j. The resulting output value y_j is computed according to the presented equations.*

3.4.3.2 Predictive ANN development

3.4.3.2.1 Data processing

Generally, different variables span different ranges. In order to ensure that all variables receive equal attention during the training process, they should be standardised. In addition, the variables have to be scaled in such a way as to be commensurate with the limits of the activation functions used in the output layer (Maier and Dandy, 2000). Several authors (Wagner et al., 2000; Chon et al., 2001, 2002; Gabriels et al., 2002; Obach et al., 2001; Park et al., 2003a, 2003b; Schleiter et al., 1999, 2001) proportionally normalised the data between zero and one [0 1] in the range of the maximum and minimum values. Dedecker et al. (2004a, c) and Gabriels et al. (2002) on the other hand, rescaled the variables to be included within the interval [−1 1]. Maier and Dandy (2000) mentioned if values are scaled to the extreme limits of the transfer function, the size of the weight updates during training is extremely small and flat spots are likely to occur.

3.4.3.2.2 Bandwidth

Lek and Guégan (1999) stated that ANN models are built solely from the examples presented during the training phase, which are together assumed to implicitly contain the information necessary to establish the relation between input and output. As a result, ANNs are unable to extrapolate beyond the range of the data used for training. Consequently, poor predictions can be expected when the validation data contain values outside of the range of those used for training (Maier and Dandy, 2000). Dedecker et al. (2004b) tested the sensitivity and robustness of the ANN models when data, containing variables beyond the range of the data for training, were added. Therefore, a virtual dataset based on ecological expert knowledge to introduce ‘extreme’ values to the model was created. The obtained results indicated that the output in the validation set was predicted significantly better when the number of ‘extreme’ examples in the training set increased. However, the overall predictive power of the ANN models decreased when a relatively large virtual dataset in the training set was applied.

3.4.3.2.3 Learning method

Neural network algorithms can be divided into supervised and unsupervised training methods. Unsupervised learning methods do not require output values for the training process and are mainly used for classification problems, which is beyond the scope of this review.

The suitability of a particular method is often a trade-off between performance and calculation time. The majority of the ANNs used for prediction are trained with the backpropagation method (e.g. Cherkassky and Lari-Najafi, 1992; Maier and Dandy, 2000). Because of its generality (robustness) and ease of implementation, backpropagation is the best choice for the majority of ANN systems. Backpropagation is the superior learning method when a sufficient number of relatively noise-free training examples are available, regardless of the complexity of the specific domain problem (Walczak and Cerpa, 1999). Although backpropagation networks can handle noise in the training data (and may actually generalise better if some noise is present in the training data), too many erroneous training values may prevent the ANN from learning the desired model. When only a few training examples or very noisy training data are available, other learning methods should be selected instead of backpropagation (Walczak and Cerpa, 1999). Radial basis funcion networks perform well in domains with limited training sets (Barnard and Wessels (1992) in Walczak and Cerpa (1999)) and counterpropagation networks perform well when a sufficient number of training examples is available, but may contain very noisy data (Fausett and Elwasif (1994) in Walczak and Cerpa (1999)).

In order to optimise the performance of backpropagation networks, it is essential to note that the performance is a function of several internal parameters including the transfer function, error function, learning rate and momentum term. The most frequently used transfer functions are sigmoidal ones such as the logistic and hyperbolic tangent functions (Maier and Dandy, 2000). However, other transfer functions may be used, such as hard limit or linear functions (Hagan et al., 1996). The error function is the function that is minimised during training. The most commonly used error function is the mean squared error (MSE) function. However, in order to obtain optimal results, the errors should be independently and normally distributed, which is not the case when the training data contain outliers (Maier and Dandy, 2000). To overcome this problem, Liano (1996) proposed the least mean log squares (LMLS) error function. The learning rate is directly proportional to the size of the steps taken in weight space. Traditionally, learning rates remain fixed during training (Maier and Dandy, 2000) and optimal learning rates are determined by trial and error. However, heuristics have been proposed which adapt the learning rate as training progresses to keep the learning step size as large as possible while keeping learning stable (Hagan et al., 1996). A momentum term is usually included in the training algorithm in order to improve learning speed (Qian, 1999) and convergence (Hagan et al., 1996). The momentum term should be less than 1.0, otherwise the training procedure does not converge (Dai and Macbeth, 1997). Dai and Macbeth (1997) suggest a learning rate of 0.7 with a momentum term of at least 0.8 and smaller than 0.9 or a learning rate of 0.6 with a momentum term of 0.9. Qian (1999) derived the bounds for convergence on learning rate and momentum parameters, and demonstrated that the momentum term can increase the range of learning rates over which the system converges.

3.4.3.2.4 Model architecture

According to Haykin (1999), generalisation capability of a neural network is influenced by three factors: the size of the training set and how representative it is of the environment of interest, the architecture of the neural network, and the complexity of the problem studied. The architecture is the only of these three factors that can be influenced in the modelling process, making it a crucial step, which should be considered carefully.

Walczak and Cerpa (1999) distinguish four design criteria for artificial neural networks which should be decided upon in subsequent steps: knowledge-based selection of input values, selection of a learning method, design of the number of hidden layers and selection of the number of hidden neurons for each layer. The selection of the learning method was already described earlier (see 3.4.3.2.3).

Input variable selection

Data driven approaches, such as ANN models, have the ability to determine which model inputs are critical. However, presenting a large number of inputs to ANN models, and relying on the network to determine the critical model inputs, usually increases network size. This has a number of disadvantages, for example decreasing processing speed and increasing the amount of data required to estimate the network parameters efficiently (Maier and Dandy, 2000). In this way, selection of input variables can be stated as an important task. It can considerably reduce the necessary labour of data collection. Complex systems can be reduced to easily surveyed models with low measuring and computing effort. Therewith they are particularly suitable for (bio-)indication in aquatic ecosystems (Schleiter et al., 2001).

Several methods can be followed to determine the optimal set of input variables. The first one is to perform standard knowledge acquisition. Typically, this involves consultation with multiple domain experts. Walczak (1995) has indicated the requirement for extensive knowledge acquisition utilizing domain experts to specify ANN input variables. The primary purpose of the knowledge acquisition phase is to guarantee that the input variable set is not under-specified, providing all relevant domain criteria to the ANN. Once a base set of input variables is defined through knowledge acquisition, the set can be pruned to eliminate variables that contribute noise to the ANN and consequently reduce the ANN generalisation performance. ANN input variables should not be correlated. Correlated variables degrade ANN performance by interacting with each other as well as with other elements to produce a biased effect. From an ecological point of view, relationships between environmental variables and taxonomic richness should be considered with caution, as these analyses, based on correlation, do not necessarily involve relevant ecological processes. However, the only way to establish reliable causal relationships between input and output, is to use experimental designs (Beauchard et al., 2003). A first pass filter to help identify 'noise' variables is to calculate the correlation of pairs of variables. If two variables are strongly correlated, then one of these two variables may be removed without adversely affecting the ANN performance. The cut-off value for variable elimination is a heuristic value and must be determined separately for every ANN application, but any correlation absolute value of 0.20 or higher indicates a probable noise source to the ANN (Walczak and Cerpa, 1999).

In addition, there are distinct advantages in using analytical techniques to help determine the inputs for ANN models (Maier and Dandy, 2000). Schleiter et al. (1999, 2001), Obach et al. (2001) and Beauchard et al. (2003) used a stepwise procedure to identify the most influential variables. In this approach, separate networks are trained for each input variable. The network performing best is retained and the effect of adding each of the remaining inputs in turn is assessed. This process is repeated for three, four, five, etc. input variables, until the addition of extra variables does not result in a significant improvement in model performance. On the other hand, one can start with all the available variables and remove one by one the least important ones (e.g. Beauchard et al., 2003). Disadvantages of these approaches are that they are computationally intensive and that they are unable to capture the importance of certain combinations of variables that might be insignificant on their own. Schleiter et al. (1999, 2001), Wagner et al. (2000) and Obach et al. (2001) applied a special variant of the backpropagation network type, the so-called senso-net, to determine the most important input variables (sensitivity analysis). Senso-nets include an additional weight for each input neuron representing the relevance (sensitivity) of the corresponding input parameter for the neural model. The sensitivities are adapted during the training process of the network. Appropriate subsets of potential input variables can be selected according to these sensitivities. A third technique which is frequently used is genetic algorithms (e.g. Obach et al., 2001; Schleiter et al., 2001; D'heygere et al., 2004). This technique automatically selects the relevant input variables (Goldberg, 1989).

So far, in most cases for the prediction of macroinvertebrates, merely ecological expert knowledge is used to select the input variables. The transformation of the variables is only used for continuous variables given the relative high amount of zero's that are typical for the ecological datasets. For this, commonly a log (abundance+1) transformation is applied as advised by Legendre and Legendre (1998).

The contribution of input variables, is another very important aspect that needs more research. Many variables are not part of the dataset, while others have a high variability that can be caused by measurement difficulties, but also by the natural changes in the river systems. Therefore, also the effect of monitoring methods needs more research, in particular the incorporation of 'new' variables which are less straightforward to be used in a model. This is in particular the case for structural and morphological variables that often need to be visually monitored, but also for heavy metals and other potential toxicants, since their effects are often related to the environment where they are released (bio-availability, accumulation, etc.). These toxicants may be a new challenge in the field of soft computing models to predict river communities, in particular macroinvertebrates.

Number of hidden layers

A greater number of hidden layers enables an ANN to improve its closeness-of-fit, while a smaller quantity improves the smoothness or extrapolation capabilities of the ANN (Walczak and Cerpa, 1999). Theoretically, an ANN with one hidden layer can approximate any function as long as sufficient neurons are used in the hidden layer (Hornik et al., 1989). Flood and Kartam (1994) suggest using two hidden layers as a starting point. However, it must be stressed that optimal network geometry is highly problem dependent.

Number of hidden neurons

The number of neurons in the input layer is fixed by the number of model inputs, whereas the number of neurons in the output layer equals the number of model outputs. The critical factor however is the choice of the number of neurons in the hidden layer. More hidden neurons result in a longer training period, while fewer hidden neurons provide faster training at the cost of having fewer feature detectors (Bebis and Georgiopoulos, 1994). For two networks with similar errors on training sets, the simpler one (the one with fewer hidden units) is likely to produce more reliable predictions on new cases, while the more complex model implies an increased chance of overfitting on the training data and reducing the model's ability to generalise on new data (Hung et al., 1996; Özesmi and Özesmi, 1999). Hecht-Nielsen (1987) showed that any continuous function with N_i inputs in the range [0 1] and N_o outputs can be represented exactly by a feedforward network with $2N_i+1$ hidden neurons.

Various authors propose rules of thumb for determining the number of hidden neurons. Some of these rules are based on the number of input and/or output neurons, whereas others are based on the number of training samples available. Walczak and Cerpa (1999) warn that these heuristics do not use domain knowledge for estimating the quantity of hidden nodes and may be counterproductive. Table 3.5 shows the rules that suggest the number of hidden neurons based on the number of input (N_i) and/or output (N_o) nodes.

Some authors suggest rules to determine the necessary number of training samples (S) based on the number of connection weights. Since the number of training samples is fixed, inverting these rules provides an indication of the maximum number of connection weights to avoid overfitting (Table 3.6).

Table 3.5 Rules suggesting the number of hidden neurons based on the number of input (N_i) and/or output (N_o) nodes

Rule	**Reference**
$(2/3)\ N_i$	Wang, 1994
$0.75\ N_i$	Lenard et al., 1995
$0.5\ (N_i + N_o)$	Piramuthu et al., 1994
$2\ N_i + 1$	Fletcher and Goss, 1993; Patuwo et al., 1993
$2\ N_i$ or $3\ N_i$	Kanellopoulos and Wilkinson, 1997

Table 3.6 Indication of the maximum number of connection weights to avoid overfitting based on the number of training samples (S)

Maximum number of connection weights	**Reference**
S	after Rogers and Dowla, 1994
S/2	after Masters, 1993
S/4	after Walczak and Cerpa, 1999
S/10	after Weigend et al., 1990
S/30	after Amari et al., 1997

The number of hidden neurons necessary can be calculated given the number of connection weights and the number of input and output neurons.

Rules of thumb are clearly divergent and when selecting the number of hidden neurons, one should take both S and N_i into account. Assuming only one hidden layer is used, the number of connection weights should not exceed, say, S/10 and the number of hidden neurons should be at least, roughly, $(N_i + N_o)/2$. Evidently, in order to be able to meet both constraints, the number of training samples has to be sufficiently large.

According to Walczak and Cerpa (1999), the number of hidden neurons in the last layer should be set equal to the number of decision factors used by domain experts to solve the problem. Decision factors are the distinguishable elements that serve to form the unique categories of the input vector space. The number of decision factors is equivalent to the number of heuristic rules or clusters used in an expert system (Walczak and Cerpa, 1999).

Alternatively, techniques for automatically selecting ANN architecture with the required number of hidden units may be used. Such techniques were proposed by e.g. Bartlett (1994), Nabhan and Zomaya (1994) and Anders and Korn (1999).

3.4.3.2.5 Model interpretation

Although in many studies ANNs have been shown to exhibit superior predictive power compared to traditional approaches, they have also been labelled as a ‘black box’ because they provide little explanatory insight into the relative influence of the independent variables in the prediction process (Olden and Jackson, 2002). This lack of explanatory power is a major concern to ecologists since the interpretation of statistical models is desirable for gaining knowledge of the causal relationships driving ecological phenomena (Karul et al., 2000). As a consequence, various authors have explored this problem and proposed several algorithms to illustrate the role of variables in ANN models. Sensitivity analysis is frequently used (Mastrorillo et al., 1997a; Guégan et al., 1998; Laë et al., 1999; Chon et al., 2001; Hoang et al., 2001, 2002; Dedecker et al., 2002, 2004c; Marshall et al., 2002; Olden and Jackson, 2002; Brosse et al., 2003) and is based on a successive variation of one input variable while the others are kept constant at a fixed value (Lek et al., 1995, 1996a, b). Gevrey et al., (2003), Dedecker et al. (2004b, d) and Beauchard et al. (2003) used the 'PaD' method (Dimopoulos et al, 1995, 1999) which consists in a calculation of the partial derivatives of the output according to the input variables. Several authors (Mastrorillo et al., 1997b; Brosse et al., 1999, 2001, 2003; Olden and Jackson, 2002; Gevrey et al., 2003; Dedecker et al., 2004b, d) applied Garson's algorithm (Garson, 1991; Goh, 1995). This algorithm is based on a computation using the connection weights. The 'Perturbation' method (Yao et al., 1998; Scardi and Harding, 1999) assesses the effect of small changes in each input on the neural network output (e.g. Park et al., 2003a; Gevrey et al., 2003; Dedecker et al., 2004b, d). Gevrey et al. (2003) and Dedecker et al. (2004b, d) applied the 'Stepwise' procedure, as discussed earlier, to identify the most influential variables. Özesmi and Özesmi (1999) described the neural interpretation diagram (NID) to provide a visual interpretation of the connection weights among neurons. The relative magnitude of each connection weight is represented by line thickness and line shading represents the direction of the weight. Olden and Jackson (2002) proposed a randomisation test for input-hidden-output connection weight selection in ANN models. By eliminating connection weights that do not differ significantly from random, they simplified the interpretation of neural networks by reducing the number of axon pathways that have to be examined for direct and indirect (i.e. interaction) effects on the response variable, for instance when using NIDs. Olden et al. (2004) compared these methodologies using a Monte Carlo simulation experiment with data exhibiting defined numeric relationships between a response variable and a set of independent predictor variables. By using simulated data with known properties, they could accurately investigate and compare the different approaches under deterministic conditions and provide a robust comparison of their performance.

3.4.3.3 Predictive ANN development studies of aquatic macroinvertebrates

Table 3.7 gives an overview of articles discussing case studies on the prediction of macroinvertebrates by means of ANNs. A total of 26 cases were found in literature. These papers were however produced by a far smaller number of research groups, since most of the research groups published more than one paper on the subject. Among them, there is a French, Belgian, German, British, South-Korean and Australian research group, counting up to 6 groups although this number is debatable because the groups do not work completely independently, as some cooperative papers clearly testify. All papers are very recent, the oldest one being from 1998.

About one out of two papers mentioned the software package used for modelling. Three different packages were cited: MATLAB, WEKA and NNEE. Many of the modellers who did not mention the software package implemented their own code in an existing modelling environment such as MATLAB. Evidently, the software package used should not influence modelling results although neither using own programming nor existing software is an absolute guarantee that small errors will not occur, so any system should be used with care.

The number of input variables ranged from 3 to 39, usually between 5 and 15. Among these variables were geographical and seasonal variables and habitat quality parameters (sinuosity, vegetation, etc.) as well as physical-chemical properties (dissolved oxygen, water temperature, pH, nutrient concentrations, COD, etc.) and characteristics of toxicity. The performance of neural networks with more input variables is not necessarily higher, as is shown in some studies (e.g. Walley and Fontama, 1998). The target variables were usually presence/absence (8 cases) or abundance (7 cases) of macroinvertebrate taxa or derived properties such as taxa richness, ASPT score or exergy.

The neural networks were almost in all cases of the feedforward connection type, in some cases combined with a self organising map (SOM). Exceptions included real-time recurrent neural networks, an Elman recurrent neural network and a forward only neural network. Most self organising maps were trained with the Kohonen learning rule, one was trained with a radial basis function. Most feedforward neural networks were trained with backpropagation or a modification of it. In some cases the Levenberg-Marquardt algorithm, general regression, a linear neural network and/or counterpropagation were tested. The real-time recurrent neural networks were trained with recurrent learning and the Elman recurrent neural network was trained with backpropagation.

Network architecture was reported in most cases. The number of hidden layers was usually one and in none of the reported cases higher than two. The number of hidden neurons was usually of the same order of magnitude as the number of input nodes. Network architecture was determined, if stated, by ‘trial and error’ (6 cases), ‘empirically’ (2 cases) or ‘arbitrarily chosen’ (1 case). In the majority of the cases, the choice of network architecture was not discussed at all. Clearly, this crucial step in the modelling process is poorly documented for this type of applications. In general, rules of thumb were not (explicitly) used while trial and error was applied without a clear strategy. However, it is recommended to use rules of thumb as a starting point for a trial and error process in order to refine and validate the choice of neural network architecture. In addition, techniques for model optimisation were hardly used to optimise network geometry.

The transfer functions, where specified, were of the sigmoid transfer function type. The data was generally rescaled to the interval [-1 1] or [0 1]. Maier and Dandy (2002) recommend avoiding the extreme limits of the transfer function when rescaling the outputs. However, in only one study (Park et al., 2001) an interval smaller than the transfer function allows was used.

A variety of performance measures was used, strongly related to the type of output parameter. For predictions of presence/absence, the percentage of CCI was the most frequently used performance measure. In some cases Cohen’s kappa (K) was calculated and in one case also the RMSE. When predicting continuous variables such as abundance or taxa richness, a variety of criteria were calculated in the cited case-studies: r, r^2, MSE, RMSE. Also the cross-validation error (CVE) and the percentage or the proportion of predictions within a specified distance of the observed value were applied as alternatives to these more common performance measures. Two other measures were used after transforming the abundance outputs into abundance classes: CCI and Cohen’s kappa (K). Some of the performance criteria used may however result in a biased representation of performance, e.g. CCI (e.g. Fielding and Bell, 1997; Manel et al., 2001). A good recommendation would be to use several performance measures to acquire a more reliable model evaluation.

Table 3.7 Overview of ANN models used to predict macroinvertebrates in rivers and lakes.

Reference	Software package	Input variables	Output	Location(s)	Connection type	Training algorithm	Network architecture	No. train. samples – No. test samples	Determination of network architecture	Transfer functions	Scaling of variables	Perf. measure
Beauchard et al. 2003	N/S	A, P, Lo, R, DISTs, SDA	richn	Morocco, Algeria, Tunisia	FF	BP	7-4-1	210-1 (leave-one-out)	Empirically	STF	N/S	*r*
Brosse et al. 2001	MATLAB	A, SDA, SO, CA, VEG, AE, D, W, S	div	Taieri river (New Zealand)	FF	BP	10-4-1	96-1 (leave-one-out)	N/S	STF	N/S	*r*, PI
Brosse et al. 2003	MATLAB	LU, SDA, A, CA, PR, SO, W, D, S	div	Taieri river (New Zealand)	FF	BP	(10, 8)-4-1	96-1 (leave-one-out)	N/S	STF	N/S	*r*, MSE
Céréghino et al. 2003	N/S	A, SO, DISTs, T	richn	Adour-Garonne river basin (France)	FF	BP	4-5-1	130-25	Trial and error	N/S	N/S	*r*
Chon et al. 2001	N/S	MI, FV, D, OM, S	comm	Yangjae stream (Korea)	RTRC	RL	(7+4)-13-7	N/S	Trial and error	N/S	[0 1]	*r*
Chon et al. 2002	N/S	MI MI, FV, D, OM, S	dens comm	Yangjae stream (Korea)	FF ERC RTRC	BP BP RL	(5-25)-(8-30)-5 5-30-5 (7+4)-13-7	N/S	Empirically	STF	[0 1]	*r*
Dedecker et al. 2002	MATLAB	T, pH, DO, Cond, SS, D, W, S, Sh, VEG, FV, Me, HRB, PR, AE	pr/ab	Zwalm river basin (Belgium)	FF	BP	15-10-1	40-20 (3-fold) 45-15 (4-fold)	Trial and error	STF	N/S	CCI
Dedecker et al. 2004a	MATLAB	T, pH, DO, Cond, SS, D, W, S, Sh, VEG, FV, Me, HRB, PR, AE	pr/ab	Zwalm river basin (Belgium)	FF	BP, LM	15-(2, 5, 10, 20, 25)-(5, 10)-1	108-12 (10-fold)	Trial and error	STF	[-1 1]	CCI, *K*
Dedecker et al. 2004b	MATLAB	T, pH, DO, Cond, SS, D, W, S, FV, Me, HRB, PR, AE, NO_3^-, PO_4^{3-}, NH_4^+, COD, Ph, Ni, SO, DISTm	abu	Zwalm river basin (Belgium)	FF	BP	24-10-1	119-60 (3-fold)	N/S	STF	IN N/S OUT[log (abu+1)]	*r*
Dedecker et al. 2004c	MATLAB	T, pH, DO, Cond, SS, D, W, S, Sh, VEG, FV, Me, HRB, PR, AE	pr/ab	Zwalm river basin (Belgium)	FF	BP, LM	15- N/S -1	108-12 (10-fold)	Trial and error	STF	[-1 1]	CCI, *K*

Dedecker et al. 2004d	MATLAB	T, pH, DO, Cond, SS, D, W, S, FV, Me, HRB, PR, AE, NO$_3^-$, PO$_4^{3-}$, NH$_4^+$, COD, Ph, Ni, SO, DISTm	abu	Zwalm river basin (Belgium)	FF	BP	24-10-1	119-60 (3-fold)	N/S	STF	IN N/S OUT[log (abu+1)]	*r*
D'heygere et al. 2004	WEKA	Day, W, D, FV, S, T, pH, DO, Cond, TOX, TOC, OM, Ni, Ph	pr/ab	Flemish river sediment (Belgium)	FF	BP	(6-17)-10-2	324-36 (10-fold)	N/S	N/S	N/S	CCI, *K*, RMSE
Gabriels et al. 2002	MATLAB	S, DM, T, pH, DO, Cond, TOC, OM, Ni, Ph	abu	Flemish river sediment (Belgium)	FF	BP	20-10-92	250-95	Arbitrarily chosen	N/S	IN[-1 1], OUT[0, 1]	*r*, CCI
Gabriels et al. 2004	MATLAB	Day, W, D, FV, S, pH, DO, Cond, Ni, Ph	abu	Flemish river sediment (Belgium)	FF	BP	12- N/S -10	294-49 (7-fold)	N/S	N/S	IN N/S OUT[log (abu+1)], [0, 1]	CCI, *K*, D
Goethals et al. 2002	MATLAB	T, pH, DO, Cond, SS, D, W, S, Sh, VEG, FV, Me, HRB, PR, AE	pr/ab	Zwalm river basin (Belgium)	FF	BP	15-10-52	40-20	Trial and error	STF	N/S	CCI
Hoang et al. 2001	N/S	A, SO, R, SoilT, VEG, S, T, …	pr/ab	Queensland streams (Australia)	FF	BP	39-15-37	650-167	N/S	STF	N/S	CCI
Hoang et al. 2002	N/S	SO, Lo, Ni, …	pr/ab	Queensland streams (Australia)	FF	BP	N/S	N/S	N/S	STF	N/S	CCI
Marshall et al. 2002	N/S	A, SO, R, SoilT, VEG, S, T, …	pr/ab	Queensland streams (Australia)	FF	BP	39-15-37	650-167	N/S	STF	N/S	CCI
Obach et al. 2001	N/S	T, DI, P	abu	Hesse (Germany)	FF FF FF SOM	Mod BP GRNN LNN RBF	N/S N/S -120	N/S	N/S	N/S N/A	[0 1]	*r*², RMSE
Park et al. 2001	N/S	Ex Comm	Ex	Suyong river (Korea)	SOM FF	KLR BP	N/S N/S -5- N/S	N/S	N/S	N/A STF	[0.01 0.99]	*r*
Park et al. 2003a	N/S	EPTC A, SO, DISTs, T	EPTC	Adour-Garonne river basin (France)	SOM FF	KLR BP	N/S -140 4- N/S -1	130-25	N/S	N/A N/S	[0 1]	*r*
Park et al. 2003b	N/S	S, VEG, DO, W, Cond, T, FV, D, NO$_3^-$, PO$_4^{3-}$, NH$_4^+$, …	richn, SH	The Netherlands	Forward only	CPN	N/S	500-164	N/S	N/S	[0 1]	*r*

Schleiter et al. 1999	N/S	T, P, pH, DO, Cond, D, W, S, DI, NO_3^-, NO_2^-, NH_4^+, COD, BOD, Ph, …	abu	Kuhbach, Lahn and Breitenbach (Germany)	FF	BP	N/S	150-150 200-100 225-75	N/S	N/S	[0 1]	MSE, r^2
Schleiter et al. 2001	NNEE	pr/ab, abu	BOD, Cond, NH_3, NO_3^-, NO_2^-, NH_4^+, Ni, pH, Ph, T, DO, SI	Hesse (Germany)	FF FF FF	Mod BP GRNN LNN	N/S	45-6	N/S	N/S	[0 1]	r^2, RMSE, CVE
Wagner et al. 2000	N/S	T, P, DI	abu	Breitenbach (Germany)	FF	BP	N/S	216-54	N/S	N/S	[0 1]	r^2
Walley and Fontama 1998	N/S	Coord, DISTs, SL, Alk, DI, A, S, W, D	ASPT, NFAM	UK	FF	BP	13-6-6-1	307-307 (2-fold)	N/S	N/S	No, log (DISTs), log(SL)	*r*

Applied symbols and abbreviations: N/S=not specified; N/A=not applicable; NNEE=Neural Network Experimental Environment; **Input and output variables** A=altitude; abu=abundance; Alk=alkalinity; AE=artificial embankment structures; ASPT=average score per taxon; BOD=biological oxygen demand; CA=catchment area; COD=chemical oxygen demand; Cond=conductivity; Comm=community data; Coord=X and Y coordinates; D=depth; Day=day; dens=density; DI=discharge; DISTs=distance from river source; DISTm=distance to mouth; div=diversity; DM=dry matter; DO=dissolved oxygen; EPTC=richness of Ephemeroptera, Plecoptera, Trichoptera and Coleoptera; Ex=exergy from the MI communities; FV=flow velocity; HRB=hollow river banks; Lo=longitude; LU=land use; Me=meandering; MI=macroinvertebrates; NFAM=number of families; NH_3=ammonia; NH_4^+=ammonium; Ni=nitrogen; NO_2^-=nitrite; NO_3^-=nitrate; OM=organic matter; P=precipitation; Ph=phosphorus; PO_4^{3-}=phosphate; PR=pool/riffle; pr/ab=presence/absence; richn=species richness; S=substrate; SDA=surface of the drainage area; SH=Shannon diversity index; Sh=shadow; SI=saprobic index; SL=slope; SO=stream order; SoilT=soil type; SS=suspended solids; T=water temperature; TOC=total organic carbon; TOX=toxicity; VEG=vegetation; W=width; **Connection type** ERC=Elman recurrent neural network; FF=feedforward; RTRC=real-time recurrent neural network; SOM=Kohonen self-organizing mapping; **Training algorithm** BP=backpropagation; CPN=counterpropagation network; GRNN=general regression neural network; KLR=Kohonen learning rule; LM=Levenberg-Marquardt; LNN=linear neural network; Mod BP=modified backpropagation; RBF=radial basis function; RL=recurrent learning; **Transfer functions** STF=sigmoid transfer function; **Scaling of variables** IN=input; OUT=output; **Performance measure** CCI=percentage of correctly classified instances; *K*=Cohen's kappa; CVE=cross-validation error; D=percentage of instances with the estimated values within a distance of 1 from the observed values; MSE=mean squared error between observed and estimated values; PI=performance index (proportion of predictions within 10% of the observed value); *r*=correlation coefficient between observed and predicted values; r^2=determination coefficient between observed and predicted values; RMSE=root mean squared error between observed and estimated values

Among the articles that specify the number of samples used for training, the number ranges from 40 to 650. The ratio of the number of training samples versus the number of hidden neurons ranges from 4.5 to 52.5 with an average of 16.8, when all specified combinations are taken into account.

3.4.3.4 Research needs on predictive ANN development studies of aquatic macroinvertebrates

Till now, there is almost no insight in the practical usefulness of ANN models in decision support systems, most articles only discuss the development of the models and evaluate them by means of one or more performance measures. The choice of an evaluation measure however should be driven primarily by the goals of the study. This may possibly lead to the attribution of different weights to the various types of prediction errors (e.g. omission, commission or confusion). Testing the model in a wider range of situations (in space and time) will permit one to define the range of applications for which the model predictions are suitable. In turn, the qualification of the model depends primarily on the goals of the study that define the qualification criteria and on the applicability of the model, rather than on statistics alone (Guisan and Zimmermann, 2000).

The contribution of input variables, is another very important aspect that needs more research. Many variables are not part of the dataset, while others have a high variability, that can be caused by measurement difficulties, but also by the natural changes in the river systems. Therefore, also the effect of monitoring methods needs more research, in particular the incorporation of ‘new’ variables which are less straightforward to be used in a model. This is in particular the case for structural and morphological variables that often need to be visually monitored, but also for heavy metals and other potential toxicants, since their effects are often related to the environment where they are released (bio-availability, accumulation, etc.). These toxicants may be a new challenge in the field of soft computing models to predict river communities, in particular macroinvertebrates.

So far, several rules of thumb for determining model geometry have been proposed. Alternatively, techniques for automatically selecting model architecture are suggested. However, in most of the studies discussing the prediction of macroinvertebrates in aquatic systems, model geometry was decided either arbitrarily or with trial and error. Consequently, there is a need to develop guidelines to clearly identify the circumstances under which particular approaches should be used and how to optimise the parameters that control neural network architecture.

3.4.4 Classification and regression trees

3.4.4.1 Description of model development method

One well-studied data soft-computing method, i.e. the induction of classification and regression trees (often referred to as decision trees when discussing both methods) has been shown to be useful in modeling complex datasets (Breiman et al., 1984). The common way to induce rules in the form of decision trees is the so-called 'Top-Down Induction of Decision Trees' (Quinlan, 1986). Tree construction proceeds recursively, starting with the entire set of training examples. At each step, the most informative input variable is selected as the root of the sub-tree and the current training set is split into subsets, according to the values of the selected input variable. In this manner, rules are generated that relate the values of input variables with the presence/absence of macroinvertebrate taxa. For discrete input variables (classification trees), a branch of a tree is typically created for each possible value of that particular variable. For continuous input variables (regression trees), a threshold is selected and two branches are created based on that threshold. Tree construction stops when all examples in a node are of the same class (or if some other stopping criterion is satisfied). Such nodes are called leaves. The C4.5 algorithm (Quinlan, 1993) is one of the most well-known and widely used classification tree induction methods. The M5 program (Witten and Frank, 2000) is on the other hand often applied for regression tree induction.

In order to reduce the noise in the data and to improve the predictive results with regard to complexity and accuracy of the predictions, several optimisation methods can be applied. Major examples are pruning, bagging and boosting. These methods will be briefly explained as they were already applied in model development studies for macroinvertebrates.

An important aspect in decision tree learning is the amount of branches. When there are many branches, the decision trees are difficult to interpret, and often these last branches do not contribute significantly to the reliability of the tree. Splitting the data has the effect that for decisions deeper in the tree, ever fewer examples are available. Therefore, a pruning method is needed to avoid that too detailed trees are trained. There are two types of tree-pruning:

forward pruning and post-pruning. When forward pruning is applied, the expansion of the tree is stopped when a certain criterion is met. For example, every leave should contain a minimum number of instances or no branching is allowed. Post-pruning on the other hand, means that first a highly branched tree is constructed. Afterwards, some of the ending subtrees are replaced by leaves based on their reliability. The reliability of the subtrees is evaluated by comparing the classification error estimates before and after replacing a subtree by a leave. The confidence factor, which is often used for this purpose, is a parameter that has an effect on the error rate estimate in each node. When the confidence factor is increased, the difference between the error estimate of a parent node and its splits decreases. In this way, it is less likely that the split will be pruned. The smaller the value of the confidence factor is, the larger is the difference between the error rate estimates of a parent node and its potential splits. Thus, the chance that splits will be replaced by leaves is higher. Optimal pruning is an important mechanism as it improves the transparency of the induced trees by reducing their size, as well as enhances their classification accuracy by eliminating errors that are present due to noise in the data.

Bagging (Bootstrap aggregation) and AdaBoost (Adaptive Boosting) (Witten and Frank, 2000) are voting classification algorithms. Bagging and boosting are used in combination with the base classifier that creates ‘child’ datasets from a single ‘parent’ dataset that is originally used for training. This allows taking advantage of the inherent instability of the base classifier. The instability of a classifier is defined as the tendency to find large changes in the predicted values caused by minor changes in the dataset. In bagging, the ‘child’ datasets are created by duplicating some of the instances of the ‘parent’ dataset randomly and deleting others. From each ‘child’ dataset, a different tree is constructed that leads to a different prediction. The different predictions of the ‘child’ datasets are combined by a majority vote to give the final prediction. Boosting also creates ‘child’ datasets from a single ‘parent’ dataset, but the difference is that each new ‘child’ dataset is influenced by the previous one, as the instances that are duplicated are not selected randomly. The instances that are incorrectly predicted in a dataset are included in the next dataset as duplicated ones, so that the chance of a correct prediction of these previously misclassified instances improves. These duplicated instances will affect the training of the model and therefore also the resulting classification tree. This procedure continues until a pre-defined number of iterations is reached, but it stops earlier in case the error estimate is less than a certain threshold value.

3.4.4.2 Predictive classification and regression tree development studies of aquatic macroinvertebrates

In contrast to ANNs, the application of classification and regression trees in ecological modelling, in particular related to macroinvertebrates, is rather limited and hardly described in literature. In the following paragraphs an overview of the major examples is presented.

Kompare et al. (1994) described some general possibilities of machine learning in the field of ecology. Dzeroski et al. (1997) were among the first to describe applications of classification trees in river community analysis. These include the biological classification of British rivers based on bioindicator data, the analysis of the influence of physical and chemical parameters on selected bioindicator organisms in Slovenian rivers and the biological classification of Slovenian rivers based on physical and chemical parameters as well as bioindicator data. In all three cases, valuable models (knowledge) in the form of rules were extracted from data acquired through environmental monitoring and/or expert interpretation of the acquired samples. The applied algorithm was CN2 (Clark and Niblett, 1989).

Blockeel et al. (1999a) applied TILDE to predict an ecological index (Saprobic Index) for Slovenian rivers. The used input variables were biological data, physical-chemical characteristics (actual and time-series) as well as combinations. Additionally, also macroinvertebrate communities were successfully predicted on the basis of physical-chemical variables. In Blockeel et al. (1999b), physical-chemical variables were predicted on the basis of biological communities. Innovative in this article is the use of a single tree to predict all these variables at once, what eases the use of this relative simple information in river management.

In Dzeroski et al. (2000), the prediction of physical-chemical variables was established on the basis of biological data. The research revealed that certain taxa occurred in many trees, what makes them useful to be selected as indicator taxa. The research proved as well that when compared to linear regression, the model seemed to give the same performance.

Dzeroski and Drumm (2003) applied regression trees (M5’) program (a Java implementation of the M5 algorithm in WEKA (Witten and Frank, 2000)), to predict sea cucumbers (*Holothuria leucospilota*) in lagoons around the Cook Islands. Based on these trees they were able to retrieve the preferred habitat of this species and found out that the dominant variables are rubble and sand.

Recently, a study was elaborated by Dakou et al. (2004a) on the use of rule induction techniques for prediction of macroinvertebrate taxa in the river Axios (Northern Greece). In this study, decision tree models were induced to predict the habitat suitability of six macroinvertebrate taxa. Rules relating the presence/absence of benthic macroinvertebrate taxa with the 15 physical-chemical and structural river characteristics and the seasonal variable were induced using the J48 algorithm (a Java implementation of the C4.5 algorithm) in WEKA (Witten and Frank, 2000). In order to improve the performance and the interpretability of the induced models, three optimisation techniques were applied: tree-pruning, bagging and boosting. The predictive performance of the decision tree models was assessed on the basis of the percentage of Correctly Classified Instances (CCI) and the Cohen's kappa statistic (*K*). The results of the present study demonstrated that although the models had a relatively high predictive performance, noise in the dataset and inappropriate input variables prevented to some extend, the models from making reliable predictions. Although tree pruning did not improve significantly the reliability of the induced models, it reduced considerably the tree complexity and in this way increased the transparency of the trees. Consequently, the induced models allowed for a correct ecological interpretation. Bagging and boosting were capable of improving the predictive performance of ecological models, but the chance exists to overfit the data. The latter study illustrated however, that quite some knowledge is missing on how to develop classification tree models and how to optimize the predictive performance and transparency of the induced models.

Regarding practical applications of classification trees in water management, the set of studies related to macroinvertebrates is very limited. Practical studies were established by D'heygere et al. (2002) and Goethals et al. (2002). Both studies can only be called 'preliminary', because of the small datasets that were available for the studies.

D'heygere et al. (2002) researched the use of classification trees to set up a monitoring network in the Dender river (Flanders, Belgium) for the implementation of the WFD. In particular the effect of seasonality was analyzed. In this manner, the trees could help to reduce the sampling costs, seen not for all stream types, a multi-seasonal sampling seemed to be necessary due to the very poor ecology present in the Dender as a result of pollution and other types of impacts.

The study of Goethals et al. (2002) aimed at analyzing the ecological niches of macroinvertebrates in the Zwalm river basin (Flanders, Belgium) and checking the convenience of these models to make predictions on river restoration projects. Classification trees were constructed for all 52 taxa collected during the 60 samplings in the headwaters of the Zwalm river basin. The poor performance of most induced trees had probably its origin in the small size of the dataset. The reliability of the predictions differed dramatically between the macroinvertebrate taxa and the frequency of occurrence of the taxa in the different sites was likely to be one of the major explanations of this phenomenon (Table 3.8). Especially when the taxa were very common or extremely rare, the number of correctly classified instances was very high during the validation process, but this can mainly be explained by the high reliability to make a good prediction even without making use of information from the data. Therefore, this study illustrated the need for other evaluation indices, such as *K* (Cohen's kappa) to cope with this type of problems. The J48 algorithm did not induce a meaningful tree in these cases, as can be seen in Table 3.8 for *Aplexa* and Tubificidae. The trees did not use any variables characterizing the river, but simply state that '*Aplexa* is always absent' and 'Tubificidae are always present'. In other words, there is no ecological information extracted from the data. The J48 algorithm is mainly interesting for moderately frequent taxa, such as *Asellus* and *Gammarus* (Figure 3.5). Based on tenfold cross validation, the CCI score was 63 % for predicting *Gammarus*.

Table 3.8 Prediction of three different macroinvertebrate taxa by means of classification trees (CCI calculation based on tenfold cross validation, the database consisted of 48 instances).

Taxa	**Frequency of occurrence in the Zwalm (%)**	**CCI (%)**	**Number of variables in the model**	**Number of leafs (model complexity)**
Aplexa	2	100	0	1
Asellus	43	63	2	3
Tubificidae	93	94	0	1

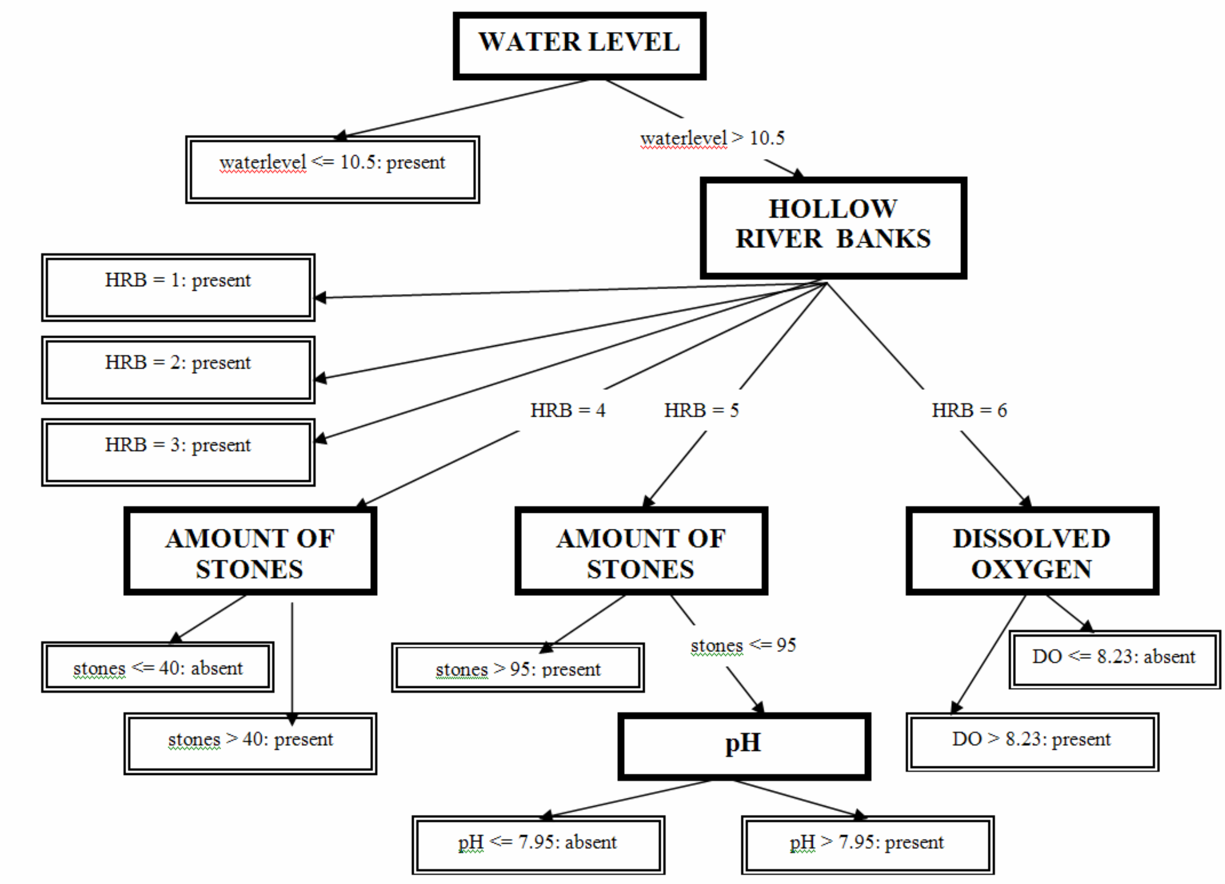


Figure 3.5 Example of a classification tree model of Gammarus in the Zwalm river basin. The single bold frames (nodes) contain the classification variables, while the double frames (leaves) contain the final prediction of the Gammarus presence or absence. 'Hollow river banks' (HRB) is a categorical variable (six classes: 1= very good hollow banks under trees; 2 = good hollow banks; 3 = hollow banks by erosion under vegetation; 4 = moderate cavities; 5 = hollow banks not probable; 6 = no hollow banks as a result of artificial embankments).

The tree also revealed interesting information concerning the variables that are important to predict this taxon. The main variables for the prediction of *Gammarus* (according to this study) are water level, amount of hollow river banks, amount of stones, dissolved oxygen and pH. From the values in the leafs of the tree one can conclude that the *Gammarus* mainly prefer the upstream parts of the river basin. The taxon is present in undeep waters (water level lower than 10.5 cm). It also prefers hollow banks and cavities, which nearly only occur in fast running waters, thus also the higher and steeper upstream parts. *Gammarus* also prefer more stony material, which means quickly running streams where sediments do not settle. In cases of deeper waters without cavities (due to artificial embankments) the *Gammarus* are only present when the dissolved oxygen percentage is sufficiently high. The pH value plays a role under specific circumstances, but this classification is of the lowest importance in the tree.

Also the effect on the *Asellus* population of the removal of the 6 weirs in the Zwalm river basin was simulated by classification trees in this study by Goethals et al. (2002). According to the classification tree model *Asellus* only colonizes the broader river sites (the only rule generated by the J48-model using tenfold cross validation on all sixty instances was:'width more than 3.5 meters: *Asellus* present, while absent in the more narrow streams'). Since change in width itself is not enough to reflect the change on the taxon (but also depth, flow velocity) the model simulations were not concordant with what ecological experts expected. Therefore the conclusion was that in that stage, even the models that seemed to be reliable based on a performance indicator such as the CCI, were not yet useful for practical predictions in water management. This study therefore illustrated that at least three types of model validation are necessary to make sure that this type of models can be used in water management: theoretical validation based on well chosen performance indicators (thus also taking care of the prevalence of the taxa), comparison with existing ecological knowledge and practical simulation exercises. On top of this, also the willingness of river managers to use these models has to be checked to solve problems that are convenient for this type of tools.

3.4.4.3 Research needs on predictive classification and regression tree development studies of aquatic macroinvertebrates

As mentioned above, so far, no crisp guidelines to support the selection of learning settings do exist. This makes the use of this type of methods less attractive, since optimal settings need to be found via trial and error D'heygere et al. (2001). Often, studies are based on general standard settings of the software, based on experiences in a variety of applications. Also no study confirms or rejects these settings. This means that there is a high need for integrated tests on several datasets (and for exercises on different species), to get insight in basic rules that are useful for the construction (and in particular the parameter settings of the learning methods).

On top of this, there are no in-depth studies that focus on the practical applicability of these methods for decision support in water management. The usefulness of the methods can only be checked and proven by such applications and will be crucial to convince river managers of the convenience of these methods to support decision making and information collection from datasets.

3.5 Comparison of different soft computing methods

Actually there is a lack of comparative papers (Manel et al., 1999) in which more than two statistical methods are applied to the same data set. Most published ecological modelling studies use only one of the many techniques that may properly be used, and little information is available on the respective predictive capacity of each approach. The debate is usually restricted to the intrinsic suitability of a particular method for a given data set. When starting a static modelling study the choice of an appropriate method would be much facilitated by having access to publications of comparative papers that show the advantages and disadvantages of using different methods in a particular context (Guisan and Zimmermann, 2000).

Therefore, based on the rather limited set of case studies in which several methods were compared, it is until now nearly impossible to have clear insight in when to use what method. On top of this, also several methods such as BBNs and fuzzy logic were rarely applied yet, which means that these methods need further elaboration. Important to note is that many different methods are applied to evaluate model performance, which makes the comparison between studies very difficult. It is additionally worth mentioning that the practical applicability of the models for decision support (ecological knowledge extraction, predictions) is also a crucial quality aspect, which has so far been evaluated in none of the studies.

3.6 Discussion

When looking at the different soft computing techniques, it seems that they all have particular strengths and weaknesses. ANNs can give for instance well performing models, but generate black box systems and the integration of expert knowledge is difficult. Fuzzy logic can be used to develop models merely on expert knowledge, but the amount of input variables is very limited, because the rule sets become very complex when more than five input variables are used. BBNs have the interesting characteristics to be able to make networks in which can be seen how different variables affect each other. On the other hand, the needed information to set up these networks and the distribution is also huge.

Presently, there is also no insight in the practical usefulness of such models in decision support systems. Most articles only discuss the development of the models and evaluate them on the basis of one or more performance indices. The choice of an evaluation measure should be driven primarily by the goals of the study. This may possibly lead to the attribution of different weights to the various types of prediction errors (e.g. omission, commission or confusion). Testing the model in a wider range of situations (in space and time) will permit one to define the range of applications for which the model predictions are suitable. In turn, the qualification of the model depends primarily on the goals of the study that defines the qualification criteria and on the usfulness of the model, rather than on statistics alone. (Guisan and Zimmermann, 2000)

The contribution of input variables, is another very important aspect that needs more research. Many variables are not making part of the dataset, some have a high variability, that can be caused by measurement difficulties, but also by the natural changes in the river systems. Therefore, also the effect of monitoring methods needs more research, in particular the incorporation of 'new' variables which are less straightforward to make predictions on. This is in particular the case for structural and morphological variables that often need to be visually monitored, but also for heavy metals and other potential toxicants, seen their effects are often related to the environment where they are released (bio-availability, accumulation, etc.). Also the latter aspects can be very innovative research goals in the field of soft computing models to predict river communities. For this purpose, the methods described in the recent paper of Gevrey et al. (2003) published in Ecological Modelling, that conducted a comprehensive comparison of eight different methodologies that have been widely used in the ecological literature, will also be a starting point of this research. In particular some of these methods are useful to get insight in the applied variables during the ANN calculations and can in this manner reveal the ecological and practical relevance of this so-called black box technique.

3.7 Conclusions

Although there is quite some experience gained with data driven models to predict macroinvertebrates, several key-questions remain with regard to the optimal architecture, the input variables, the ecological relevance and practical use of the models for river management. Therefore, this thesis is comparing different methods to develop predictive models for macrobenthos communities. In particular the ecological relevance of the models will be focussed on as well as their performance in practical river management applications.

Given the many questions that remain on the effect of input variables on the model performance and the need to get insight in the habitat preferences of macroinvertebrates, different input variable contribution methods (in combination with ANN) and pruning settings (in combination with classification trees) will be tested as well.

Chapter 4
Data and information collection to develop and validate models predicting macroinvertebrates in rivers

Redrafted after:

Goethals, P. and De Pauw, N. (2001). Development of a concept for integrated river assessment in Flanders, Belgium. Journal of Limnology 60: 7-16.

4.1 Data collection on river sediments in Flanders (1996-1998)

4.1.1 Introduction

Contaminated river sediments can have direct adverse impacts on bottom fauna. Contaminated sediments can also be a long-term source of toxic substances to the environment and can impact wildlife and humans through the consumption of food or water or through direct contact. These impacts may be present even though the overlying water is meeting the required quality criteria. This chapter describes the data collection of river sediments in Flanders during the period 1996-1998. These samples were taken with a Van Veen grab sampler within a stretch of 50 m in 360 sites of unnavigable river stretches in Flanders. These sites were sampled within the framework of the TRIAD assessment methodology for sediments of Flemish watercourses (de Deckere et al., 2000; De Pauw and Heylen, 2001). The database consists of physical-chemical, ecotoxicological and biological results.

4.1.2 The TRIAD approach for the assessment of river sediments in Flanders

In recent years, contaminated sediments have become an important environmental issue. Contamination of river sediments is often identified as high risk to the environment. Management requires a specific approach concerning sampling and analysis of the sediments. For this reason, the TRIAD approach is being applied in Flanders (De Cooman et al., 1999) on the basis of the principles described by Chapman (1992) and Van de Guchte (1992). In the TRIAD approach three categories of data are linked: observations demonstrating effects occurring in the field (biological data), results of bioassays linking field effects to sediment toxicity (ecotoxicological data) and concentrations of contaminants in the sediment (physical-chemical data) (den Besten et al., 1995).

4.1.3 Monitoring strategy and sites

From 1996 till 1998, 360 sites were sampled within the framework of the TRIAD assessment methodology for sediments of Flemish unnavigable watercourses (De Cooman et al., 1999) (Figure 4.1).

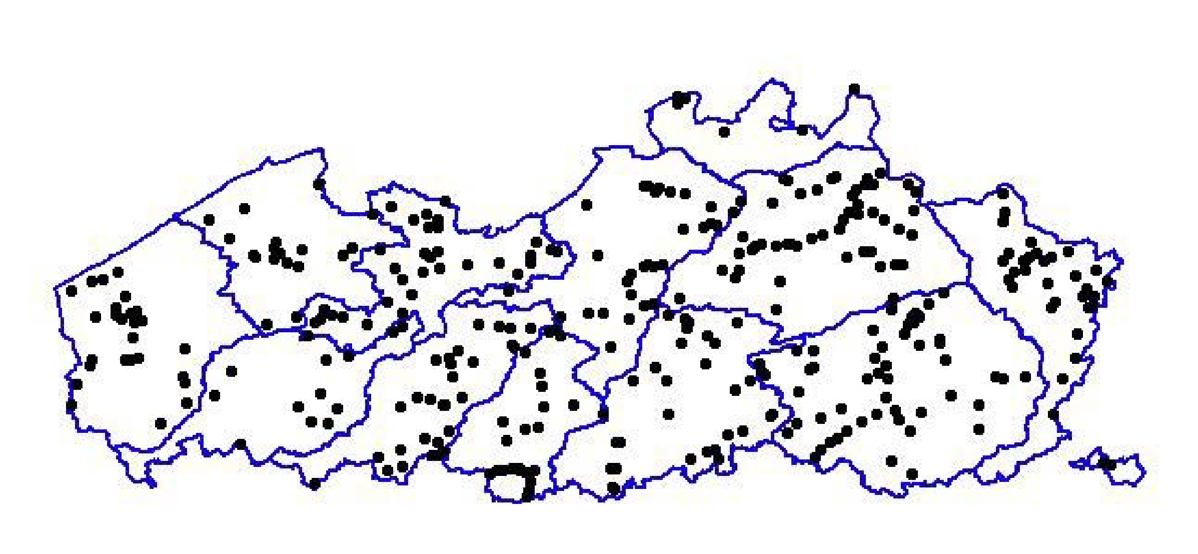


Figure 4.1 Sampling sites (360) in Flanders (Belgium) analysed during the period 1996-1998.

The selection of the study sites was based on three principles (De Cooman et al., 1996):

- Scientific principle (stratified site selection): the characterisation of the sediments is based on a ratio to reference principle, which implicates that several potential reference sites had to be selected. With this information, data on natural, unpolluted situation of different kinds of sediments as well as heavily polluted ones were collected to obtain a pollution gradient over the different sampling sites. This is also of major importance to guarantee that induced data driven models can be applied to make proper classifications and predictions for various river conditions:

- Ecological principle: in order to guarantee the standstill principle (no further degradation), data on the current quality of (rare) ecological valuable sites was collected;

- Pragmatic principle: sites were selected with the intention of an eventual remediation and/or protection. These intended actions however only make sense in rivers where all major sources of pollution are already disconnected. Therefore, major part of the sites had to fulfil to this requirement.

4.1.4 Sampling methods

Sediment samples were taken by means of a Van Veen grab sampler (2 l volume, see Figure 4.2), zigzagging across the watercourse over a length of 50 m. Between 25 and 40 sub-sample grabs (up to a total volume of approximately 40 l) were collected and mixed together homogeneously (Figure 4.2). From this mixture about 13 l were kept apart for studying the macroinvertebrate community (De Pauw and Heylen, 2001). The samples were rinsed over a sieve (mesh-size 0.5 mm) and all macroinvertebrates were sorted out carefully.

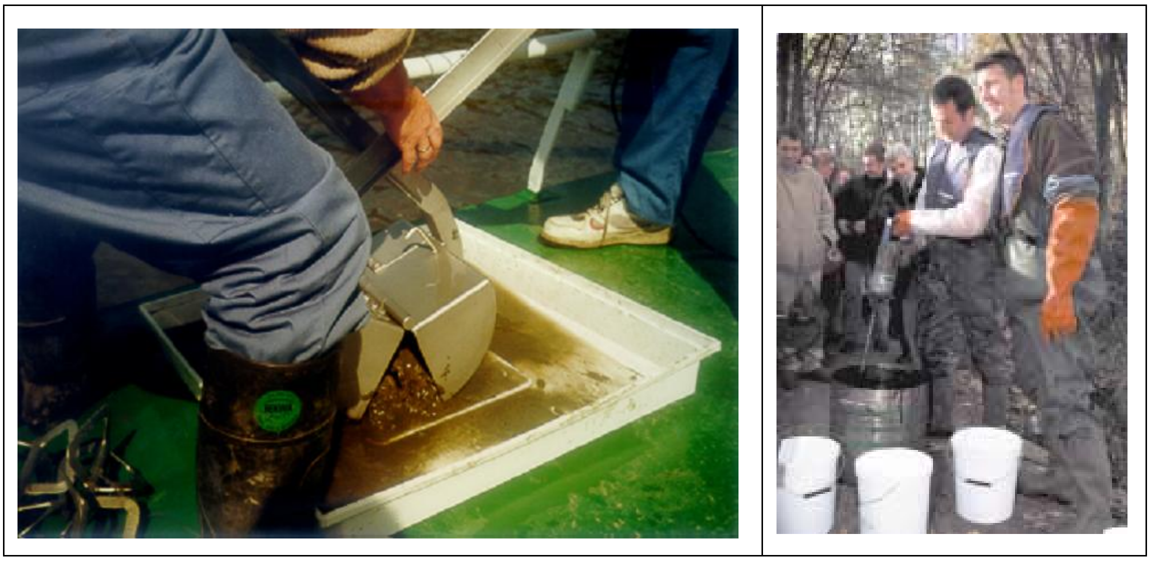


Figure 4.2 Van Veen grab sampler (left) and mixing to obtain homogeneous subsamples for the different analyses (right).

4.1.5 Abiotic and ecotoxicological river sediment characteristics

In total, fifteen physical-chemical sediment characteristics were measured and used in the analyses of this thesis: temperature, pH, dissolved oxygen concentration (mg/l) and conductivity (µS/cm), all four measured in the water column combined with organic matter (kg OM/kg DM), Kjeldahl nitrogen (mg Kj-N/kg DM) and total phosphorus concentrations (mg P/kg DM) in the sediment, in combination with the metal concentrations of Cr, Pb, As, Cd, Cu, Hg, Ni and Zn. The ecotoxicological evaluation was based on two acute tests on pore water of the sediments: a 24h growth-inhibition-test with the alga *Raphidocelis subcapitata* (TOXR) and a 72h growth-inhibition-test with the crustacean *Tamnocephalus* *platyurus* (TOXT). Also 6 structural variables were incorporated in the selection process, namely width, depth and flow velocity of the river together with the percentage of clay (0 – 2 µm), loam (2 – 50 µm) and sand (50 – 2000 µm). The ranges of all the input variables were continuous, except for the two ecotoxicological variables for which false (0) or true (1) was used and the flow velocity that was divided into 6 classes (0, 1, 2, 3, 4, 5) from stagnant (0) up to very fast (5). Seasonality was included by means of day number.

4.1.6 Macroinvertebrate community analysis

In total 92 different taxa were found. The identification (Table 4.1) was performed as described in De Pauw and Heylen (2001).

Table 4.1 Taxa identification level for the river sediments in Flanders dataset collected during the period 1996-1998. (De Pauw and Heylen, 2001)

Taxa	**Identification level**
Plathelminthes	genus
Oligochaeta	presence
Hirudinea	genus
Mollusca	genus
Crustacea	family
Plecoptera	genus
Ephemeroptera	genus
Trichoptera	family
Odonata	genus
Megaloptera	genus
Hemiptera	genus
Coleoptera	family
Diptera	family (exception: fam. Chironomidae: differentiate beween group thummi-plumosis / group non thummi-plumosis
Hydracarina	presence

4.1.7 Database setup

The database consisted of 360 instances about 24 environmental variables (day number included). The macroinvertebrate abundances were available as such, but also transformations were made to presence/absence variables and log(abundance + 1) to permit a broader range of analyses (classification as well as regression models) and model validations on the basis of different performance indicators.

4.2 Data collection in the Zwalm river basin (2000-2002)

4.2.1 General characteristics of the Zwalm river basin

The Zwalm river basin is part of the Scheldt river basin (Carchon and De Pauw, 1997). The Zwalm river basin has a total surface of 11.650 ha and the Zwalm river has a length of 22 km (Figure 4.4). The average water flow (at Nederzwalm, very near the Scheldt) is about one m^3s^{-1}. It has a very irregular regime, with low values in the summer (minima lower than 0.3 m^3s^{-1}) and relatively high values in rainy periods (maximums up to 4.7 m^3s^{-1}) (Lauryssen et al., 1994). The water quality in the Zwalm river basin improved a lot during the year 1999, due to investments in sewer systems and wastewater treatment plants during the last years (VMM, 2000). Nevertheless, most parts of the river are still polluted by untreated urban wastewater discharges and by diffuse pollution originating from agricultural activities. Although Flanders is in general rather flat, the Zwalm river basin is characterized by several height differences, resulting in a very unique river ecosystem within Flanders. However, due to the agricultural activities on several slopes, soil erosion is the most important geo-morphological process resulting in an import transport of (contaminated) sediments in the river. In addition to this, numerous structural and morphological disturbances such as weirs for water quantity control and artificial embankments affect the river ecology. A description of the major sources of stress to the ecology due to human activities in the Zwalm river basin is given in Table 4.2 (adapted from Goethals and De Pauw, 2001).

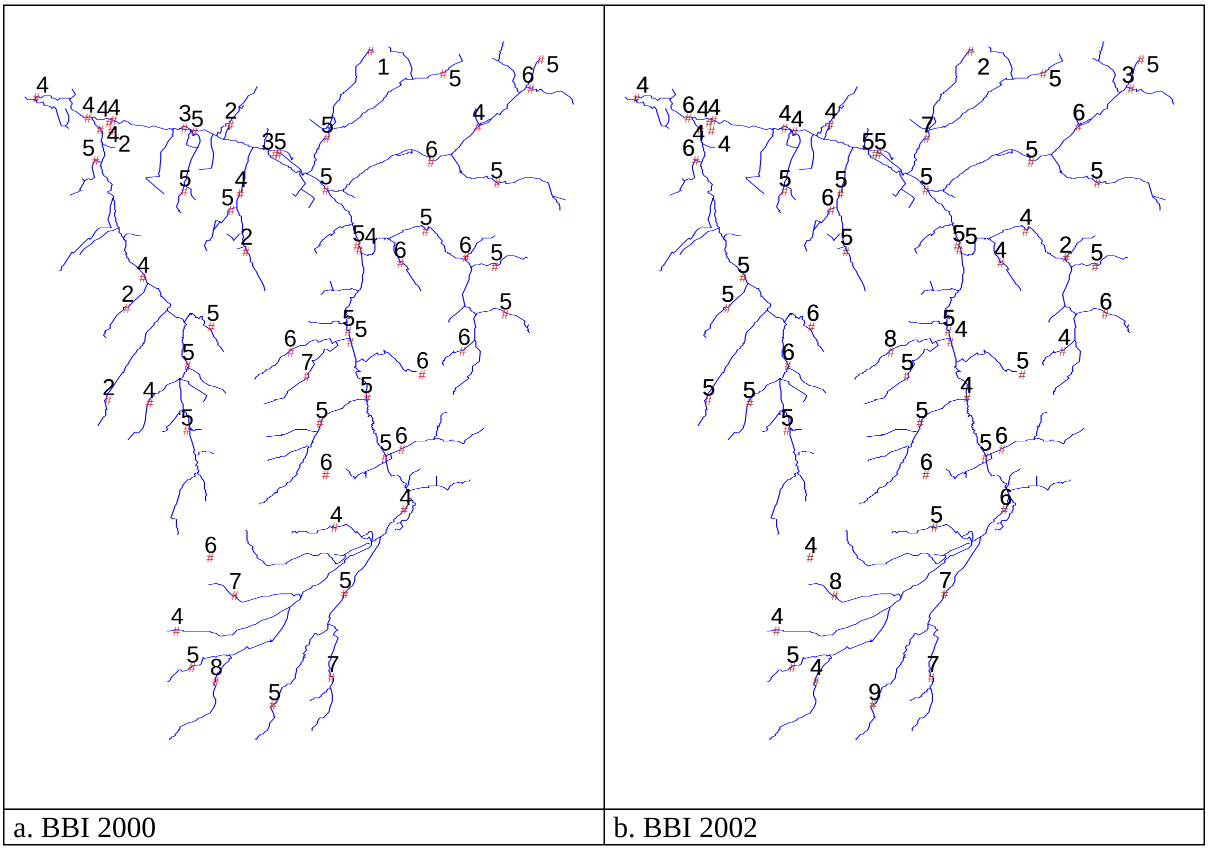
In Figure 4.3 the water quality in the Zwalm river basin during the years 2000 and 2002 is illustrated on the basis of the Belgian Biotic Index. The BBI method uses macro-invertebrates as indicators for the level of pollution (De Pauw and Vannevel, 1993). The methodology is based on the theorem that increasing pollution will result in a loss of diversity and a progressive elimination of certain pollution-sensitive groups. The BBI-system is interpreted as follows: 1 - 2 = very heavily polluted or red, 3 – 4 = heavily polluted or orange, 5 – 6 = moderately polluted or yellow, 7 – 8 = slightly polluted or green, 9 – 10 = unpolluted or blue. The map illustrates that most sites (except some very good ones in the southern part of the basin) do not meet a good ecological status, despite all restoration activities. In particular structural modifications and diffuse pollution are probably responsible for this.

Table 4.2 Description of the main factors responsible for the spatial and temporal diversity in the ecosystem in the Zwalm river basin (Goethals and De Pauw, 2001).

Physical-chemical disturbances	**Structural and morphological features**	**Direct biological 'disturbances'**
Point sources: - Effluent WWTP (urban and industrial) - Combined sewer overflows Sewer systems - Accidents (fuel storage tanks) - Feeding of animals, fishing Diffuse sources: - Agriculture - Traffic - Scattered housings	- Water quantity management (weirs, artificial embankment) - Transport infrastructure - Physical pollution (wood debris, large wastes)	- Fishing - Rat traps - Sampling related to monitoring - Fish stock modification (angling management, pond overflows) - Game (e.g. birds: wild ducks)

4.2.2 Monitoring strategy and sampling sites

In spite of the fact that a lot of data have been gathered in Flanders on the numerous river systems, still many gaps have to be filled before these data will meet the requirements of our modelling objectives. First of all, the data are scattered over different institutes in Flanders using various format types, other co-ordinate systems, etc. Although a database of the Zwalm river basin was developed to be implemented in a GIS as an exercise in which data of several institutes were integrated, it was never put into practice (Carchon and De Pauw, 1997). This methodology was based on the use of charts in which coloured lines and symbols give information on the ecological quality, types of disturbances, their causes and also on potential solutions. The river quality lines represented the most recent data on the physical and chemical, biological and structural and morphological status of the river. A major problem with these data was however that they were too scarce and not enough variables were measured at each site. Similar problems were encountered with more recent data of the Flemish Environment Agency (VMM).



a. BBI 2000

b. BBI 2002

Figure 4.3 Illustration of the water quality in the Zwalm river basin in the 60 monitored sites of this study in the years 2000 and 2002.

Therefore, methods to improve sampling for the development of habitat suitability models of macroinvertebrates were searched for. In this context, Hirzel and Guisan (2002) proposed the following considerations to improve the sampling strategy for habitat suitability modelling:

- increase sample size;
- prefer systematic to random sampling;
- include environmental information in the design of the sampling strategy.

Additionally, to be efficient, a sampling strategy needs to be based on those gradients that are believed to exercise major control over the distribution of species. These gradients should be considered primarily to sampling, because otherwise vital information will limit model accuracy, in particular when data driven model development methods are used, such as artificial neural networks and classification trees. Random sampling could lead to truncated response curves for some species if the extremities of the main environmental gradients are under-sampled. Stratifying along these gradients and sampling the extremities can assure an efficient sampling of these outer limits (Hirzel and Guisan, 2002). This is why it is important not only to sample sites which are degraded to identify point and diffuse pollution sources based on physical-chemical characteristics. Also the more pristine sites in the upper reaches need to be sampled. These sites will reveal what is feasible from an ecological point of view. Therefore, prior to model development, for each variable the variability was visually analysed, to get insight in the maximum and minimum values, the average as well as the type of distribution.

The above considerations were taken into account during the development of a new monitoring network. Based on the experience with the river sediments in Flanders database, it was clear that the amount of impacts of the selected river system needed to be reduced, because otherwise too many processes had to be taken into consideration and the required dataset would have to be very large to develop reliable models handling all the natural as well as man induced variability. Nearly simultaneously, two river systems were analysed to set up a monitoring network.

The Dender river basin was selected as one of both, because a lot of studies were done on this river during the mid nineties. Based on a set of preliminary sampling campaigns on the Dender river basin during the period 1998-2000 (e.g. D’heygere et al, 2002), it seemed necessary to be very well familiar with the river system, to know what kind of variables needed to be measured, which local (small scale) aspects could have a dramatic effect on the ecology, etc. When too many factors interfere, as was the case in the Dender river (e.g. Vandenberghe et al., 2004), often very dynamic patterns were encountered and the allocation of the effects on the ecology would become very difficult. Therefore, it was concluded that the selected river system needed to be characterized by a limited complexity to develop successful models, but on the other hand needed to contain enough variability (combination near natural sites as well an interesting mixture of impacts) to make interesting models for river managers (to predict reference conditions, to simulate the effect or restorations actions, etc.). The Dender river did not meet these requirements and from 2000 onwards, the focus was put on the Zwalm river basin as further described.

So the second river basin considered was the Zwalm. The study of Carchon and De Pauw (1997) in combination with a lot of practical field experience in the Zwalm river basin gathered during training courses delivered valuable insights in the river system. Also the interest of local river managers to set up restoration plans (e.g. Flemish Environment Agency was setting up investment plans for wastewater treatment, Aminal Division Water was doing studies to improve the flood control and link it with nature conservation, …) was a good argument to consider this system for further research. During the years 1998 and 1999 several preliminary monitoring exercises were done to get insight in the major impact sources and natural variability, while also existing monitoring sites of the Flemish Environment Agency (VMM) were analysed, to check the feasibility to develop habitat suitability models. A major issue was to increase the amount of monitoring sites compared to the VMM, but on the other hand also to take care of time constraints (e.g. leaving time for data analysis, model development, calculating simulations and discussing results with river managers). Also budget limitations played a key role in the selection of variables and measurement frequency, e.g. no metal and organic micro-pollutants could be part of the analysis (but because of the limited industrial activities, these compounds probably played a minor role in the river ecology and could be excluded from the data collection).

In total, 60 sites were selected in the Zwalm river basin in which samples for physical, chemical and biological analysis were taken (Figure 4.4). Observations regarding the structural characteristics were also made. These sites were examined each summer over a three year period (2000-2002). In this way, 180 sets of observations were available. At one site, however, the artificial substrates got lost which means that no biological data were available for that year. The database consisted thus only of 179 instances.

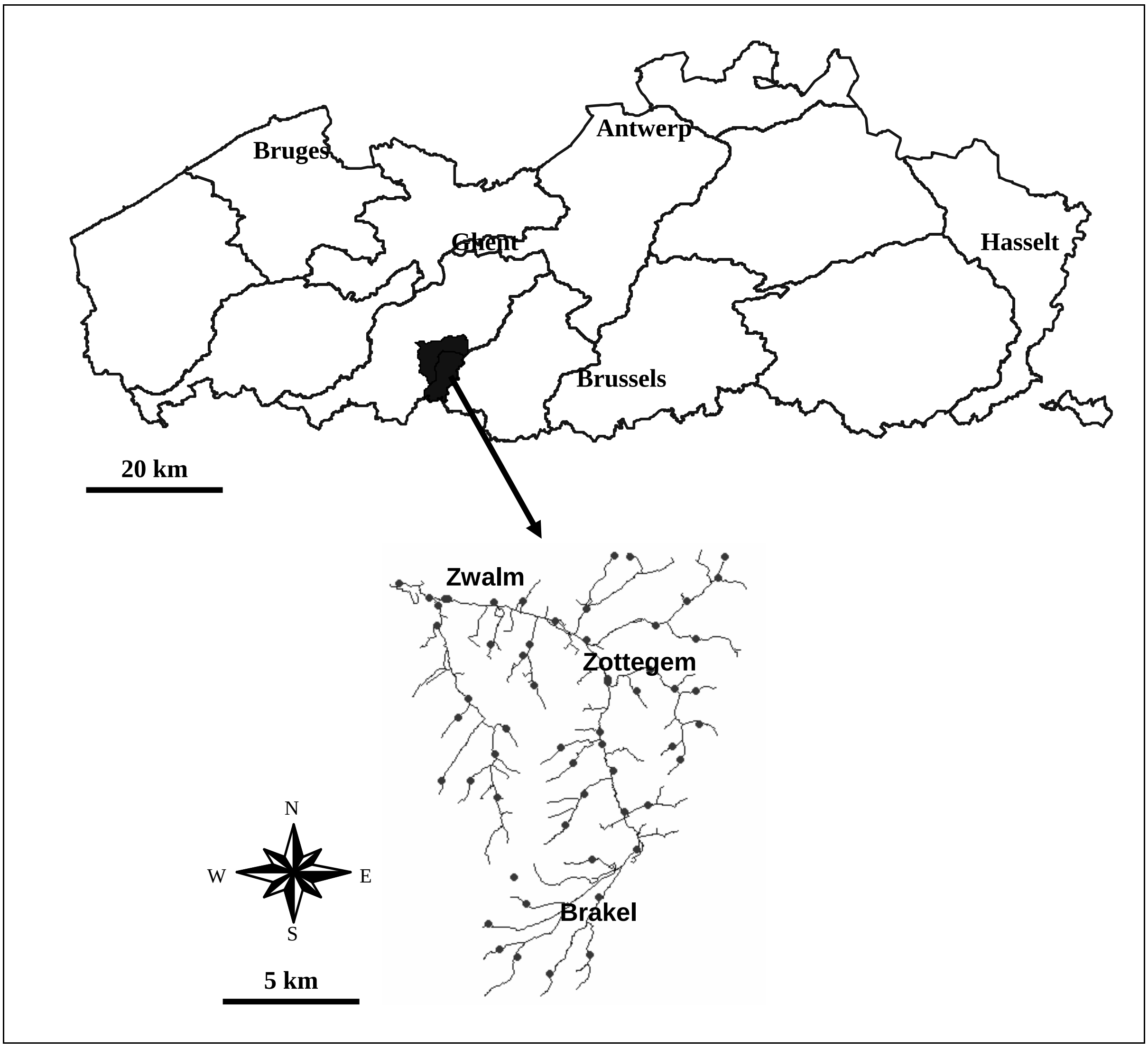


Figure 4.4 Location of the Zwalm catchment in Flanders, Belgium, and the selected sampling sites in this catchment.

4.2.3 Abiotic river characteristics

At each site, 24 environmental variables were recorded. Field measurements were made for temperature (°C) and dissolved oxygen (mg/l), pH and conductivity (µS/cm). Suspended solids (mg/l), nitrate (mg NO_3^-/l), phosphate (mg PO_4^{3-}/l), ammonium (mg NH_4^+/l), COD (mg O_2/l), total phosphorus (mg P/l) and total nitrogen (mg N/l) were measured spectrophotometrically in the laboratory. Flow velocity (m/s) was measured by means of a propeller device. Water level (cm) and width (cm) were determined with a measuring tape. Meandering, hollow river banks, deep/shallow variation and artificial embankment structures (banks) were monitored visually (Dedecker et al., 2002). To illustrate the meaning of the different classes of these variables, a description in combination with some pictures clarifying their meaning is presented in Tables 4.3, 4.4, 4.5 and 4.6. The fractions of boulders, pebbles, sand, loam and clay (%) were determinded granulometrically in the laboratory. Distance to mouth (m) was calculated using ArcView GIS 3.2a. A topographic map was used to determine the stream order (scale 1/25000).

4.2.4 Macroinvertebrate community monitoring and analysis

The macroinvertebrates were collected by means of a standard handnet during five minute kick sampling within a river stretch of 10 m (IBN, 1984) and by *in situ* exposure of artificial substrates (De Pauw et al., 1994) (Figure 4.5). The objective of the sampling was to collect the most representative diversity of the macroinvertebrates at the examined site (De Pauw and Vanhooren, 1983).

In contrast to previous study (river sediments in Flanders), the biological data collection was done in a less quantitative manner, but on the other hand more taxa could be collected, because the hand net sampling methodology allows to collect taxa in more habitats, what is in particular important to be able to explain the effects of modifications of the physical habitat.

4.2.5 Database setup

The database consisted of 179 instances about 24 environmental variables. The macroinvertebrate abundances were available as such, but also transformations were made to presence/absence variables and log(abundance + 1) to permit a broader range of analyses and model validations on the basis of different performance indicators.

Table 4.3 Meandering pattern.

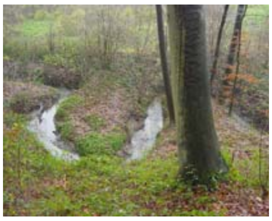
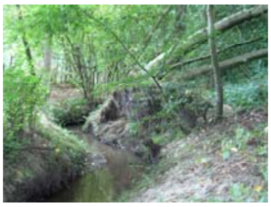
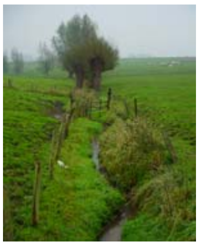
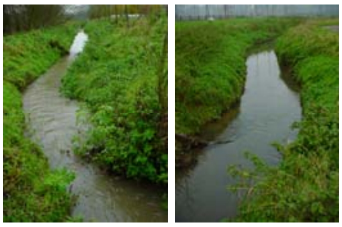
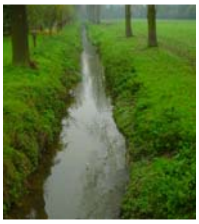
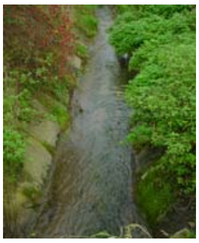
	Meandering pattern is (nearly) pristine: sinuously meandering pattern, continuous presence of big curves. Class 1
	Meandering pattern is well developed: presence of big curves, not continuous. Class 2
	Meandering pattern is moderately developed: slightly meandering pattern, continuously. Class 3
	Meandering pattern is poorly developed: slightly meandering pattern, not continuously. Class 4
	Meandering pattern is absent: straight river channel (without artificial embankments). Class 5
	Meandering pattern is absent due to structural changes: straight river channel (artificial embankments). Class 6

Table 4.4 Pool-riffle pattern.

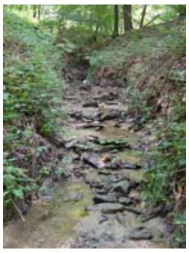
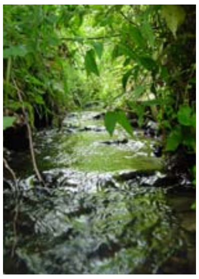
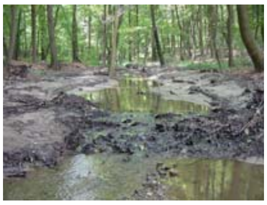
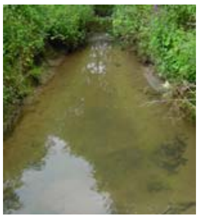
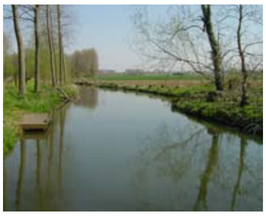
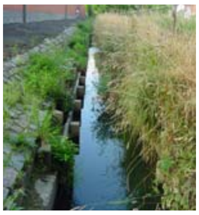
	Pool-riffle pattern is (nearly) pristine: extensive sequences of pools and riffles. Class 1
	Pool-riffle pattern is well developed: high variety in pools and riffles. Class 2
	Pool-riffle pattern is moderately developed: variety in pools and riffles but locally. Class 3
	Pool-riffle pattern is poorly developed: low variety in pools and riffles. Class 4
	Pool-riffle pattern is absent: uniform pool-riffle pattern. Class 5
	Pool-riffle pattern is absent due to structural changes: uniform pool-riffle pattern due to reinforced bank and bed structures. Class 6

Table 4.5 Hollow river banks.

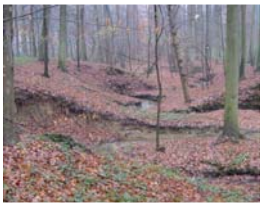
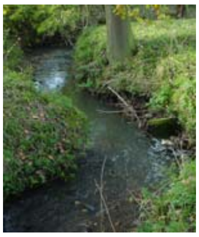
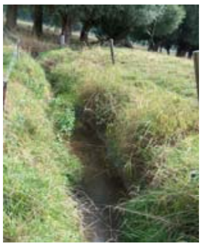
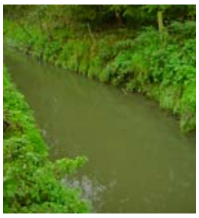
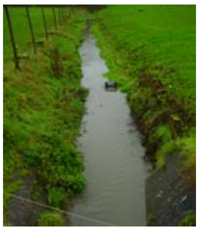
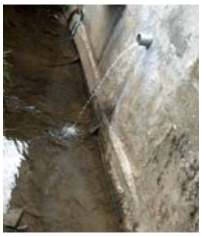
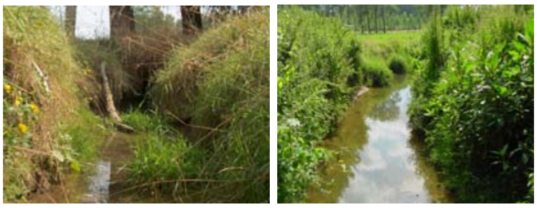
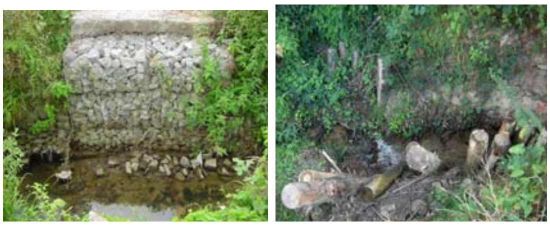
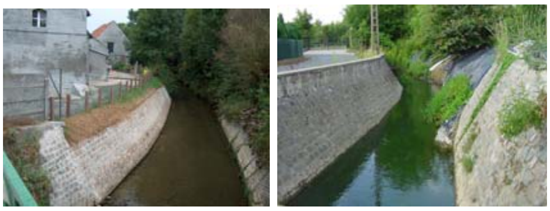
	Hollow river banks are (nearly) pristine: cavities under trees and in the outside curves. Class 1
	Hollow river banks are well developed: cavities merely in the outside curves. Class 2
	Hollow river banks are moderately developed: cavities under vegetation due to erosion. Class 3
	Hollow river banks are poorly developed: shallow bank erosion. Class 4
	Hollow river banks are absent: no cavities expected due to low dynamics. Class 5
	Hollow river banks are absent due to structural changes: absent due to reinforced bank structures. Class 6

Table 4.6 Bank structure.

	Natural/unmodified: no artificial bank reinforcement structures present. Class 0 (absent)
	Moderately and/or partial artificial/modified: part of the banks are reinforced with wood, stones, brick, concrete, … Class 1 (partial)
	Completely artificial/modified: banks are reinforced with wood, stones, brick, concrete, … Class 2 (total)

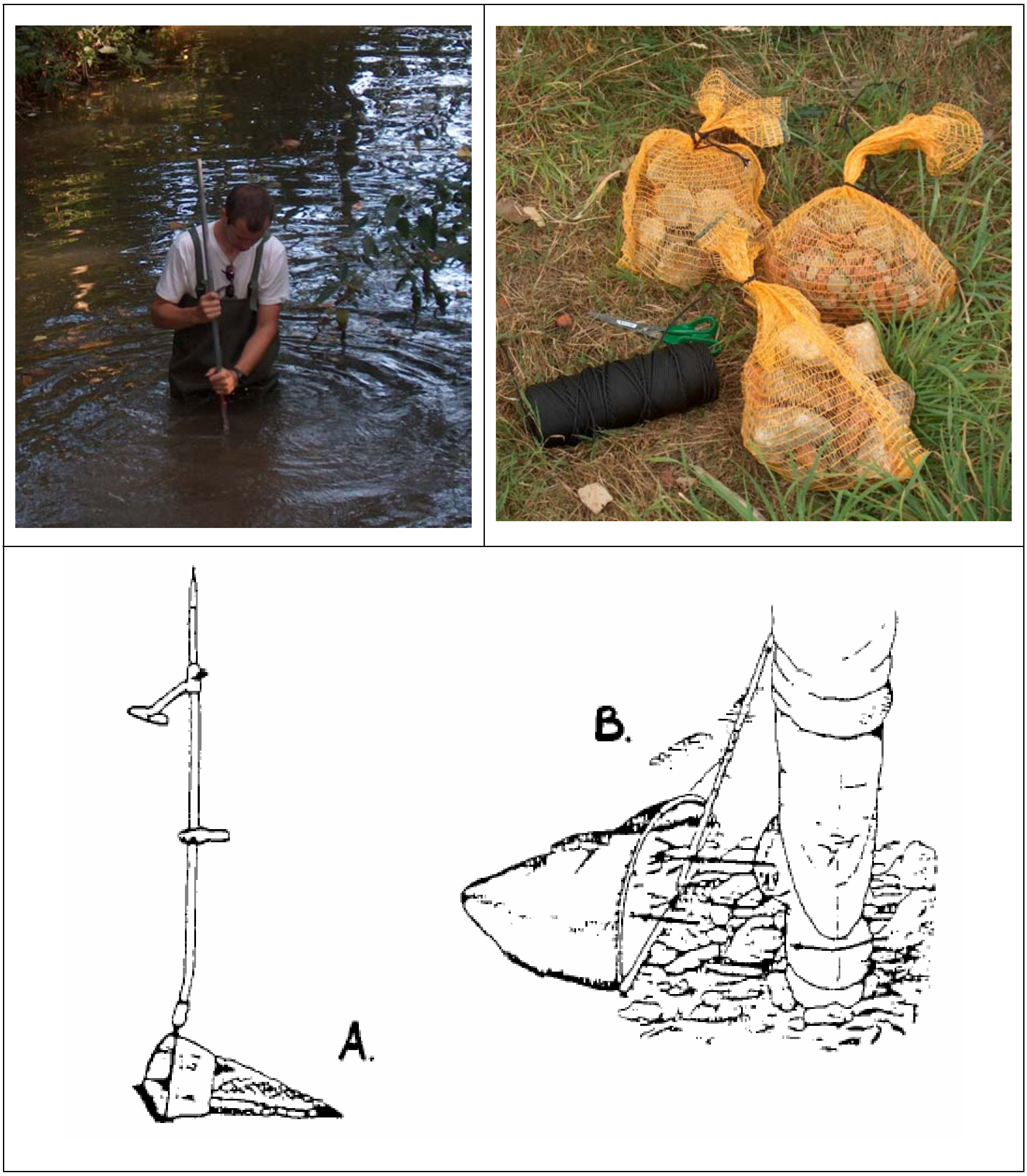


Figure 4.5 Illustrations concerning the kick sampling technique and artificial substrates (top) and details related to applied hand net to collect macroinvertebrates: A. handnet with handles; B. kick method (bottom).

4.3 Information collection on the habitat preferences of *Gammarus* and *Asellus*

4.3.1 Introduction

Gammarus and *Asellus* were chosen as representative taxa because of their highly variable presence in both ecological databases, their use as bio-indicators in river quality assessment (MacNeil et al., 2002), the relative high amount of ecological studies on this type of organisms (ecological as well ecotoxicological, e.g. Peeters, 2001; de Haas, 2004) and their importance in the food web (e.g. food source for many fish species (e.g. for Bullhead in streams) and break down of organic materials such as leaves). Both taxa are part of the subphylum of the crustaceans. However *Gammarus* is an amphipod (Amphipoda order), while *Asellus* is part of the Isopoda order (Table 4.7).

Table 4.7 ITIS taxonomic reports of Gammarus and Asellus (Linnaeus, 1758) (source: http://www.itis.usda.gov).

<table>
<tr><td colspan="2">Kingdom Animalia
Phylum Arthropoda
Subphylum Crustacea
Class Malacostraca
Subclass Eumalacostraca
Superorder Peracarida</td></tr>
<tr><td><b><i>Gammarus</i></b></td><td><b><i>Asellus</i></b></td></tr>
<tr><td>Order Amphipoda (amphipods)
Suborder Gammaridea

Family Gammaridae
Genus <i>Gammarus</i>
Species e.g. <i>Gammarus pulex</i></td><td>Order Isopoda (isopods, sowbugs)
Suborder Asellota
Superfamily Aselloidea
Family Asellidae
Genus <i>Asellus</i>
Species e.g. <i>Asellus aquaticus</i></td></tr>
</table>

In the following chapters a general desciption is given of the habitat preferences of both taxa. This information will be used for the practical ecological validation of the data driven models. Major difficulties were encountered to find consistent expert knowledge. Many descriptions did not explicitly mention numerical ranges or regression curves, identification was often done at different levels, studies were performed on data sets of all kinds (e.g. Peeters (2001) who studied both groups in lakes, estuaries and rivers, in combination with toxicity tests and laboratory setups), etc. Nevertheless, some concordant characteristics where found and are mentioned below.

4.3.2 Knowledge base on Gammarus

For the *Gammarus* genus, the species *Gammarus pulex* is of major importance in Flanders. *Gammarus pulex* appears in all kind of types of waters: lakes, headwaters, river tributaries, canals, etc… (Karaman and Pinkster, 1977; Hawkes, 1979; Verdonschot, 1990; Peeters, 2001), but prefers rather fast running streams (Bayerisches Landesamt für Wasserwirtschaft, 1996), since it has very good swimming abilities (Brehm and Meijering, 1990). Illustrations of the organism and a scheme of its typical habitat is presented in Figure 4.6. *Gammarus pulex* is almost non-tolerant for low oxygen conditions (Wesenberg-Lund, 1982), but it can tolerate low oxygen concentrations when water temperatures are low (Gledhill et al., 1993). It generally prefers localities with a temperature well below 20°C (Gledhill et al., 1993). *Gammarus pulex* is suppressed by high organic conditions (Hawkes, 1979), but can stand organic pollution (Gledhill et al., 1976; Gledhill et al., 1993). Generally, *Gammarus pulex* is less tolerant to inorganic pollutants and to organic sewage (Whitehurst and Lindsey, 1990). *Gammarus pulex* prefers substrate heterogeneity (Tolkamp, 1980), especially detritus substrates or detritus mixed with sand or gravel or leaf material (Tolkamp, 1982). Gammaridae are sensitive to high conductivity values. At conductivity values above 1000 µS/cm, they experience negative influences (Macrofauna-atlas of North Holland, 1990). *Gammarus pulex* is normally absent from acid waters where the pH is below 5.7 (Gledhill et al., 1993), this was confirmed by Peeters (2001), who described via logistic regression the habitat niche for several environmental variables (Table 4.8). The latter author found out that *Gammarus pulex* occurs in ranges between pH 4.7 to 11.6, while for the related species *Gammarus fossarum* a more narrow range was described (pH 6.9 to 9.9). The order of importance of the variables in the logistic regression model by Peeters (2001) were: current velocity, Kjeldahl nitrogen, pH and depth.

Table 4.8 Values for the environmental variables at which the maximum probability of presence of Gammarus pulex was reached and the total range of occurrence (probability larger than one percent). These values were based on a logistic regression model. The < or > signs mean respectively that these model values are lower or higher than the observations (Peeters, 2001).

Variable	**Maximum probability of presence value**	**Range of occurrence**
Current velocity (cm/s)	71	0 - 198
Width (m)	0.1	0.1 - >40.0
Depth (m)	0.1	0.01 - >5.00
BOD (mg/l)	0.1	0.1 - 37.0
Chloride (mg/l)	6	<6 - >498
Conductivity (µS/cm)	398	<88- >7942
Ammonium nitrogen (mg/l)	0.01	0.01-57.00
Kjeldahl nitrogen (mg/l)	0.20	0.10 - >68.00
Oxygen (mg/l)	14.0	<0 - >27
Oxygen saturation (%)	90	1 – 220
Total phosphorus (mg/l)	0.13	0.01 - >18.00
pH	8.1	4.7 - >11.6
Water temperature (°C)	9.8	<0 - >30

4.3.3 Knowledge base on Asellus

Two *Asellus* species were found (*A. aquaticus* and *A. meridianus*) in the Zwalm database, while no species-level information was available about the river sediments database (1996-1998). These species have almost no differences in ecological preferences, although *Asellus aquaticus* is thought to be a little bit more resistant against pollution than *Asellus meridianus* (Gledhill et al., 1976; Chambers, 1977; Gongrijp, 1981; Verdonschot, 1990). *Asellus aquaticus* is on the other hand very resistant against low oxygen conditions (Hawkes, 1979; Verdonschot, 1990). *Asellus aquaticus* is tolerant against organic conditions, and often replaces *Gammarus* species under high organic conditions (Hawkes, 1979; Verdonschot, 1990). *Asellus aquaticus* lives in waters especially when there is a varied detritus layer. Asellidae are mentioned to behave as indifferent along a water velocity gradient according to the Bayerisches Landesamt für Wasserwirtschaft (1996), while Tachet et al. (2002) mention the preference for downstream sections characterized by low flow velocities. Also Peeters (2001) mentions that *Asellus aquaticus* attempts to escape from sites with higher flow stress or that repeated passive drift took place. Asellidae also have a preference for water courses with higher width (Macrofauna atlas of North Holland, 1990). Peeters (2001) mentions a moderate sensitivity towards metal contamination, in comparison to other macroinvertebrate taxa.

A typical habitat is presented in Figure 4.7, in addition with a picture and a drawing. Because of the nearly similar ecological preferences of *Asellus aquaticus* and *Asellus meridianus*, *Asellus* prediction models were constructed for both species together. For bio-assessment, this generalization could be of an important practical use, reducing the number of models for prediction of macroinvertebrate taxa in rivers.

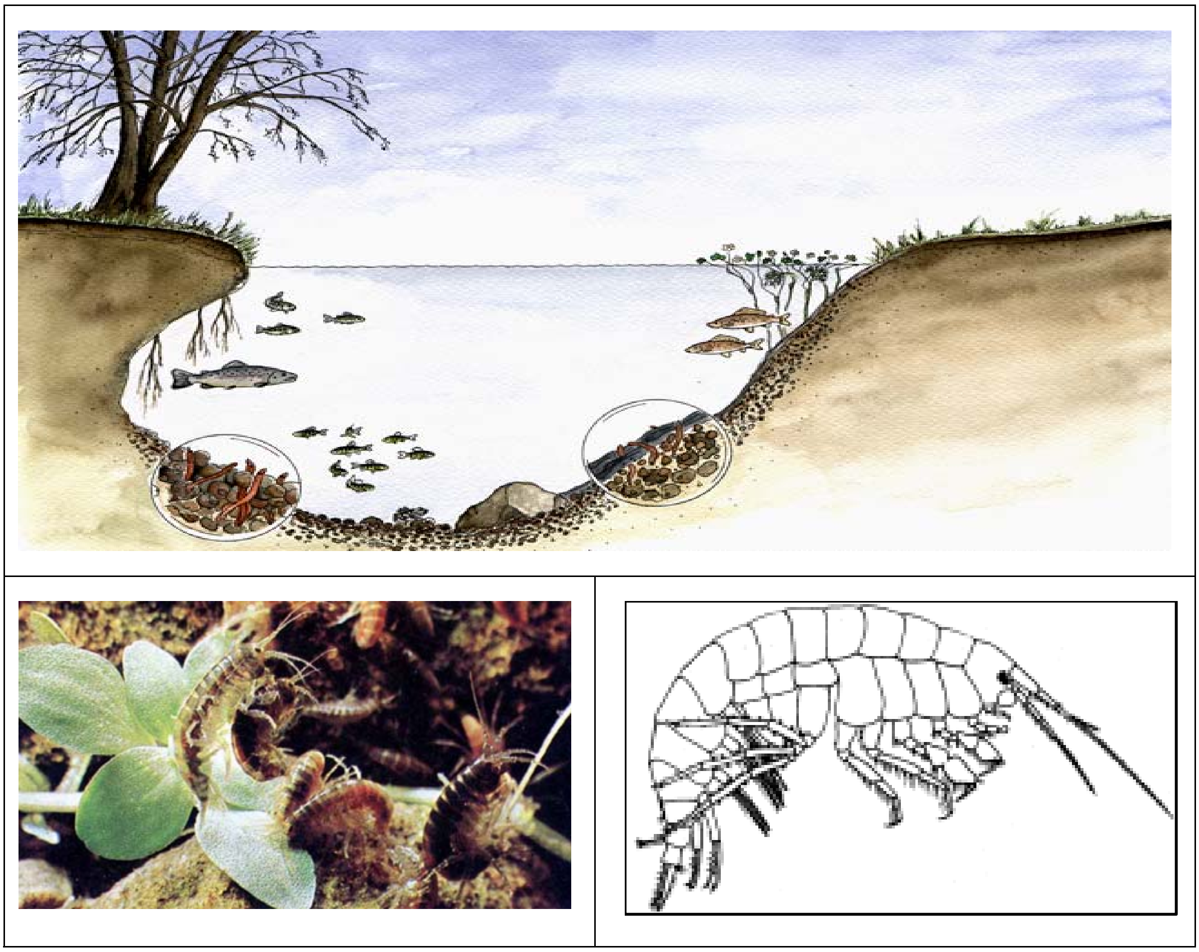


Figure 4.6 Scheme of a typical habitat of Gammarus (top) and in situ picture and drawing of Gammarus pulex (bottom).

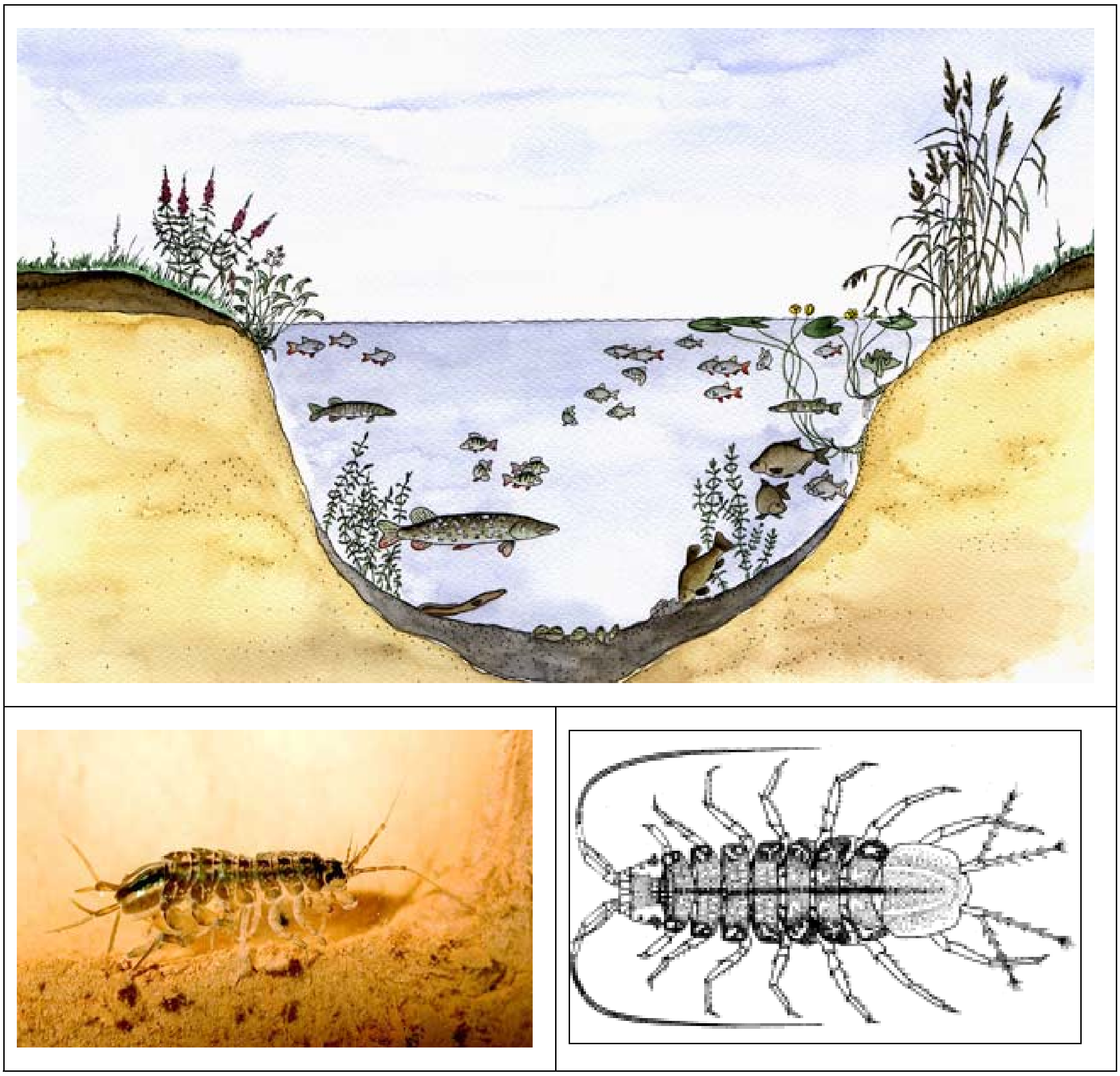


Figure 4.7 *Scheme of a typical habitat of Asellus (top) and in situ picture and drawing of Asellus aquaticus (bottom).*

Chapter 5
Data driven development of habitat suitability models based on classification trees and artificial neural networks to predict *Gammarus* and *Asellus* in rivers

5.1 Introduction

This chapter is devoted to the development of predictive ecological models based on data driven methods (classification trees and artificial neural networks). These model development and habitat suitability studies focus on two taxa, the Crustaceans *Gammarus* and *Asellus*. In addition and in combination with the ANN models, several input contribution methods were applied to detect the major river characteristics to describe the habitat suitability for the two taxa. For the classification tree method, the major variables are selected and visualised based on pruning procedures and the outcomes can be compared to the ANN model input variable contribution methods. The ecological relevance of the models is analysed on the basis of expert knowledge from scientific literature.

This chapter consists of seven components. The first component describes the methodologies behind the data analysis and preparation, development of predictive models based on data driven methods and the model validation methods. The second one deals with the application of data analysis and preparation methods. The other five components are presenting the following results:

- development of habitat suitability models based on classification trees to predict *Gammarus* and *Asellus* in river sediments in Flanders;
- application of backpropagation artificial neural networks predicting *Gammarus* and *Asellus* in river sediments in Flanders;
- development of habitat suitability models based on classification trees to predict *Gammarus* and *Asellus* in the Zwalm river basin;
- application of backpropagation artificial neural networks predicting *Gammarus* and *Asellus* in the Zwalm river basin;
- a comparative discussion of the obtained results.

5.2 Data driven model development methods

5.2.1 Overview of the data analysis, model development and validation procedures

This component describes the model development and validation procedures that were selected for this study. Prior to the description of these methods, a set of data analysis procedures is presented. These procedures were applied to get insight in the used datasets, needed for a reliable development of data driven models, but also useful for the interpretation of the model results afterwards.

5.2.2 Data analysis procedures

5.2.2.1 Bandwidth and distribution of input and output variables

Data driven models are built solely from the examples presented during the training phase, which are together assumed to implicitly contain the information necessary to establish the relation between input and output. As a result, these models are unable to extrapolate beyond the range of the data used for training. Consequently, poor predictions can be expected when the validation data contain values outside of the range of those used for training (Maier and Dandy, 2000). Insight in the range of inputs and outputs, which determine also the maximum application range of data driven models, is therefore a first and basic step before model development and application.

5.2.2.2 Correlation between input variables

A first pass filter to help identify ‘noise’ variables is to calculate the correlation of pairs of variables. If two variables are strongly correlated, then one of these two variables may be removed without adversely affecting the model performance. The cut-off value for variable elimination is a heuristic value and must be determined separately for every model development application. But any correlation with absolute value of 0.20 or higher indicates a probable noise source to in particular ANN models according to Walczak and Cerpa (1999). However, the removal of input variables can be overruled for practical reasons (river managers interest for particular simulations) or by use of ecological expert knowledge.

5.2.2.3 Visual relation analysis between input and output variables

A visual relation analysis between the input and output variables can be beneficial to get insight in outliers, the data clusters, missing or scarce variable combinations in certain ranges, … As such, these methods can be very interesting in delivering insight in the difficulty to develop well performing models, why models perform weakly, whether some data can be classified as outliers (even check whether it involves errors of all sort, e.g. measurement uncertainty, data digitalization errors, …). For this, data visualisation methods can be very interesting to get a better understanding of the model performance in the end and also to reveal what type of measurements should be undertaken in the future to enhance the data set.

These analyses have gained a lot of popularity during the last years and became standard tools in most data mining and analysis software packages. These analyses were performed in Weka (Witten and Frank, 2000).

5.2.3 Model development methods

5.2.3.1 Dataset construction for model training and validation

Three fold cross validation was used for the model training and evaluation of the predictive performance, as well as to compare the different methods to test the contributions of the input variables of the ANN models. In this manner one gets insight in the stability of the model development and the input variable contribution methods, while also limiting the work load (data preparation, model training and validation). Much more intensive cross validation, such as ten fold cross validation (what is described as a standard value by Witten and Frank (2000)) seemed also less suitable seen the relative small datasets, and the related unreliably small validation sets it would involve.

The training dataset consisted of 228 instances and the validation set of 114 instances in case of the river sediments in Flanders (1996-1998), while similarly 119 instances and 60 instances were used respectively in the training and validation sets of the Zwalm river (2000-2002). The prevalence of the taxon was similar in all training and validation sets (thus different combinations were constructed for *Gammarus* and *Asellus*), as is illustrated for *Asellus* in Figure 5.1. Before training of the decision trees and neural networks, the data were randomly shuffled in the training datasets, to avoid biased training.

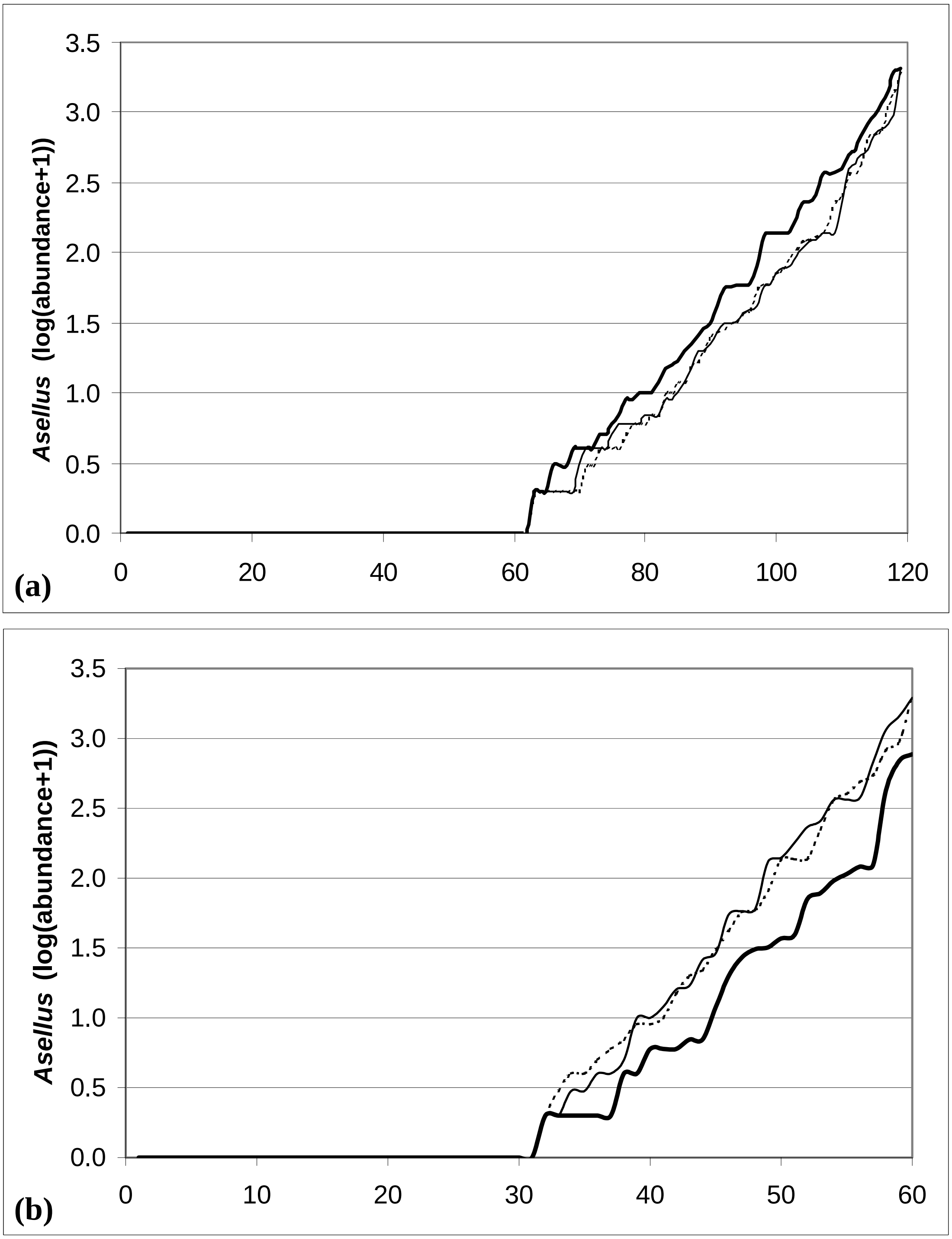


Figure 5.1 Distribution plots of the log(abundance+1) of Asellus in the three training (a) and validation (b) data subsets derived from the Zwalm river basin database. The training dataset consisted of 119 instances, while the validation set comprised 60 instances. The amount of instances in which Asellus is present is similar in all training and validation sets. Before training of the neural network, the data were randomly shuffled in the training datasets, to avoid biased training.

The abundances of the macroinvertebrates were transformed to the class absent defined as [0, 0.5[and present as [0.5, 1[for the presence/absence models (both classification trees and ANNs). For the regression ANN models, a log(abundance+1) transformation was executed, as is often applied in ecological databases consisting of a lot of zero's and a relatively small amount of data in the larger range (Legendre and Legendre, 1998).

5.2.3.2 Classification trees

Classification trees were constructed on the basis of the Weka software (Witten and Frank, 2000). Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

The applied algorithm to grow classification trees is 'weka.classifiers.trees.J48'. This is an algorithm to construct pruned or unpruned C4.5 classification trees (Witten and Frank, 2000). Default settings were applied (based on trial and error these gave relative good results that are difficult to improve in a general manner for all used datasets), except for the confidence factor (default 0.25), which had a very important effect on the selected variables and the number that were used for the classification. This factor was tested at four levels: 0.5, 0.25, 0.1 and 0.01. This led to the following settings to develop the classification trees:

- 'binary splits' (whether to use binary splits on nominal attributes when building the trees): false;
- 'confidence factor' (the confidence factor is used for pruning; smaller values incur more pruning): 0.5, 0.25, 0.1 and 0.01;
- 'minimum number of instances per leaf': 2;
- 'number of folds' (determines the amount of data used for reduced-error pruning: one fold is used for pruning, the rest for growing the tree): (3);
- 'reduced error pruning' (whether reduced-error pruning is used instead of C.4.5 pruning): false;
- 'seed' (the seed used for randomizing the data when reduced-error pruning is used): 1;
- 'subtree raising' (whether to consider the subtree raising operation when pruning): true;
- 'unpruned' (whether pruning is performed): false;
- 'use Laplace' (whether counts at leaves are smoothed based on Laplace): false.

5.2.3.3 Artificial neural networks

5.2.3.3.1 Artifical neural network parameter settings and architecture

A multi-layer feed-forward neural network was trained using an error backpropagation training algorithm (Rumelhart et al., 1986). The structure of the applied Artificial Neural Network is presented in Figure 5.2. The network consisted of 24 input neurons, each representing an environmental variable, 10 hidden neurons and one output neuron, indicating the probability of presence (in case of training with presence/absence classes) or log(abundance+1) of the taxon. The transfer functions were of the logistic sigmoid type. In case of presence/absence model predictions, the model output ('probability of presence') was transformed to the class absent defined as [0, 0.5[and present as [0.5, 1[. In this manner the CCI and *K* could be calculated for comparisons between the models.

The training was stopped when the error in the validation set started to increase in order to avoid overfitting (cf. Gevrey et al., 2003). The neural network models were implemented in the software package MATLAB 6.1 for MS Windows™ (according to Gevrey et al., 2003). The settings in this toolbox were all default and were determined based on experience: learning rate = 0.001, incremental learning rate = 1.05, decreasing learning rate = 0.75, momentum = 0.95, transfer functions in hidden and output layer = logsig, error ratio = 1.04 and weight coefficient = 0.3.

5.2.3.3.2 Input variables contribution methods

Although many methods (and terms) exist for attribute selection and sensitivity analyses (e.g. Witten and Frank, 2000), only a limited set, consisting of six methods that had already proven to be convenient in ecological modelling studies, was applied during this study. These methods were selected and integrated in a MATLAB toolbox by Gevrey et al. (2003) at the Université Paul-Sabatier (Toulouse, France) as part of the European PAEQANN-project (EVK1-CT1999-00026): 'Predicting Aquatic Ecosystem Quality using Artificial Neural Networks: Impact of Environmental Charateristics on the Structure of Aquatic Communities (Algae, Benthic and Fish Fauna)' (http://quercus.cemes.fr/paeqann). These methods are briefly described underneath.

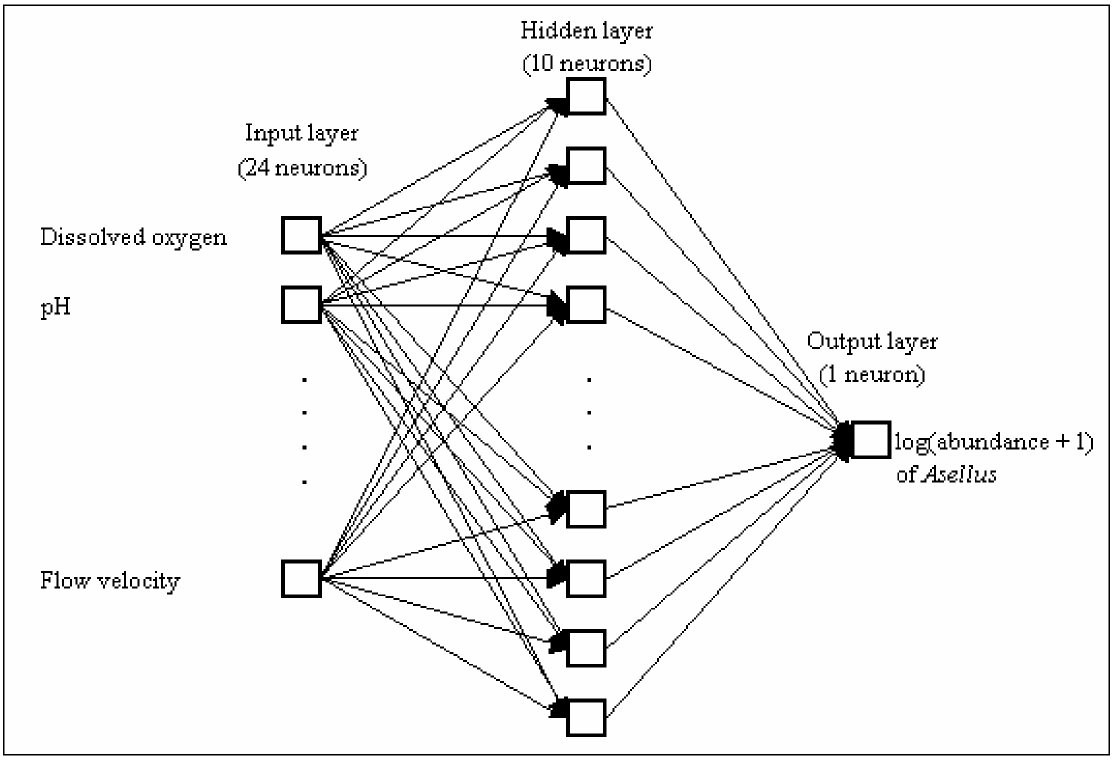


Figure 5.2 Example of the structure of the Artificial Neural Network used in this study. The input layer consisted of the 24 input variables, the hidden layer comprised 10 neurons and the output of the network was the log(abundance+1) or probability of presence of the taxon (in this case Asellus).

‘PaD’ method

Two results can be obtained by this method. The first one is a profile of the output variations for small changes of each input variable and the second one a classification of the relative contributions of each variable to the network output. Only the second method is presented. Therefore, the partial derivatives of the ANN output with respect to the input are calculated (Dimopoulos et al., 1995, 1999). For a network with n_i inputs, one hidden layer with n_h neurons and one output (i.e. n_o = 1), the partial derivatives of the output y_j with respect to input x, (with j = 1, …, N and N the total number of observations) are:

$$d_{ji} = S_j \sum_{h=1}^{n_h} w_{ho} I_{hj} (1 - I_{hj}) w_{ih}$$

(on the assumption that a logistic sigmoid function is used for the activation). When S_j is the derivative of the output neuron with respect to its input, I_{hj} is the response of the hth hidden neuron, w_{ho} and w_{ih} are the weights between the output neuron and hth hidden neuron, and between the ith input neuron and the hth hidden neuron.

The result of the second method concerns the relative contribution of the ANN output to the dataset with respect to an input. It is calculated by a sum of the square partial derivatives obtained per input variable:

$$SSD_i = \sum_{j=1}^{N} (d_{ji})^2$$

One SSD (Sum of Square Derivatives) value is obtained per input variable. The SSD values allow classification of the variables according to their increasing contribution to the output variable in the model. The input variable that has the highest SSD value is the variable, which influences the output most.

‘Weights’ method

The procedure for partitioning the connection weights to determine the relative importance of the various inputs was proposed first by Garson (1991) and repeated by Goh (1995). The method essentially involves partitioning the hidden-output connection weights of each hidden neuron into components associated with each input neuron. This algorithm is simplified but gives results identical to the algorithm initially proposed:

(1) For each hidden neuron h, divide the absolute value of the input-hidden layer connection weight by the sum of the absolute value of the input-hidden layer connection weight of all input neurons, i.e.

```
For h = 1 to $n_h$,
        For i = 1 to $n_i$,
```

$$Q_{ih} = \frac{|W_{ih}|}{\sum_{i=1}^{n_i} |W_{ih}|}$$

```
    end,
end.
```

(2) For each input neuron i, divide the sum of the Q_{ih} for each hidden neuron by the sum for each hidden neuron of the sum for each input neuron of Q_{ih}, multiplied by 100. The relative importance of all output weights attributable to the given input variable is then obtained.

```
For i = 1 to $n_i$

    $RI(\%)_i = \frac{\sum_{h=1}^{n_h} Q_{ih}}{\sum_{h=1}^{n_h} \sum_{i=1}^{n_i} Q_{ih}} \times 100$

end.
```

‘Perturb’ method

This method aims to assess the effect of small changes in each input on the neural network output. The algorithm adjusts the input values of one variable while keeping all the others untouched. The responses of the output variable against each change in the input variable are noted. The input variable whose changes affect the output most is the one that has the most relative influence. In fact, the mean square error (MSE) of the ANN output is expected to increase if a larger amount of noise is added to the selected input variable (Yao et al., 1998; Scardi and Harding, 1999). These changes can take the form of $x_i = x_i + \delta$ where x_i is the selected input variable and δ is the change. δ can be increased in steps of 10% of the input value up to 50% (commonly used values). The aim is to assess the effect of small changes in each input on the ANN output. We can then obtain a classification of the input variables by order of importance.

‘Profile’ method

This method was proposed by Lek et al. (1995, 1996a, b). The general idea is to study each input variable successively when the others are blocked at fixed values. The principle of this algorithm is to construct a fictitious matrix pertaining to the range of all input variables. In greater detail, each variable is divided into a certain number of equal intervals between its minimum and maximum values. The chosen number of intervals is called the scale. All variables, except one, are set initially (as many times as required for each scale) at their minimum values, then successively at their first quartile, median, third quartile and maximum. For each variable studied, five values for each of the scale's points are obtained. These five values are reduced to the median value. Then the profile of the output variable can be plotted for the scale's values of the variable considered. The same calculations can then be repeated for each of the other variables. For each variable a curve is then obtained. This gives a set of profiles of the variation of the dependent variable according to the increase of the input variables. In this work, a scale of 12 was used between the minimum and maximum of the input variables.

'Stepwise' method

This method is the 'Classical Stepwise' method that consists of adding or rejecting step by step one input variable and noting the effect on the output result. Based on the changes in MSE, the input variables can be ranked according to their importance in several different ways, depending on different arguments. For instance the largest changes in MSE due to input deletions can allow these inputs to be classified by order of significance. Another approach is that the largest decrease in MSE can identify the most important variables, i.e. the most relevant to the construction of a network with a small MSE (Sung, 1998). In this study, the backward stepwise modelling approach (one by one elimination of the input variables) was adopted to assess the effect of the 24 input variables used. Therefore, 24 models were generated, each using only 23 of the available variables as inputs. The 24th missed out variable for which the resulting models gave the largest error, is the most important. Then, 23 models were generated, combining 22 variables, i.e. all the variables minus that one eliminated just before and one of the other available inputs was eliminated in each model. This procedure was repeated using models with 21 input variables, 20, etc. until the 23 variables were all eliminated. The order of elimination of the input variables in the network is the order of the importance of their contribution.

‘Improved Stepwise’ method

The major drawback of the ‘Classical Stepwise’ method is that at each step a new model is generated and requires training. An improvement of this method consisted of building another called ‘Improved Stepwise’ method where only one model is used. In methods that use a single trained model, each variable in turn is processed and the MSE examined. The variable that gives the largest MSE when eliminated is the most important one. A classification of the variables can thus be made. For the ‘Improved Stepwise’ method used in this study, all the values of one input are transformed to the same value, i.e. its mean.

Method stability evaluation

In order to check the stability of each method, the training of the network was repeated three times, according to the three fold cross validation, and the relative contributions of the input variables on the output obtained evaluated for all methods and each trained network. Then, the mean contribution of each variable for the different methods was calculated. The three training sessions allowed for calculating the standard deviation what gave an indication of the stability of each method.

5.2.4 Model validation methods

Model validation in this study is based on performance indicators, comparison with ecological expert knowledge and their convenience for practical applications (this work is presented in Chapter 6). Depending on the type of output, different performance measures are convenient to evaluate and compare models.

When presence/absence of the macroinvertebrates is predicted, most of the reviewed papers (cf. Chapter 3) applied the percentage of Correctly Classified Instances (CCI) to assess model performance. There is however clear evidence that this CCI is affected by the frequency of occurrence of the test organism(s) being modelled (Fielding and Bell, 1997; Manel et al., 1999). Among the different measures, which are based on a confusion matrix (Table 5.1), proposed to assess the performance of presence/absence models (Table 5.2), Fielding and Bell (1997) and Manel et al. (1999) recommended the Cohen’s kappa (*K*) as a reliable performance measure, since the effect of prevalence on the *K* appeared to be negligible (e.g. Dedecker et al., 2004a, c, D'heygere et al., 2004). A CCI of at least 70% and K higher than 0.4 were considered as good classifications.

Table 5.1 The confusion matrix as a basis for the performance measures with true positive values (TP), false positives (FP), false negatives (FN) and true negative values (TN).

		Observed	
		+	-
Predicted	+	a (TP)	b (FP)
	-	c (FN)	d (TN)

Table 5.2 Selected measures based on the confusion matrix to assess the performance of presence/absence models (after Fielding and Bell (1997)).

Performance measure	**Calculation**
CCI	*(a+d)/N*
Cohen's kappa	$\frac{[(a+d)-(((a+c)(a+b)+(b+d)(c+d))/N)]}{[N-(((a+c)(a+b)+(b+d)(c+d))/N)]}$

When the output of a model consists of the species abundance, commonly used performance measures are the correlation (r) or determination (r^2) coefficient and the (root) mean squared error ((R)MSE) or a derivative between observed (O) and predicted (P) values. For this study, the r is selected for the evaluation of the model performance. An r value larger than 0.4 is considered as a good model.

Correlation coefficient (r) =

$$\frac{\sum(P\times O)-\frac{(\sum P\times\sum O)}{N}}{\sqrt{(\sum P^2-\frac{(\sum P)^2}{N})\times(\sum O^2-\frac{(\sum O)^2)}{N})}}$$

5.2.5 Model development scheme

As an overview of the delivered results in this study, Table 5.3 presents an overview how the data driven model development methods were applied on the river sediments of Flanders (1996-1998) and Zwalm river basin (2000-2002) databases. To ease the interpretation and comparison between the model development methods, input variable contribution techniques, the two taxa and the databases, all methods were applied in an identical manner over both databases and taxa.

Table 5.3 Model development scheme as applied to both databases.

Classification trees (four pruning levels)
Evaluation based on performance indicators ***Selected variables and ranking of importance*** ***Ecological interpretation and discussion of practical use***
Artificial neural networks
Presence/absence classification of taxa Performance based on performance indicators Application of input variable contribution methods Weights PaD Perturb Stepwise Reg Stepwise Imp Profile ***Prediction taxa abundance*** Evaluation based on performance indicators Application of input variable contribution methods Weights PaD Perturb Stepwise Reg Stepwise Imp Profile ***Selected variables and ranking of importance*** ***Ecological interpretation and discussion of practical use of the ANN models***

5.3 Data analysis results

5.3.1 Bandwidth and distribution of input and output variables

A first step in the data analysis consisted of the analysis of the minima, maxima, averages and standard deviations (Table 5.4 and 5.5). Preferably, these analyses can be combined with visualisation graphs as presented in Figures 5.3-5.7. By doing so, one can directly see whether high standard deviations are a result of a wide span of most data or more related to some outliers (or 'strange' distributions). The use of the median (and compare it with the average) can as well give a good indication in this context. Since the visualisation graphs were available, this seemed not to give an added value.

Table 5.4 Minima, maxima, averages and standard deviations of the input and output variables that were used for the river sediments in Flanders database (1996-1998).

Variable	**Minimum**	**Maximum**	**Average**	**Standard deviation**
Gammarus (abundance)	0	2000	9	111
Gammarus (log(abundance+1))	0.0	3.3	0.2	0.4
Asellus (abundance)	0	257	6	24
Asellus (log(abundance+1))	0.0	2.4	0.3	0.5
Day	20	338	175	107
Width (m)	0.4	15.0	3.9	2.9
Depth (m)	0.0	3.0	0.6	0.5
Flow velocity (class variable)	0.0	4.0	-	-
Clay (%)	0.0	65.0	10.6	10.7
Loam (%)	0.0	80.0	20.2	19.5
Sand (%)	0.0	100.0	69.1	26.3
Temperature (°C)	0.0	24.5	10.9	5.5
pH	3.4	9.1	7.4	0.6
Dissolved oxygen (mg/l)	0.1	13.2	5.7	2.5
Conductivity (µS/cm)	110	16660	907	1183
Organic matter (mg OM/kg DM)	0.4	113.0	4.9	7.3
TOXT (class variable)	0	1	-	-
TOXR (class variable)	0	1	-	-
Total phosphorus (mg P/kg DM)	17	42200	1747	3399
Kjeldahl nitrogen (mg N/kg DM)	100	11200	2010	1799
Cr (mg /kg DM)	0	7020	64	406
Pb (mg /kg DM)	0	1780	46	132
As (mg /kg DM)	0	120	11	16
Cd (mg /kg DM)	0	53	1	4
Cu (mg /kg DM)	0	3740	41	220
Hg (mg /kg DM)	0	21100	88	1208
Ni (mg /kg DM)	0	300	14	23
Zn (mg /kg DM)	8	4440	222	381

Based on this analysis of the river sediments in Flanders dataset, one can observe already several outliers (very high values and standard deviations) in the variables conductivity, total phosphorus, Kjeldahl nitrogen and most of the metals. However, these observations were probably correct measurements, and concerned very contaminated sites. This is clearly a major disadvantage of not being involved in the data collection. In case of the Zwalm river basin this doubt about data reliability was not encountered.

Table 5.5 Minima, maxima, averages and standard deviations of the input and output variables that were used for the Zwalm river basin database (2000-2002).

Variable	Minimum	Maximum	Average	Standard deviation
Gammarus (abundance)	0	2850	164	401
Gammarus (log(abundance+1))	0.0	3.5	1.2	1.0
Asellus (abundance)	0	2040	92	287
Asellus (log(abundance+1))	0.0	3.3	0.7	1.0
Width (cm)	39	950	231	233
Banks (class variable)	0.0	2.0	-	-
Meandering (class variable)	1.0	6.0	-	-
Pool/Riffle (class variable)	1.0	6.0	-	-
Hollow beds (class variable)	1.0	6.0	-	-
Depth (cm)	0	170	32	34
Flow velocity (m/s)	0.0	1.9	0.4	0.3
pH	6.7	8.1	7.6	0.3
Temperature (°C)	10.6	20.9	14.3	2.1
Dissolved oxygen (mg/l)	0.1	10.8	7.1	2.0
Conductivity (µS/cm)	10	1414	741	183
Suspended solids (mg/l)	0	949	46	90
Ammonium (mg NH_4^+/l)	0.0	6.0	0.9	1.1
Nitrate (mg NO_2^-/l)	0.2	15.2	5.8	2.9
Total nitrogen (mg N/l)	2.5	75.7	11.3	7.2
Phosphate (mg PO_4^{3-}/l)	0.0	5.0	0.4	0.5
Total phosphorus (mg P/l)	0.1	4.7	0.4	0.5
COD (mg O_2/l)	7.0	52.0	19.0	7.4
Boulders (%)	0.0	100.0	37.3	38.9
Gravel (%)	0.0	67.7	12.4	17.5
Sand (%)	0.0	87.8	20.3	21.2
Loam/clay (%)	0.0	100.0	26.7	26.7
Distance to mouth (m)	1541	19864	9959	5081
Stream order	1	4	2	1

This led directly to a though decision, whether to leave these measurements (instances) in the database or not, because for sure they can lead to less reliable models, as the broad range of some variables results in a relative compression of the majority of the measurements. In other words, the choice between sensitivity and bandwidth of the models had to be made. In this context, Dedecker et al. (2004b) tested the sensitivity and robustness of the ANN models when data, containing variables beyond the range of the data for training, were added. Therefore, the authors created a virtual dataset based on ecological expert knowledge to introduce 'extreme' values to the model. According to this study, the overall predictive power of the ANN models only decreased significantly when a relatively large virtual dataset in the training set was applied. Seen the limited set of 'extreme' values according to the data visualisation plots in 5.3.3, this study by Dedecker et al. (2004b) could be an argument to keep the outliers in. But also to make the models applicable in the widest span of cases and to make a tryout on data that are as natural as possible these 'outliers' were kept in the dataset (as such it was possible to check whether these data driven model development methods can deal themselves with outliers as is sometimes referred to by ANN experts). The latter has to do with testing the objectivity of the method and user-convenience as well. When too much needs to be prepared on the dataset, the methods will probably become less attractive.

The dataset of the Zwalm river basin consisted as well of variables with a high standard deviation. However, few potential outliers could be detected and by rechecking in the field one could be convinced that in most cases it concerned indeed very good or very bad sites. Especially the very good ones are necessary for the prediction of the restoration options. In this manner these data analyses can also be helpful to check what kind of additional data are needed. In case of the Zwalm it are in particular very good sites that are missing and that could make the dataset better balanced. As a result, the prediction of very good conditions will be rather difficult for the derived data driven models. Also here, no data (instances) were removed.

5.3.2 Correlation between the input variables

The next step consisted of checking how related some variables might be on the basis of their correlation coefficient (*r*). According to Walczak and Cerpa (1999), any *r* with an absolute value of 0.20 or higher indicates a probable noise source to in particular ANN models and they advice to consider the removal of one of these variables. However, there might be practical reasons to leave these correlated variables in, such as ecologically not relevant correlations (merely on the bases of coincidence), but also for practical applications where both variables might be altered in a different manner to simulate restoration options (e.g. metal pollution, specific channel modifications). This exercise is in this respect very interesting, because it means that the models are not trained to deal with these independent alterations of the highly correlated variables and might be characterized by an ill performance as they are ‘not trained for this job’. Therefore the validation with practical simulations is also necessary.

In the sediments database, increased correlation values are identified for physical habitat variables (width, depth, flow velocity), sediment characteristics (clay, loam, sand), organic matter (organic matter, Kjeldahl nitrogen) and most of the metals (except As and Hg). There might have been as well a higher *r* observed between the metals and toxicity tests. The latter may indicate that toxicity can be related to other pollutants, that the bio-availability is an additional factor that has to be included to link the metal concentrations with their effects on organisms or that the available toxicity tests are not representative.

In the Zwalm, quite a high *r* can be observed between the physical habitat variables. This is rather logic, because many artificial structures are combined (e.g. channel straightening with bank fortification). However, this is not always the case, and many exceptions exist, and models without this set of variables might lead to practical limitations of the models. Meandering is however clearly related to pool/riffle structures and hollow river banks. Perhaps in future analyses these three variables could be reduced to one. Also an expected good relation existed between width, depth, distance to mouth and stream order. Also here a variable reduction might be interesting. A better relation on the other hand was expected between the latter two sets of highly correlated variables (such as meandering and flow velocity).

So, although there are several variables characterized by a high r, all variables were kept in both databases. In most cases there is a practical reason to keep them in, such as to prevent the limitation of simulations that can be done. Also the effect on the data driven model development and the variable contribution methods is interesting: will the data driven model development methods succeed or not to ‘remove’ these highly redundant variables and how will they be ranked by the methods. If so, it would again be advantageous from a user friendliness perspective. On the other hand, several variables that were expected to be correlated were not. This is also an indication that one has to be careful during the selection phase of variables before the data collection. Better to monitor some extra variables… In addition, one has to be aware of the mere effect related to the manner how variables are calculated or expressed (e.g. classes, or transformations combining several variables, such as combination of metals with organic matter fraction to calculate bio-available metal fraction). Simply the latter 'manipulation' can already have a significant effect on the observed correlation and probably on the derived models as well.

5.3.3 Visual relation analysis between input and output variables

The third type of data analysis was based on visualisation graphs as presented in Figures 5.3-5.7. In addition to insight in the distributions of the input variables (as already discussed in the first data analysis paragraph), also directly the distribution of the observed output classes is plotted. Two types of plots were used. The first type was based on the use of classes to look at the distributions. The second type (only a small part is presented in Figure 5.7 as an example), scatter plots were used as well. These go more into detail, but are on other hand much more complex and difficult to interpret. Based on both types of graphs, one can get directly some idea of the influence of the individual variables on the output variable. Therefore, these graphs are very interesting to compare with the model outcomes as well and will be part of the discussions in the next paragraphs when the results are evaluated.

For most variables a logic relation can be observed, e.g. for the pollution variables a relation characterized by a reduction of the presence class when the concentration of the pollutant increases is presented. Also one can observe that several combinations of input variable ranges are less represented. Therefore in addition to removing variables due to correlations, and instances because of outliers, one can also consider to remove instances to make the distributions over all classes and values of the input variables more even. Also a transformation of the input variables can be considered. Also here, the dataset was kept as natural as possible, mainly to see whether the techniques can also cope with this bottleneck or not.

Table 5.6 Correlation matrix of the 24 environmental variables in the river sediments of Flanders dataset. Correlation coefficients with an absolute value of at least 0.20 and lower than 0.50 are marked yellow, higher values are marked in orange.

	Day	**Width**	**Depth**	**Flow velocity**	**Clay**	**Loam**	**Sand**	**Temperature**	**pH**	**Dissolved oxygen**	**Conductivity**	**Organic matter**	**TOXT**	**TOXR**	**TotalP**	**KjeldahlN**	**Cr**	**Pb**	**As**	**Cd**	**Cu**	**Hg**	**Ni**	**Zn**
Day	1.00	0.00	0.06	0.00	-0.23	-0.08	0.15	0.17	-0.19	-0.18	-0.06	-0.16	0.09	-0.13	-0.12	-0.19	-0.03	0.02	0.06	-0.06	0.06	-0.10	0.06	0.04
Width		1.00	0.66	-0.22	0.18	0.10	-0.15	0.19	0.04	0.06	0.00	0.01	0.02	-0.03	0.25	0.11	-0.04	-0.04	0.10	0.11	-0.04	0.21	0.05	0.02
Depth			1.00	-0.20	0.25	0.07	-0.15	0.17	0.06	-0.11	0.02	0.05	0.04	-0.01	0.11	0.18	0.05	0.02	0.15	0.05	0.06	0.07	0.12	0.07
Flow velocity				1.00	-0.34	0.00	0.14	0.11	0.05	0.07	-0.08	-0.14	-0.14	-0.02	-0.10	-0.30	-0.06	-0.01	-0.01	-0.08	0.11	-0.08	0.07	0.04
Clay					1.00	0.48	-0.76	0.01	0.22	-0.04	0.15	0.41	0.10	0.14	0.26	0.72	0.11	0.12	-0.01	0.07	0.09	0.12	0.21	0.17
Loam						1.00	-0.93	0.05	0.38	-0.08	0.04	0.31	0.06	0.14	0.17	0.45	0.01	0.12	-0.16	-0.01	0.10	0.10	0.19	0.10
Sand							1.00	-0.04	-0.37	0.08	-0.09	-0.40	-0.09	-0.16	-0.23	-0.62	-0.05	-0.13	0.12	-0.02	-0.11	-0.12	-0.23	-0.14
Temperature								1.00	0.15	-0.26	0.06	-0.11	0.09	-0.06	-0.05	-0.06	0.00	-0.02	-0.04	-0.06	0.04	-0.08	0.07	0.06
pH									1.00	-0.07	0.09	0.05	0.12	0.05	-0.02	0.20	0.01	-0.03	-0.22	-0.07	-0.03	-0.01	-0.01	-0.05
Dissolved oxygen										1.00	-0.07	-0.13	-0.26	-0.00	-0.22	-0.16	-0.10	-0.11	0.03	-0.05	-0.05	-0.14	-0.08	-0.12
Conductivity											1.00	0.08	0.03	0.00	0.13	0.15	0.15	0.04	0.27	0.04	0.03	0.05	0.04	0.11
Organic matter												1.00	0.04	0.04	0.24	0.68	0.10	0.15	0.04	0.09	0.17	0.11	0.22	0.24
TOXT													1.00	0.20	0.01	0.14	-0.01	0.00	-0.06	-0.04	-0.01	-0.03	0.03	0.01
TOXR														1.00	-0.01	0.08	-0.02	0.01	-0.05	-0.02	-0.01	-0.02	0.00	0.01
TotalP															1.00	0.39	0.11	0.20	0.24	0.45	0.20	0.81	0.34	0.41
KjeldahlN																1.00	0.19	0.23	0.00	0.13	0.24	0.15	0.32	0.35
Cr																	1.00	0.11	0.03	0.04	0.14	0.02	0.14	0.27
Pb																		1.00	0.12	0.32	0.72	0.08	0.60	0.66
As																			1.00	0.22	0.04	0.15	0.03	0.15
Cd																				1.00	0.19	0.19	0.25	0.44
Cu																					1.00	0.05	0.80	0.74
Hg																						1.00	0.24	0.26
Ni																							1.00	0.73
Zn																								1.00

Table 5.7 Correlation matrix of the 24 environmental variables of the Zwalm river basin dataset. Correlation coefficients with an absolute value of at least 0.20 and lower than 0.50 are marked yellow, higher values are marked in orange.

	Width	**Banks**	**Meandering**	**Pool/Riffle**	**Hollow banks**	**Depth**	**Flow velocity**	**pH**	**Temperature**	**Dissolved oxygen**	**Conductivity**	**Suspended solids**	**Ammonium**	**Nitrate**	**Total nitrogen**	**Phosphate**	**Total phosphorus**	**COD**	**Boulders**	**Gravel**	**Sand**	**Loam/clay**	**Distance to mouth**	**Stream order**
Width	1.00	-0.14	0.12	0.28	0.19	0.74	0.05	0.20	-0.07	0.07	0.03	0.04	-0.18	0.01	-0.15	0.02	0.03	-0.04	-0.02	-0.14	-0.06	0.22	-0.51	0.76
Banks		1.00	0.42	0.40	0.52	-0.13	0.01	-0.01	-0.04	0.03	0.00	-0.10	0.09	-0.11	0.05	0.12	0.12	-0.03	0.05	0.01	-0.24	-0.19	-0.07	-0.05
Meandering			1.00	0.55	0.67	0.11	-0.01	-0.04	-0.27	0.05	0.25	-0.17	-0.02	-0.02	0.09	0.05	0.06	-0.13	0.10	-0.10	-0.31	-0.02	-0.38	0.03
Pool/Riffle				1.00	0.52	0.30	0.02	-0.06	-0.17	-0.05	-0.13	-0.18	-0.11	-0.17	-0.14	-0.02	-0.02	-0.10	-0.05	-0.14	-0.05	-0.04	-0.17	0.30
Hollow banks					1.00	0.12	-0.04	0.01	-0.30	0.05	0.15	-0.22	0.07	0.08	0.08	0.06	0.06	-0.12	-0.01	0.02	-0.19	-0.02	-0.38	0.09
Depth						1.00	-0.05	0.04	-0.16	0.04	0.01	-0.04	-0.12	-0.03	-0.11	-0.02	-0.01	-0.03	-0.14	-0.12	0.09	0.27	-0.40	0.57
Flow velocity							1.00	0.31	-0.46	0.43	-0.06	0.16	-0.03	0.04	-0.11	-0.10	-0.09	-0.01	0.43	-0.10	-0.31	-0.25	-0.20	0.23
pH								1.00	-0.27	0.47	0.08	0.04	-0.12	0.15	-0.08	-0.12	-0.12	-0.15	0.11	-0.01	-0.10	-0.05	-0.27	0.16
Temperature									1.00	-0.80	-0.18	-0.24	-0.16	-0.24	-0.15	-0.06	-0.08	-0.12	-0.10	0.00	0.04	0.07	0.21	0.03
Dissolved oxygen										1.00	-0.16	0.12	-0.19	0.29	-0.24	-0.36	-0.39	-0.10	0.12	0.11	-0.19	-0.20	-0.21	-0.09
Conductivity											1.00	-0.26	0.30	0.41	0.52	0.28	0.26	-0.18	0.08	-0.30	-0.21	0.22	-0.63	-0.12
Suspended solids												1.00	0.05	-0.07	0.00	0.19	0.21	0.55	-0.03	0.08	0.01	0.02	0.09	-0.12
Ammonium													1.00	0.09	0.43	0.40	0.44	0.26	-0.10	-0.18	-0.06	0.17	-0.13	-0.27
Nitrate														1.00	0.29	-0.10	-0.13	-0.17	0.04	-0.09	-0.17	0.08	-0.44	-0.15
Total nitrogen															1.00	0.80	0.77	0.10	0.06	-0.14	-0.18	0.09	-0.23	-0.29
Phosphate																1.00	0.96	0.39	0.02	-0.10	-0.10	0.06	-0.13	-0.06
Total phosphorus																	1.00	0.40	0.01	-0.11	-0.10	0.09	-0.14	-0.06
COD																		1.00	-0.07	0.13	0.02	-0.04	0.16	-0.10
Boulders																			1.00	-0.22	-0.61	-0.53	-0.10	0.18
Gravel																				1.00	-0.10	-0.18	0.27	-0.15
Sand																					1.00	0.16	0.22	-0.07
Loam/clay																						1.00	-0.23	-0.01
Distance to mouth																							1.00	-0.34
Stream order																								1.00

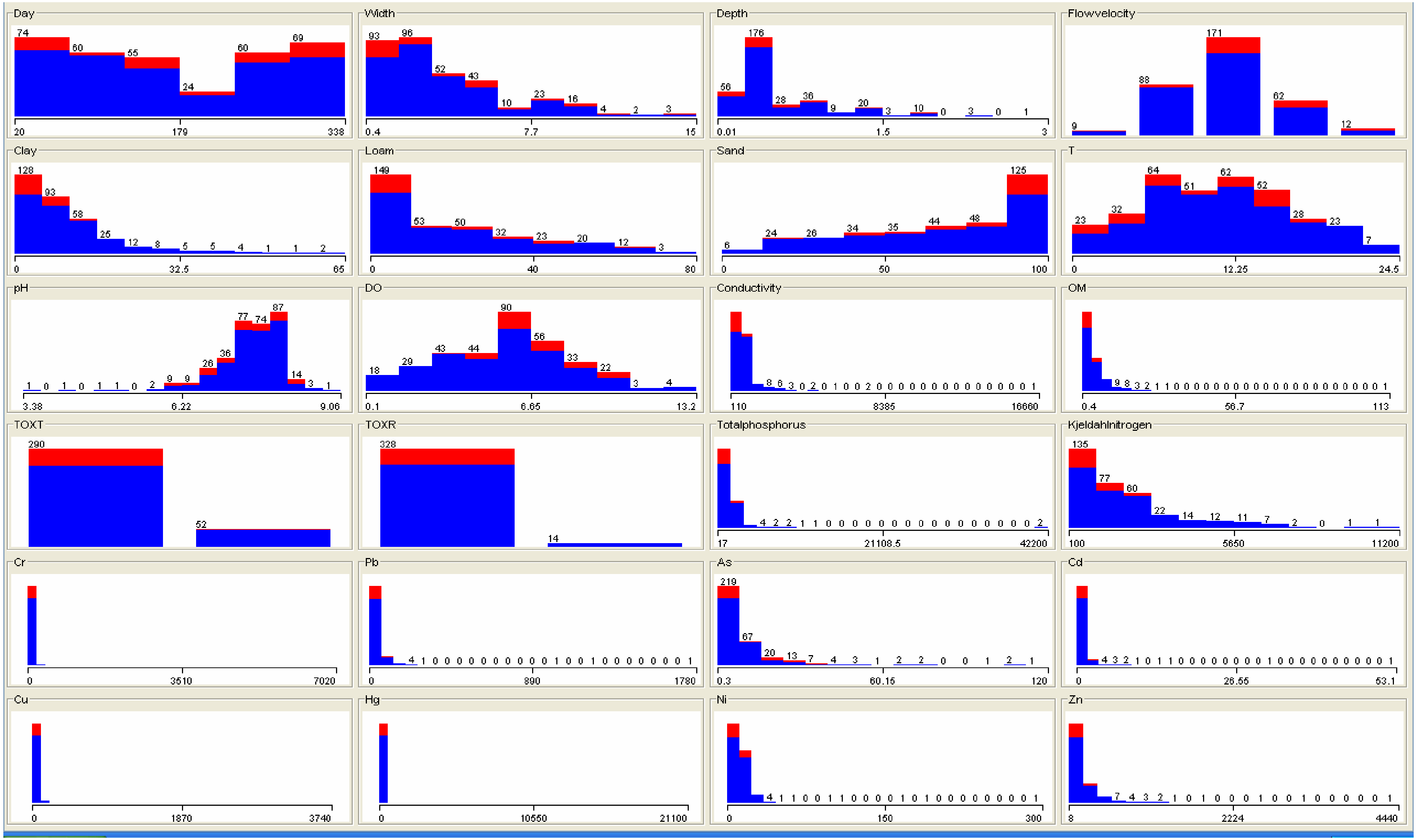


Figure 5.3 Data relation visualisation graphs for Gammarus presence/absence in the river sediments of Flanders (in total 342 instances) in relation to the 24 environmental variables (Gammarus absent in 290 instances (blue), Gammarus present in 52 instances (red)).

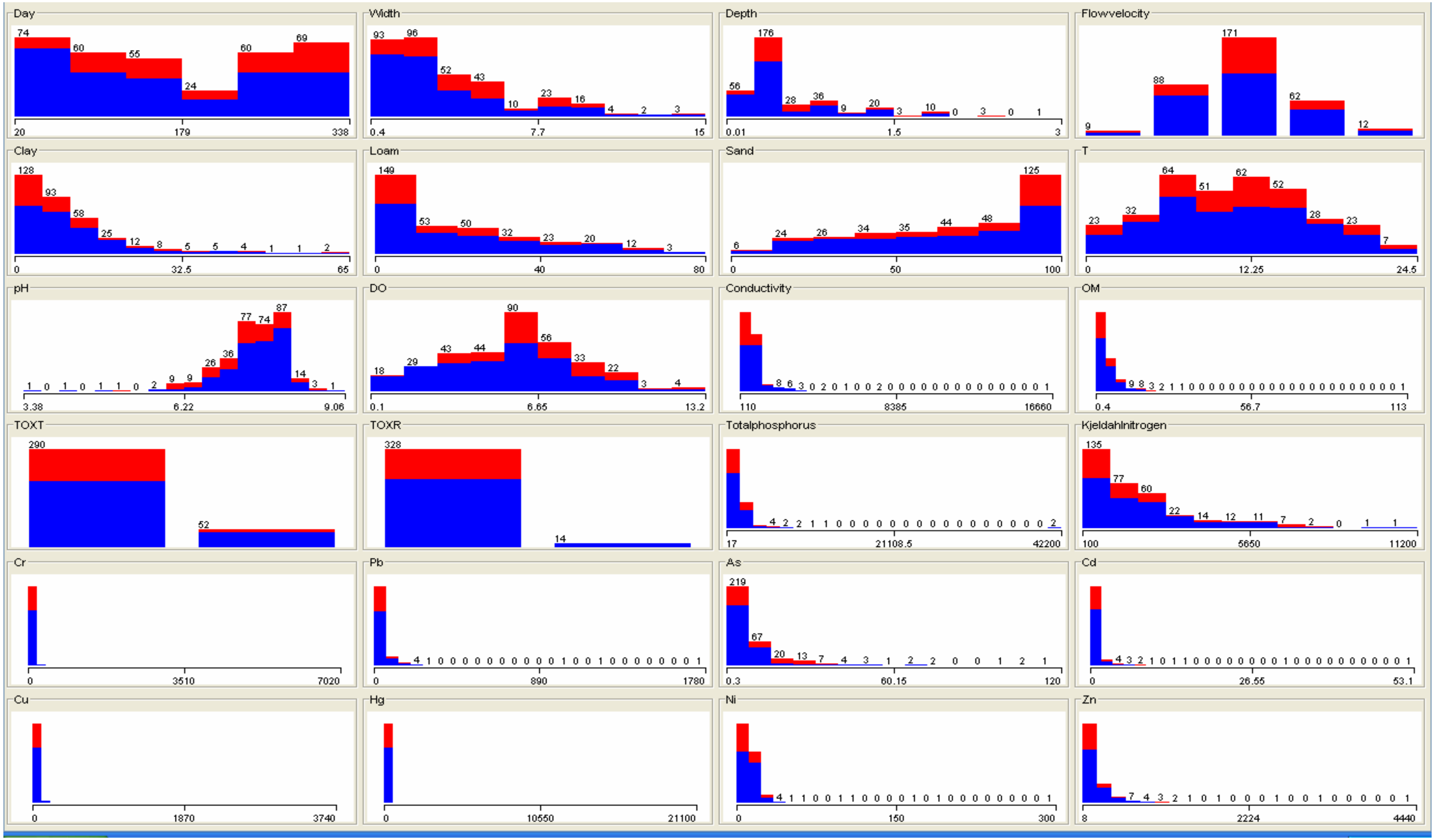


Figure 5.4 Data relation visualisation graphs for Asellus presence/absence in the river sediments of Flanders (in total 342 instances) in relation to the 24 environmental variables (Asellus absent in 239 instances (blue), Asellus present in 103 instances (red)).

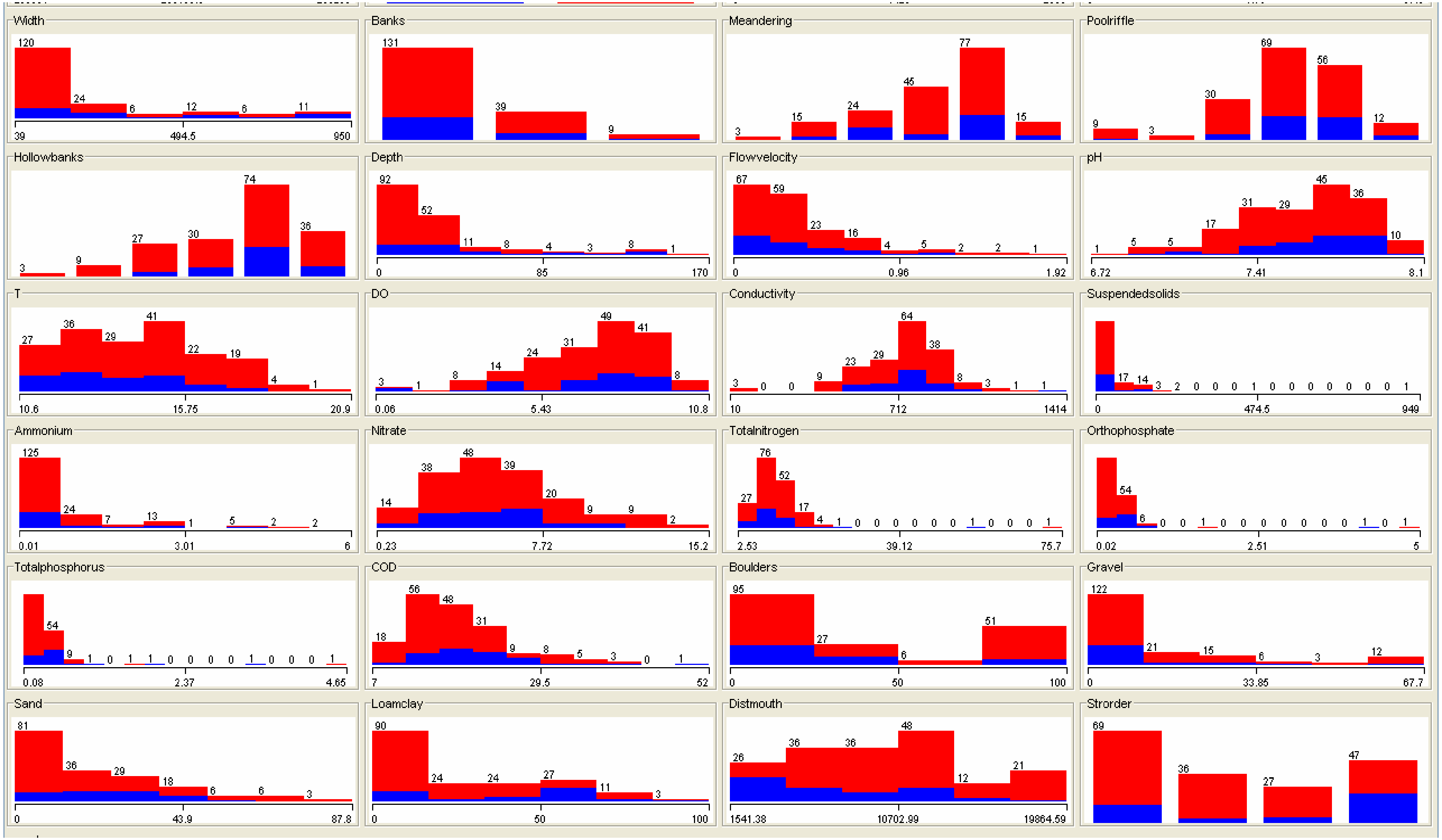


Figure 5.5 Data relation visualisation graphs for Gammarus presence/absence in Zwalm river basin (in total 179 instances) in relation to the 24 environmental variables (Gammarus absent in 43 instances (blue), Gammarus present in 136 instances (red)).

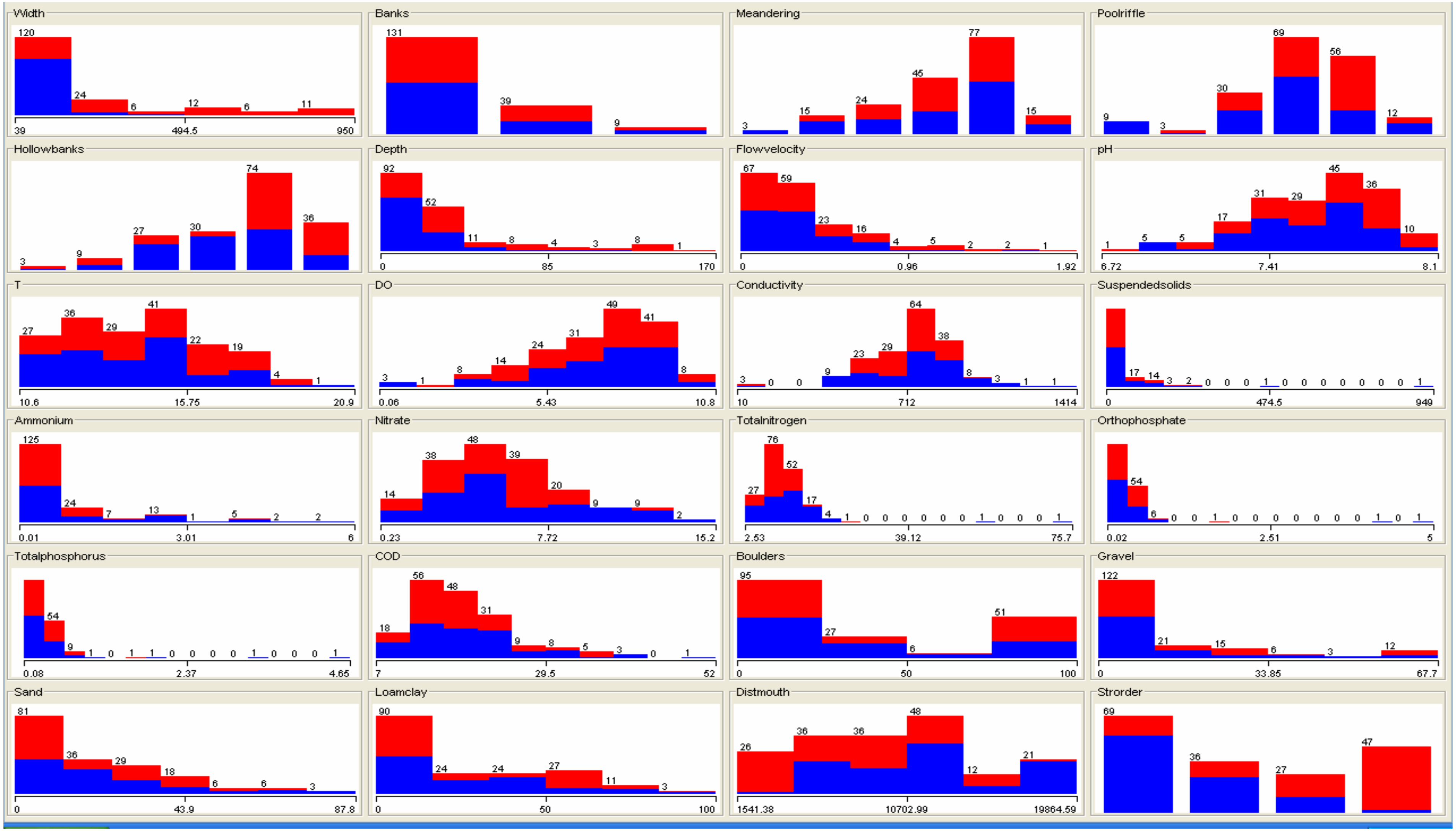


Figure 5.6 *Data relation visualisation graphs for Asellus presence/absence in Zwalm river basin (in total 179 instances) in relation to the 24 environmental variables (Asellus absent in 93 instances (blue), Asellus present in 86 instances (red)).*

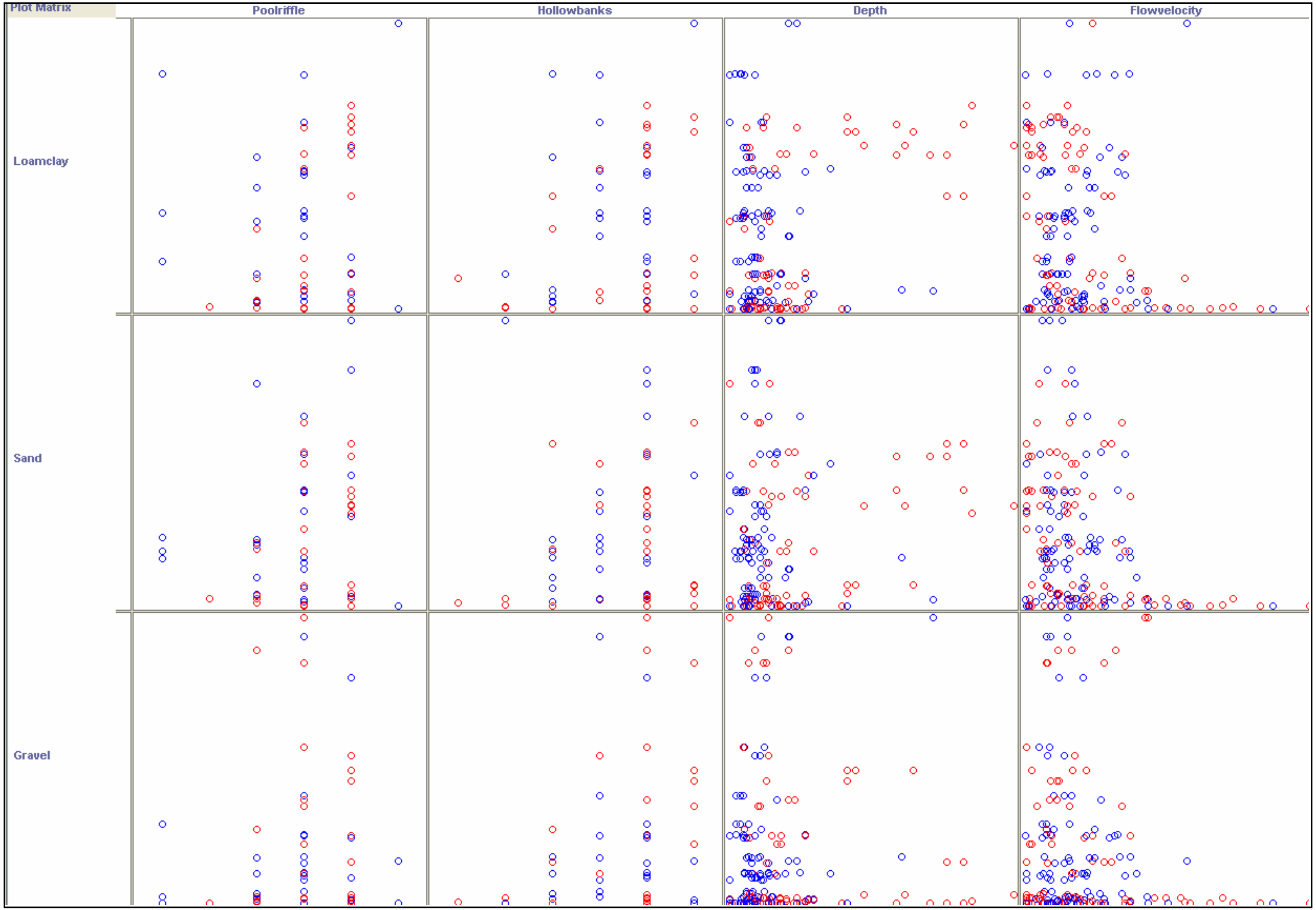


Figure 5.7 Example of a set of detailed data relation graphs (scatter plots) for Asellus (Asellus present (red), Asellus absent (blue)) in the Zwalm river basin (in total 179 instances) for a selection of physical habitat variables.

5.3.4 Discussion

A prior and very important selection not earlier mentioned in this part of the data analysis and variables/instances selection is the data collection itself. Missing crucial values resulted in the elimination of some instances. For the river sediment database, in 18 instances crucial variables were missing. Therefore only 342 out of 360 instances were used for the data driven model development. In case of the Zwalm, at one site the artificial substrates disappeared, and no biological measurement was available. So 179 instead of 180 instances could be used. But also the reason behind the data collection played a key role. As mentioned in the previous chapter on data collection, both data bases were constructed in a very different manner and for another purpose. The first one was built to develop an indicator system (TRIAD methodology). Several interesting variables were missing (e.g. variables on physical habitat) and some variables (e.g. toxicity tests) were probably not representative (recently another toxicity test was therefore introduced, directly done on the sediment and not on the extracted pore water). The sampling strategy differed a lot from the second one in the Zwalm river basin. The latter database and sampling strategy was developed with the purpose of building habitat suitability models, and also *a priori* knowledge from field campaigns played a major role in the selected variables (and even the selection of the river basin). Nevertheless, also financial and time limits played a major role why certain variables (metals, organic micropollutants, variables important for bioavailability calculations, certain hydraulic measurements) were not included in the latter database (or on the manner the measurements were performed). In addition, this also influenced the amount of instances (60 field observations per year was the absolute maximum with this set of variables). Field knowledge can be very helpful to remove variables afterwards, as was seen during this exercise, when one had to decide what to do with the ‘outliers’. Field knowledge also helps to identify what variables play a major role on the ecosystems and can be important for river managers.

No standard procedures for preliminary data analyses were described in most articles reviewed in Chapter 3. Nevertheless, this analysis and related filtering of data is probably very important for the performance of the models, from a theoretical (performance indicators) and practical point of view (such as ecological relevance of the models and their use for different types of simulations).

The number of variables used in both databases (24) is relatively high compared to most articles in the review presented in Chapter 3, where the number of input variables ranged from 3 to 39, usually between 5 and 15. Several theoretical reasons to remove variables and instances can be given, but also quite some practical reasons to keep them in (as part of the research on the data driven techniques, but also for the practical simulations). In this PhD research they are all remained, mainly as part of the research to see how the data driven methods cope with these obstacles and are able to overcome related problems. According to several authors (e.g. Maier and Dandy (2000)), data driven approaches, such as ANN models, have the ability to determine which model inputs are critical. However, the question remains whether they can cope with outliers and redundant variables in the meantime.

In addition, no transformations were made on the input variables, not for numerical reasons (distributions), nor for ecological reasons (e.g. calculation of bio-available fraction of metals by compensation for clay and organic materials). In this PhD research, it is analysed whether these classification trees and ANN themselves can make the necessary inferences, because two major variables for bioavailability were provided as input variables along with the metals.

5.3.5 Conclusions

Several factors played a key role in the final set of variables (and the amount of instances) that are presented to the data driven model development methods. The first set of (practical) factors was the purpose of the data collection, the knowledge on how to measure different aspects of the ecosystem, financial (and time) constraints and also measurement problems. The dataset on the river sediments in Flanders was not developed for habitat suitability modelling, resulting in a lack of some interesting variables (however, during the last years these are included and will lead to interesting new data driven model development studies) for that purpose. The second dataset was built with the aim of model development, but here time and financial budget constraints were encountered. Also knowledge on particular measurement methods for this dataset (e.g. new methods for hydraulic measurements were included, taken maxima and minima into account for instance, as will be shown in the next chapter on the applications) increased during the years. A fifth (theoretical) factor was related to the numerical characteristics of the variables and these were tested with some analysis techniques. Several theoretical arguments appeared to remove variables and instances, but also some practical reasons to keep them in (as part of the research on the data driven techniques, but also for the practical simulations) as well. In this PhD research they were finally all kept in and no transformations of the input variables were done, mainly as part of the research to see how the data driven methods cope with these numerical obstacles and are able to overcome related problems.

5.4 Development of predictive habitat suitability models based on classification tree methods for *Gammarus* and *Asellus* in river sediments in Flanders

5.4.1 Introduction

This component of the results deals with the development of predictive habitat suitability models based on classification tree methods for *Gammarus* and *Asellus* in river sediments in Flanders. First the evaluation based on performance indicators is presented, followed by an overview of the selected variables and their ranking of importance. Finally also an ecological interpretation and discussion of practical use of the induced classification trees is given.

The trees are all presented in the Appendices 1-24. These are ordered per taxon (*Gammarus* and *Asellus*). For each of the three subsets, several pruning confidence factors (0.5, 0.25, 0.1 and 0.01) were tested.

5.4.2 Evaluation based on performance indicators

In Figures 5.8 and 5.9, the best performing trees (according to the CCI and *K* as marked in yellow in Table 5.9) are presented respectively for *Gammarus* and *Asellus*. As can be deducted from Table 5.8, these trees do merely take a small set of the 24 input variables into account, and trees with an intermediary pruning confidence factor seem to perform best. This is a well known general phenomenon being amongst others described by Witten and Frank (2000). These authors confirm that simple classification trees can perform better than complex ones and can in the meantime make more sense as well. This will also be further discussed in the ecological interpretation and practical application part.

```
Clay <= 11
|  DO <= 6: 0 (82.0/8.0)
|  DO > 6
|  |  Width <= 0.75: 1 (4.0)
|  |  Width > 0.75
|  |  |  Day <= 135: 0 (29.0/2.0)
|  |  |  Day > 135
|  |  |  |  Flowvelocity = 0: 0 (2.0/1.0)
|  |  |  |  Flowvelocity = 1: 0 (6.0/1.0)
|  |  |  |  Flowvelocity = 2
|  |  |  |  |  Pb <= 0: 1 (9.0/1.0)
|  |  |  |  |  Pb > 0
|  |  |  |  |  |  T <= 5.1: 1 (3.0)
|  |  |  |  |  |  T > 5.1: 0 (10.0)
|  |  |  |  Flowvelocity = 3
|  |  |  |  |  pH <= 7.13: 0 (3.0)
|  |  |  |  |  pH > 7.13: 1 (5.0/1.0)
|  |  |  |  Flowvelocity = 4: 1 (4.0)
Clay > 11: 0 (71.0)
```

Figure 5.8 Classification tree for Gammarus based on the river sediments in Flanders (Subset 3, PCF=0.25). (0 = Gammarus absent; 1 = Gammarus present; values between brackets indicate instances in which rules are true/false)

```
DO <= 2.7: 0 (35.0)
DO > 2.7
|  As <= 16.7
|  |  Day <= 273: 0 (122.0/27.0)
|  |  Day > 273
|  |  |  Clay <= 12
|  |  |  |  Width <= 3
|  |  |  |  |  Depth <= 0.2: 1 (3.0)
|  |  |  |  |  Depth > 0.2
|  |  |  |  |  |  Day <= 280: 1 (6.0/1.0)
|  |  |  |  |  |  Day > 280: 0 (15.0/2.0)
|  |  |  |  Width > 3: 1 (11.0)
|  |  |  Clay > 12: 0 (5.0)
|  As > 16.7
|  |  Conductivity <= 939.999998
|  |  |  Conductivity <= 250: 0 (3.0)
|  |  |  Conductivity > 250: 1 (23.0/2.0)
|  |  Conductivity > 939.999998: 0 (5.0)
```

Figure 5.9 Classification tree for Asellus based on the river sediments in Flanders (Subset 3, PCF=0.1). (0 = Asellus absent; 1 = Asellus present; values between brackets indicate instances in which rules are true/false)

In case of *Gammarus*, twelve leaves and eight nodes (tree size of 20) gave the best result (CCI = 90.4 and K = 0.64). These values indicate that reliable models were learned (CCI > 70 and K > 0.40). For *Asellus*, ten leaves and nine nodes (tree size of 19) was the best outcome among the three subsets. Based on the CCI this results seems satisfying (CCI = 71.9), however the Cohen's kappa (*K* = 0.28) indicates that this high CCI is for a major part related to the relatively low prevalence (*Asellus* absent in 239 instances, *Asellus* present in 103 instances) in the dataset, and the related ease to make good qualifications, even without the extraction of information from the environmental variables. This directly illustrates the convenience of using two performance indicators.

Table 5.8 illustrates the effect of the pruning algorithm on the tree size, and Table 5.9 the relation with the tree performance based on CCI and *K*. The pruning has a tremendous effect on the size and related complexity of the trees: in case of *Gammarus* the tree size drops from 28 to 5 in average, and for *Asellus* from 45 to 19 (SS Average row indicated in blue in Table 5.8). Nevertheless the CCI and *K* seem to stay rather constant under different pruning levels, with one striking exception at PCF = 0.01 in subset 3 (indicated in orange in Tables 5.7 and 5.8) where the *K* drops to zero in subset 3, and no tree is induced (tree size is equal to one, meaning that the merely the 'nonsense' rule '*Gammarus* is absent' was induced). This means that a CCI of 85.1 % can be obtained without using any information from the environmental variables. This is again a consequence of the relative high absence of *Gammarus* in most sites (in 290 instances out of 342 *Gammarus* is absent).

Table 5.8 Tree size of the induced classification trees (for different pruning confidence factors (PCFs)) for Gammarus and Asellus based on the river sediments in Flanders database.

Gammarus						
Tree size	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	28	28	11	7	19	11
Subset 2	32	15	7	7	15	12
Subset 3	24	20	20	1	16	10
SS Average	28	21	13	5	17	10
SS Stdev	4	7	7	3		
Asellus						
Tree size	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	50	38	34	19	35	13
Subset 2	36	36	32	19	31	8
Subset 3	48	46	19	19	33	16
SS Average	45	40	28	19	33	12
SS Stdev	8	5	8	0		

When analyzing the three different folds, the effect of the pruning seems to vary quite a lot per subset, however, when comparing the averages (PCF Average columns), the average tree size and their average reliability expressed as CCI and *K* seem to be very stable (except for the *K* in subset 3 of *Gammarus*). Based on the average CCI, the trees for both *Gammarus* and *Asellus* seem to be reliable, however when analysing the *K*, the trees do not meet the threshold value of 0.4, indicating that the trees are not that reliable (average values indicated in green). The best pruning is obtained at a level of 0.1 for both taxa, resulting in a *K* = 0.34 for *Gammarus* and *K* = 0.28 for *Asellus*.

Table 5.9 Performance of the induced classification trees (for different pruning confidence factors (PCFs)) for Gammarus and Asellus based on the river sediments in Flanders database.

Gammarus						
CCI	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	81.6	81.6	83.3	86.0	83.1	2.1
Subset 2	75.4	76.3	82.5	82.5	79.2	3.9
Subset 3	89.5	90.4	90.4	85.1	88.9	2.5
SS Average	82.2	82.8	85.4	84.5	83.7	1.5
SS Stdev	7.1	7.1	4.3	1.8		
K	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	0.22	0.22	0.21	0.32	0.24	0.05
Subset 2	0.16	0.13	0.19	0.19	0.17	0.03
Subset 3	0.62	0.63	0.63	0.00	0.47	0.31
SS Average	0.33	0.33	0.34	0.17	0.29	0.08
SS Stdev	0.25	0.27	0.25	0.16		
Asellus						
CCI	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	65.8	67.5	70.2	71.9	68.9	2.7
Subset 2	69.3	69.3	66.7	64.9	67.6	2.2
Subset 3	59.6	58.8	71.9	71.9	65.6	7.3
SS Average	64.9	65.2	69.6	69.6	67.3	2.6
SS Stdev	4.9	5.6	2.7	4.0		
K	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	0.10	0.12	0.21	0.18	0.15	0.05
Subset 2	0.18	0.18	0.13	0.16	0.16	0.02
Subset 3	0.15	0.14	0.28	0.28	0.21	0.08
SS Average	0.14	0.15	0.21	0.21	0.18	0.04
SS Stdev	0.04	0.03	0.08	0.06		

5.4.3 Ranking of importance of the input variables combined with ecological interpretation and discussion of practical use of the classification tree models

As a first step towards the ecological interpretation and to ease the comparison with the ANN input variable contribution methods, a ranking of the input variables for both *Gammarus* and *Asellus* was made in Table 5.10. These are based on the trees presented in the Appendices 1-24.

The applied procedure comes down to the following: the variables met at each splitting level are written down in each column (only the first five levels are considered) for each subset separately (this is performed on the highest PCF level, however the major variables stay in general constant over the different PCF values, only the amount of considered input variables shrinks by reducing the PCF). When a variable occurs at several levels, only the first level at which the variable is used is considered (in other words, each variable occurs only once in each row of the table). Finally, the outcomes of the three subsets are combined to see how stable the result is. For this, the variables occurring one, two and three times at each level are noted in Table 5.10.

Table 5.10 Major variables of the induced classification trees for Gammarus and Asellus based on the river sediments in Flanders database.

Gammarus					
Variables	**First**	**Second**	**Third**	**Fourth**	**Fifth**
Subset 1	Clay	EC	Width	Day	FV
Subset 2	Pb	Day, Clay	Depth, As	TOXT, Pb, FV	T
Subset 3	Clay	DO	Loam, Width	Ni, Day	FV
Three times	-	-	-	-	-
Twice	Clay	-	Width	Day	FV
Once	Pb	EC, Day, Clay, DO	Depth, As, Loam	TOXT, Pb, FV, Ni	T
Asellus					
Variables	**First**	**Second**	**Third**	**Fourth**	**Fifth**
Subset 1	DO	TOXR	Day	Cd, Width	FV, Clay
Subset 2	DO	Day	Width	Clay	KjeldahlN, Cd
Subset 3	DO	TOXR	As	Day, EC	Depth, Clay
Three times	DO	-	-	-	-
Twice	-	TOXR	-	-	Clay
Once	-	Day	Day, Width, As	Cd, Width, Clay, Day, EC	FV, KjeldahlN, Cd, Depth

Based on the outcomes of Table 5.10, one can conclude that the major variables for *Gammarus* are clay, Pb, conductivity, day and dissolved oxygen, while for *Asellus* these are dissolved oxygen, toxicity test TOXR, day, width and As. The results over the three subsets seem to be very instable. Only dissolved oxygen is used three times as major variable for *Asellus*. So although the results regarding the reliability seemed rather constant over the three folds, the used information to obtain the classification trees was very different. This means that very dissimilar information (input variables) can be used to explain the presence/absence of both taxa. In general, when analyzing the trees in Figures 5.8 and 5.9, the rules are in general confirming existing ecological knowledge. But several exceptional variables and relation popped up as well.

From an ecological point of view the dissolved oxygen for both *Gammarus* (Wesenberg-Lund, 1982) and *Asellus* (although the latter being relatively tolerant according to Verdonschot (1990), but the extreme low concentrations in some of the very polluted waters were probably too stressful) seems logical and was confirmed by D’heygere et al. (2003 and 2004) as well when applying genetic algorithms in combination with classification trees and ANNs. Rather strange is the effect of the day on the presence/absence. In case of *Asellus* it seems that the populations like to peak during the warmer periods. At first it looked strange that the toxicity test with alga (TOXR) was selected above the one with the crustacean (TOXT) for *Asellus*, but when discussing with toxicological experts involved in the analyses, it seemed that the TOXT tests were often giving a ‘false’ toxic signal as a result of ammonium formation in the sediments. This was not the case for the TOXR test, as such this test is probably more reliable, even although it concerns another type of organism. The important effect of clay is not easy to explain, nor the impact of the two metals. Metal accumulation in many benthic organisms appears to correlate more with concentrations in overlying water rather than those in the solid phase according to several authors (Deaver and Rodgers, 1996; Warren et al., 1998; Hare et al., 2001). However, this type of measurements is lacking in the dataset. Metal concentrations in overlying water are, therefore, another important indicator of metal bioavailability and potential effects. A full assessment of the environmental impact of metals should include measurement of metal bioaccumulation and metal concentrations in overlying water in carefully conducted sediment toxicity tests. When combined with more traditional measures of sediment toxicity, sediment chemistry and benthic community structure, this can provide a clearer picture of metal contamination, metal bioavailability, toxic effects and the causative agent(s) (Borgmann et al., 2001b).

5.4.4 Conclusions

Based on the average CCI (over all pruning levels), the trees for both *Gammarus* and *Asellus* seem to be reliable (respectively 83.7% and 67.3%), however when analysing the average *K* (respectively 0.29 and 0.18), the trees are not meeting the threshold value of 0.4, indicating that the trees' performance is mainly related to the relative low prevalence of both taxa in the database and related 'easy' classification, even without using environmental information.

The effect of the pruning seems to vary quite a lot per subset, however, the average tree size and their average reliability expressed as CCI and *K* seem to be very stable. The best pruning is obtained at a level of 0.1 for both taxa, resulting in a *K* = 0.34 for *Gammarus* and *K* = 0.28 for *Asellus*.

When searching for the crucial variables to induce the classification trees, the results seem to be very instable. Based on the three subsets, the major variables for *Gammarus* were clay, Pb, conductivity, day and dissolved oxygen, while for *Asellus* these were dissolved oxygen, toxicity test TOXR, day, width and As. Although the results regarding the reliability seemed rather constant over the three folds, the used information to obtain the classification trees was very different. This means that very dissimilar information (input variables) can be used to explain the presence/absence of both taxa.

5.5 Application of backpropagation artificial neural networks predicting *Gammarus* and *Asellus* in river sediments in Flanders

5.5.1 Introduction

This study aims at analysing the relationship between river sediment characteristics and the presence/absence (and abundance) of the two macroinvertebrate taxa *Asellus* and *Gammarus*. Table 5.11 gives a scheme on how the models were applied on the databases and the results are presented in Appendix 49-60. Six input variable contribution methods were applied on *Gammarus* and *Asellus*, first presence/absence models were used (to allow a comparison with the outcomes of the classification trees), followed by abundances models (actually these models use log(abundance+1) output transformations).

Table 5.11 Overview of the applied contribution methods to the ANN models.

Presence/absence classification of taxa
Performance based on performance indicators Application of input variable contribution methods Weights PaD Perturb Stepwise Reg Stepwise Imp Profile
Prediction taxa abundance
Evaluation based on performance indicators Application of input variable contribution methods Weights PaD Perturb Stepwise Reg Stepwise Imp Profile
Selected variables and ranking of importance ***Ecological interpretation and discussion of practical use of the ANN models***

5.5.2 Predictive performance of the classification and regression ANN models

In Table 5.12 the CCI and *K* of the classification (presence/absence) models is presented. For both *Gammarus* and *Asellus*, the CCI and *K* are strikingly constant over the three folds, with one major exception (indicated in yellow), where the *K* drops to 0.10 in subset 3. The CCI value is good for both taxa (86.3 % and 76.3 % for respectively *Gammarus* and *Asellus*), while the *K* values are more or less acceptable (in the neighbourhood of about 0.4), without the exception mentioned earlier that results in a large standard deviation of the *K* over the three folds (indicated in orange).

Table 5.12 CCI and K for the classification ANN models of Gammarus and Asellus in the Zwalm river basin.

Gammarus	**CCI**	***K***
Subset 1	86.8	0.49
Subset 2	86.0	0.38
Subset 3	86.0	0.10
Average	86.3	0.32
Standard deviation	0.5	0.20
Asellus	**CCI**	***K***
Subset 1	77.2	0.36
Subset 2	77.2	0.44
Subset 3	74.6	0.34
Average	76.3	0.38
Standard deviation	1.5	0.05

In Table 5.13 the performance of the abundance models (regression) trained and validated on the same subsets are presented. In average the *r* values in the validation sets are about 0.4, indicating that the models are rather good. Based on the much higher values in the training sets, one has to conclude that the method seems to over train on the training set. The problems with the *K* in subset three of *Gammarus* were not translated to performance problems with this type of models (yellow row).

Table 5.13 Correlation coefficient r for the regression ANN models of Gammarus and Asellus in the Zwalm river basin.

Gammarus	***r* (validation set)**	***r* (training set)**
Subset 1	0.43	0.87
Subset 2	0.46	0.77
Subset 3	0.46	0.60
Average	0.45	0.75
Standard deviation	0.02	0.14
Asellus	***r* (validation set)**	***r* (training set)**
Subset 1	0.41	0.43
Subset 2	0.45	0.75
Subset 3	0.30	0.42
Average	0.39	0.53
Standard deviation	0.08	0.19

5.5.3 Ranking of importance of the input variables combined with ecological interpretation and discussion of practical use of ANN models

In Tables 5.14 to 5.17, the outcomes of the six different input variable contribution methods for the presence/absence and abundance models of *Gammarus* and *Asellus* are presented. These tables were made on the basis of the results presented in detail in the Appendices 49-60. Based on the figures in the appendices and the four tables, one can deduce that it is difficult to find major trends over the two taxa, the six contribution methods and the three subsets. The first two can be explained by different ecological preferences of the taxa (cf. ecological expert knowledge in previous chapter) and by the different aspects the six contribution methods deal with (e.g. one method analyzes the effect of small changes of the input variables, such as the perturb method, while others like the profile method makes a similar analysis over the whole range). The instability over the different folds is perhaps related with the relative small size (342 instances) of the dataset in combination with the high variability of the sites (whole Flanders) and high number of input variables.

To make an overall ranking of the variables, the sum was made of the ranks per contribution method, and based on this sum, the overall rank was determined in Tables 5.14 to 5.17. As was found by D'heygere et al (2003 and 2004) as well, conductivity and dissolved oxygen play a major role in this database to explain the presence/absence but also the abundance of both taxa. This can be explained by the mixture of sites with and without wastewater treatment during the years of sampling. Based on this, several sites were still in oxygen deficit, while others were already in relatively good condition (also some reference sites were included in the dataset as well). As a result, these oxygen and conductivity gradient were of major influence for both taxa (cf. expert knowledge in previous chapter), but for *Asellus*, the conductivity plays a less important role than for *Gammarus*.

Nutrients and organic matter (total phosphorus, Kjeldahl nitrogen and organic matter variables) seemed to play a key role for both the abundance and presence/absence of *Gammarus*. *Asellus* is indeed known to be less influenced by these variables, and rather the dimension of the streams (in particular) width were recognized by the ANN and contribution methods as major variables. It was however strange that also day played an important role for both types of models. For the abundance models it seems rather logical that seasonal patterns affect the abundance, but that even the presence/absence is dictated by this was rather unexpected, but also appeared in the classification tree models. The effect of the substrate loam and clay had in several cases a major effect. It is not clear whether it concerns a direct effect of the substrate, or rather an indirect one (e.g. relation with bio-availability of metals and other toxic compounds).

The outcome of the heavy metals was anyway much less straightforward and rather instable. Pb and As where both selected as intermediary important for Gammarus and Asellus in all model types, while Ni and Zn appeared solely in the abundance models of Asellus. This could mean that the latter only play a significant role on the abundance levels (e.g. on reproduction, limited mortality, etc.), but do not seem to have a significant lethal effect to extinct a whole population. However, this kind of rules are to be taken with careful consideration. As Borgmann et al. (2004) mention, total metal concentrations in sediments are poor indicators of potential toxic effects because metal bioavailability can vary considerably between sediments (Chapman et al., 1998; Borgmann et al., 2001a). Furthermore, metals are often present in mixtures with other metals and other non-metal contaminants. Determining if toxic effects are due to a metal, and if so which metal, is not possible from simple chemical analysis of sediment. Quantifying metal bioavailability and relating this bioavailability to effects are, therefore, essential for assessing metal impacts. Metal bioaccumulation has been shown to correlate well with toxicity for non-regulated metals in the amphipod *Hyalella azteca* (Borgmann et al., 1991; Borgmann et al., 1998), but body concentrations of Cu and Zn are often independent of environmental concentrations (Borgmann et al., 2001b). Unfortunately, this type of measurements was not routinely done during data collection. Therefore, this can only be checked on more recent databases of river sediments in which *Hyalella* tests were performed. The TOXT and TOXR tests were only selected by the presence/absence model of *Asellus*, giving an indication that in part of the database, toxicity is useful to explain the presence or absence of this taxon. As however indicated, these tests are not that representative, as they are performed on pore water. In addition, the TOXT test is biased due to ammonium influences.

Figures 5.10 and 5.11 present the outcomes of the application of the profile methods on the four models. As could be expected, the results from the two taxa and the two types of models are very different.

Table 5.14 Comparison of the outcome of the six different input variable contribution methods for the Gammarus presence/absence models based on the river sediments in Flanders database.

Gammarus P/A	Weights	PaD	Perturb	Stepwise Reg	Stepwise Imp	Profile	Overal rank	Average rank	Standard deviation
Day	9	9	20	18	9	17	16	14	5.2
Width	10	11	13	17	5	18	12	12	4.8
Depth	16	21	17	24	13	19	21	18	3.9
Flow velocity	8	8	12	22	11	14	14	13	5.2
Clay	7	5	3	21	2	8	5	8	6.9
Loam	15	7	14	10	4	22	11	12	6.4
Sand	13	12	15	16	14	16	17	14	1.6
Temperature	18	14	18	23	15	23	22	19	3.8
pH	4	17	5	12	3	3	4	7	5.8
Dissolved oxygen	3	4	8	8	7	13	3	7	3.5
Conductivity	1	1	1	20	1	1	1	4	7.8
Organic matter	12	3	6	19	6	20	10	11	7.2
TOXT	21	23	21	5	19	15	19	17	6.6
TOXR	20	16	24	1	20	9	18	15	8.5
Total phosphorus	2	2	2	7	8	10	2	5	3.6
Kjeldahl nitrogen	6	10	16	3	12	12	8	10	4.7
Cr	24	20	22	15	23	21	23	21	3.2
Pb	11	6	4	14	16	6	7	10	4.9
As	5	15	10	4	10	5	6	8	4.3
Cd	14	13	7	2	21	2	9	10	7.5
Cu	22	22	19	11	22	11	20	18	5.4
Hg	23	24	23	13	24	24	24	22	4.4
Ni	17	18	11	6	18	4	13	12	6.3
Zn	19	19	9	9	17	7	15	13	5.6

Table 5.15 Comparison of the outcome of the six different input variable contribution methods for the Asellus presence/absence models based on the river sediments in Flanders database.

Asellus P/A	**Weights**	**PaD**	**Perturb**	**Stepwise Reg**	**Stepwise Imp**	**Profile**	**Overal rank**	**Average rank**	**Standard deviation**
Day	2	3	6	9	14	4	2	6	4.5
Width	4	4	2	4	10	18	4	7	6.0
Depth	8	5	9	17	21	23	13	14	7.5
Flow velocity	19	22	17	3	6	6	9	12	8.1
Clay	16	13	11	22	20	13	22	16	4.4
Loam	15	24	3	16	2	14	10	12	8.4
Sand	23	20	22	11	13	24	24	19	5.5
Temperature	6	11	14	2	7	1	3	7	5.0
pH	13	17	18	13	15	8	16	14	3.6
Dissolved oxygen	1	1	7	6	8	12	1	6	4.3
Conductivity	3	2	23	23	19	16	17	14	9.5
Organic matter	10	9	12	24	24	3	12	14	8.5
TOXT	9	8	4	12	5	19	5	10	5.5
TOXR	7	6	5	15	23	15	7	12	7.1
Total phosphorus	11	16	21	20	9	11	20	15	5.1
Kjeldahl nitrogen	14	18	24	5	1	21	14	14	9.1
Cr	21	14	13	1	4	10	6	11	7.2
Pb	22	10	16	18	3	2	8	12	8.2
As	5	12	8	14	18	20	11	13	5.7
Cd	12	7	10	21	11	22	15	14	6.2
Cu	18	21	1	7	22	17	18	14	8.4
Hg	17	19	19	10	16	9	21	15	4.4
Ni	20	15	15	19	12	5	19	14	5.4
Zn	24	23	20	8	17	7	23	17	7.4

Table 5.16 Comparison of the outcome of the six different input variable contribution methods for the Gammarus abundance models based on the river sediments in Flanders database.

Gammarus abundance	**Weights**	**PaD**	**Perturb**	**Stepwise Reg**	**Stepwise Imp**	**Profile**	**Overal rank**	**Average rank**	**Standard deviation**
Day	11	9	16	9	9	23	12	13	5.7
Width	8	22	20	15	15	16	17	16	4.9
Depth	13	15	23	14	14	13	16	15	3.8
Flow velocity	7	6	10	3	3	20	7	8	6.4
Clay	4	5	6	5	5	6	3	5	0.8
Loam	15	10	19	4	4	17	11	12	6.5
Sand	10	19	17	13	13	19	15	15	3.7
Temperature	12	23	21	12	12	21	18	17	5.3
pH	17	11	8	10	10	9	9	11	3.2
Dissolved oxygen	6	4	9	2	2	24	6	8	8.4
Conductivity	1	1	1	1	1	2	1	1	0.4
Organic matter	9	13	5	11	11	5	8	9	3.3
TOXT	19	16	24	18	18	15	21	18	3.1
TOXR	21	12	18	19	19	14	19	17	3.4
Total phosphorus	3	2	2	6	6	1	2	3	2.2
Kjeldahl nitrogen	5	3	4	7	7	7	4	6	1.8
Cr	23	24	22	24	24	18	24	23	2.3
Pb	16	8	3	16	16	8	10	11	5.6
As	2	7	11	8	8	4	5	7	3.2
Cd	22	21	14	22	22	22	23	21	3.2
Cu	20	18	13	21	21	10	20	17	4.6
Hg	24	20	15	23	23	12	22	20	4.9
Ni	18	17	12	20	20	3	14	15	6.6
Zn	14	14	7	17	17	11	13	13	3.8

Table 5.17 Comparison of the outcome of the six different input variable contribution methods for the Asellus abundance models based on the river sediments in Flanders database.

Asellus abundance	**Weights**	**PaD**	**Perturb**	**Stepwise Reg**	**Stepwise Imp**	**Profile**	**Overal rank**	**Average rank**	**Standard deviation**
Day	1	1	4	2	1	1	1	2	1.2
Width	15	13	19	3	8	17	10	13	6.0
Depth	8	20	13	12	19	15	14	15	4.5
Flow velocity	19	16	16	5	24	20	19	17	6.4
Clay	6	5	3	10	4	8	4	6	2.6
Loam	3	4	7	11	10	2	5	6	3.8
Sand	11	23	23	13	16	21	21	18	5.2
Temperature	17	14	14	22	18	14	18	17	3.2
pH	18	8	9	21	6	19	13	14	6.5
Dissolved oxygen	13	7	1	23	7	12	8	11	7.5
Conductivity	14	15	10	14	15	7	11	13	3.3
Organic matter	24	19	22	16	20	13	22	19	4.0
TOXT	12	9	12	20	17	9	12	13	4.4
TOXR	22	12	11	17	14	11	15	15	4.3
Total phosphorus	21	24	20	7	23	23	23	20	6.4
Kjeldahl nitrogen	23	21	24	6	21	5	20	17	8.7
Cr	16	22	18	18	22	24	24	20	3.1
Pb	10	10	8	15	13	6	7	10	3.3
As	5	11	17	9	11	16	9	12	4.5
Cd	20	17	21	4	9	22	17	16	7.3
Cu	9	6	5	24	5	10	6	10	7.3
Hg	7	18	15	19	12	18	16	15	4.6
Ni	4	3	2	8	3	4	3	4	2.1
Zn	2	2	6	1	2	3	2	3	1.8

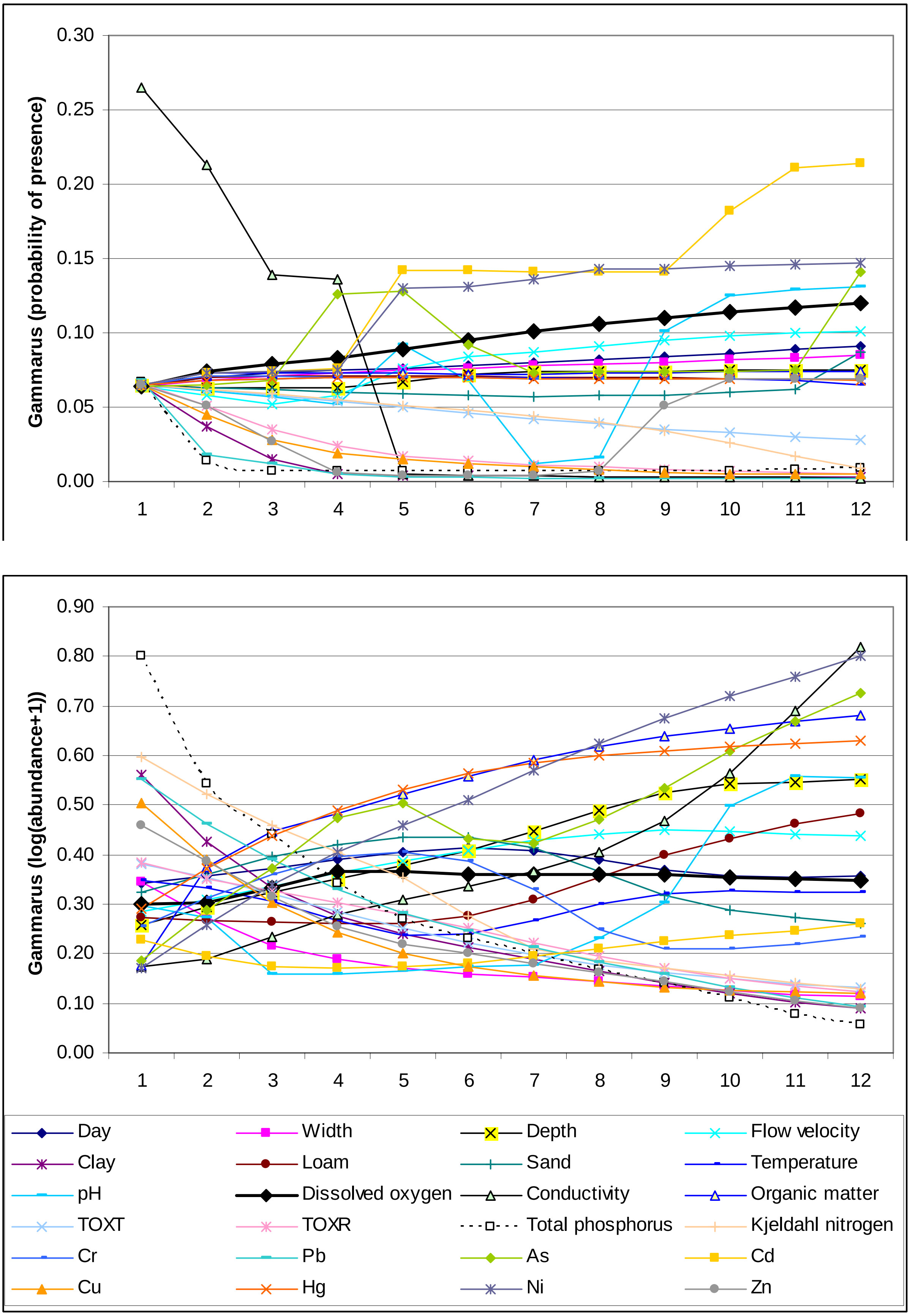


Figure 5.10 Curves for the 24 input variables in relation to the probability of presence (top) and abundance (bottom) of Gammarus based on the profile method and ANNs.

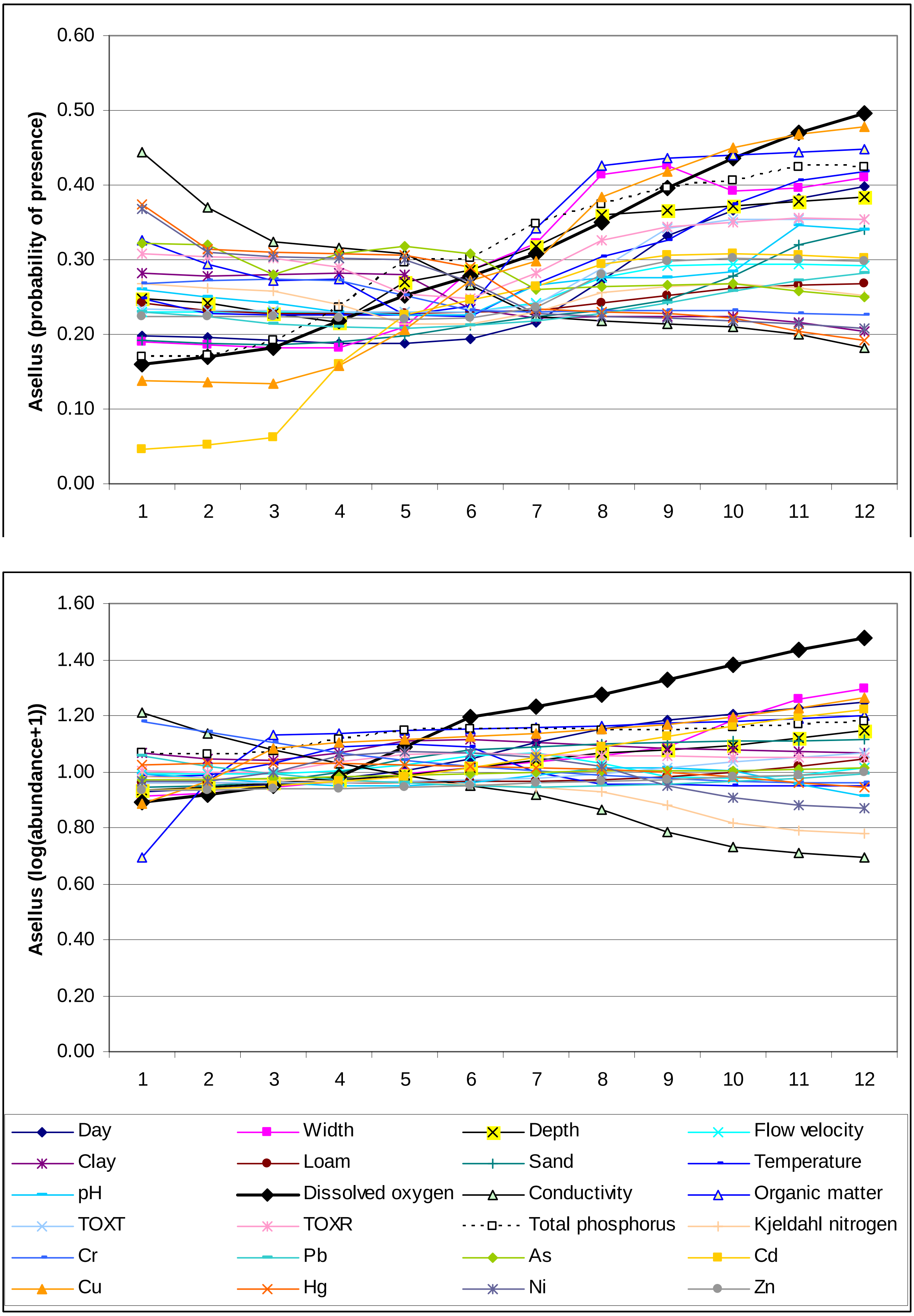
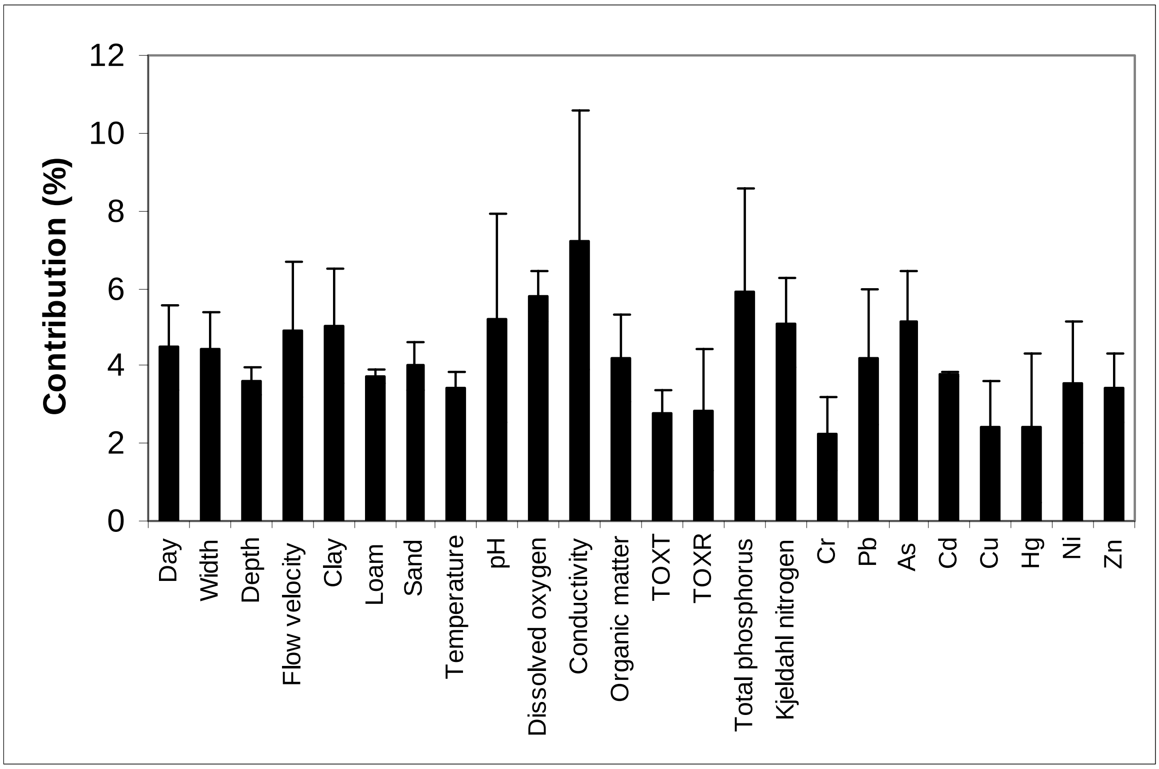


Figure 5.11 Curves for the 24 input variables in relation to the probability of presence (top) and abundance (bottom) of Asellus based on the profile method and ANNs.

The impact of minimum and maximum values of the input variables is very important of course for this curves in particular. The effect of conductivity seems very logic in the presented graphs (a clear negative effect at higher levels, e.g. very well presented in the *Gammarus* presence/absence ANN model graph in Figure 5.10 at the top). However in the bottom chart of the same figure, this relation is inverse and seems very unlikely from a ecological point of view. A major outcome of these figures is that only a few variables (about five) really seem to play a role in the models, and the effect of the other variables seems almost nihil for the predictions (very horizontal curves). This type of graphs is therefore crucial to know how ecologically sound the models are and what their meaning can be for practical simulations. Only when the variables of interest for managers take a crucial part in the predictions, the models are useful and reliable for decision support in river management.

When comparing the other methods, it seems that the ‘PaD’ method distinguishes more clearly minor and major contributing environmental variables in comparison to the ‘Weights’, ‘Perturb’ and ‘Stepwise Reg’ methods (Figures 5.12 to 5.15). Similar results were found by Gevrey et al. (2003). In that study, the relation between environmental variables and trout density was predicted. This is also rather logical, because each method expresses a different aspect of sensitivity or importance of the environmental variables to the presence/absence of the taxa. In case of the weights method, the overall importance of the input variables is taken into account, while for the perturb method only the sensitivity over a part of the range (in this case 50%) is considered. The PaD method makes a classification of the relative contributions of each variable to the network output. The input variable that has the highest SSD value is the variable, which influences the output most. This method is therefore very good to detect the variables of major concern. From these graphs one can see that as concluded from the profile method, only a few variables are really taken into consideration during the calculations. Thus the combination of the PaD method with the profile method gives a very good idea of the ecological meaning of the models and their practical relevance for decision support in river management. The Stepwise Reg approach gives more an idea of the importance of the variables for the overall theoretical reliability of the models (based on the used performance indicator). As such, it is rather not surprising that the outcomes are quite different for the six methods.

However, when looking at the graphs, one can see that dissolved oxygen and conductivity are nearly always the major variables in all graphs. These outcomes are also similar to the previous study with genetic algorithms by D'heygere et al. (2003 and 2004).



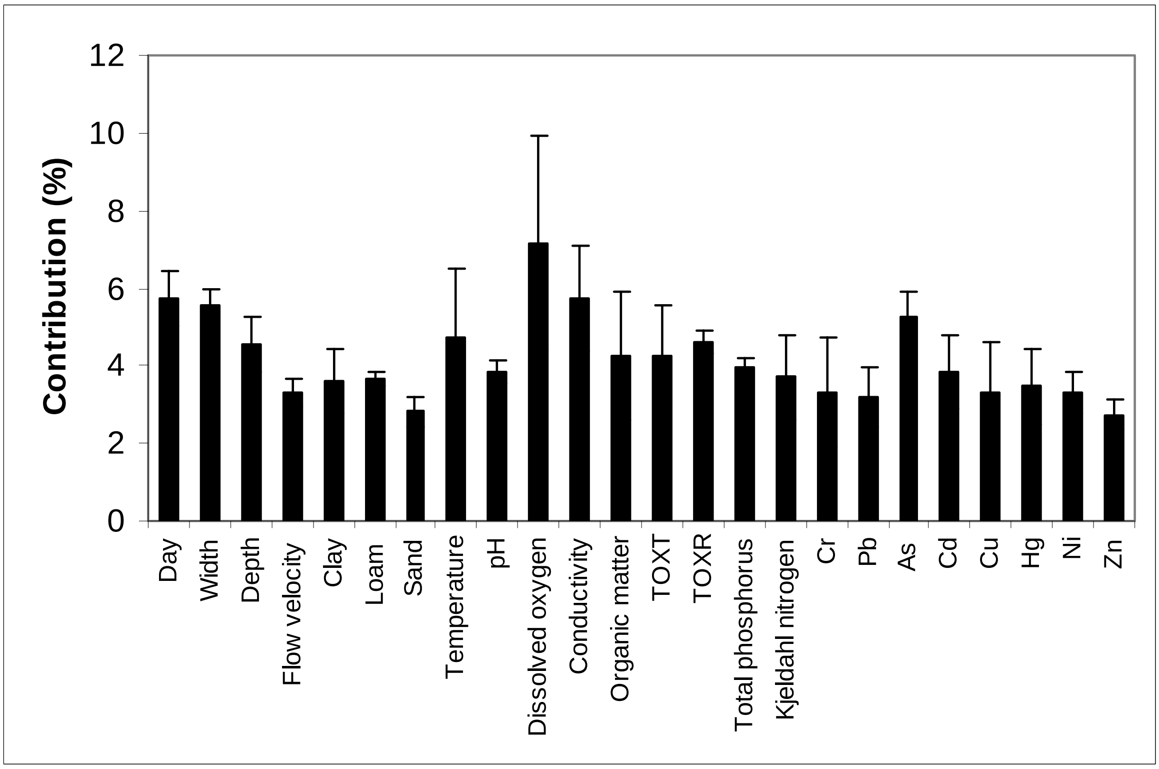
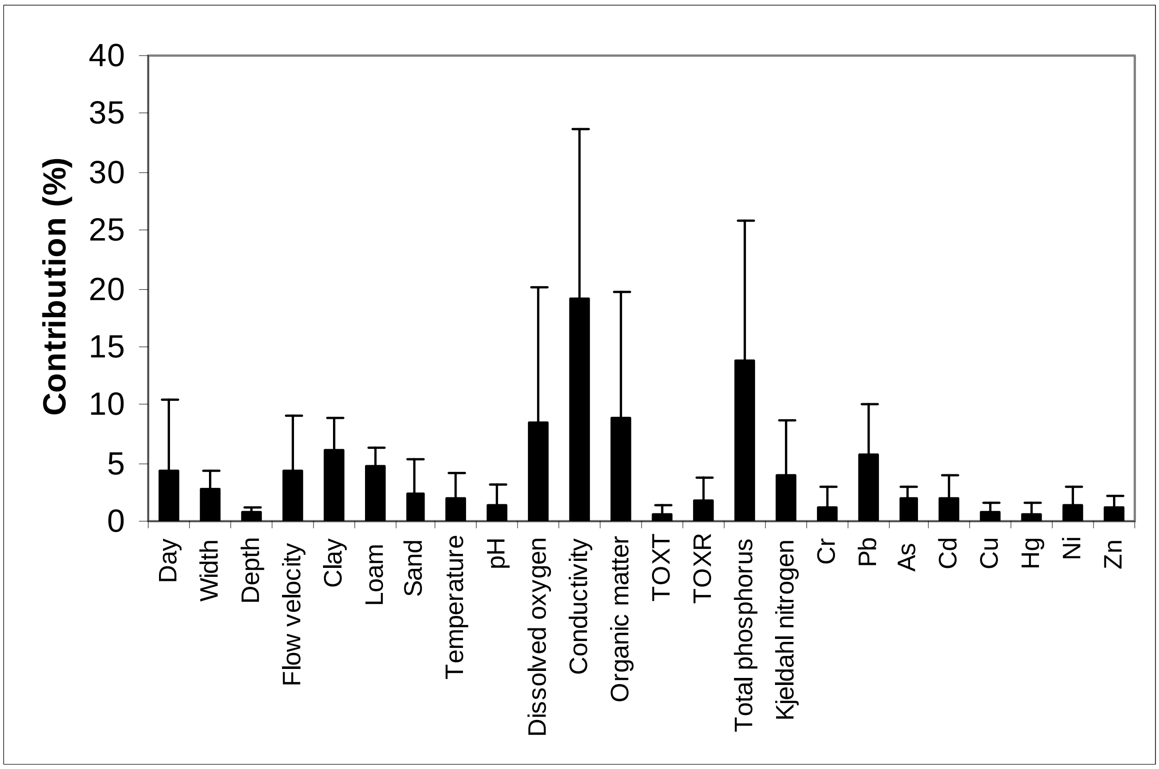


Figure 5.12 Results of the weights method applied on the presence/absence models for Gammarus and Asellus of the river sediments in Flanders dataset.

For *Gammarus*, also total phosphorus and Kjeldahl nitrogen are often important. While the other variables only play a very variable role in general. However, the outcomes are not very constant over all the different subsets, as can be deduced from the high standard deviation flags in the graphs.



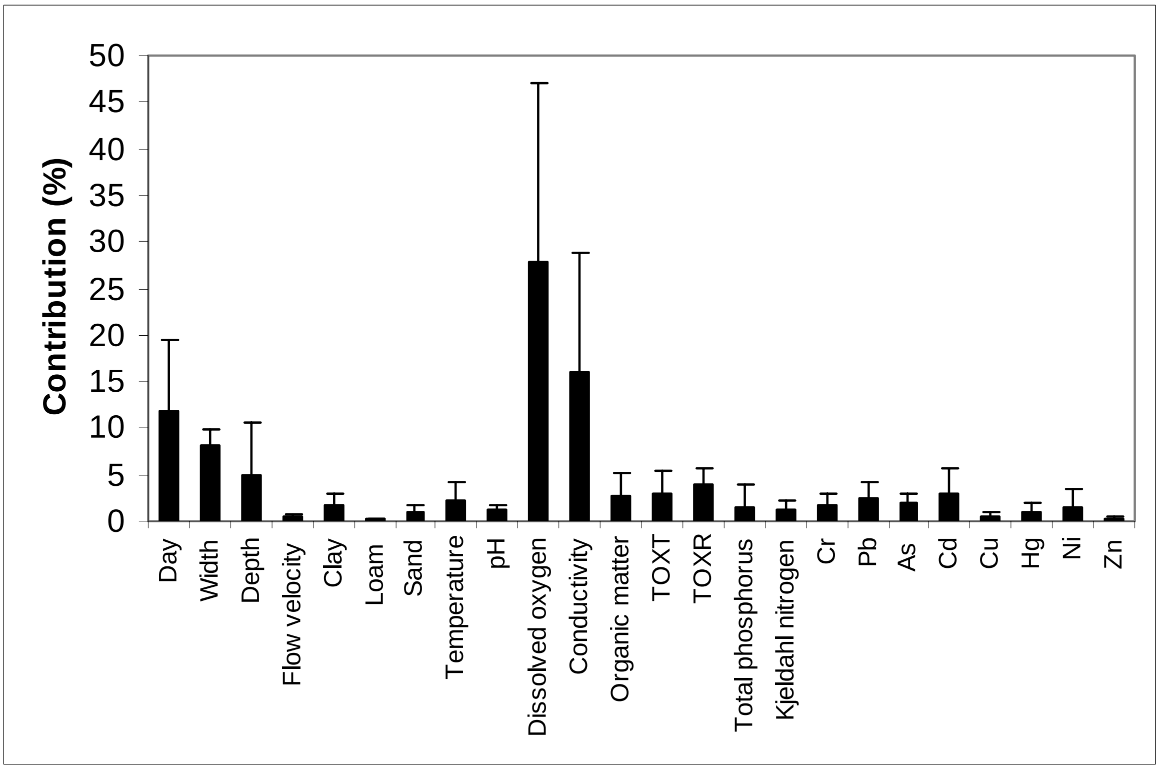
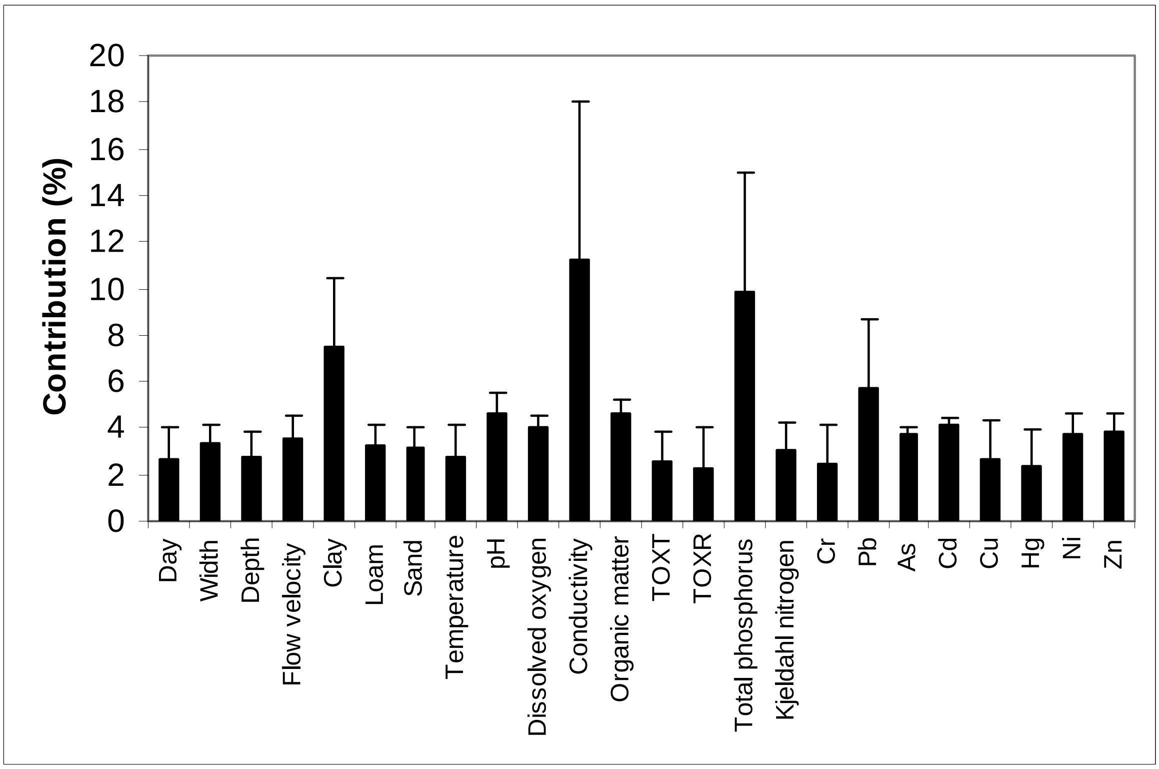


Figure 5.13 Results of the PaD method applied on the presence/absence models for Gammarus and Asellus of the river sediments in Flanders dataset.

These large standard deviations can be a result of outliers. Perhaps these could also be reduced by making stratified subsets based on these major variables (in addition to the output variable).



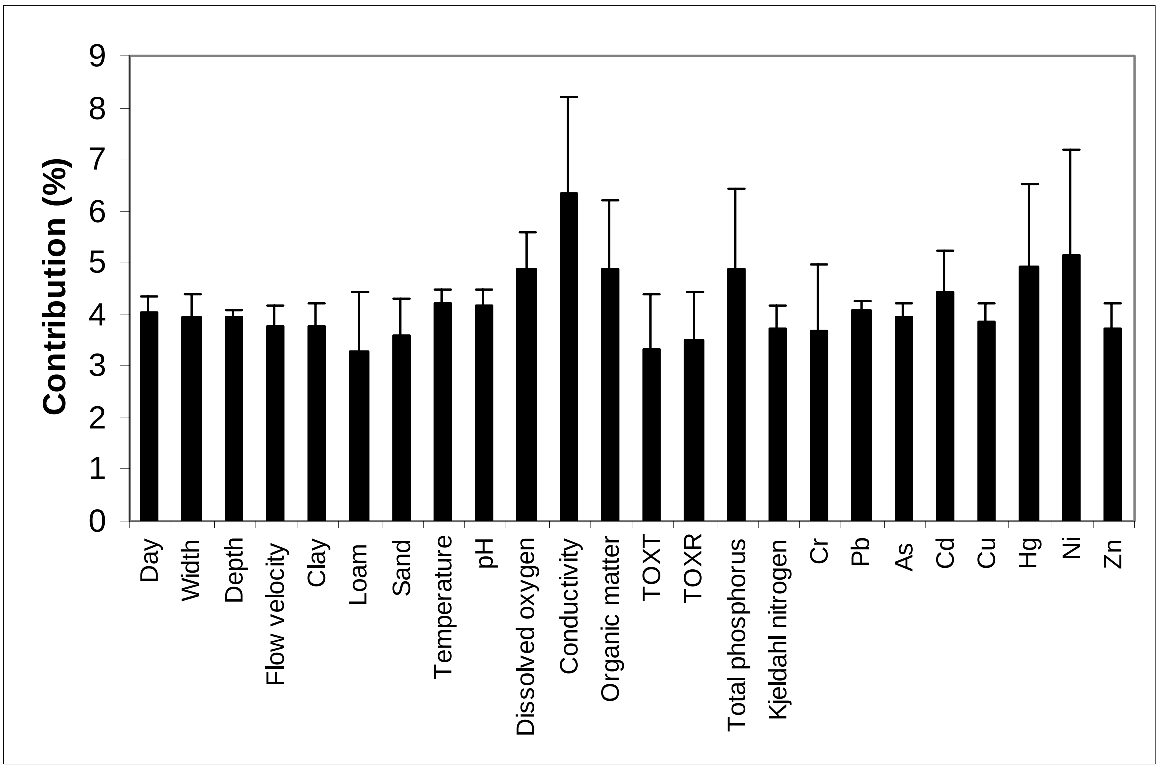


Figure 5.14 Results of the Perturb method applied on the presence/absence models for Gammarus and Asellus of the river sediments in Flanders dataset.

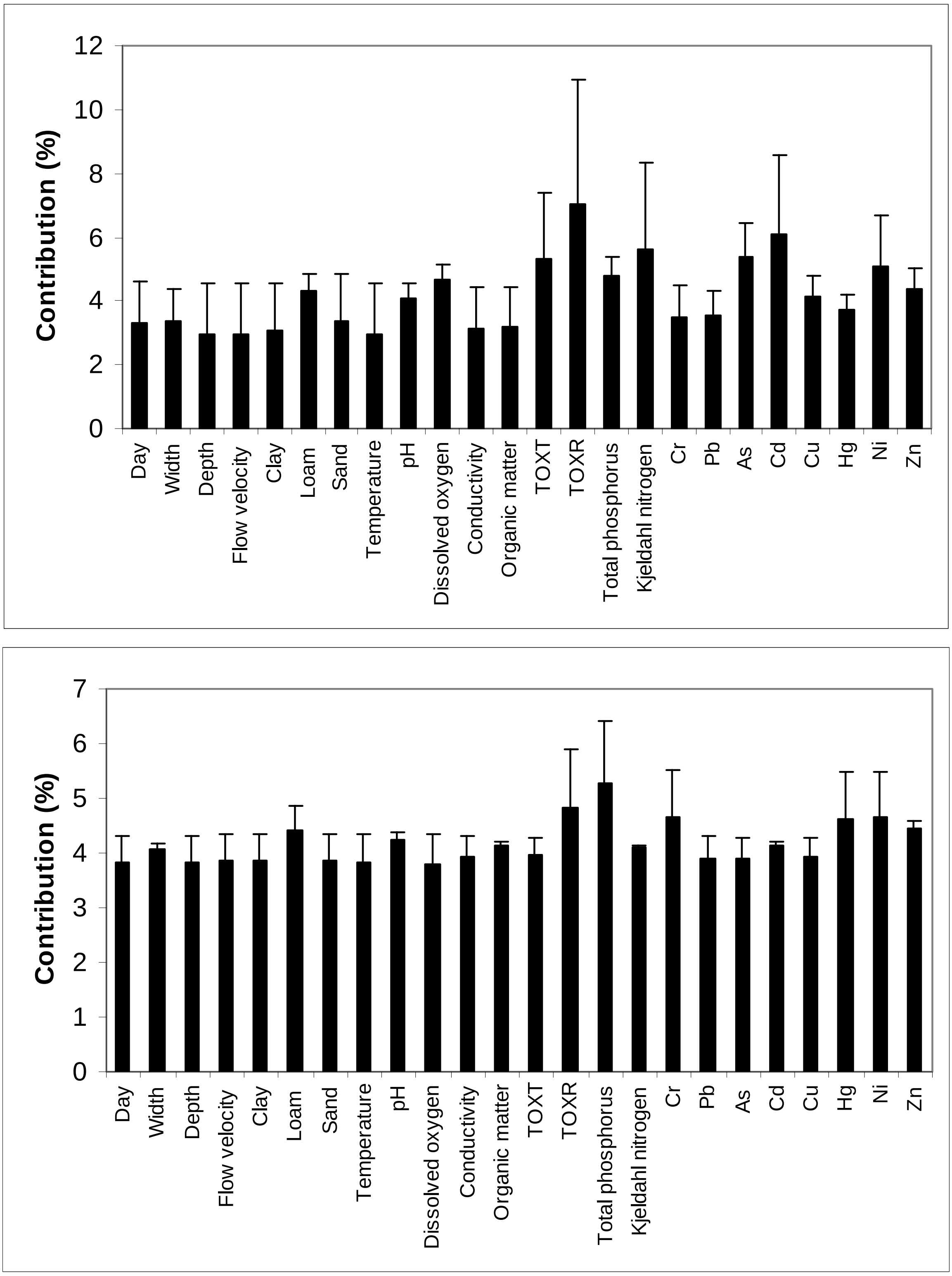


Figure 5.15 Results of the 'Stepwise Reg' method applied on the presence/absence models for Gammarus and Asellus of the river sediments in Flanders dataset.

Based on the results it seems in particular interesting to make at least these input contribution analyses to identify the major variables affecting the output (such as this PaD method), and combine it with the profile method to see how they affect these output and whether this is ecological logically or not. As such they are valuable instruments to analyse the convenience of the models for decision support in river management and also they contribute to the generation of expert knowledge and help to bring clarity in these often called black box models.

5.5.4 Conclusions

This study aimed at analysing the relationship between river sediment characteristics and the presence/absence (and abundance) of the two macroinvertebrate taxa *Asellus* and *Gammarus* based on ANN models. The CCI value of the classification models was good for both taxa (86.3 % and 76.3 % for respectively *Gammarus* and *Asellus*), while the *K* values were more or less acceptable (in the neighbourhood of about 0.4). The performance of the abundance models (regression) trained and validated on the same subsets had *r* values in the validation sets of about 0.4, indicating that the models were as well rather good. However, based on the much higher values in the training sets, one has to conclude that the latter models were probably overtrained on the data, leaving thus room for further improvement of the generalisation capacity of these models.

To study the effect of the environmental variables on the two taxa, six input variables contribution methods were applied on the models. It was difficult to find major trends over the two taxa, the six contribution methods and the three subsets. The first two can be explained by different ecological preferences of the taxa and by the different aspects the six contribution methods deal with. The instability over the different folds is perhaps related with the relative small size (342 instances) of the dataset in combination with the high variability of the sites (whole Flanders), the high number of input variables or outliers in the measurements. This will therefore need further research based on larger datasets and sub-sampling methods. This study also revealed that new variables should be included to give reliable predictions in future. A new toxicity test based on *Hyalella* for instance, but also information about metal concentrations in the water column. Also salts concentrations and other compounds affecting the bio-availability of the metals need to be included in the measurement campaigns in the future.

5.6 Development of predictive habitat suitability models based on classification tree methods for *Gammarus* and *Asellus* in the Zwalm river basin

5.6.1 Introduction

This component of the results deals with the development of predictive habitat suitability models based on classification tree methods for *Gammarus* and *Asellus* in the Zwalm river basin. First the evaluation based on performance indicators is presented, followed by an overview of the selected variables and their ranking of importance. Finally also an ecological interpretation and discussion of practical use of the induced classification trees is given.

The trees are all presented in the Appendices 25-48. These are ordered per taxon (*Gammarus* and *Asellus*). For each of the three subsets, several pruning confidence factors (0.5, 0.25, 0.1 and 0.01) were tested.

5.6.2 Evaluation based on performance indicators

In Figures 5.16 and 5.17, the best performing trees (according to the CCI and *K* as marked in yellow in Table 5.19) are presented respectively for *Gammarus* and *Asellus*. In case of *Gammarus*, a tree size of 11 gave the best results, for *Asellus* this was 18 (Table 5.18).

```
Loamclay <= 49.1: 1 (93.0/13.0)
Loamclay > 49.1
|  Orthophosphate <= 0.235: 1 (5.0)
|  Orthophosphate > 0.235
|  |  Ammonium <= 0.6: 0 (9.0)
|  |  Ammonium > 0.6
|  |  |  COD <= 16: 1 (3.0)
|  |  |  COD > 16
|  |  |  |  Conductivity <= 583: 1 (3.0)
|  |  |  |  Conductivity > 583: 0 (6.0)
```

Figure 5.16 Classification tree for Gammarus based on the Zwalm river basin data set (Subset 3, PCF=0.25; values between brackets indicate instances in which rules are true/false).

```
Width <= 250
|  Width <= 123
|  |  Banks = 0: 0 (44.0/5.0)
|  |  Banks = 1
|  |  |  T <= 15.1: 0 (8.0)
|  |  |  T > 15.1: 1 (4.0)
|  |  Banks = 2
|  |  |  Conductivity <= 738: 1 (2.0)
|  |  |  Conductivity > 738: 0 (3.0)
|  Width > 123
|  |  Distmouth <= 15778.284
|  |  |  Width <= 144: 1 (9.0)
|  |  |  Width > 144
|  |  |  |  Ammonium <= 0.23: 1 (3.0)
|  |  |  |  Ammonium > 0.23: 0 (7.0/1.0)
|  |  Distmouth > 15778.284: 0 (4.0)
Width > 250: 1 (35.0/2.0)
```

Figure 5.17 Classification tree for Asellus based on the Zwalm river basin data set (Subset 3, PCF=0.5; values between brackets indicate instances in which rules are true/false).

Table 5.18 Tree size of the induced classification trees (for different pruning confidence factors (PCFs)) for Gammarus and Asellus based on the Zwalm river database.

Gammarus						
Tree size	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	19	19	13	13	16	3
Subset 2	36	3	3	3	11	17
Subset 3	29	11	11	11	16	9
SS Average	28	11	9	9	14	9
SS Stdev	9	8	5	5		
Asellus						
Tree size	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	26	15	3	3	12	11
Subset 2	29	29	3	3	16	15
Subset 3	18	18	18	11	16	4
SS Average	24	21	8	6	15	9
SS Stdev	6	7	9	5		

Table 5.18 illustrates the effect of the pruning algorithm on the tree size, and Table 5.19 the relation with the tree performance based on CCI and *K*. The pruning has a dramatic effect on the size and related complexity of the trees, but this seems to take place within a small interval of the PCF. In case of *Gammarus* the average tree size (over the three subsets) drops from 28 to 11 between PCF 0.5 and 0.25, while for *Asellus* from 21 to 8 for PCF values between 0.25 and 0.1 (SS Average row indicated in blue in Table 5.18). Nevertheless, the CCI and *K* seem to stay rather constant under different pruning levels (Table 5.19).

Table 5.19 Performance of the induced classification trees (for different pruning confidence factors (PCFs)) for Gammarus and Asellus based on the river Zwalm river database.

Gammarus						
CCI	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	70.0	70.0	73.3	73.3	71.7	1.9
Subset 2	71.7	76.7	76.7	76.7	75.5	2.5
Subset 3	73.3	80.0	80.0	80.0	78.3	3.3
SS Average	71.7	75.6	76.7	76.7	75.1	2.4
SS Stdev	1.7	5.1	3.4	3.4		
K	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	0.16	0.16	0.17	0.17	0.17	0.01
Subset 2	0.30	0.13	0.13	0.13	0.17	0.09
Subset 3	0.22	0.35	0.35	0.35	0.32	0.07
SS Average	0.23	0.21	0.22	0.22	0.22	0.01
SS Stdev	0.07	0.12	0.12	0.12		
Asellus						
CCI	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	71.7	75.0	76.7	76.7	75.0	2.4
Subset 2	73.3	73.3	75.0	75.0	74.2	1.0
Subset 3	86.7	86.7	86.7	86.7	86.7	0.0
SS Average	77.2	78.3	79.5	79.5	78.6	1.1
SS Stdev	8.2	7.3	6.3	6.3		
K	**PCF=0.5**	**PCF=0.25**	**PCF=0.1**	**PCF=0.01**	**PCF Average**	**PCF Stdev**
Subset 1	0.43	0.50	0.53	0.53	0.50	0.05
Subset 2	0.46	0.46	0.49	0.49	0.48	0.02
Subset 3	0.73	0.73	0.73	0.73	0.73	0.00
SS Average	0.54	0.56	0.58	0.58	0.57	0.02
SS Stdev	0.17	0.15	0.13	0.13		

Based on the results of Table 5.19, no reliable trees could be developed for *Gammarus* based on the K (PCF Average of 0.22, only in subset 3 exceptionally good trees were induced for PCF values between 0.01 and 0.25). For *Asellus* on the contrary, over all PCF ranges and subsets a stable and good set of trees was developed (PCF Average of 0.57 with standard deviation of only 0.02).

5.6.3 Ranking of importance of input variables, ecological interpretation and discussion of practical use of classification tree models

The ranking of the input variables for both *Gammarus* and *Asellus* is presented in Table 5.20. The applied ranking procedure is similar to the one explained in 5.4.3. These results are based on the trees presented in the Appendices 25-48.

Table 5.20 Major variables of the induced classification trees for Gammarus and Asellus based on the Zwalm river database.

Gammarus					
Variables	**First**	**Second**	**Third**	**Fourth**	**Fifth**
Subset 1	TotalP	Depht, Distm	FV, T	pH, Loamclay	-
Subset 2	Width	AmmN	T, PR	FV, HB, EC, SS	Banks, pH
Subset 3	Loamclay	HB, OrthoP	TotalN, AmmN	Meandering, COD	Depth, EC
Three times	-	-	-	-	-
Twice	-	-	T	-	-
Once	TotalP, Width, Loamclay	Depth, Distm, AmmN, HB, OrthoP	FV, PR, TotalN, AmmN	pH, Loamclay, FV, HB, EC, SS, Meandering, COD	Banks, pH, Depth, EC
Asellus					
Variables	**First**	**Second**	**Third**	**Fourth**	**Fifth**
Subset 1	Width	Banks, Strorder	PR, pH, OrthoP	-	-
Subset 2	Width	HB	FV, Depth, Strorder, Gravel	TotalP, SS	Nitrate
Subset 3	Width	-	Banks, Distm	T, EC	AmmN
Three times	Width	-	-	-	-
Twice	-	-	-	-	-
Once	-	Banks, Strorder, HB	PR, pH, OrthoP, FV, Depth, Strorder, Gravel, Banks, Distm	TotalP, SS, T, EC	Nitrate, AmmN

The major variables for *Gammarus* are total phosphorus, width, loam/clay, depth, distance to mouth, ammonium, hollow banks and orthophosphate. For *Asellus* these are width, banks, stream order, hollow banks, pool/riffles, pH, orthophosphate, flow velocity, depth, stream order, gravel, banks and distance to mouth. Similar to the results on the river sediments in Flanders, the results over the three subsets seem to be very instable as well. Only for *Asellus* the river variable ‘width’ is three times used as major variable.

In contrast to the river sediments database, most of the determining variables are related to physical habitat. The major reason for this is that in general the water quality in the Zwalm river basin is rather good, and as such these pollution related variables are not that much affecting the presence/absence of the two taxa. The major pollution is originating from agricultural activities, and therefore it is logical that the nutrients are the only dominating water quality variables. In general, the major impacts are related to physical habitat deterioration, together with the natural habitat aspects (reflected in the distance to mouth and stream order according to the river continuum concept of Vannote (1980)).

5.6.4 Conclusions

For *Gammarus*, no reliable trees could be developed based on the *K* (maximum value of 0.35 as a positive exception in subset 3, while only 0.22 in average). For *Asellus* on the contrary, well performing trees were developed at all pruning levels (average *K* of 0.57 with standard deviation of only 0.02). The pruning had a dramatic effect on the size and related complexity of the trees, but this took place within a small interval of the PCF.

The major variables for *Gammarus* were total phosphorus, width, loam/clay, depth, distance to mouth, ammonium, hollow banks and orthophosphate. For *Asellus* these were width, banks, stream order, hollow banks, pool/riffles, pH, orthophosphate, flow velocity, depth, stream order, gravel, banks and distance to mouth. In general it concerns nutrient variables related to impacts of mainly agricultural activities and physical habitat variables (both natural as well as artificial modifications to streams).

5.7 Application of backpropagation artificial neural networks predicting *Gammarus* and *Asellus* in the Zwalm river basin

5.7.1 Introduction

This study aims at analysing the relationship between the stream characteristics and the presence/absence (and abundance) of the two macroinvertebrate taxa *Asellus* and *Gammarus* in the Zwalm river basin. Table 5.21 gives a scheme on how the models were applied on the databases and the results are presented in Appendix 61-72. Six methods were applied on *Gammarus* and *Asellus*, first presence/absence models were used (to compare with the outcomes of the classification trees), followed by abundances models (actually these models use log(abundance+1) output transformations).

Table 5.21 Overview of the applied contribution methods to the ANN models.

Presence/absence classification of taxa
Performance based on performance indicators Application of input variable contribution methods Weights PaD Perturb Stepwise Reg Stepwise Imp Profile
Prediction taxa abundance
Evaluation based on performance indicators Application of input variable contribution methods Weights PaD Perturb Stepwise Reg Stepwise Imp Profile
Selected variables and ranking of importance ***Ecological interpretation and discussion of practical use of the ANN models***

5.7.2 Predictive performance of classification and regression ANN models

In Table 5.22 the CCI and *K* of the classification (presence/absence) models in the Zwalm river basin are presented. For both *Gammarus* and *Asellus*, the CCI and *K* are relatively constant over the three folds, with one major exception (indicated in yellow), where the *K* drops to 0.10 in subset 3. The CCI value is good both for both taxa (75.0 % and 81.1 % for respectively *Gammarus* and *Asellus*), while the *K* values are good for Asellus (0.62), but not for *Gammarus* (0.15 in average). The latter means that the model of subset three does not make use of the environmental variables to predict *Gammarus*. This indicates the danger of the training procedure without a performance index such as the *K*, to compensate for the prevalence to the taxon.

Table 5.22 CCI and K for the classification ANN models of Gammarus and Asellus in the Zwalm river basin.

Gammarus	**CCI**	***K***
Subset 1	76.7	0.23
Subset 2	73.3	0.22
Subset 3	75.0	0.00
Average	75.0	0.15
Standard deviation	1.7	0.13
Asellus	**CCI**	***K***
Subset 1	81.7	0.63
Subset 2	76.7	0.53
Subset 3	85.0	0.70
Average	81.1	0.62
Standard deviation	4.2	0.09

In Table 5.23 the performance of the abundance models (regression) trained and validated on the same subsets are presented. In average the *r* values in the validation sets are very good (0.73 and 0.80) and constant over the three folds.

Table 5.23 Correlation coefficient r for the regression ANN models of Gammarus and Asellus in the Zwalm river basin.

Gammarus	***r* (validation set)**
Subset 1	0.75
Subset 2	0.73
Subset 3	0.71
Average	0.73
Standard deviation	0.02
Asellus	***r* (validation set)**
Subset 1	0.80
Subset 2	0.81
Subset 3	0.73
Average	0.78
Standard deviation	0.05

5.7.3 Ranking of importance of input variables combined with ecological interpretation and discussion of practical use of ANN models

The lack of illustrative power of ANN models is a major concern to ecologists since the interpretation of statistical models is desirable for gaining knowledge of the causal relationships driving ecological phenomena (Olden and Jackson, 2002). To dispose of this disadvantage, six contribution methods (Tables 5.24 to 5.27) were applied to the ANN models in this study. These techniques helped to identify environmental factors influencing the presence/absence and abundance of both taxa. Among the six methods applied, the 'Profile' method was the only technique that provided two elements of information on the contribution of the variables. On the one hand, this method presented the order of contribution of the different environmental variables, on the other hand, gave direct interpretation of the effect of river characteristics on the abundance or presence/absence of the taxa. The investigation of the sensitivity curves could enhance the understanding of the effects of impacts of various types on individual macroinvertebrate taxa (Marshall et al., 2002). In this way, the 'Profile' method would enable impact-specific indicator taxa to be readily identified and would enhance the capacity to monitor and mitigate the effects of human activities on river ecosystems. The other methods were merely able to classify the variables by order of their importance, in other words, to reveal their contribution to the output.

Table 5.24 Comparison of the outcome of the six different input variable contribution methods for the Gammarus presence/absence models based on the river sediments in the Zwalm river database.

Gammarus P/A	**Weights**	**PaD**	**Perturb**	**Stepwise Reg**	**Stepwise Imp**	**Profile**	**Overal rank**	**Average rank**	**Standard deviation**
Width	1	1	24	10	5	2	3	7	8.9
Banks	20	21	22	1	20	19	19	17	8.0
Meandering	22	18	5	4	13	22	17	14	8.1
Pool/Riffle	8	19	6	23	17	13	18	14	6.6
Hollow beds	23	20	16	11	16	18	21	17	4.1
Depth	10	6	23	12	6	5	6	10	6.8
Flow velocity	13	9	17	7	10	9	10	11	3.6
pH	24	10	10	3	21	11	14	13	7.8
Temperature	3	3	3	16	2	3	2	5	5.4
Dissolved oxygen	7	15	2	24	18	15	16	14	7.9
Conductivity	16	22	8	22	19	16	20	17	5.2
Suspended solids	15	11	7	8	14	8	7	11	3.4
Ammonium	12	7	14	14	8	4	5	10	4.1
Nitrate	4	24	13	19	22	24	22	18	7.9
Total nitrogen	18	23	18	17	23	23	24	20	2.9
Phosphate	17	16	4	2	15	21	13	13	7.7
Total phosphorus	6	5	9	20	3	6	4	8	6.1
COD	5	13	12	9	7	20	11	11	5.3
Boulders	19	12	11	6	12	12	12	12	4.1
Gravel	11	14	19	15	11	10	15	13	3.4
Sand	21	17	20	18	24	14	23	19	3.5
Loam/clay	14	4	21	13	4	7	8	11	6.7
Distance to mouth	2	2	1	21	1	1	1	5	8.0
Stream order	9	8	15	5	9	17	9	11	4.5

Table 5.25 Comparison of the outcome of the six different input variable contribution methods for the Asellus presence/absence models based on the river sediments in the Zwalm river database.

Asellus P/A	**Weights**	**PaD**	**Perturb**	**Stepwise Reg**	**Stepwise Imp**	**Profile**	**Overal rank**	**Average rank**	**Standard deviation**
Width	3	3	7	14	6	7	3	7	4.0
Banks	11	14	23	1	10	21	12	13	8.0
Meandering	2	2	1	13	2	2	2	4	4.6
Pool/Riffle	5	7	4	8	7	10	4	7	2.1
Hollow beds	6	8	3	21	5	15	8	10	6.9
Depth	7	6	11	9	16	12	9	10	3.7
Flow velocity	17	22	22	15	12	20	22	18	4.0
pH	15	19	10	19	18	24	21	18	4.7
Temperature	13	17	12	24	11	19	18	16	5.0
Dissolved oxygen	14	18	8	18	17	5	13	13	5.6
Conductivity	23	23	18	3	23	11	20	17	8.3
Suspended solids	22	20	19	4	22	9	19	16	7.6
Ammonium	8	5	5	22	4	3	5	8	7.1
Nitrate	9	9	6	23	3	6	6	9	7.1
Total nitrogen	12	10	9	17	19	4	10	12	5.5
Phosphate	24	24	17	11	24	16	24	19	5.5
Total phosphorus	21	21	20	20	20	8	23	18	5.1
COD	18	15	13	12	15	18	15	15	2.5
Boulders	16	13	21	16	9	17	16	15	4.0
Gravel	19	12	24	5	21	14	17	16	6.9
Sand	20	16	16	6	8	23	14	15	6.6
Loam/clay	10	11	15	2	13	22	11	12	6.6
Distance to mouth	1	1	2	10	1	1	1	3	3.6
Stream order	4	4	14	7	14	13	7	9	4.9

Table 5.26   Comparison of the outcome of the six different input variable contribution methods for the Gammarus abundance models based on the river sediments in Zwalm river database.

Gammarus abundance	**Weights**	**PaD**	**Perturb**	**Stepwise Reg**	**Stepwise Imp**	**Profile**	**Overal rank**	**Average rank**	**Standard deviation**
Width	5	3	15	2	10	7	4	7	4.9
Banks	21	22	22	5	17	21	21	18	6.6
Meandering	14	16	13	11	9	15	13	13	2.6
Pool/Riffle	12	19	11	23	13	20	17	16	5.0
Hollow beds	10	9	2	3	4	11	3	7	3.9
Depth	9	8	16	8	14	5	10	10	4.1
Flow velocity	24	24	24	7	22	23	23	21	6.7
pH	22	17	14	15	12	24	20	17	4.7
Temperature	20	12	10	20	6	17	14	14	5.7
Dissolved oxygen	17	14	6	24	19	10	16	15	6.4
Conductivity	18	20	8	16	21	18	18	17	4.7
Suspended solids	15	13	12	1	18	6	11	11	6.2
Ammonium	1	1	1	18	1	2	1	4	6.9
Nitrate	16	18	18	21	16	12	19	17	3.0
Total nitrogen	19	21	21	19	24	16	22	20	2.7
Phosphate	7	7	5	6	20	3	8	8	6.1
Total phosphorus	8	4	3	12	15	1	5	7	5.5
COD	3	6	4	22	7	4	7	8	7.2
Boulders	13	15	20	14	8	19	15	15	4.4
Gravel	6	5	17	4	3	9	6	7	5.2
Sand	23	23	23	13	23	22	24	21	4.0
Loam/clay	11	10	19	9	11	8	12	11	3.9
Distance to mouth	4	11	7	17	5	14	9	10	5.2
Stream order	2	2	9	10	2	13	2	6	4.9

Table 5.27 Comparison of the outcome of the six different input variable contribution methods for the Asellus abundance models based on the river sediments in the Zwalm river database.

Asellus abundance	**Weights**	**PaD**	**Perturb**	**Stepwise Reg**	**Stepwise Imp**	**Profile**	**Overal rank**	**Average rank**	**Standard deviation**
Width	1	1	4	2	1	1	1	2	1.2
Banks	15	13	19	3	8	17	10	13	6.0
Meandering	8	20	13	12	19	15	14	15	4.5
Pool/Riffle	19	16	16	5	24	20	19	17	6.4
Hollow beds	6	5	3	10	4	8	4	6	2.6
Depth	3	4	7	11	10	2	5	6	3.8
Flow velocity	11	23	23	13	16	21	21	18	5.2
pH	17	14	14	22	18	14	18	17	3.2
Temperature	18	8	9	21	6	19	13	14	6.5
Dissolved oxygen	13	7	1	23	7	12	8	11	7.5
Conductivity	14	15	10	14	15	7	11	13	3.3
Suspended solids	24	19	22	16	20	13	22	19	4.0
Ammonium	12	9	12	20	17	9	12	13	4.4
Nitrate	22	12	11	17	14	11	15	15	4.3
Total nitrogen	21	24	20	7	23	23	23	20	6.4
Phosphate	23	21	24	6	21	5	20	17	8.7
Total phosphorus	16	22	18	18	22	24	24	20	3.1
COD	10	10	8	15	13	6	7	10	3.3
Boulders	5	11	17	9	11	16	9	12	4.5
Gravel	20	17	21	4	9	22	17	16	7.3
Sand	9	6	5	24	5	10	6	10	7.3
Loam/clay	7	18	15	19	12	18	16	15	4.6
Distance to mouth	4	3	2	8	3	4	3	4	2.1
Stream order	2	2	6	1	2	3	2	3	1.8

Although each method expresses a different aspect of sensitivity or importance of the environmental variables to the presence/absence (or abundance) of the taxa (see previous result analysis on the river sediments in Flanders), the average results of the six methods are rather stable per model type. The PaD method makes a classification of the relative contributions of each variable to the network output. The input variable that has the highest SSD value is the variable, which influences the output most. This method is therefore very good to detect the variables of major concern. It also seems to correlate very well with the overall rank. The convenience of this method seems as such to be confirmed in this dataset.

When analyzing the ecological relevance of the models (Tables 5.24 to 5.27), it seems that the river continuum concept is very well confirmed by the data (distance to mouth and stream order variables). Also width plays a major role for *Asellus*. The physical habitat variables (meandering, pool/riffle and hollow beds) seem of major concern as well for *Asellus*, but the results are not very stable. However, these were highly correlated and can be seen as one set of variables. For *Gammarus*, the nutrient variables ammonium and total phosphorus seemed to be crucial. The effect of dissolved oxygen seemed to be less important in the Zwalm river basin. Although Wesenberg-Lund (1982) stated that *Gammarus pulex* is almost non-tolerant for low oxygen conditions, these conditions were almost not present anymore in the Zwalm and pollution problems are more related to nutrients. And the same holds for conductivity, as the Macrofauna-atlas of North Holland (1990) mentioned that *Gammarus pulex* is sensitive to high conductivity values, also this was not reflected in the models. So probably the investments in wastewater treatment during the nineties are responsible for this improvement of the water quality. Contrary to what was expected, flow velocity was not considered as an important variable to predict the abundance of *Gammarus* and *Asellus* in the Zwalm river basin. Based on the ‘Bayerisches Landesamt für Wasserwirtschaft’ (1996) however, *Gammarus pulex* prefers rather fast running streams, since having very good swimming abilities (Brehm and Meijering, 1990). This study revealed that *Gammarus* can be considered as an indicator of nutrient related pollution, which is of major importance in Flanders. The selection of the Gammaridae as an indicator taxon in the Belgian Biotic Index method can therefore be motivated on the basis of these results.

Based on the different contribution methods applied on the Zwalm catchment, habitat characteristics seemed to be more important than the impact of physical and chemical variables for *Asellus*. In this way, *Asellus* only has potential as a good indicator organism in broader streams, since the habitat of the headwaters is less suited by nature for this species. As a consequence, *Asellus* is less abundant in those streams. In the end, these methods for testing the contributions of the different input variables facilitate the selection of the suitable habitats in which certain species can or can not act as an indicator organism for the assessment and management of rivers.

However, one can question to what extent these conclusions can be extrapolated. Data mining and modelling studies were made in Flanders on the basis of datasets from several catchments during other periods. Adriaenssens (2004) for example, found that the environmental variable conductivity explains a major part of the abundance based on Fuzzy knowledge-based models. This variable describes pollution caused by agricultural activities and treated or untreated wastewater effluents. The same results were obtained when decision trees were used in combination with input variable selection by means of genetic algorithms (D'heygere et al., 2003) applied on the river sediments in Flanders dataset.

Concerning their value as indicator organisms, the Gammaridae/Asellidae ratio is used in running waters in the U.K. (Hawkes and Davies, 1971; Whitehurst, 1988). This ratio is able to detect subtle changes in organic pollution level, because the change in organic load alters the relative abundance of Asellidae and Gammaridae rather than the total species composition (MacNeil et al., 2002). In this way, one might conclude that besides the habitat characteristics indicated by the different contribution methods, also the pollution related physical-chemical variables can be important to explain the abundance of *Asellus*. This underlines the need for relevant datasets for habitat preference studies. It also underpins that predictive ecological models developed with data driven techniques should be used with enough care for practical decision support in river management as illustrated in Goethals et al. (2002). To use of the PaD method and the profile graphs are very good instruments for this as was earlier mentioned. Figures 5.18 shows the results obtained with the 'Profile' method for 12 scale intervals between the minimum and maximum of the input variables. The major variables for the two taxa are quite different. The dominating variables for the abundance models of *Gammarus* were total phosphorus, ammonium, phosphate and COD were best expressed (see also Table 5.26). For *Asellus* these were distance to mouth, stream order, width and depth. The abundance values for *Asellus* were also much higher in general (the average value is 92 organisms per sampling site, while for *Gammarus* only 6).

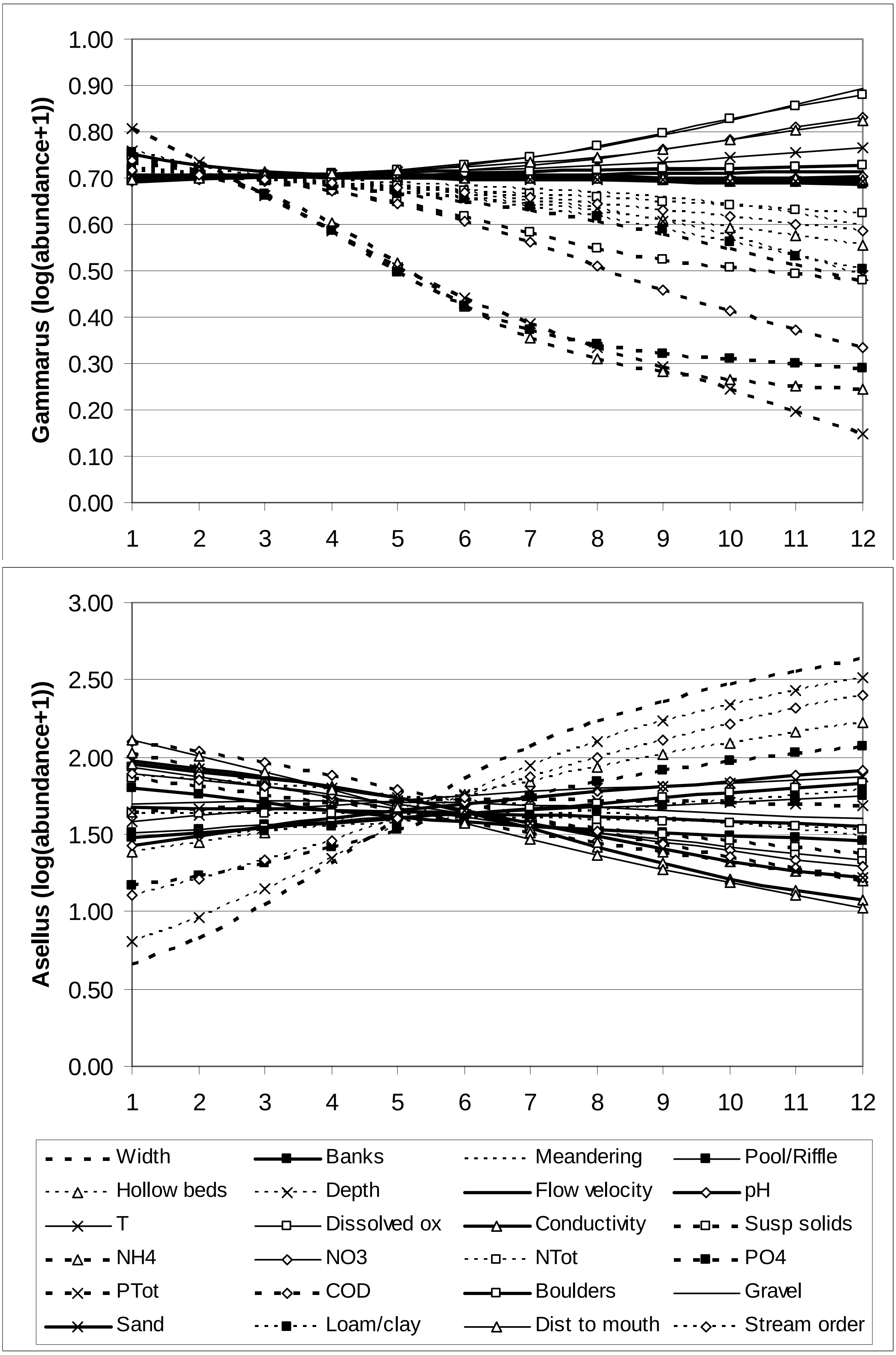


Figure 5.18 Curves for the 24 input variables in relation to the abundance of Gammarus (top) and Asellus (bottom) based on the profile method and ANNs.

Additionally, it seemed interesting to check the effect of variable pruning on the model performance and compare it with the experience on classification trees. The mean predictive performances of the three testing sets for *Gammarus* after removing step by step the least contributing input variables in the abundance model are shown in Figure 5.20. As mentioned before, the correlation coefficient for the ANN model including the 24 environmental variables was 0.73. However, elimination of the less important variables gave an increase of the correlation coefficient. If eight variables (flow velocity, sand, banks, total nitrogen, pH, nitrate, boulders and conductivity) were removed, then the highest performance (0.76) was obtained. Excluding more variables led to a decrease of the predictive performance. If only one variable remained (ammonium was the most contributing variable), a correlation coefficient of only 0.57 was achieved. At the same time, the standard deviation was increasing, indicating that the stability of the models over the three folds decreased.

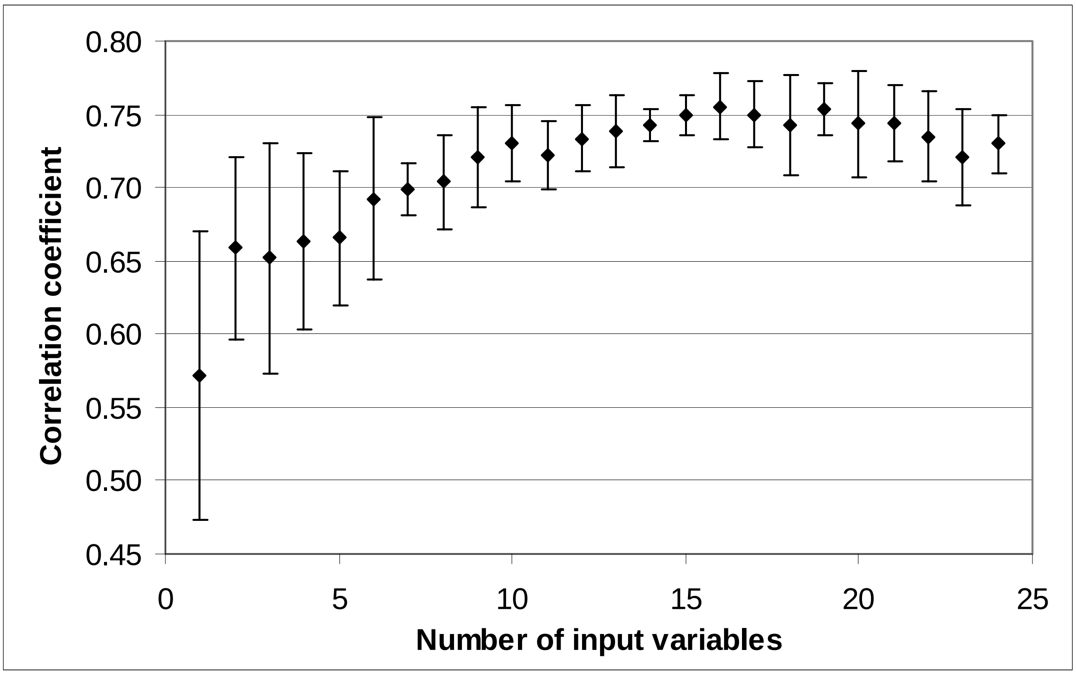


Figure 5.20 Mean predictive performances (correlation coefficient) of the three fold cross validation for Gammarus after stepwise removal of the least contributing variables.

As such, similar to the previous results with the classification trees in this PhD study, the variables that were not selected could be seen as irrelevant for the modelled taxon. In other words, these variables were less important to describe the habitat of *Gammarus*.

5.7.4 Conclusions

This study aimed at analysing the relationship between the stream characteristics and the presence/absence (and abundance) of the two macroinvertebrate taxa *Asellus* and *Gammarus* in the Zwalm river basin. Regarding the presence/absence models, the CCI value was good for both taxa (75.0 % and 81.1 % for respectively *Gammarus* and *Asellus*), while the *K* values are good for *Asellus* (0.62), but not for *Gammarus* (0.15 in average). The latter means that the model of subset three does not make use of the environmental variables to predict *Gammarus*. In average the *r* values of the abundance models were very good (respectively 0.73 and 0.80 for *Gammarus* and *Asellus*) and constant over the three folds.

Input variable contribution methods applied to ANN models can be useful to select the essential environmental variables for macroinvertebrate taxa. In this way, the choice of ecologically significant variables to describe the species' habitat(s) and to include in monitoring campaigns for river assessment can be well-founded.

On the other hand, the prediction of abundance of species or populations based on habitat characteristics can be of high interest to ecologists, managers or engineers, who are dealing with river assessment and restoration management. In particular the insight in the sensitivity curves can be useful to select meaningful indicator taxa. These curves can support the decisions related to river restoration and protection, by showing how the environmental variables affect the biological communities.

5.8 A comparative discussion of the obtained results

5.8.1 Performance of classification tree and ANN models based on numerical indicators

Based on the river sediments in Flanders database, the average CCI (over all pruning levels), the trees for both *Gammarus* and *Asellus* seem to be reliable (respectively 83.7% and 67.3%), however when analysing the average *K* (respectively 0.29 and 0.18), the trees do not meet the threshold value of 0.4, indicating that the trees' performance is mainly related to the relatively low prevalence of both taxa in the database and related 'easy' classification, even without using environmental information. The CCI value of the classification ANN models was good for both taxa (86.3 % and 76.3 % for respectively *Gammarus* and *Asellus*), while the *K* values were more or less acceptable (in the neighbourhood of about 0.4), thus slightly higher than those of the classification trees. The performance of the abundance models (regression) trained and validated on the same subsets had *r* values in the validation sets of about 0.4, indicating that the models were as well rather good. As such, all trained ANN models were more or less on the edge of good and not good, while the classification trees performed a little less, especially when the *K* values are taken into account. The ANN model for *Asellus* was in particular much better than the one based on classification trees.

Based on the data of the Zwalm river basin, the model performance was very different for both taxa, and also between the presence/absence and abundance models. For *Gammarus*, no reliable trees could be developed based on the *K* (maximum value of 0.35, while only 0.22 on average). For *Asellus* on the contrary, well performing trees were developed at all pruning levels (average *K* of 0.57 with standard deviation of only 0.02). Regarding the presence/absence ANN models, the CCI value was good for both taxa (75.0 % and 81.1 % for respectively *Gammarus* and *Asellus*), while the *K* values are good for *Asellus* (0.62), but not for *Gammarus* (0.15 on average). It seems that the K values for both classification trees and ANN models were very similar (*K* values for *Gammarus* respectively 0.22 and 0.15 and for *Asellus* respectively 0.57 and 0.62). Both methods are thus performing very similar. On average the *r* values of the abundance models were very good (respectively 0.73 and 0.80 for *Gammarus* and *Asellus*) and constant over the three folds. As such, it seemed ‘easier’ to built good abundance models for *Gammarus* than to make presence/absence predictions in the Zwalm river basin.

When comparing the overall results, it seems that the ANN are performing slightly better than the classification trees. On the other hand, the calculation time of the classification trees is very short and the results (used variables, threshold values) can be directly seen, and allow as such to detect immediately the ecological relevance of the models. For ANN the calculation time is higher and it takes some extra efforts to analyse the ecological meaning of the models.

5.8.2 Importance of different input variables and ecological relevance of the models

The distribution of macroinvertebrates in rivers as well as the inter-relationships of all the different factors which influence this distribution have been widely studied (e.g. Bournaud and Cogerino, 1986). Nevertheless, investigation of this area of river ecology is complicated by the difficulty of separating the effects of competing variables (Rabeni and Minshall, 1977). Our knowledge is still far insufficient to completely understand the habitat preferences of the river macroinvertebrates. This fact is, naturally, a strong handicap when using macroinvertebrate communities for surveillance purposes (Fontoura and De Pauw, 1994). Therefore, this type of studies performed in this PhD are of major concern.

However, when bringing all the rankings of the input variables in the ANN models together (Table 5.28), not many general and robust conclusions can be made. In addition, for the classification trees the major variables in the sediments database for *Gammarus* were clay, Pb, conductivity, day and dissolved oxygen, while for *Asellus* these were dissolved oxygen, toxicity test TOXR, day, width and As. In the Zwalm river basin these were quite different. The major variables for *Gammarus* were total phosphorus, width, loam/clay, depth, distance to mouth, ammonium, hollow banks and orthophosphate. For *Asellus* these were width, banks, stream order, hollow banks, pool/riffles, pH, orthophosphate, flow velocity, depth, stream order, gravel, banks and distance to mouth. Similar to the classification tree results on the river sediments in Flanders, the results over the three subsets seem to be very instable as well. Only for *Asellus* the river variable ‘width’ is three times used as major variable.

In the river sediments of Flanders the basic pollution variables seemed to play a major role according to the models. Metal and toxicity were not that important. Probably these contaminants are important, but only at a limited amount of sites. In the Zwalm river basin physical habitat and nutrients played a major role.

To study the effect of the environmental variables on the two taxa, six input variables contribution methods were applied on the ANN models (Table 5.28). It was difficult to find major trends over the two datasets, taxa, and the six contribution methods. These differences can be explained by dissimilar input variable sets in both databases, the different ecological preferences of the taxa and by the particular aspects the six contribution methods deal with.

Table 5.28 General overview of the ranking of all variables in the two databases for both taxa.

Sediments	***Gam p/a***	***Gam abun***	***Asel p/a***	***Asel abun***
Day	16	12	2	1
Width	12	17	4	10
Depth	21	16	13	14
Flow velocity	14	7	9	19
Clay	5	3	22	4
Loam	11	11	10	5
Sand	17	15	24	21
Temperature	22	18	3	18
pH	4	9	16	13
Dissolved oxygen	3	6	1	8
Conductivity	1	1	17	11
Organic matter	10	8	12	22
TOXT	19	21	5	12
TOXR	18	19	7	15
Total phosphorus	2	2	20	23
Kjeldahl nitrogen	8	4	14	20
Cr	23	24	6	24
Pb	7	10	8	7
As	6	5	11	9
Cd	9	23	15	17
Cu	20	20	18	6
Hg	24	22	21	16
Ni	13	14	19	3
Zn	15	13	23	2
Zwalm river basin	***Gam p/a***	***Gam abun***	***Asel p/a***	***Asel abun***
Width	3	4	3	1
Banks	19	21	12	10
Meandering	17	13	2	14
Pool/Riffle	18	17	4	19
Hollow beds	21	3	8	4
Depth	6	10	9	5
Flow velocity	10	23	22	21
pH	14	20	21	18
Temperature	2	14	18	13
Dissolved oxygen	16	16	13	8
Conductivity	20	18	20	11
Suspended solids	7	11	19	22
Ammonium	5	1	5	12
Nitrate	22	19	6	15
Total nitrogen	24	22	10	23
Phosphate	13	8	24	20
Total phosphorus	4	5	23	24
COD	11	7	15	7
Boulders	12	15	16	9
Gravel	15	6	17	17
Sand	23	24	14	6
Loam/clay	8	12	11	16
Distmouth	1	9	1	3
Stream order	9	2	7	2

The instability over the different folds, that was encountered, is perhaps related to the relative small size (342 instances) of the dataset in combination with the high variability of the sites (whole Flanders), the high number of input variables or outliers in the measurements. This will therefore need further research based on larger datasets and sub-sampling methods. This study also revealed that new variables should be included to give reliable predictions in future.

5.9 Conclusions

The dependence of a species or a community on its habitat is a crucial hypothesis in ecology (Wagner et al., 2000). Thus, the prediction of abundance of species or populations based on habitat characteristics is an interesting task in basic and applied ecology (Baran et al., 1996; Whitehead et al., 1997) and can be of high interest for managers and engineers dealing with rivers and channels (Lek et al., 1996a; Mastrorillo et al., 1997a; Guégan et al., 1998).

Classification trees and ANN models can in this context play an interesting role to find general trends on habitat suitability of macroinvertebrate taxa. The methods revealed to be able to find ecological meaningful relations in the two databases. It is however interesting to see that different analysis procedures (data collection, modelling and model evaluation) often result in dissimilar conclusions. Therefore, it is of major importance to look at results with enough consideration. In particular the use of models for practical applications needs to be preceded with careful validation of these instruments, also in a practical perspective (see next chapter).

Nevertheless, as this research revealed, a major clue to the model development is the data collection. The outcome of this research is as such not only the developed models, but also advice regarding new data that need to be collected and how this should be done to be able to collect meaningful information for river management afterwards.

Chapter 6
Application of predictive macroinvertebrate habitat suitability models for information and decision support in river management: case studies in the Zwalm river basin

6.1 Introduction

This chapter will illustrate and validate the application of the data driven habitat suitability models to support decision making in river management. All presented applications are based on models developed as in the previous chapter, in particular the presence/absence and log(abundance+1) ANN models of the Zwalm river basin (2000-2002). The performance of the models is presented in Appendix 73.

Three types of model applications will be presented:
- prediction of the effect of restoration options;
- selection of monitoring sites based on model uncertainty;
- delivering insight in the habitat preferences of taxa.

6.2 Prediction of the effect of restoration options

6.2.1 Site selection

The first sites for calculating the effect of a planned restoration option are based on a report by Belconsulting (2003). Several restoration options within an integrated water management perspective in the Zwalm river basin (Figure 6.1 left) were described in that report. This restoration action at the Traveinsbeek (sites 56, 120, 121 in Figure 6.1) makes part of it as one of the proposed actions to improve the river ecology and flood control by reintroducing a meandering pattern in the river in combination with a natural flooding area.

The second set of monitoring sites, near the Boembekemolen, consisting of a flood control weir (between sites 25 and 55 in Figure 6.1), are proposed as an interesting virtual action to undertake, to get insight in the reference conditions as is necessary for the implementation of the WFD. By revealing what would be the ecological shifts, one is able to get insight in the reference communities, but also whether it is really worth to consider such a restoration towards a near natural condition and increasing the risks for flooding downstream during intensive rain events as a potential consequence. Most probably this situation with a weir for flood control cannot be altered drastically, and the attribution of this site as a strongly modified water body will probably be necessary and be defended from a social-economical perspective broader than nature conservation.

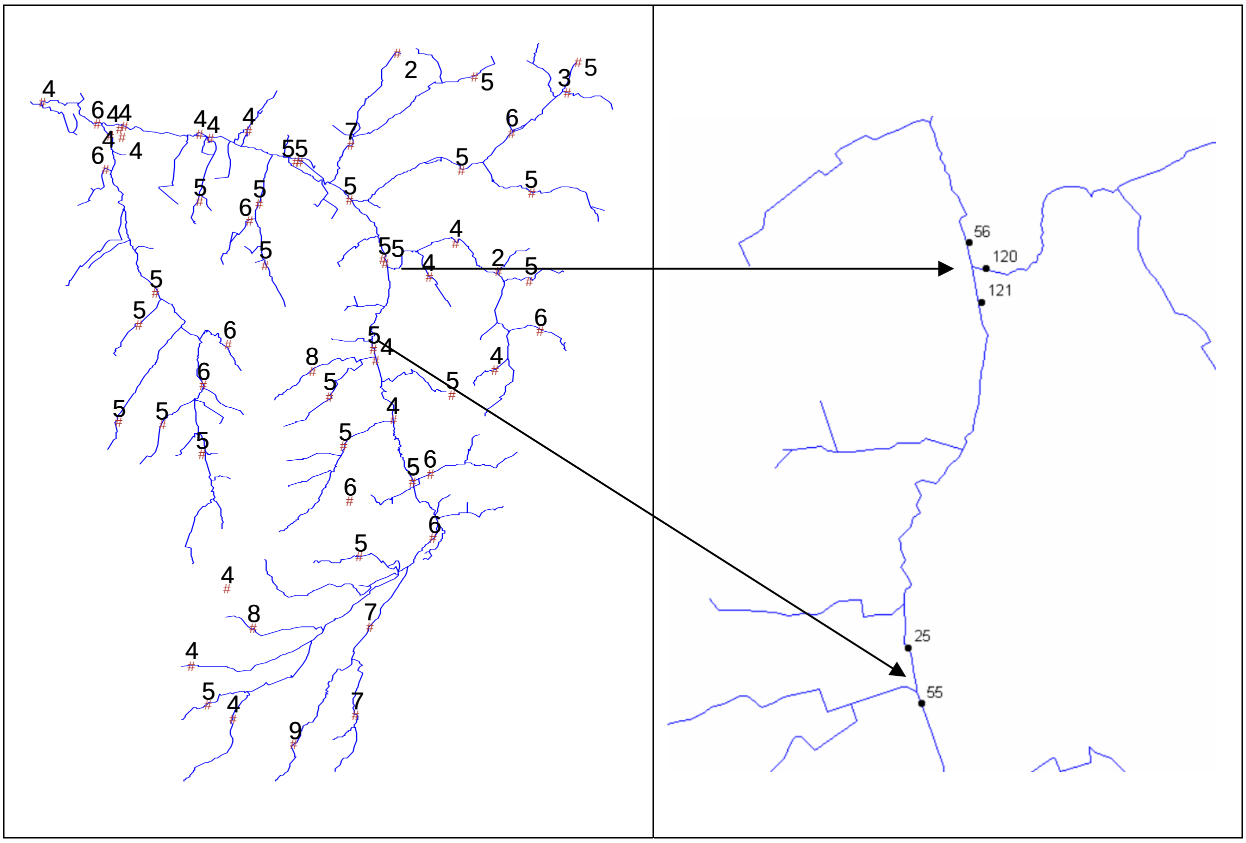


Figure 6.1 The Zwalm river basin and BBI scores at the sixty monitoring sites (during the year 2002) is presented at the left, while at the right a detailed map is presented of the two selected sites: the Traveinsbeek site is in the northern part (monitoring sites 56, 120, 121) and the Boembeke.weir is between the sites 25 and 55).

6.2.2 *Remeandering project of the Traveinsbeek*

The Traveinsbeek brook, a tributary of the Zwalm river (site 120 in Figure 6.1) was straightened to be used as a ‘natural’ border between two meadows (Figure 6.2). Recently there were several projects to investigate sustainable flood control measures in the Zwalm river basin. The construction of several natural flooding areas is among these, and is often combined with nature development. Also this restoration action is a combination of both (Figure 6.3). The planned remeandering has an impact on several river characteristics as shown in Table 6.1. In this table, only the altered variables are presented. In total 30 variables were presented to the artificial neural network (in contrast to the previous chapter also maxima and minima values of width, depth and flow velocity were used, because these are values that are seriously affected by the remeandering works and as well be important from an ecological point of view for most of the taxa considered here). The expected values of the river characteristics in three studied monitoring sites after remeandering were based on an upstream sampling site, characterised by a well-developed meandering pattern and could therefore be seen as the expected situation after remeandering works. In the future, hydrological/hydraulic models could be used in order to calculate the exact values of the river characteristics after remeandering.

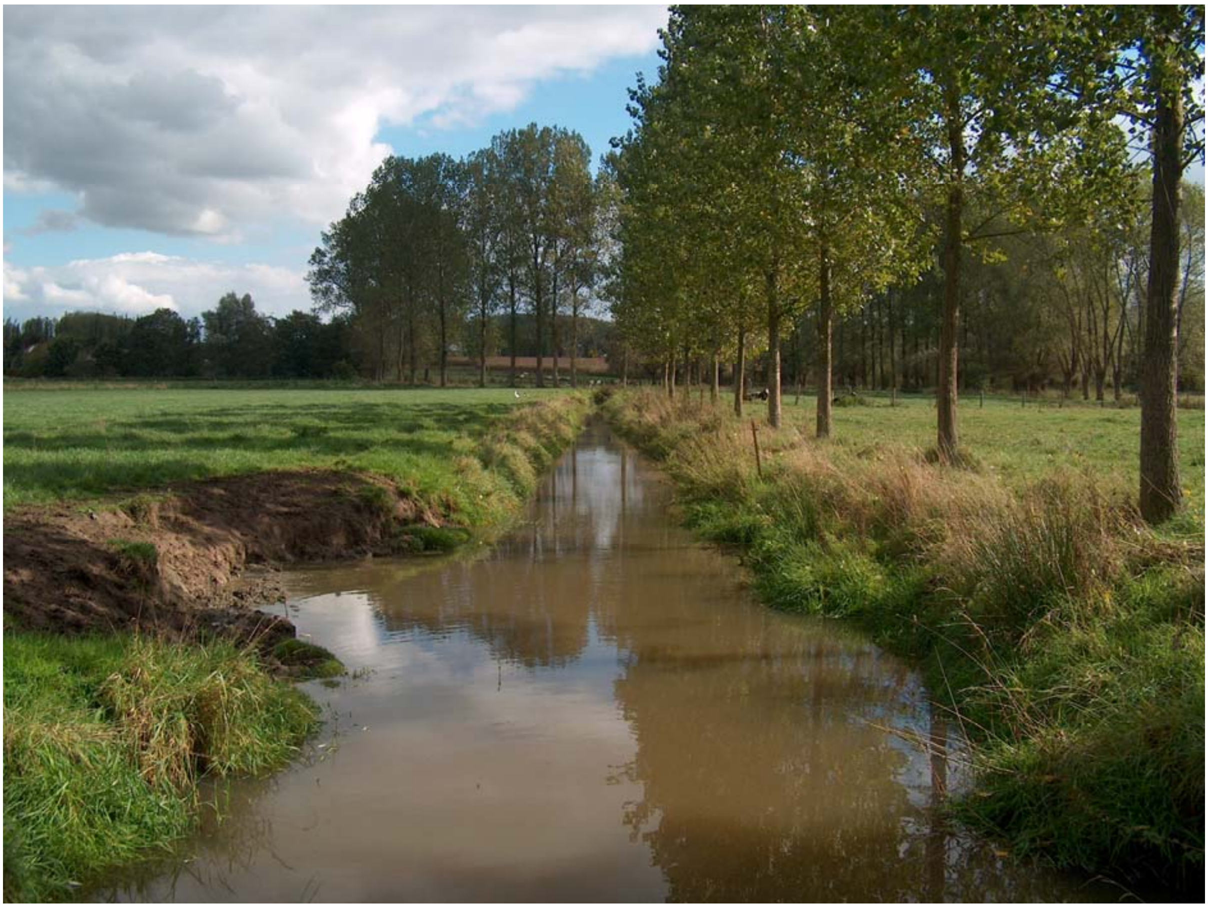


Figure 6.2 Picture of the Traveinsbeek in its actual condition (August 2003). This river part was straightened to use it as a 'natural' border between two meadows.

Based on the changes of the habitat characteristics after remeandering, the prediction of the habitat suitability of four macroinvertebrate taxa was tested (Table 6.2): *Asellus*, *Gammaru*s, *Erpobdella* and *Baetis*. *Baetis* is an indicator of good water and habitat quality, while *Erpobdella* occurs in more impacted streams. For the four taxa, the following conclusions could be made: remeandering had no significant effect on the probability of presence of *Asellus* and *Gammarus*, while on the contrary, an increase of habitat suitability was detected for *Baetis*, whereas a decrease was predicted for *Erpobdella*. It can be concluded that this remeandering project could be valuable for improving ecological quality as well as for enlarging the water bearing capacity of the concerning systems. In this manner, this project is advantageous for the ecological value and the safety of the housings and fields against flooding, while not much agricultural area has to be sacrified for this type of works.

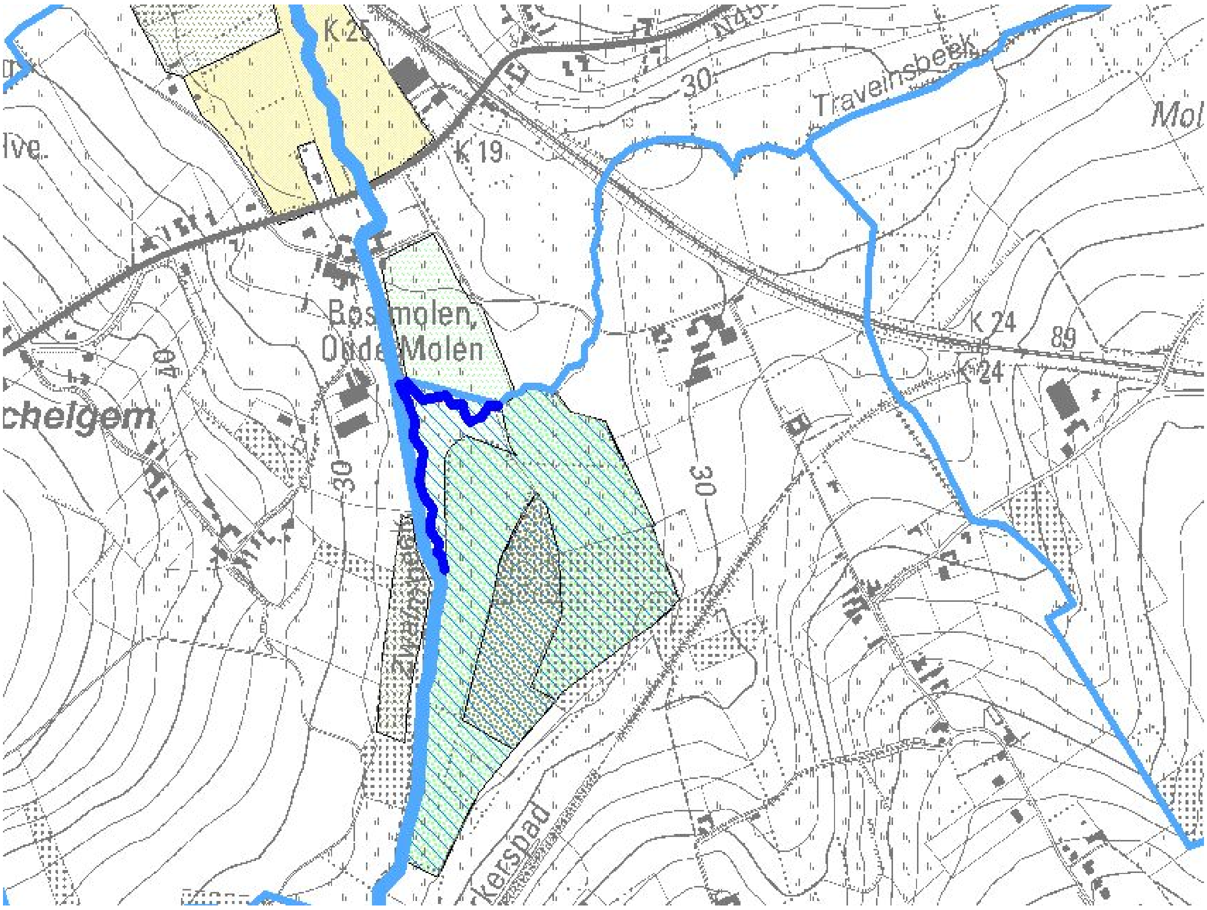


Figure 6.3 *Scheme of the planned remeandering action at Traveinsbeek. Near the Zwalm this river part was straightened to use it as a 'natural' border between two meadows (light blue). Belconsulting (2003) proposed to remeander the river and to install a natural flooding area (green zone) to minimise the use of a weir for water quantity control.*

In Table 6.2, the outcomes of the predictions of the different folds are mentioned. The sites 120 and 121 were only monitored during the year 2003 (with the purpose to make this type of restoration modelling exercises). Table 6.2 reveals that the neural networks perform well to predict the four taxa under the actual conditions (very similar to the training conditions in other words). However, when predictions were made for the restoration conditions, the model outcomes seemed to give the expected results compared to the similar site for *Asellus* and *Gammarus*, but not for *Erpobdella* and *Baetis*. Although only slight differences existed with a site used as example for 14 out of 30 variables (while 16 were nearly similar of that example site), the model performance decreased: several folds contained wrong predictions, illustrating quite some uncertainty on the used models. However, when bringing the results of the three models together, they seem to perform in general well when voting is applied. Most probably, when predictions have to be made on conditions of which no more or less similar sites are available in the database, the predictions will probably be less reliable. In particular when no physical habitat variables are involved (e.g. width), because for this variable most techniques seemed to take this as an important variable into account as was presented in the previous chapter.

Table 6.1 Expected values of the altered stream characteristics (in total 30 variables are used as input, only the altered variables are presented in this table). These were obtained on the basis of monitored conditions about 1 km upstream site 56 in the Traveinsbeek in combination with expert knowledge. All three sites will probably look similar regarding most physical habitat aspects after restoration.

Variables	**SP 56 actual**	**SP 56 restored**	**SP 120 actual**	**SP 120 restored**	**SP 121 actual**	**SP 121 restored**
Average width (cm)	900	144	270	144	600	144
Meandering	5	4	5	4	5	4
Pool/Riffle pattern	5	4	5	4	5	4
Hollow banks	6	5	3	5	4	5
Average depth (cm)	22	50	73	50	72	50
Average flowvelocity (m/s)	0.17	0.28	0.03	0.28	0.03	0.28
% Boulders	0.0	24.0	0	24	0	24
% Gravel	29.1	68.0	0	68	0	68
% Sand	3.9	2.0	15.5	2	15.5	2
% Loamclay	67.0	6.0	84.5	6	84.5	6
Minimum width (cm)	900	120	250	120	600	120
Maximum width (cm)	900	170	290	170	600	170
Minimum depth (cm)	10	10	40	10	5	10
Maximum depth (cm)	40	15	110	15	155	15
Minimum flowvelocity (m/s)	0.01	0.11	0.03	0.11	0.03	0.11
Maximum flowvelocity (m/s)	0.27	0.45	0.03	0.45	0.03	0.45

Table 6.2 Actual and expected taxa presence/absence values according to the 30-10-1 ANN models trained with measurements of 30 river characteristics in 60 sites during the period 2000-2002 in the Zwalm river basin. As input values of the ANN models the measured conditions in combination with the altered input values of the variables in Table 6.1 were used. 'Observed at similar site' means: Observed at site similar as will be expected based on environmental variables.

Asellus				
Site (year)	**Observed**	**Predicted (under actual condition)**	**Predicted (under altered condition)**	**Observed at similar site**
56 (2000)	present	present (3/3)	present (2/3)	present
56 (2001)	present	present (3/3)	present (2/3)	present
56 (2002)	present	present (3/3)	present (3/3)	present
120 (2003)	present	absent (3/3)	absent (3/3)	present
121 (2003)	present	absent (2/3)	present (3/3)	present
Gammarus				
Site (year)	**Observed**	**Predicted (under actual condition)**	**Predicted (under altered condition)**	**Observed at similar site**
56 (2000)	absent	present (3/3)	present (2/3)	present
56 (2001)	present	present (2/3)	present (3/3)	present
56 (2002)	present	present (3/3)	present (3/3)	present
120 (2003)	present	present (2/3)	present (3/3)	present
121 (2003)	absent	absent (3/3)	absent (2/3)	present
Erpobdella				
Site (year)	**Observed**	**Predicted (under actual condition)**	**Predicted (under altered condition)**	**Observed at similar site**
56 (2000)	present	present (3/3)	present (3/3)	absent
56 (2001)	present	present (3/3)	present (3/3)	absent
56 (2002)	present	present (3/3)	present (3/3)	absent
120 (2003)	absent	present (2/3)	absent (2/3)	absent
121 (2003)	absent	present (3/3)	absent (2/3)	absent
Baetis				
Site (year)	**Observed**	**Predicted (under actual condition)**	**Predicted (under altered condition)**	**Observed at similar site**
56 (2000)	absent	absent (3/3)	absent (3/3)	present
56 (2001)	absent	absent (3/3)	absent (2/3)	present
56 (2002)	absent	absent (3/3)	absent (2/3)	present
120 (2003)	present	absent (3/3)	absent (3/3)	present
121 (2003)	present	absent (3/3)	absent (3/3)	present

6.2.3 Removal of a weir for water quantity control in the Zwalmbeek

The sites on the Zwalmbeek near the Boembekemolen (Figure 6.4) are characterized by a modification of the flow channel due to a flood control weir (between sites 25 and 55 in Figure 6.1). The site 55 in particular is drastically deepened. Just in front (upstream) of the weir, the depth can be nearly around 2 meters (depending on the control level of the weir and the amount of sediments accumulated at the site). During the measurements in site 55 an average depth of around 1 meter was recorded, what is much deeper than under natural conditions (Table 6.3). Also the flow velocity is reduced drastically, resulting in direct and indirect impacts on the ecology. A direct impact is that the shear stress is quite different, being an advantage for *Asellus* for instance, but for some animals like *Baetis* who can profit from a continuous water flow over their gills on the back of their body, these artificially induced conditions are less optimal. The indirect effects of the lower flows can play a crucial role for the river biology as well. As a result of the lower flow, there is a serious accumulation of sediments (as a result of the erosion problems in that area), containing organic materials and probably also toxic materials (such as pesticides from agricultural soils). The organic compounds are degraded by the local microbiota, reducing the amount of oxygen in the water in particular near the bottom of this deepened system. This oxygen increase is not taken into account in case of weir removal in the presented simulations, because it is very difficult to predict this value on the basis of expert knowledge without a water quality model. Additionally this variable played not an important role in the ANN models in the Zwalm for *Asellus* and *Gammarus* as was presented in the previous chapter.

In Table 6.4, the actual and expected taxon presence/absence values (between brackets the amount of folds out of a total of three that support the outcome are indicated) are presented. These expected values are calculated according the 30-10-1 ANN models trained with measurements of 30 river characteristics in 60 sites during the period 2000-2002 in the Zwalm river basin (similar to the model development methods in Chapter 5). As input values of the ANN models, the measured conditions in combination with the altered input values of the variables in Table 6.3 were used. In total the values of 14 variables were altered compared to the actual conditions. ‘Observed at similar site’ means: observed at a similar site about 1 km upstream of the Boembeke weir, as will be expected based on environmental variables. On that site, the effect of the weir is nihil, and therefore it is a good site as basis for comparison and validation (also under the unaltered water quality conditions presented in this simulation exercise).

Table 6.3 Expected values of the altered stream characteristics (in total 30 variables are used as input, only the altered variables are presented in this table). These were obtained on the basis of monitored conditions about 1 km upstream the weir in combination with expert knowledge. All three sites will probably look similar regarding most physical habitat aspects after restoration.

Variables	**SP 25 actual**	**SP 55 actual**	**SP 55 restored**
Average width (cm)	323	630	400
Pool/Riffle pattern	4	5	5
Hollow banks	2	3	6
Average depth (cm)	18	110	70
Average flowvelocity (m/s)	0.87	0.04	0.05
% Gravel	98.0	0.0	11.0
% Sand	0.4	6.5	20.0
% Loamclay	0.3	62.0	37.5
Minimum width (cm)	280	630	400
Maximum width (cm)	380	630	410
Minimum depth (cm)	6	50	45
Maximum depth (cm)	30	170	90
Minimum flowvelocity (m/s)	0.77	0.04	0.04
Maximum flowvelocity (m/s)	1.02	0.04	0.07

In Table 6.4 the outcomes of the models are presented over the three folds (also here voting is used to label the altered sites with the presence or absence of the taxa). In between brackets the number of folds supporting this presence/absence label is indicated. For *Asellus* the models seem to perform very poorly. Even the predictions of the actual conditions seem to be very difficult for the trained models. Probably, this has partly to do with the relatively low number of sites affected by weirs in the database. For *Gammarus* the actual conditions can be classified well on the basis of the environmental variables, but the model does not support the observed absence in the site similar to the expected conditions. For *Erpobdella* and *Baetis*, the model outcomes are very instable. Overall, only a shift in *Gammarus* is predicted based on the used reference sites, according to the models on the contrary, *Gammarus* becomes present and *Asellus* disappears. The latter is not so unlikely based on the expert knowledge in Chapter 4, that also described quite some controversy about the effect of flow velocity on *Asellus*.

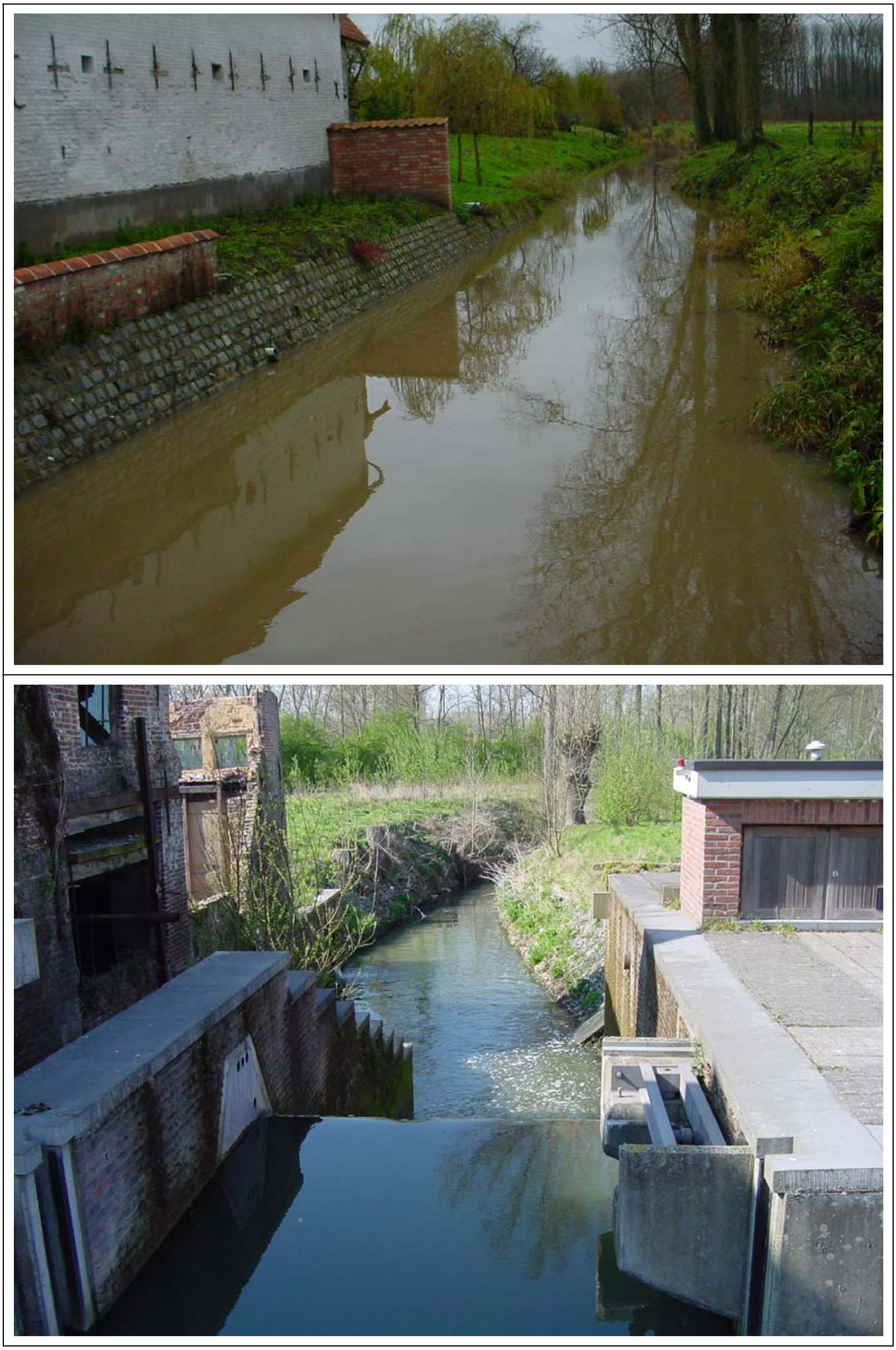


Figure 6.4 Picture of the upstream (top) and downstream (bottom) part of the weir in the Zwalmbeek in its actual condition (August 2003).

Table 6.4 Observed and expected taxon presence/absence values (between brackets the amount of folds out of a total of three that support the outcome) according the 30-10-1 ANN models. As input values of the ANN models the measured conditions in combination with the altered input values of the variables in Table 6.3 were used. 'Observed at similar site' means: Observed at a site similar about 1 km upstream the Boembeke weir, as will be expected based on environmental variables.

Asellus				
Site (year)	Observed	Predicted (under actual condition)	Predicted (under altered condition)	Observed at similar site
25 (2000)	present	absent (2/3)	absent (2/3)	present
25 (2001)	present	present (2/3)	present (2/3)	present
25 (2002)	present	absent (2/3)	present (2/3)	present
55 (2000)	present	absent (3/3)	absent (3/3)	present
55 (2001)	present	absent (2/3)	absent (2/3)	present
55 (2002)	present	absent (3/3)	absent (3/3)	present
Gammarus				
Site (year)	Observed	Predicted (under actual condition)	Predicted (under altered condition)	Observed at similar site
25 (2000)	present	present (3/3)	absent (2/3)	absent
25 (2001)	present	present (3/3)	present (2/3)	absent
25 (2002)	present	present (3/3)	present (3/3)	absent
55 (2000)	present	present (3/3)	present (3/3)	absent
55 (2001)	present	present (3/3)	present (3/3)	absent
55 (2002)	absent	present (2/3)	present (3/3)	absent
Erpobdella				
Site (year)	Observed	Predicted (under actual condition)	Predicted (under altered condition)	Observed at similar site
25 (2000)	present	absent (3/3)	absent (3/3)	present
25 (2001)	present	absent (2/3)	present (2/3)	present
25 (2002)	present	present (3/3)	present (2/3)	present
55 (2000)	present	absent (2/3)	absent (2/3)	present
55 (2001)	present	present (3/3)	present (3/3)	present
55 (2002)	present	present (3/3)	present (3/3)	present
Baetis				
Site (year)	Observed	Predicted (under actual condition)	Predicted (under altered condition)	Observed at similar site
25 (2000)	absent	absent (3/3)	absent (3/3)	absent
25 (2001)	absent	present (2/3)	present (2/3)	absent
25 (2002)	absent	present (2/3)	absent (2/3)	absent
55 (2000)	absent	absent (3/3)	absent (3/3)	absent
55 (2001)	absent	absent (2/3)	absent (2/3)	absent
55 (2002)	absent	absent (2/3)	present (2/3)	absent

According to the Bayerisches Landesamt für Wasserwirtschaft (1996), Asellidae are mentioned to behave as indifferent along water velocity, while Tachet et al. (2002) mention the preference for downstream sections characterized by low flow velocities. Also Peeters (2001) mentions that *Asellus aquaticus* attempts to flee from sites with higher flow stress or that repeated passive drift took place. So whether the models or the selected similar site are correct is hard to decide on. Quite a lot will probably depend on the remaining water quality.

Possibly these bad predictive results are also related to the difficulty to fix the input variables. The need for water quality and quantity models seems indeed very important. Nevertheless, the exercise is a very interesting validation on how far one can get with these data driven models and show the limitations under particular practical conditions.

6.3 Selection of monitoring sites based on predictive errors of data driven models

The performance of the models in the previous simulation exercises was highly variable. To get a better insight in what type of data are needed in the future to improve the models (and where these can be collected), maps were constructed indicating where the highest predictive errors occurred according to the Zwalm database in 2000-2002 (Figure 6.5). The predictions were made with one ANN model on the three years database.

For *Gammarus,* the largest quantity of errors occurred in the main stream (the Zwalmbeek). Perhaps this is related to the complex and dynamic set of impacts that are integrated at these sites, as it is a sink of the whole basin surface water and pollution sources. Seen the taxon is vulnerable to river pollution, this can give spatial-temporal patterns that might be very difficult to explain. Perhaps the coupling with a water quality model is therefore a solution, to get better insights in the pollutant dynamics.

In the case of *Asellus,* the presence/absence predictions contained most errors in the upstream parts, scattered over the whole river basin. The cause of this is probably related to very specific conditions because the number of errors is also rather low (only at one site two errors are detected over the three years dataset).

Based on the outcomes of both taxa, it seems for *Gammarus* interesting to have more data on the main stream, while for *Gammarus* no more particular data seem to be necessary when looking at the errors. Preferentially, new sites are selected on the stem river, to confront the methods with slightly altered conditions. Perhaps also sites from other river basins, e.g. the Dender river basin, could be added to the future data collection.

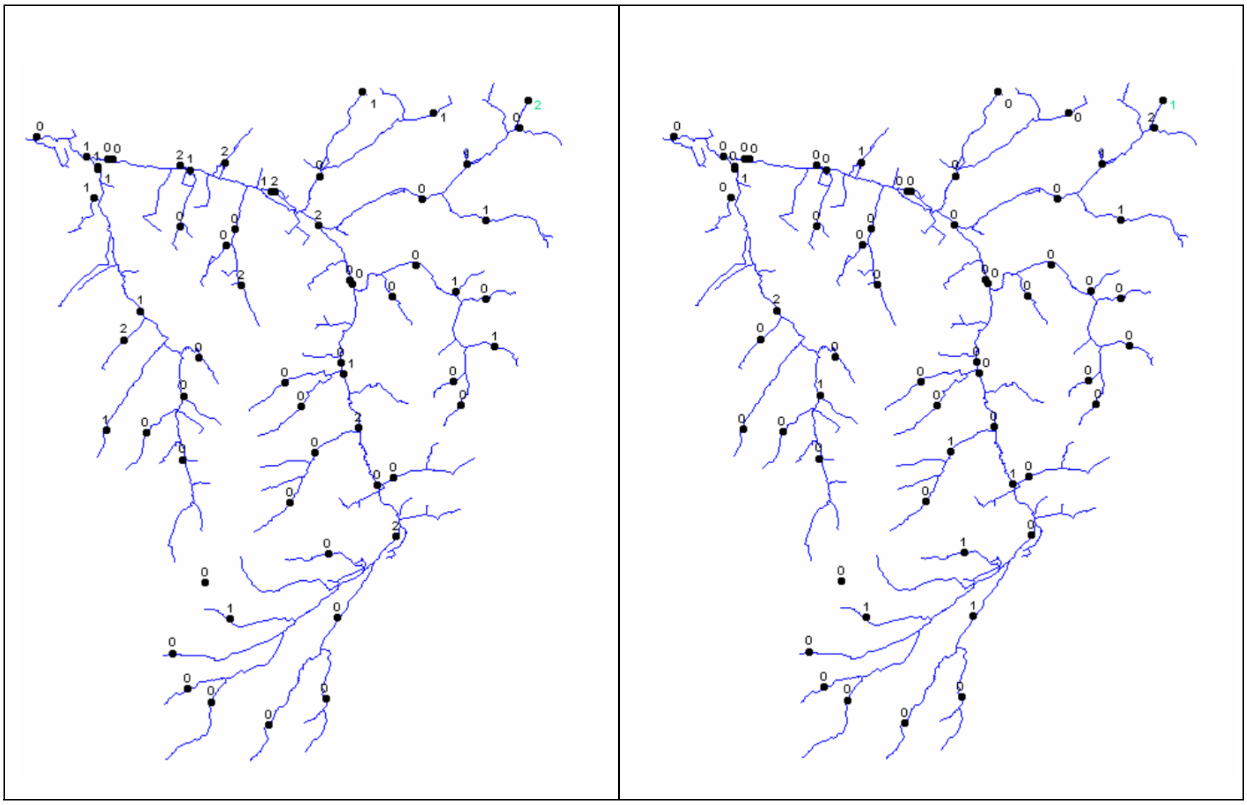


Figure 6.5 *Predictive errors (based on presence/absence ANN models) in the 179 instances, as a result of model calculations (and comparisons) on the field observations of 2000-2002 in the Zwalm river basin. The number of wrong Gammarus predictions is presented in the left map, the right one presents these of the Asellus models.*

6.4 Delivering insight in the habitat preferences of taxa

When searching for expert knowledge on the habitat preferences of taxa, some characteristics are confirmed by several authors, while others remain vague or are even contradictory. This is also the case for *Gammarus* and *Asellus*, as was experienced during the construction of the knowledge base of both taxa for Chapter 4.

Therefore, to check how the general responses of these organisms are according to the collected data in the Zwalm (2000-2002), some sensitivity curves were repeated here and the stability over the three folds was checked. The curves are based on log(abundance+1) ANN models, as were developed in the previous chapter by making use of the profile method. This method was proposed by Lek et al. (1995, 1996a, b) and the general idea is to study each input variable successively when the others are blocked at fixed values. As in Chapter 5, also here the 24 environmental variables were used as inputs of the ANN models and the four major variables (according to this profile method, determined on the basis of the difference between the minimum and maximum log(abundance+1) covered by the curves) were focussed on here.

According to several literature sources, *Gammarus pulex* appears in all types of waters: lakes, headwaters, river tributaries, canals, etc. (Karaman and Pinkster, 1977; Hawkes, 1979; Verdonschot, 1990; Peeters, 2001). Peeters (2001) developed a logistic regression model for *Gammarus pulex*. The order of importance of the variables in his study were: current velocity, Kjeldahl nitrogen, pH and depth. This order was however not confirmed in this study, although the same variables were present in the database. According to the profile method applied in the previous chapter the major variables for the Zwalm river basin data set were: total phosphorus, ammonium, phosphate and COD. Nevertheless, the curves are characterized by quite a large variability when comparing the three folds. The trends stay similar, however the intensity of the gradient in one fold is clearly much lower than for the other. When comparing the graphs to the outcomes of Peeters (2001), the same relation was found for the four variables: the lower these pollution indication variables are, the better for the *Gammarus* populations (within the range of the observations). For the nutrients this was also confirmed by the gathered expert knowledge, e.g. *Gammarus pulex* is less tolerant to inorganic pollutants and to organic sewage (Whitehurst and Lindsey, 1990). However related to the COD, the descriptions are often merely qualitative and indirect. Examples are: '*Gammarus pulex* is suppressed by high organic conditions (Hawkes, 1979), but can stand some organic pollution according to Gledhill *et al.* (1993)' and '*Gammarus pulex* prefers substrate heterogeneity (Tolkamp, 1980), especially detritus substrates or detritus mixed with sand or gravel or leaf material (Tolkamp, 1982)'.

Asellidae have a preference for watercourses with higher width according to the Macrofauna-atlas of North Holland (1990). Also Tachet et al. (2002) mention the preference of *Asellus* *aquaticus* for downstream sections (characterized by low flow and higher width velocities), what is also confirmed by the sensitivity curves c and d in Figure 6.7 in a quantitative manner. Downstream sections are indeed positively related with higher stream orders and a lower distance to mouth, and in this manner the preference for downstream sections is reflected by the log(abundance+1)-curves of *Asellus*. The curves for this taxon are much more stable over the three subsets compared to the ones of *Gammarus*.

According to these sensitivity analyses, one is able to see that the ANN models can reflect ecological relations in a quantitative manner, but that some instability is still involved in the case of the Zwalm exercises. Therefore the causes for this have to be searched for, and probably a combination of the data set and the model training configuration can be optimized to reduce the variability of the curves in future.

6.5 Discussion

Based on these data driven models, calculating the effect of future river restoration actions on aquatic ecosystems and supporting the selection of the most sustainable options seems to be part of the options. The simulation exercises at the Traveinsbeek and Boembekemolen taught us that depending on the type of problem, the added value of the models can differ significantly. These simulations revealed that validating the models for such exercises itself is also difficult, and the only manner to really validate the models is to follow up these restorations in case they are practically executed. Perhaps the use of artificial rivers might be a solution to this as well. Anyway, this type of studies needs much more models concerning other biological communities to make a full assessment of the overall ecological effects. Also the coupling with water quantity and quality models is necessary. The environmental variables of the rivers are sometimes very difficult to fix based on other sites or expert knowledge. The coupling of models will be necessary to get insight in the interactions that take place as a result of changing habitat characteristics or pollution levels.

However, making these data driven models more stable is a prerequisite. For this, more research is necessary to find the causes of wrong predictions. The predictive errors themselves can help to find the major gaps in our knowledge of river systems and help to set-up cost effective monitoring programmes to improve the models, as was shown in 6.3.

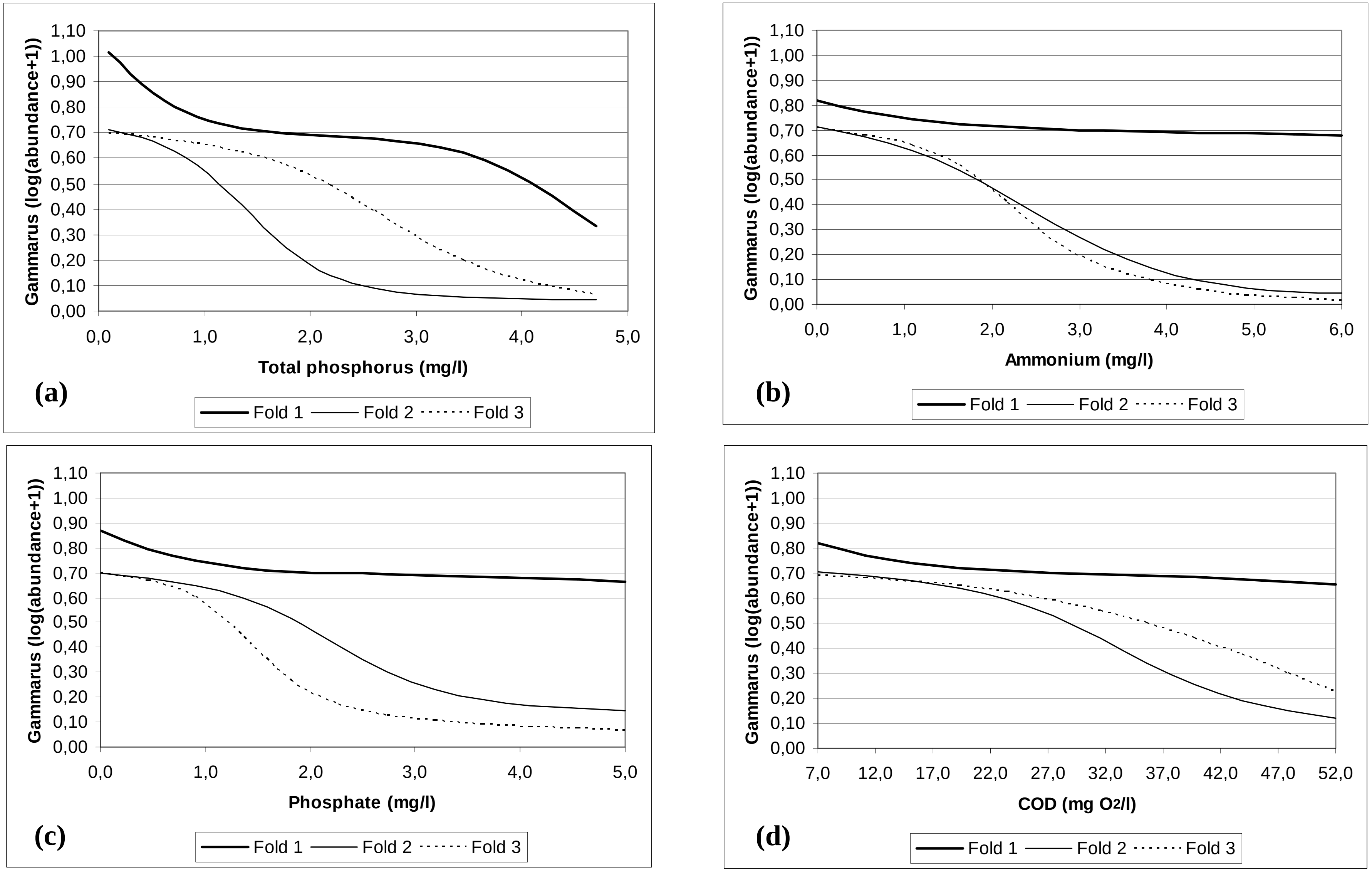


Figure 6.6 *Contribution of the three folds of the four most important environmental variables used in the 24-10-1 ANN model for Gammarus (based on the Zwalm river data) using the 'Profile' algorithm: (a)variable ' Total phosphorus'; (b) variable 'Ammonium'; (c) variable 'Phosphate'; (d) variable 'COD'.*

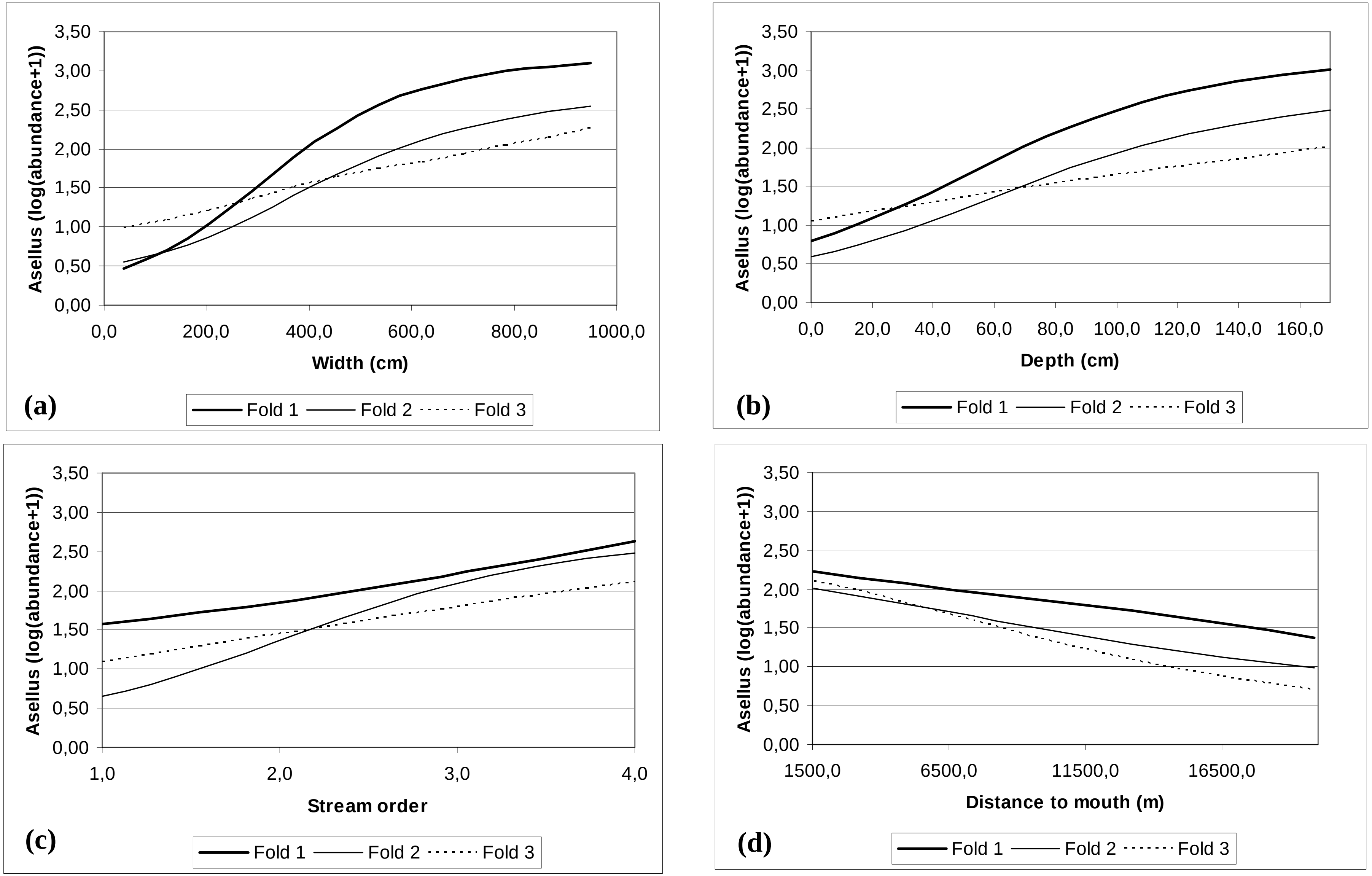


Figure 6.7 *Contribution of the three folds of the four most important environmental variables used in the 24-10-1 ANN model for Asellus (based on the Zwalm river data) using the 'Profile' algorithm: (a) variable 'Width'; (b) variable 'Depth'; (c) variable 'Stream order'; (d) variable 'Distance to mouth'.*

The research based on the sensitivity analyses with the profile method, indicates that the data driven models like ANNs can generate ecological valid information and deliver quantitative habitat information, what is in comparison with the often vague or contradictory expert knowledge from literature a serious step forward. However, when looking at all curves, sometimes striking outliers can be identified (e.g. one subset in the *Gammarus* curves). The cause of this needs further research, in particular on what is the cause (data (number of instances, outliers, natural diversity patterns, etc.), parameter settings or the techniques themselves) and what can be done to solve them (other data collection methods, more data, outlier removal, other training algorithms or parameter settings, etc.).

6.6 Conclusions

The simulation exercises at the Traveinsbeek and Boembekemolen illustrated that depending on the type of problem, the convenience and added value of the data driven models can differ significantly. Based on the results of the previous chapter, it is obvious that the datadriven models need to be drastically improved to be practically used for decision support in river management. However, compared to the actual decision making process, the use of the models is providing already interesting new insights. However, one has to be aware of the simplifications that were made during the development process, what is also probably responsible for the limited performance of the models.

These simulations also revealed that validating the models for such exercises itself is difficult, because it is hard to fix the environmental variables of the rivers based on other sites or expert knowledge. The coupling with water quantity and quality models is therefore necessary.

Extension to other taxa and optimisation of these data driven models and more research are necessary to find the causes of wrong predictions. The predictive errors themselves can help to find the major gaps in our knowledge of river systems and help to set up cost effective monitoring programmes to improve the models.

The results also illustrated that data driven models like ANNs can generate ecological valid information and deliver quantitative habitat information. Nevertheless, also here the stability needs to be increased and the set of used environmental variables in the models needs to be enlarged.

Chapter 7
General discussion and further research

Parts of this chapter are redrafted after:

D'heygere, T., Goethals, P.L.M. and De Pauw, N. (in press). Genetic algorithms for optimisation of predictive ecosystem models based on decision trees and neural networks. Ecological Modelling.

Vandenberghe, V., Goethals, P.L.M., van Griensven, A., Meirlaen, J., De Pauw, N., Vanrolleghem, P. and Bauwens, W. (accepted). Application of automated measurement stations for continuous water quality monitoring of the Dender river in Flanders, Belgium. Environmental Monitoring and Assessment.

7.1 Introduction

This PhD study consisted of three major activities: data collection, data driven model development and the application of models to support decision making in river management. The aim of this chapter is to link the results and discussions in the previous chapters and present some general and practical manners with regard to the development and application of predictive models for decision support in water management. This chapter consists of the following parts:

- data collection;
- model development;
- model applications for decision support in water management;
- further research;
- general conclusions.

In other words, this chapter comes down to the set of typical evaluation questions: what has been done, what can be improved, how this can be achieved and what might be expected in the future.

7.2 Data collection

The type of variables that are collected have clearly a major effect on the derived data driven models. This study illustrated that for the same taxon, clearly different models are built by the same techniques when different databases are used. So, first, one has to know what type of variables that need to be collected in the field. Therefore, model development studies need to be based on questions from managers. Once these are identified and the necessary models and variables are known, a relevant data acquisition has to be set up. Too many static modelling exercises are still based on field data from observational studies, lacking a designed sampling strategy (Guisan and Zimmerman, 2000).

The results from previous chapters show that more variables are necessary (because several issues such as what is the effect of pesticide used in the Zwalm river basin on the stream ecology cannot be coped with so far) and that further standardization and quality control is needed to guarantee a convenient data driven model development. Important to know is also which variables need to be included in the models and what effect they have on the predictive performance and how the habitat suitability can be extracted from data. Although, data driven approaches, such as ANN models, have the ability to determine which model inputs are critical according to several literature sources mentioned in the Chapter 3, they clearly showed to be able to make ecologically valuable inferences of only about three to five variables, according to the results in this work (this can be related to the size of the dataset, but therefore more research is needed). In other words, presenting a large number of inputs to ANN models, and relying on the network to determine the critical model inputs, usually increases network size and after all only a limited set of variables are really used for the inferences (cf. results with input variable contribution methods). This has a number of disadvantages, for example decreasing processing speed and increasing the amount of data required to estimate the connection weights efficiently (Maier and Dandy, 2000), while having no additive effect to the reliability or practical applicability of the models. Therefore the appropriate selection of input variables is crucial, as well as the way they are presented to the neural network or other data driven method.

The use of structural class variables seemed to be crucial in the case studies in the Zwalm river basin, but is probably becoming relatively a more and more important issue in Flanders because of the improving water quality due to the implementation of wastewater treatment during the past fifteen years. In particular for *Asellus* these variables seemed to play a major role, probably also because this taxon is less vulnerable to (organic) pollution, but is rather fleeing from higher flow velocities (cf. ecological expert knowledge in Chapter 4). However, further research is needed on the scale at which these habitat variables should be measured and whether the other analyses (e.g. biology, stream velocity) should not be measured at microhabitat or pool/riffle scale to make a better description of the ecological processes going on in the river. However, practical constraints (like time and money) will be hard to overcome if more detailed habitat monitoring has to be established, especially when one likes to cover a large area (e.g. Flanders or complete Scheldt river basin). Therefore, from a practical point of view, instead of going into more detail to develop better performing models, it might be interesting to work directly at community level and predict whole communities at once, characteristics of communities (e.g. expected biodiversity) or even ecological indices. For example, Goethals et al. (2002) predicted the BBI for several restoration options in the Zwalm. Seen the experience river managers have with these indices, in this manner they also get a better understanding of the use of these models.

The use of several variables is not always that straightforward. The inclusion of compounds with a potential toxic effect is more difficult as was illustrated on the river sediments database of Flanders (1996-1998). Probably, the use of bio-availability models or direct toxicity (as was applied in the Chapter 5) is the most easy, but entails high sampling and analysis costs... Regarding bio-availability of metals, several components could be taken into account. The binding of metals to clay and organic matter is well known (e.g. as applied by Peeters, 2001), but in sediments also the amount of acid volatile sulfides (AVS) is important to know the bio-available amount of metals (often called SEM, what means simultaneously extractable metals) (Di Toro et al., 1990; Ankley et al., 1994; Ankley et al., 1996 in Vangheluwe et al., 2002). Attempts have been made to account for the bioavailable fraction by using normalization procedures based on key processes controlling the bioavailability of the chemicals of concern. For this, a bio-availability model was developed, called SEM-AVS Vangheluwe et al., 2002). In applying the SEM-AVS model for a specific metal, such as nickel, it has to be taken into consideration that ΣSEM represents the sum of different metals acting in a competitive manner when binding to AVS. Acknowledging the existence of competitive displacement kinetics the SEM-AVS model can be made metal-specific. The procedure that is used is to assign the AVS pool to the metals in the sequence of their solubility products. Ranked from the lowest to the highest solubility product the following sequence is observed for the following five metals for instance: SEM_{Cu}, SEM_{Pb}, SEM_{Cd}, SEM_{Zn} and SEM_{Ni}. This means that copper has the highest affinity for AVS, followed by lead, cadmium etc until the AVS is exhausted. The remaining SEM is that amount present in excess of the AVS. In this manner, the bio-availability can be much better predicted. Therefore, in recent analyses of the VMM, these components needed to calculate the bio-availability were also incorporated and the resulting database can support the development of in situ ecotoxicity models that can be compared to laboratory tests, and vice versa. In this context, it is probably interesting to check retrieved relations via data driven models by means of spiking tests in the river and laboratory tests e.g. artificial rivers (thus setting up experiments as indicated by Beauchard et al. (2003)). In this manner causal relations can be confirmed, what is crucial because datadriven models cannot distinguish between correlations and causal relations, and this can be dangerous when the models are applied as such without causal relation validation. The use of dynamical models will probably also become necessary in this context. AQUATOX (http://www.epa.gov/waterscience/models/aquatox), an example of a model that is already applied in the context of water management (but in particular on lake systems), is such a process-based or mechanistic ecosystem model that simulates the transfer of biomass, energy and chemicals from one compartment of the ecosystem to another. It includes different types of plants, invertebrates and fish, and also treats the biota as interacting with the chemical/physical system. This is done by simultaneously computing each of the most important chemical or biological processes for each day of the simulation period. AQUATOX can predict not only the environmental fate of chemicals in aquatic ecosystems, but also their direct and indirect effects on the resident organisms (Park and Clough, 2004). Therefore it has the potential to establish causal links between chemical water quality and biological response and aquatic life uses.

On top of the extra variables that need to be measured to model all kind of relevant ecological processes to be able to cover different aspects in river management, it is extremely important that these data cover a broad enough range of all variables and that enough instances are collected. As such it does not make sense to include new variables when not enough data can be presented for the model development. Concerning the range of the data, in particular in Flanders there is a major lack of good river ecosystems, what makes it difficult to develop well performing models for restoration options and reference condition prediction. Therefore, probably more data from different river basins (e.g. international data) will be needed as well in future to enhance the models described in this study.

7.3 Model development

Based on the river sediments in Flanders database, the induced classification trees for both *Gammarus* and *Asellus* did not meet the requirement of K higher than 0.4, indicating that the trees' performance (relatively high CCI) is mainly related to the relatively low prevalence of both taxa in the database and related 'easy' classification, even without using environmental information. The CCI value of the classification ANN models was good for both taxa, while the *K* values were more or less acceptable (and slightly higher than those of the classification trees). The performance of the abundance models (regression) trained and validated on the same subsets had *r* values in the validation sets of about 0.4, indicating that the models were as well rather good. As such, all trained ANN models were more or less on the edge of good and not good, while the classification trees performed a little less, especially when the *K* values were taken into account. The ANN model for *Asellus* was in particular much better than the one based on classification trees.

Based on the data of the Zwalm river basin, the model performance was very different for both taxa, and also between the presence/absence and abundance models. For *Gammarus*, no reliable trees could be developed. For *Asellus* on the contrary, well performing trees were developed. Regarding the presence/absence ANN models, the CCI value was good for both taxa, while the *K* values were good for *Asellus* (0.62), but not for *Gammarus* (0.15 on average). It seems that the K values for both classification trees and ANN models were very similar. On average the *r* values of the abundance models were very good and constant over the three folds. As such, it seemed ‘easier’ to built good abundance models for *Gammarus* than to make presence/absence predictions in the Zwalm river basin.

When comparing the overall results, it seems that the ANN models were performing slightly better than the classification trees. On the other hand, the ecological meaning of the classification trees can be directly seen, whereas it takes some extra efforts to analyse the ecological meaning of the ANN models.

Actually, a lot of uncertainty exists on the development of data driven models. For example, in Dedecker et al. (2004a), different architectures, training methods etc. were compared, but the outcome was not easy to summarize in a set of simple rules of thumb. But also during the preparation of the data several options are open (e.g. remove outliers or keep them in, what is an outlier for a particular database, remove variables with high correlation, etc.). Therefore, presently it remains rather unclear how to develop well performing data driven models based on a general set of rules. The existing rules of thumb are often not working well and trial and error is in most cases the only solution to find the most optimal model training (and input variables) for data driven techniques such as classification trees and ANNs. Also in this study, standard data mining software, consulting of data mining experts and a lot of practical experience (based on trial and error) was the major factor to develop well performing models. The performance expressed by the indices was sometimes highly variable for the different techniques and taxa in the applied databases. Also the extracted ecological knowledge did often not present new insights, but in most cases a confirmation of basic ecological expert knowledge (e.g. relations with physical habitat such as width, distance to mouth, etc.). On one hand it confirms that these techniques can extract sound ecological knowledge, but on the other hand the added value seems rather limited in the two databases, in particular for the two taxa on which quite a lot of studies are done and expert knowledge is available. However, for taxa for which no or limited knowledge is available, this can already mean an important step forward.

Automated development (via stochastic searches and other optimization algorithms), eventually followed by rule extraction might look like a valuable option for the future. This is already partially worked out by D’heygere et al. (2003, 2004) for classification trees and ANN models. However, so far only input variable selection was performed, and optimization of the parameter settings will be necessary in the future. Also larger databases will for sure be necessary to be able to use these optimization techniques in a valid manner.

An important step forward could be made if a good model development guideline would exist. There is a high demand for this, but the offers stay out… However, such documents could make these methods much more popular and increase the practical application and validation of the methods in ecology. For this purpose, a multi-disciplinary approach will be crucial: bringing river managers, mathematicians, applied informatics specialists, ecosystem scientists and data collectors together… and make them communicate!

7.4 Model applications for decision support in water management

To improve the reliability and efficiency of management actions in the future, decision support systems enabling cost-benefit analyses can play an important role. However, when constructing such systems, careful attention needs to be paid to the possibilities and limitations of the available water system models and how they can be integrated in a user-friendly simulation shell. For the latter aspect, the analysis of the requirements from the managers’ and stakeholders’ point of view is of crucial importance. This is probably why the development of models needed for water management already has a fairly long history, e.g. Young and Beck (1974), without a standard tendency to use models for river management, in particular in Flanders. Too many modelling studies are not validated in practice and as such do not prove to work properly or be able to give the information of interest to managers. van der Molen (1999) underlined that the acceptability of models by water managers is mainly defined by their perception of the practical value of the model. The latter involves more than a theoretical reliability, but also criteria such as user convenience, model transparency, etc.

Therefore, some practical simulations were set up in the last results chapter. This clearly showed that there are several needs for improvement, and not only the habitat suitability models themselves… For example, there is a need to make simulations of all taxa (or whole communities), but also other models are needed to predict the input variables of these habitat suitability models, such as land use, water quantity and quality models, etc. This type of practical studies also delivers insights in how to improve the data collection (e.g. inclusion of river morphology, toxicants, etc.) to develop models that are of practical use for decision support in river management. Also which sites (monitoring network) need to be monitored to reduce uncertainty can be one of the practical outputs of the models. How to decide which restoration options to choose based on the presented models is not yet possible. These exercises are rather interesting as validation instruments than practical tools for river managers in this stage (the models can be seen as prototypes to conduct some tests on). The models can give insights in shifts of indicator species, but first the stability of the models needs improvement and a link with other models is needed.

Involving users (river managers) in the model development process is not easy. This was worked out in the COST 626 project ‘European Aquatic Modelling Network’. The main objective of this project was to define and develop integrated methods and models to assess the interactions between aquatic flora and fauna and riverine habitats on reach scale and provide transferability to a catchment scale. Therefore a timetable (Figure 7.1) was worked out. Based on a first phase of making state of the art studies on data collection, model development and applications (user needs), several integrated key topics were defined to work on. Based on these more practical and focused working groups the discussions and products became more interesting and fruitful. The overall outcome of the project is a model user internet framework that enables river managers to define their problems based on a set of keywords (e.g. scale, stakeholders, etc.). The internet framework guides the user to a set of models that are described and also to who can be contacted to use these models. In addition, also a data internet framework is elaborated, to bring databases on river fish and macroinvertebrates together, but also to guide people to what kind of environmental characteristics need to be measured in addition to the biological samplings to make sensible ecological models. Also the related equipment and how to get it is part of this subproject.

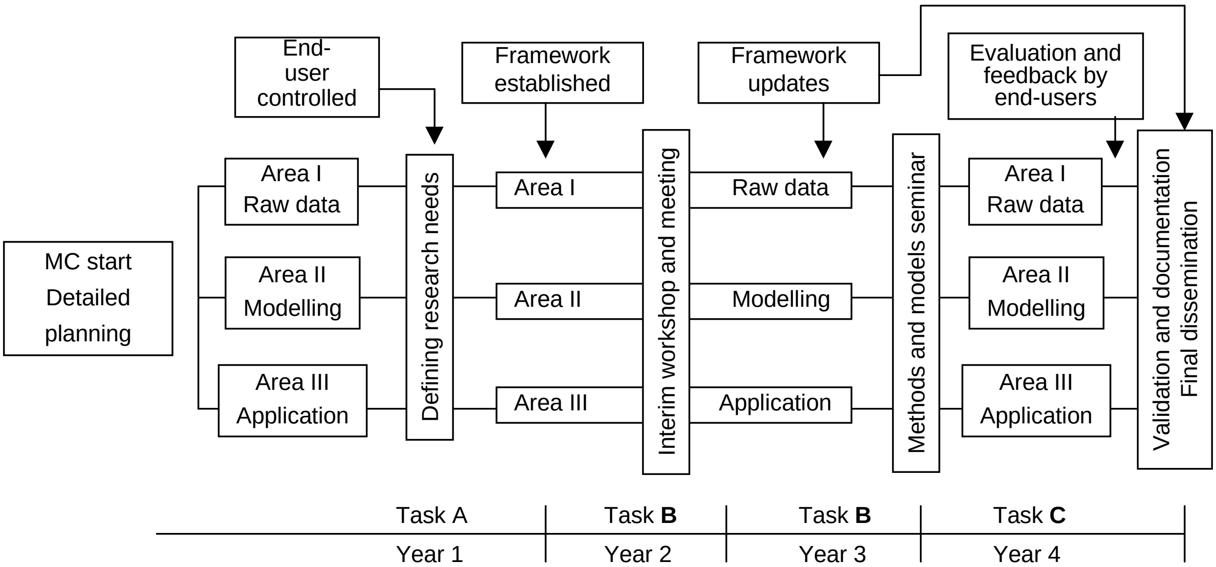


Figure 7.1 *Scheme of the COST626 project (European Aquatic Modelling Network) aiming at the stimulation of the use of models for decision support in water management.*

7.5 Further research

7.5.1 Overview of potential research

This component contains some features for further research. The following topics are discussed and suggestions how such research can be set up and some preleminar results are presented:

- improvement and extension of the model inputs via automated data collection procedures;
- extension of the habitat preference models with migration models;
- model evaluation and optimisation methods;
- linking ecological models to climate, land-use, discharges, river water quality and other physical models;
- linking ecological models to social-economical models and stakeholder information needs.

7.5.2 Improvement and extension of the model inputs via automated data collection procedures

In addition to earlier mentioned ecotoxicological and physical habitat variables (mentioned and discussed in the previous part of this discussion), also continuous measurement sets of some water quality and quantity variables might be interesting in several cases, seen the highly dynamical processes that can occur in rivers. The use of on-line measurement instrumentation is becoming more and more popular. In most cases however, the use of on-line instrumentation for water quality monitoring is applied in the field of wastewater treatment (e.g. Vanrolleghem and Lee, 2003). Only a limited set of studies describe the experiences of this type of systems for analysing water quality variables in rivers and lakes. According to these published experiences on on-line analysers in surface waters, it seems that the applications, the type of sensors as well as the concepts to combine several measurements are rather diverse. Several studies describe rather specialised measurements, e.g. TNT (2,4,6-trinitrotoluene) in Wang and Thongngamdee (2003), while in others the combination of different variables is of major importance to gain new insights in aquatic processes, e.g. Beck et al. (1998). In some studies the use of a minimal setup with low maintenance is promoted, e.g. Edwards (1998) and Johnson (1998), while in others rather complex and automated systems seem to give good results (Beck et al., 1998; Du Preez et al., 1998). Also the data handling process can differ a lot and is still in full development. In particular the use of the Internet for this purpose seems to be a very promising solution, as described by Toran et al. (2001). In particular, the reliability of the measurements needs further research, because most on-line sensors are less precise and accurate in comparison with laboratory instruments. Therefore it is necessary to first investigate the required reliability of the field measurements before selecting the type of instrumentation (Schlegel and Baumann, 1996; Leeks et al., 1997; Mohapl, 2000) and decide whether continuous measurements can have an added value for the purpose of the user.

The results of the research of Vandenberghe et al. (2004) at the Dender river in Flanders (Belgium) illustrated that automated measurement stations can reveal the limitations of contemporary water quality assessment networks and monitoring strategies. In view of the variability of the observations of the physical and chemical variables in time and space, as shown in Vandenberghe et al. (2004), it is clear that classical sampling with a monthly periodicity on a restricted number of locations by the Flemish Environment Agency (VMM) should be questioned for the determination of a physical-chemical water quality assessment (and the usefulness of their measurement to develop ecological models). Clearly, a lot of attention should be paid to the sampling location(s) and the timing of the sampling (day vs. night; flow regime; hydro-meteorological conditions, etc.). In another study about the Dender by Vandenberghe et al. (2001) the usefulness of continuous measurements to reduce the model parameter uncertainty and thus to increase the reliability of the model predictions was demonstrated. The latter authors extended the research to the optimal design of measurement campaigns, in view of maximising the model reliability (Vanrolleghem et al., 1999), given the logistic and financial constraints. D'heygere et al. (2002) optimised the monitoring strategy for macroinvertebrates in the Dender river, making use of the insights in the seasonality of the biological as well as chemical processes. In this manner, this automated river quality monitoring can help to define which variables are useful and necessary to predict biological communities, because this is still one of the major difficulties in the development of ecological models for decision support in water management as was also illustrated in this PhD research.

In addition to the automated measurement stations, also the use of aerial photographs and digital maps (e.g. for physical habitat characteristics) can be valuable to improve the data input of such models as well as their performance and their set of applications. In this manner, the field work could be drastically reduced and also a more straightforward approach could be used to collect the data (e.g. more independent of the differences and uncertainties of field observation done by different people), and even allow to make variables to work at different scales. The link of river ecology with land use could be another aspect related to this.

7.5.3 Extension of the habitat preference models with migration models

The upstream and downstream movement of animals within and between habitats is of particular ecological significance. In running waters, the drift of benthic invertebrates is a well-studied phenomenon (Waters, 1965; Brittain and Eikeland, 1988). Upstream migration in many different taxa on the other hand, has been reviewed by Söderström (1987). Drift enables organisms to escape unfavourable conditions and gives them the potential to colonize new habitats (Brittain and Eikeland, 1988). Unlike drift, upstream movements are always active (Söderström,1987). Search for new habitat and food (Bishop and Hynes, 1969), avoidance of unfavourable abiotic conditions (Hayden and Clifford, 1974; Olsson and Söderström, 1978) and a compensation for drift in order to maintain a basal population within a certain habitat (Hayden and Clifford, 1974; Goedmaker and Pinkster, 1981) can be reasons for upstream movement. Discussing the downstream drift and upstream movement of macroinvertebrates without mentioning the 'drift paradox' is inappropriate. It is a well-known and frequently discussed concept in ecological literature. The 'drift paradox' arises because the upper reaches of streams remain colonized by aquatic insects despite an apparently considerable reduction in their numbers. This reduction is due to the tendency of aquatic invertebrates to drift downstream with the current (Brittain and Eikeland, 1988; Allan, 1995). The compromise has generally been that upstream flight (whether directed or as part of random, undirected dispersal) by some pre-ovipositing adult females may be the key factor that resolves the paradox (Hershey et al., 1993; Allan, 1995). But how do we explain the persistence of the many species that are commonly found to drift but do not have an aerial adult stage, such as for example *Gammarus pulex* (Humphries and Ruxton, 2002)? Speirs and Gurney (2001) and Humphries and Ruxton (2002) demonstrated, long-distance flight by aerial adults is not required to prevent extinction of upstream reaches. It can be achieved with very small movement of individuals along the substrate.

Dedecker et al. (2004e) developed such a migration model for *Gammarus pulex*. In preliminary studies, data driven models were tested and optimised to obtain the best model configuration for the prediction of the habitat suitability of for instance *Gammarus pulex* based on the abiotic characteristics of their aquatic environment in the Zwalm river basin (Flanders, Belgium). This migration model, implemented in a Geographical Information System (GIS), was used for the simulation of a practical river restoration scenario. After removing a weir, the negative effect on the habitat suitability of *Gammarus pulex* disappeared. The ANN models predicted that the habitat was suitable again for *Gammarus pulex*. The migration model indicated the restored parts of the river would be colonized within about two months. In this way, decision makers have an idea whether and when the restoration option has the desired effect.

The next step is the development of migration models for other macroinvertebrates. Indicator species for good water quality, such as the EPT taxa (Ephemeroptera, Plecoptera and Trichoptera), will get special attention. In preliminary studies, the factors affecting the migration capacity of these organisms will be examined. In contrast to the present migration model for *Gammarus pulex*, an additional layer for the migration over land/through the air will be necessary because most of these macroinvertebrates have an aerial adult stage. Finally, the scale of the developed models will be extended to the whole Zwalm river basin. Extension of the intensive monitoring campaign as done for the selected part of the river basin, should be very costly and time consuming. Therefore, using aerial photographs or digital maps to extract the necessary information is recommended.

Dynamic models might also be needed to describe these migration processes (and other ecological interactions such as competition, food web interactions, etc.), because some of them are steered by seasonal influences, others by the interaction in and between the different communities or by escaping activities (e.g. from pollution or other suboptimal habitat conditions such as light penetration). Nevertheless, the amount of knowledge and data to develop, train and validate this type of models is probably a major difficulty for their successful application, at least in the near future. For this, also coupling with other models will be necessary (see further).

7.5.4 Model evaluation and optimization methods

The choice of an evaluation measure should be driven primarily by the goals of the study (Guisan and Zimmerman, 2000). Therefore, the single use of performance indices is insufficient for setting up data driven models for use in decision support. So a set of different measures is necessary. This may possibly lead to the attribution of different weights to the various types of prediction errors (e.g. omission, commission or confusion). Testing the model in a wider range of situations (in space and time) will permit one to define the range of applications for which the model predictions are suitable. In turn, the qualification of the model depends primarily on the goals of the study that define the qualification criteria and on the usability of the model, rather than on statistics alone. (Guisan and Zimmerman, 2000). So as described in another part in this further research component, the decision support aspects and relation with users and stakeholders is also crucial in the optimization process.

Once one is able to select the appropriate methods for evaluation, a model optimization strategy can be selected. This was also done for the Zwalm river basin during the last years. Each year the data collection was analyzed and new variables were measured based on detected model shortcomings. In other words, the use of the Deming circle (continuous improvement methodology) looks like a good option to improve the data collection, model development procedures and their application step by step. This is a very well known strategy in applied data mining cf. Figure 7.2 provided by the Software Competence Centre in Hagenberg.

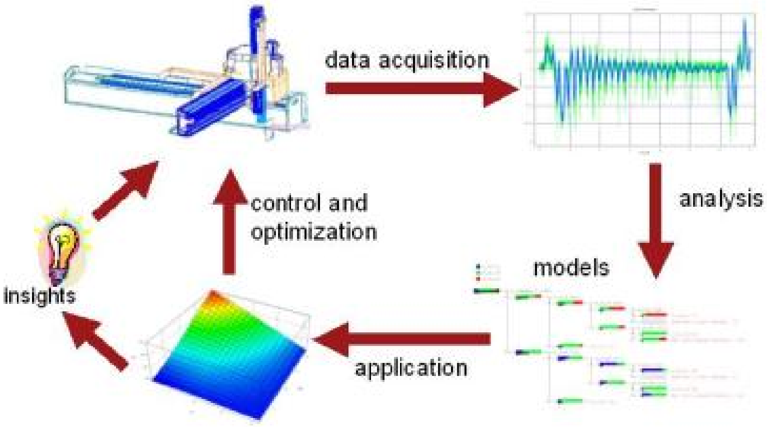


Figure 7.2 Model optimization strategy provided by the Software Competence Centre in Hagenberg.

So the goal of the research on the development of data driven models for decision support in river management can not be limited to merely comparing data mining techniques on their qualities and all technical options etc. in detail, but also how these tools can gain insight in what new variables are needed for field measurements and what type of predictions are feasible to support decision making.

The selection of appropriate input variables in predictive ecological modelling is an important issue since numerous variables can be involved. Most of the input variables cannot be omitted because it results in a significant loss of information. The collection of field data on the other hand is both time-consuming and expensive. Rigorous methods are therefore needed to decide which explanatory variables or combinations of variables should enter the model. 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 selection of the input variables is in anyway a crucial part in the optimization process. Although in the present study similarities could be detected between the different contribution methods and as well between the used variables in the classification trees, some striking differences also appeared. Therefore a more quantitative procedure is needed to compare the different contribution techniques, as is recently proposed by Olden et al. (2004), using Monte Carlo methods. But a technique which can as well be used is genetic algorithms (e.g. Obach et al. 2001; Schleiter et al. 2001; D'heygere et al. 2004). This algorithm automatically selects the relevant input variables (Goldberg, 1989). In the study of D'heygere et al. (2004), the use of genetic algorithms is explored to automatically select the relevant input variables for classification trees and artificial neural networks (ANNs), predicting the presence or absence of benthic macroinvertebrate taxa. The applied database consisted of measurements from 360 sites in unnavigable watercourses in Flanders, the same database as was used in this PhD study (but without metal variables). As is shown in Figure 7.3 (D'heygere et al., 2004), the average merit or CCI of the applied ANN models started off at about 85.3% and increased up to 92.4% at generation 40 during the optimization process. A highest CCI was detected at generation 32 with a value of 93.3%. This variable subset is therefore the best to be selected as the result of the variable selection. The low peaks in the graph of the minimum CCI reveal that the probabilities for cross-over and mutation were set high enough to ensure that the probability of being trapped in a local minimum was avoided.

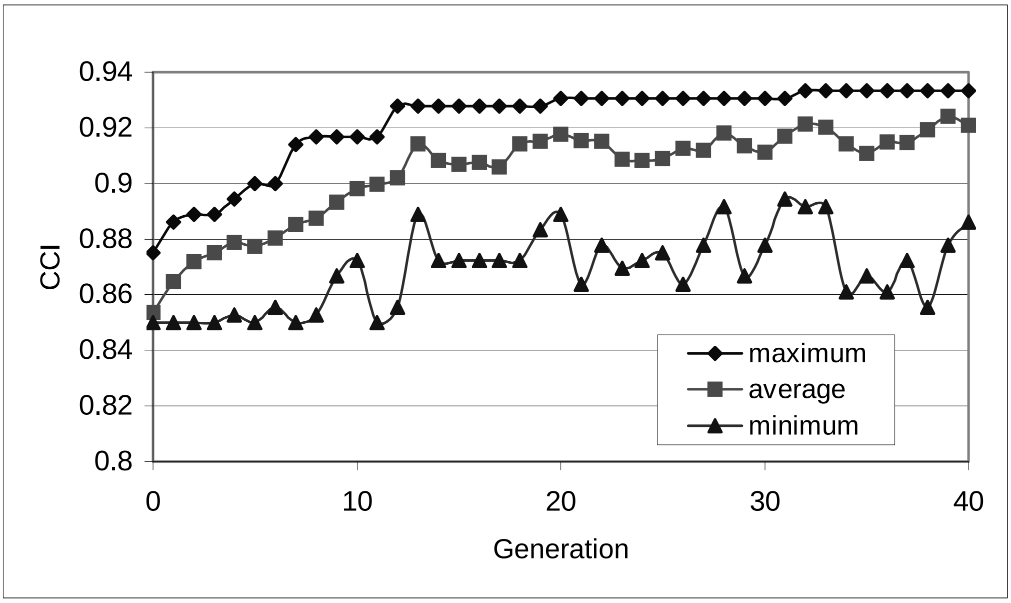


Figure 7.3 Evolution of the minimal, maximal and average CCI of a variable subset for every generation for the ANN of Gammaridae.

An assessment of the models performance based on the CCI with a statistical test revealed that except for *Pisidium* the performance increased significantly for all tested taxa with a fair model performance (based on Cohen's kappa) when the modelling was preceded by a variable selection stage (Figure 7.4). This was at first surprising because a loss in information in a data-driven approach should result in a loss in performance. However, the pruning of models seems often to give an improvement of the model performance, because noise data are removed. This was also confirmed in this PhD study. The largest classification trees never were the best performing according the CCI and *K*.

In the study of D'heygere, the most important variables for the ANN models were day, depth, current, water temperature, saturated oxygen, conductivity and the 72h growth-inhibition test with *Tamnocephalus platyurus* (TOXT) (Table 7.1).

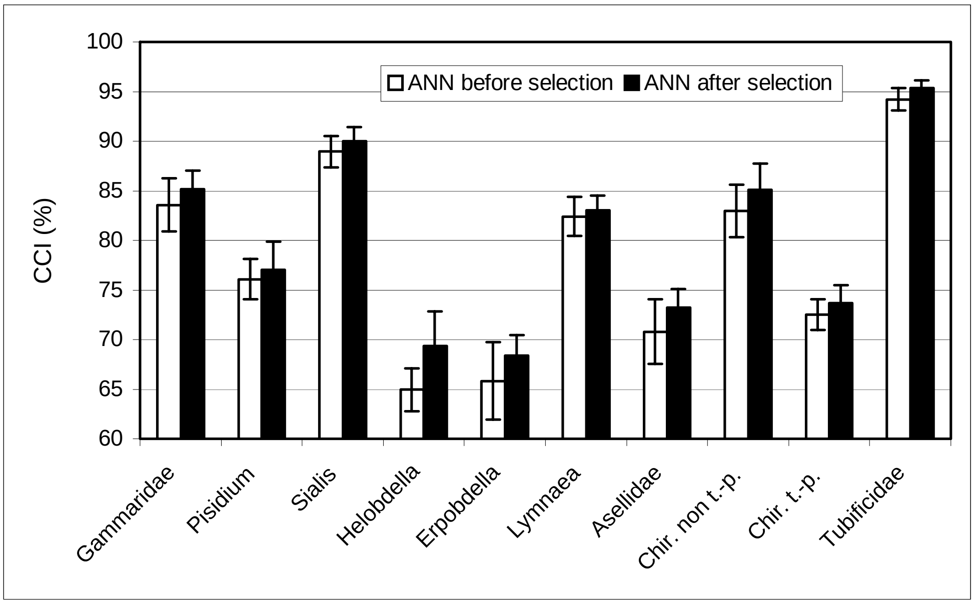


Figure 7.4 CCI of ANN models for 10 benthic macroinvertebrate taxa before and after variable selection.

However, this is only a first step in the automation process. The next part of the research involves the optimization (tuning) of the parameter settings of the data mining tools to the set of input variables. As such, also an optimization algorithm (e.g. genetic algorithm) can be used to combine the selection of input variables and the best model development technique. One step further will consist of a quality function deployment, that helps to translate the quality criteria of river managers into mathematical goal functions, that can be used for the optimization algorithms. As such, depending on the model market, the best models can be produced for the different sets of customers (river managers) in an automated manner.

Table 7.1 Selected variables by means of the Goldberg genetic algorithm with ANNs as an evaluation function for 10 benthic macroinvertebrate taxa in river sediments in Flanders (1996-1998). (based on D'heygere et al., 2004)

	Gammaridae	***Pisidium***	***Sialis***	***Helobdella***	***Erpobdella***	***Lymnaea***	**Asellidae**	**Chir. non t.-p.**	**Chir. t.-p.**	**Tubificidae**
Day	*	*	*	*	*	*	*	*	*	
Width	*			*		*	*			
Depth	*	*	*	*	*	*	*	*	*	*
Flowvelocity	*	*	*	*	*	*	*	*	*	*
Clay		*				*		*		
Loam	*	*				*		*		
Sand	*	*			*	*	*	*		
Temperature		*	*	*	*	*		*	*	
pH	*	*		*	*					
DO (%sat)	*	*	*	*	*	*	*	*	*	*
Conductivity		*	*	*	*			*	*	*
TOXT	*	*	*	*	*	*	*		*	*
TOXR			*		*	*		*		
OM										
KjeldahlN	*		*		*	*		*		*
TotalP										

7.5.5 Linking ecological models to climate, land-use, discharges, river water quality and other physical models

Recently, several practical concepts and software systems were developed related to environmental decision support, e.g. Rizolli and Young (1997); Paggio et al. (1999); Reed et al. (1999); Young et al. (2000); Argent and Grayson (2001); Booty et al. (2001); Lam and Swayne (2001); Argent (2004); Lam et al. (2004); Voinov et al. (2004); Poch et al. (in press). Ceccaroni et al. (in press) stress the importance of ontologies in this context for sharing and reusing knowledge, by careful consideration of the general and specific application areas of the applied models, to avoid a wrong extrapolation or coupling of existing knowledge or models. In particular when also knowledge from laboratory tests is to be included in the models, e.g. eco-toxicological relations as in Babut et al. (2003), the in-field relevancy needs to be checked.

From a technical point of view, one can opt to build a new model for each application or to utilize existing models where possible. The first approach has the benefit of control in the models design and linkage, but requires longer development time. The second approach saves on the development time, but requires additional work to link up existing models (Lam et al., 2004). However, when a lot of models are already available, it is probably the best option. The use of the linked models can also be a good start to gain the required knowledge in what processes are of major importance for the different simulations and which can be neglected. Thus model integration by the use of simplified and inter-tuned models can probably be a feasible option as well.

A major issue in this context is the selection of the most convenient inference technique, hereby considering a realistic process description, relevant outputs for the users and a low simulation time. More and more attention is herein paid to the link with the desired management information, in particular when also for this purpose data mining and modelling approaches (e.g. Chun and Kim, 2004) are used on that level of decision-making. The delivery of relevant data in a useful format is of major importance for the successful coupling of the model simulations with economic methods to value ecosystems and also requires an altered data collection approach for river management in Flanders.

7.5.6 Linking ecological models to socio-economical models and stakeholder information needs

Within the policy area of water management economic valuation can play an important role to analyse the costs and benefits for river restoration options. The use of models and DSS that allow a better allocation of the contribution of all stakeholders to the deterioration of the water system (water quality problems, floodings, ecosystem destruction, etc.) is an important step forward for river management (e.g. Denzer et al., 2000). These instruments can deliver the data needed for an integrated economic valuation of the water system and can help to obtain a more sustainable use of one of the most critical natural resources for mankind.

Goethals et al. (2003, 2004) presented in this context a concept to link the outcome of ecological models with economic valuation methods. The WAter ECOlogy Decision Support System (WAECO-DSS) (Figure 7.5) combines the computational strength and complementarities of different types of habitat suitability models. This can enable the user to perform reliable simulations at different spatial and temporal scales. Decision trees extract simple rules from large quantities of data, while ANNs are able to establish patterns and characteristics in situations where rules are not known. Fuzzy logic on the other hand allows to process unreliability and inaccuracy of data and to incorporate external expert knowledge (Adriaenssens, 2004), what is in particular useful to predict rare species, because the development of data driven models is not possible under data-poor conditions. To know when, how and in what sequences to use the models and data to solve specific problems is an essential part of the WAECO-DSS. This involves knowledge on how to perform spatial modelling and how to use a set of tools in combination for particular analytical purposes.

However, one has to be aware of several pitfalls and limitations when coupling DSS and ecological valuation methods for cost-benefit analyses in water management. The level of decision-making (regional or river basin) might strongly influence the economic value attached to human interventions in water systems. Clearly, at the more general and regional level only a limited overview of the effects on the ecological quality might be identified. The use of a DSS, which explicitly maps the effects for the whole water system, prevents the calculation of only a limited part of the total economic value of the human intervention. The use of the WAECO-DSS for the management of Zwalm and whole Scheldt for instance, might prevent sub-optimal decision making in the area of water management. A second important element has to do with selecting a representative set of stakeholders and their knowledge on aquatic ecosystems. During the cost-benefit analysis, only part of the set of stakeholders who are confronted with the changes due to the planned interventions might be involved in the valuation process (Hanley et al., 2003). The type of stakeholder that is involved and his relation with ecological water system quality and his relative contribution to the valuation results can generate drastic differences in the outcome, e.g. Alessa et al. (2003); Knowler et al. (2003). Fishermen will prefer a moderate (e.g. much white fish and big carps in a nutrient rich system) over a good river ecosystem quality (when expressed as an ecological fish index), because an improved quality would entail relocation costs to find another carp-abundant system. Consequently, the value they attach to further improve the river quality is negative. This also will be the case for farmers. Households that depend on the river for their drinking water supply will prefer the good quality, because this implies less investment in water purification systems.

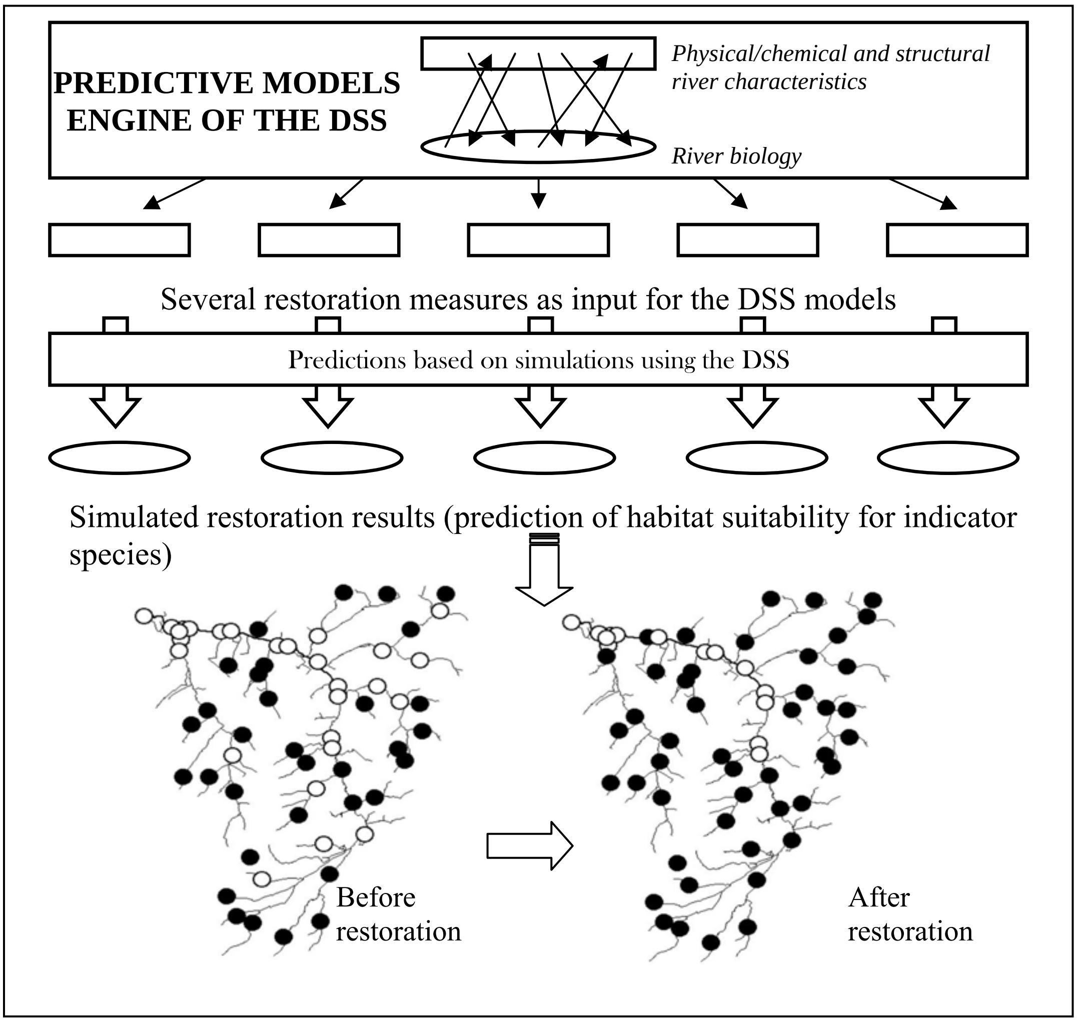

PREDICTIVE MODELS
ENGINE OF THE DSS
Physical/chemical and structural river characteristics
River biology
Several restoration measures as input for the DSS models
Predictions based on simulations using the DSS
Simulated restoration results (prediction of habitat suitability for indicator species)
Before restoration
After restoration

Figure 7.5 The WAter ECOlogy decision support system (WAECO-DSS) (Goethals et al., 2004).

However, when these stakeholders could be assisted by models explaining how different characteristics are interrelated, the values from the different stakeholders would probably be more consistent. That was also concluded by Mustajoli et al. (in press), who found out that when the stakeholders clearly understand each other's views, a consensus can be reached more easily. In this context, Myatt et al. (2003) stressed the need of public participation to facilitate acceptance of managed realignment schemes and to set up relevant criteria for this purpose. In this manner, DSS systems and valuation methods will not only be useful to calculate the outcome of potential management options, but can as well be interesting for stakeholders to gain insights in the river processes. However, one has to be aware that improving the information flow towards stakeholders could as well have an opposite effect. Most probably several river system users will be disappointed about what they have to sacrifice for particular ‘benefits’ of other stakeholders, and this can perhaps also just start new vivid discussions on the particular values. In particular when introducing such systems, (new) problems might at first arise rather than just get solved and one has to be aware of a period in which stakeholders need to get familiar with an extended view on the water system processes and an integrated management approach.

However, although the combined use of DSS and valuation techniques has certain limitations, in general one can conclude that there are also many advantages in comparison with the contemporary approach making merely use of ecological indices. In this manner, the use of DSS and economic methods to value ecosystems in cost-benefit analyses for water management can be seen as a best available technology to obtain a sustainable restoration of rivers.

7.6 General conclusions

This research dealt with the use of ecological modelling in water management. Data were collected in the field and two datadriven model development techniques, classification trees and artificial neural networks were compared. This PhD research illustrated that data driven ecological models can be interesting tools to get insights in the relations between river characteristics and the inhabiting biology. By getting more insight in the habitat preferences of different taxa, one can allocate indicator taxa for particular types of river deterioration (cause detection such as oil spills, diffuse pollution, etc.) and improve the development and optimization of ecological indices for the assessment of river quality. The models can be useful to make simulations of the potential ecological effects of river restoration options (as well as the effects of river deterioration) and can as such support the decision making process that river managers are daily facing. As such, they can mean one of the first steps to come to more integrated cost-benefit analyses in water management by providing the necessary insights in the expected effects of changing uses of rivers. In addition, the models can reduce the monitoring costs, as they can show the sites or type of streams where the highest uncertainty is present and where in particular new data should be collected.

However, up to now, still several difficulties have to be solved. These problems are related to data collection requirements, the optimisation of models and needs for new/other approaches and the communication with river managers to stimulate the practical application of water system models.

The need for better databases (more reliable and standardized data collection, inclusion of essential variables (e.g. several micropollutants, but also variables important to predict bio-availability of toxicants)) was one of the major problems that was encountered during this study. In this respect, the script illustrated that several variables are missing in the routine environmental monitoring networks and that the monitoring site selection could be drastically improved. For this purpose, a new modelling network (in the Zwalm river basin) had to be constructed to develop reliable and useful predictive habitat suitability models for macroinvertebrates. However, also this monitoring network could benefit from the inclusion of new variables to explain the presence/absence and abundance of the macroinvertebrates and for several vulnerable and rare taxa, the sites were too scarce to develop reliable models.

Therefore, several types of models will be necessary to cover all taxa. The data driven techniques used in this study are in particular interesting when a lot of data of good quality are available. Probably the size of the monitoring network needs to be increased to ensure that the derived models also work with more input variables. Therefore, it does not make sense to monitor a lot of variables, when there are not enough instances collected. For rare species, it is most likely that other modelling techniques such as fuzzy logic and Bayesian belief networks will be more convenient (because of the very limited instances where the species are present). However, also for this type of models the availability of proper and reliable expert knowledge is of crucial importance, as well as at least a good validation set. Also the inclusion of migration behaviour as well as competition and other types of interactions between the taxa needs to be incorporated in the predictive models in the future. Probably, also dynamic models can be established in this respect, having a much better performance, because all kind of feedbacks which are now neglected in the habitat suitability can be considered in that type of models. As such, (dynamic) interactions between species (e.g. predation, competition) and between the species and their environment (e.g. nutrient availability, growth in function of temperature and other climatic conditions) might be taken into account in a more convenient manner. Nevertheless, the amount of knowledge and data needed to develop, train and validate this type of models is probably a major difficulty for their successful application, at least in the near future. For this, also coupling with other models will be necessary and many water quantity and quality models are also still facing serious uncertainties up to now (perhaps just because the biological component is missing… e.g. biological components might have serious effects on degradation processes or use of nutrients). This integration of models to whole water systems will therefore be one of the key challenges in the future of water management.

Discussions with river managers (e.g. what do they want?) are as well of crucial importance. Often they are not interested in models that describe water systems in a very reliable and detailed manner. Instead of this, functionality and user-convenience are often much more important. The inclusion of a visual interface and embedding in a decision-support system that can automate the answers to optimisation problems typical for water management might in this context not be forgotten. In this respect it is also important to mention that model development should be more based on management questions that need to be solved. Based on the set of questions that can be solved with models, the type of models needs to be selected. From that point, the data collection to develop, train and validate such models has to be steered. Up to now, the process is most often the opposite: data collectors present their data to model developers, who develop often not very well working models that can solve problems that no river manager is interested in… therefore, the decision support task was always kept in mind during the model development and data collection phase, to ascertain the ability that at least some river management problems can be tackled by means of the presented models.

References

Adriaenssens, V. (2004). Knowledge-based macroinvertebrate habitat suitability models for use in ecological river management. PhD thesis, Ghent University, pp. 296.

Adriaenssens, V., De Baets, B., Goethals, P.L.M. and De Pauw, N. (2004a). Fuzzy rule-based models for decision support in ecosystem management. *The Science of the Total Environment*, 319, 1-12.

Adriaenssens, V., Goethals, P. and De Pauw, N. (2002). Assessment of land-use impact on macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium) using multivariate analysis and Geographical Information Systems. *The Scientific World Journal*, 2, 546-557.

Adriaenssens, V., Goethals, P.L.M. and De Pauw, N. (2004b). Fuzzy knowledge-based models for prediction of macroinvertebrate taxa in watercourses in Flanders, Belgium. *Ecological Modelling*. (in press)

Adriaenssens, V., Goethals, P.L.M., Charles, J. and De Pauw, N. (2004c). Application of Bayesian Belief Networks for the prediction of macroinvertebrate taxa in rivers. *Annales de Limnologie – International Journal of Limnology*, 40(3), 181-191.

Adriaenssens, V., Simons, F., Nguyen, L.T.H., Goddeeris, B., Goethals, P.L.M. and De Pauw, N. (2004d). Potential of bio-indication of Chironomid communities for assessment of running water quality in Flanders (Belgium). *Belgian Journal of Zoology*, 134, 15-24.

AFNOR (1992). Essai des eaux: determination de l'indice biologique global normalise (IBGN). Norme française NF T90-350.

Alba-Tercedor, J. and Sanchez-Ortega, A. (1988). Un metodo rapido y simple para evoluar la qualidad biologica de las aquas corrientes basado en el de Helawell (1978). *Limnetica*, 4, 51-56.

Alessa, L., Bennett, S.M. and Kliskey, A.D. (2003). Effects of knowledge, personal attribution and perception of ecosystem health on depreciative behaviors in the intertidal zone of Pacific Rim National Park and Reserve. *Journal of Environmental Management*, 68, 207-218.

Allan, J.D. (1995). *Drift. Stream Ecology: structure and function of running waters*. Chapman & Hall, London.

Alkemade, J.R.M., van Grinsven, J.J.M., Wiertz, J. and Kros, J. (1998). Towards integrated national modelling with particular reference to the environmental effects of nutrients. *Environmental Pollution*, 102, S1, 101-105.

Amari, S., Murata, N., Müller, K.-R., Finke, M. and Yang, H.H. (1997). Asymptotic statistical theory of overtraining and cross-validation. *IEEE Transactions on Neural Networks*, 8(5), 985-996.

Anders, U. and Korn, O. (1999). Model selection in neural networks. *Neural Networks*, 12(2), 309-323.

Ankley, G.T., Thomas, N.A., Di Toro, D.M., Hansen, D.J., Mahony, J.D., Berry, W.J., Swartz, R.C. and Hoke, R.A. (1994). Assessing potential bioavailability of metals in sediments : a proposed approach. *Environmental Management*, 18 (3), 331-337.

Ankley, G.T., Di Toro, D.M., Hansen, D.J. and Berry, W.J. (1996). Technical basis and proposal for deriving sediment quality criteria for metals. *Environmental Toxicology and Chemistry*, 15(2), 2056-2066.

Argent, R.M. and Grayson, R.B. (2001). Design of information systems for environmental managers: an example using interface prototyping. *Environmental Modelling & Software*, 16, 433-438.

Argent, R.M. (2004). An overview of model integration for environmental applications – components, frameworks and semantics. *Environmental Modelling & Software*, 19, 219-234.

Armitage, P.D., Moss, D., Wright, J.F. and Furse, M.T. (1983). The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research*, 17, 333-347.

Babut, M., Bonnet, C., Bray, M., Flammarion, P., Garric, J. and Golaszewski, G. (2003). Developing environmental quality standards for various pesticides and priority pollutants for French freshwaters. *Journal of Environmental Management*, 69, 139-147.

Baran, P., Lek, S., Delacoste, M. and Belaud, A. (1996). Stochastic models that predict trout population density or biomass on a mesohabitat scale. *Hydrobiologia*, 337, 1-9.

Barbour, M.T., Gerritsen, J., Snyder, B.D. and Stribling, J.B. (1992). Rapid bioassessment protocols for use in wadeable streams and rivers. Periphyton, benthic macroinvertebrates and fish. EPA U.S. Office of Water, Washington DC.

Barbour, M.T. and Yoder, C.O. (2000). The multimetric approach to bioassessment, as used in the United States of America. In: Wright, J.F., Sutcliffe, W. and Furse, M.T. (eds), *Assessing the biological quality of freshwaters – RIVPACS and other techniques*. Freshwater Biological Association, pp. 281-292.

Barnard, E. and Wessels, L. (1992). Extrapolation and interpolation in neural network classifiers. *IEEE Control Systems*, 12(5), 50-53.

Barros, L.C., Bassanezi, R.C. and Tonelli, P.A. (2000). Fuzzy modelling in population dynamics. *Ecological Modelling*, 128, 27-33.

Bartch, A.F. and Ingram, W.M. (1966). Biological analysis of water pollution in North America. *Verhandlungen Internationale Vereinigung für theoretische und angewandte Limnologie*, 16, 786-800.

Bartlett, E.B. (1994). Dynamic node architecture learning: an information theoretic approach. *Neural Networks*, 7(1), 129-140.

Bayerisches Landesamt für Wasserwirtschaft (1996). Ecological characterisation of aquatic macro fauna. Inforamtionsberichte des Bayerischen Landesamtes für Wasserwirtschaft. Heft 4/96. (In German)

Beauchard, O., Gagneur, J. and Brosse, S. (2003). Macroinvertebrate richness patterns in North African streams. *Journal of Biogeography*, 30, 1821-1833.

Bebis, G. and Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials*, 13(4), 27-31.

Beck, W.M. (1954). Studies in stream pollution biology. A simplified ecological classification of organisms. *Quart. J. Fla. Ac. Sci.*, 17, 211-277.

Beck, M.B., Watts, J.B. and Winkler, S. (1998). An environmental process control laboratory: at the interface between instrumentation and model development. *Water Science & Technology*, 37(12), 252-362.

Belconsulting (2003). Ecologische inventarisatie en visievorming in het kader van integraal waterbeheer. De Zwalm. In opdracht van Ministerie van de Vlaamse Gemeenschap, Departement Leefmilieu en Infrastructuur, AMINAL, Afdeling Water, pp. 136.

Belpaire, C., Smolders, R., Vanden Auweele, I., Ercken, D., Breine, J., Van Thuyne, G. and Ollevier, F. (2000). An Index of Biotic Integrity characterizing fish populations and the ecological quality of Flandrian water bodies. *Hydrobiologia*, 434, 17-33.

Bernardo, J. and Smith, A. (1994). Bayesian Theory. John Wiley and Sons, Chichester.

Bishop, J.E. and Hynes, H.B.N. (1969). Upstream movements of the benthic invertebrates in the Speed River, Ontario. *J. Fish. Res. Bd. Can.*, 26, 279-298.

Blockeel, H., Dzeroski, S. and Grbovic, J. (1999a). Experiments with TILDE in the river quality domain.

Blockeel, H., Dzeroski, S. and Grbovic, J. (1999b). Simultaneous prediction of multiple chemical parameters of river quality with TILDE.

Bock, W. and Salski, A.A. (1998). fuzzy knowledge-based model of population dynamics of the Yellow-necked mouse *(Apodemus flavicollis)* in a beech forest. *Ecological Modelling*, 108 (1-3), 155-161.

Booty, W.G., Lam, D.C.L., Wong, I.W.S. and Siconolfi, P. (2001). Design and implementation of an environmental decision support system. *Environmental Modelling & Software*, 16, 453-458.

Borgmann, U., Cheam, V., Norwood, W.P. and Lechner, J. (1998). Toxicity and bioaccumulation of thallium in Hyalella azteca, with comparison to other metals and prediction of environmental impact. *Environmental Pollution*, 99, 105–114.

Borgmann, U., Norwood, W.P. and Babirad, L.M. (1991). Relationship between chronic toxicity and bioaccumulation of cadmium in Hyalella azteca. *Can. J. Fish. Aquat. Sci.*, 48, 1055–1060.

Borgmann, U., Neron, R. and Norwood, W.P. (2001a). Quantification of bioavailable nickel in sediments and toxic thresholds to *Hyalella azteca*. *Environmental Pollution*, 111, 189–198.

Borgmann, U., Norwood, W.P., Reynoldson, T.B. and Rosa, F. (2001b). Identifying cause in sediment assessments: bioavailability and the Sediment Quality Triad. *Can. I Fish. Aquat. Sci.*, 58, 950–960.

Borgmann, U., Nowierski, M., Grapentine, L.C. and Dixon, D.G. (2004). Assessing the cause of impacts on benthic organisms near Rouyn-Noranda, Quebec. *Environmental Pollution*, 129, 39-48.

Borsuk, M.E., Reichert, P. and Burkhardt-Holm, P. (2002). A Bayesian network for investigating the decline in fish catch in Switzerland. In: Rizzoli, A.E. and Jakeman, A.J. (eds.), *Integrated Assessment and Decision Support*, Proceedings of the 1st biennial meeting of the International Environmental Modelling and Software Society, Lugano, Switzerland, 2, pp. 108-113.

Borsuk, M., Stow, C.A. and Reckhow, K.H. (2004). A Bayesian network of eutrophication models for synthesis, prediction and uncertainty analysis. *Ecological Modelling*, 173(2-3), 219-239.

Bouma, J.J. (1998). Environmental Management Accounting in the Netherlands. In: Bennet, M. and James, P. (eds.), *The green bottom line: environmental accounting for management, Current Practice and Future Trends*. Greenleaf Publishing, Sheffield, pp. 139-151.

Bournaud, M. and Cogerino, L. (1986). The aquatic microhabitats in stretches of a large river: a faunal approach. *Annales de Limnologie*, 23(3), 285-294.

Brabec, K., Zahradkova, S., Nemejcova, D., Paril, P., Kokes, J. and Jarkovsky, J. (2004). Assessment of organic pollution effect considering differences between lotic and lentic stream habitats. *Hydrobiologia*, 516, 331-346.

Braukmann, U. (2001). Stream acidification in South Germany – chemical and biological assessment methods and trends. *Aquatic Ecology*, 35, 207-232.

Bray, J.R. and Curtis, J.T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs*, 27, 325-349.

Brehm, J. and Meijering, M.P.D. (1990). Fließgewässerkunde. Einfürhrung in de Limnolie der Quellen, Bäche und Flüsse. Biologische Arbeitsbücher, 36; 295 S., Quelle and Meyer Verlag, Heidelberg, Wiesbaden.

Breimann, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). Classification and regression trees. Pacific Grove, Wadsworth.

Brittain, J.E. and Eikeland, T.J. (1988). Invertebrate drift – a review. *Hydrobiologia*, 166, 77-93.

Brosse, S., Arbuckle, C.J. and Townsend, C.R. (2003). Habitat scale and biodiversity: influence of catchment, stream reach and bedform scales on local invertebrate diversity. *Biodiversity and Conservation*, 12, 2057-2075.

Brosse, S., Guégan, J.F., Tourenq, J.N. and Lek, S. (1999). The use of neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. *Ecological Modelling*, 120, 299-311.

Brosse, S., Lek, S. and Townsend, C.R. (2001). Abundance, diversity, and structure of freshwater invertebrates and fish communities: an artificial neural network approach. *New Zealand Journal of Marine and Freshwater Research*, 35, 135-145.

Buffagni, A., Erba, S., Cazolla, M. and Kemp, J.L. (2004). The AQEM multimetric system for the southern Italian Apennines: assessing the impact of water quality and habitat degradation on pool macroinvertebrates in Mediterranean rivers. *Hydrobiologia*, 516, 313-329.

Cairns, J., Albaugh, D.W., Busey, F. and Chaney, M.D. (1968). The Sequential Comparison Index. A simplified method to estimate relative differences in biological diversity in stream pollution studies. *Journal of Water Pollution Control Federation*, 40, 1607-1613.

Carchon, P. and De Pauw, N. (1997). Development of a methodology for the assessment of surface waters. Study for the Flemish Environmental Agency. Ghent University, Laboratory of Environmental Toxicology and Aquatic Ecology, Gent, Belgium, pp. 55. (in Dutch)

Ceccaroni, L., Cortés, U. and Sànchez-Marrè, M. OntoWEDSS: augmenting environmental decision-support systems with ontologies. *Environmental Modelling & Software*. (in press)

Céréghino, R., Park, Y.S., Compin, A. and Lek, S. (2003). Predicting the species richness of aquatic insects in streams using a limited number of environmental variables. *Journal of the North American Benthological Society*, 22(3), 442-456.

Chambers, M.R. (1977). A comparison of the population ecology of *Asellus aquaticus* (L.) and *Asellus meridianus* Rac. in the reed beds of the Tjeukemeer. *Hydrobiologia*, 53(2), 147-154.

Chandler, J.R. (1970). A biological approach to water quality management. *Water Pollution Control*, 69, 415-422.

Chapman, D. (1992). *Water quality assessments. A guide to the use of biota, sediments and water in environmental monitoring*. Chapman and Hall, London, pp. 585.

Chapman, P.M., Wang, F.Y., Janssen, C., Persoone, G. and Allen, H.E. (1998). Ecotoxicology of metals in aquatic sediments: binding and release, bioavailability, risk assessment, and remediation. *Can. J. Fish. Aquat. Sci.*, 55, 2221–2243.

Cherkassky, V. and Lari-Najafi, H. (1992). Data representation for diagnostic neural networks. *IEEE Intelligent Systems*, 7(5), 43-53.

Chon, T.S., Kwak, I.S., Park, Y.S., Kim T.H. and Kim Y. (2001). Patterning and short-term predictions of benthic macroinvertebrate community dynamics by using a recurrent artificial neural network. *Ecological Modelling*, 146, 181-193.

Chon, T.S., Park, Y.S., Kwak, I.S. and Cha, E.Y. (2002). Non-linear approach to grouping, dynamics and organizational informatics of benthic macroinvertebrate communities in streams by Artificial Neural Networks. In: Recknagel, F. (eds.), *Ecological Informatics. Understanding ecology by biologically-inspired computation*. Springer, Berlin, pp. 127-178.

Chun, S.-H. and Kim, S.H. (2004). Data mining for financial prediction and trading: application to single and multiple markets. *Expert Systems with Applications*, 26, 131-139.

Chutter, F.M. (1972). An empirical biotic index of the quality of water in South African streams and rivers. *Water Research*, 6, 19-30.

Clark, P. and Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning*, 3(4), 261–283.

Cohen, Y. (1988). Bayesian estimation of clutch size for scientific and management purposes. *Journal Wildlife Management*, 52 (4), 787-793.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P. and van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387, 253-260.

Crome, F.H.J., Thomas, M.R. and Moore, L.A. (1996). A novel Bayesian approach to assessing impacts of rain forest logging. *Ecological Applications*, 6(4), 1095-1103.

Cummins, K.W. (1975). Macroinvertebrates. pp. 170-198. In: Whitton, B.A. (ed.) *River ecology*. Blackwell, London, pp.725.

Dai, H.C. and Macbeth, C. (1997). Effects of learning parameters on learning procedure and performance of a BPNN. *Neural Networks*, 10(8), 1505-1521.

Dakou, E., Lazaridou-Dimitriadou, M., D'heygere, T., Dedecker, A., Goethals, P.L.M. and De Pauw, N. (2004). Development of models predicting macroinvertebrate communities in Greek rivers using rule induction techniques. *Aquatic Ecology*. (submitted)

Daunicht, W., Salski, A., Nohr, P. and Neubert, C. (1996). A fuzzy knowledge-based model of annual production of skylarks. *Ecological Modelling*, 85 (1), 65-73.

Davies, P.E. (2000). Development of a national river bioassessment system (AUSRIVAS) in Australia. In: Wright, J.F., Sutcliffe, W. and Furse, M.T. (eds.), *Assessing the biological quality of freshwaters – RIVPACS and other techniques*. Freshwater Biological Association, pp. 113-124.

Deaver, E. and Rodgers Jr., J.H. (1996). Measuring bioavailable copper using anodic stripping voltammetry. *Environmental Toxicology and Chemistry*, 15, 1925–1996.

De Cooman, W., Florus, M., Vangheluwe, M., Janssen, C., Heylen, S., De Pauw, N., Rillaerts, E., Meire, P., and Verheyen, R. (1999). Sediment characterisation of rivers in Flanders. In: De Schutter, G. (ed.), *CATS 4*. PIH, Antwerp, Belgium.

De Cooman, P., Hatse, I. and Guns, M. (1996). Assessment of sediment quality. Water, 89, 162-168. (in Dutch)

de Deckere, E., De Cooman, W., Florus, M. and Devroede-Vanderlinden, M.P. (2000). Characterisation of Flemish navigable watercourses. AMINAL, Brussels. (in Dutch)

Dedecker, A., Goethals, P.L.M., Gabriels, W. and De Pauw, N. (2004a). Optimisation of Artificial Neural Network (ANN) model design for prediction of macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium). *Ecological Modelling*, 174(1-2), 161-173.

Dedecker, A.P., Goethals, P.L.M. and De Pauw, N. (2002). Comparison of artificial neural network (ANN) model development methods for prediction of macroinvertebrate communities in the Zwalm river basin in Flanders, Belgium. *The Scientific World Journal*, 2, 96-104.

Dedecker, A.P., Goethals, P.L.M. and De Pauw, N. (2004b). Sensitivity and robustness of predictive neural network ecosystem models for simulation of different management scenarios. In: Jorgensen, S., Lek, S., Scardi, M. and Verdonschot, P.F.M. (eds.), *Modelling community structure in freshwater ecosystems*. Springer-Verlag, Berlin. (in press)

Dedecker, A.P., Goethals, P.L.M., D'heygere, T., Gevrey, M., Lek, S. and De Pauw, N. (2004c). Habitat preference study of *Asellus* (Crustacea, Isopoda) by applying input variable contribution methods to Artificial Neural Network models. *Environmental Modeling & Assessment*. (submitted)

Dedecker, A.P., Goethals, P.L.M., D'heygere, T., Gevrey, M., Lek, S. and De Pauw, N. (2004d). Application of Artificial Neural Network models to study the relationship between Gammaridae (Crustacea, Amphipoda) and river conditions. *Environmental Monitoring and Assessment*. (in press)

Dedecker, A.P., Goethals, P.L.M., D'heygere, T. and De Pauw, N. (2004e). Development of an in-stream migration model for *Gammarus pulex* L. (Crustacea, Amphipoda) as a tool in river restoration management. *Aquatic Ecology*. (submitted)

De Haas, E. (2004). Persistence of benthic invertebrates in polluted sediments. PhD thesis, University of Amsterdam, FNWI, The Netherlands, pp.135.

den Besten, P.J., Schmidt, C.A., Ohm, M., Ruys, M.M., van Berghem, J.W. and van de Guchte, C. (1995). Sediment quality assessment in the delta of rivers Rhine and Meuse based on field observations, bioassays and food chain implications. *Journal of Aquatic Ecosystem Health*, 4, 257-270.

Dennis, B. (1996). Discussion: should ecologists become Bayesians? *Ecological Applications*, 6(4), 1104-1123.

Denzer, R., Güttlerr, R. and Houy, P. (2000). TEMSIS Consortium, TEMSIS – a transnational system for public information and environmental decision support. *Environmental Modelling & Software*, 15, 235-243.

De Pauw, N. (2000). Using RIVPACS as a modeling tool to predict the impacts of environmental changes. p. 311-314. In: Wright, J.F., Sutcliffe, D.W. and Furse, M.T. (eds.), *Assessing the biological quality of fresh waters: RIVPACS and other techniques*. FBA, Ambleside, UK., pp. 373.

De Pauw, N., Ghetti, P.F., Manzini, P. and Spaggiari, P. (1992). Biological assessment methods for running water. In: Newman, P., Piavaux, A. and Sweeting, R. (eds.), *River water quality – assessment and control. Commission of the European Communities*. EUR 14606 EN-FR, 1992-III. Brussels, pp. 217-248.

De Pauw, N. and Hawkes, H.A. (1993). Biological monitoring of river water quality. p. 87-111. In: Walley, W.J. and Judd, S. (eds.), *River water quality monitoring and control*. Aston University, Birmingham, UK.

De Pauw, N. and Heylen, S. (2001). Biotic index for sediment quality assessment of watercourses in Flanders, Belgium. *Aquatic Ecology*, 35, 121-133.

De Pauw, N., Lambert, V., Van Kenhove, A. and bij de Vaate, A. (1994). Performance of two artificial substrate samplers for macroinvertebrates in biological monitoring of large and deep rivers and canals in Belgium and The Netherlands. *Environmental Monitoring and Assessment*, 30, 25-47.

De Pauw, N. and Vanhooren, G. (1983). Method for biological quality assessment of water courses in Belgium. *Hydrobiologia*, 100, 153-168.

De Pauw, N. and Vannevel, R. (1993). Macroinvertebrates and water quality. Stichting Leefmilieu. Dossier N° 11, Antwerp, pp. 316. (in Dutch)

DEV (Deutsche Einheitsverfahren zur Wasser, Abwasser- und Schlammuntersuchung) (1988-1991). Methoden der biologisch-ökologischen Gewässeruntersuchung, Gruppe M: Fliessende Gewässer, DIN 38 410, Teil 1 und 2, Weinheim.

D'heygere, T., Goethals, P. and De Pauw, N. (2001). Application of genetic algorithms for input variables selection of decision tree models predicting mollusca in unnavigable Flemish watercourses. *Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen*, 66(4), 219-223.

D'heygere, T., Goethals, P. and De Pauw, N. (2002). Optimisation of the monitoring strategy of macroinvertebrate communities in the river Dender, in relation to the EU Water Framework Directive. *The Scientific World Journal*, 2, 607-617.

D'heygere, T., Goethals, P.L.M. and De Pauw, N. (2003). Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates. *Ecological Modelling*, 160, 291-300.

D'heygere, T., Goethals, P.L.M. and De Pauw, N. (2004). Genetic algorithms for optimisation of predictive ecosystems models based on decision trees and neural networks. *Ecological Modelling*. (in press)

Dietz, S. and Adger, W.N. (2003). Economic growth, biodiversity loss and conservation effort. *Journal of Environmental Management*, 68, 23-35.

Dimopoulos, I., Chronopoulos, J., Chronopoulou Sereli, A. and Lek, S. (1999). Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). *Ecological Modelling*, 120, 157-165.

Dimopoulos, Y., Bourret, P. and Lek, S. (1995). Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters*, 2, 1-4.

Di Toro, D.M., Mahony, J.D. Hansen, D.J., Scott, K.J., Hicks, M.B., Mays, S.M. and Redmond, M.S. (1990). Toxicity of cadmium in sediments: the role of acid volatile sulfides. *Environmental Toxicology and Chemistry*, 9, 1487-1502.

Dixon, P. and Ellison, A.M. (1996). Introduction: ecological applications of Bayesian inference. *Ecological Applications*, 6(4), 1034-1035.

Dunford, R.W., Ginn, T.C. and Desvousges, W.H. (2004). The use of habitat equivalency analysis in natural resource damage assessments. *Ecological Economics*, 48, 49-70.

Du Preez, L.A., Husselmann, A.J., Acton, N.R. and Lange, L. (1998). Establishing a network of on-line monitors at the purification works and in the distribution network of rand water. *Water Science & Technology*, 37(9), 65-71.

Dzeroski, S., Demsar, D. and Grbovic, J. (2000). Predicting chemical parameters of river water quality from bioindicator data. *Applied Intelligence*, 13(1), 7-17.

Dzeroski, S. and Drumm, D. (2003). Using regression trees to identify the habitat preference of sea cucumber (*Holothuria leucospilota*) on Rarotonga, Cook Island. *Ecological Modelling*, 170, 219-226.

Dzeroski, S., Grbovic, J. and Walley, W.J. (1997). Machine learning applications in biological classification of river water quality, p. 429-448. In: Michalski, R.S., Bratko, I. and Kubat, M. (eds.), *Machine learning and data mining: methods and applications*. John Wiley and Sons Ltd., New York, USA.

Edwards, H.E. (1998). An instrument for the measurement of colour and turbidity in natural waters. *Water Science & Technology*, 37(12), 263-267.

Ellison, A.M. (1996). An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications*, 6(4), 1036-1046.

EU (2000). Directive of the European Parliament and of the Council 2000/60/EC establishing a framework for community action in the field of water policy, Rep. No. PE-CONS 3639/1/00 REV 1. European Union, Luxembourg.

Failing, L. and Gregory, R. (2003). Ten common mistakes in designing biodiversity indicators for forest policy. *Journal of Environmental Management*, 68, 121-132.

Farber, S.C., Costanza, R. and Wilson, M.A. (2002). Economic and ecological concepts for valuing ecosystem services. *Ecological Economics*, 41, 375-392.

Fausett, L. and Elwasif, W. (1994). Predicting performance from test scores using backpropagation and counterpropagation. In: Proceedings of IEEE International Conference on Neural Networks, pp. 3398-3402.

Fielding, A.H. and Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38-49.

Fleishman, E., Mac Nally, R., Fay, J.P. and Murphy, D.D. (2001). Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology*, 15, 1674–1685.

Fleishman, E., Mac Nally, R. and Fay J.P. (2002). Validation tests of predictive models of butterfly occurrence based on environmental variables. *Conservation Biology*, 17(3), 806-817.

Fletcher, D. and Goss, E. (1993). Forecasting with neural networks: an application using bankruptcy data. *Information and Management*, 24(3), 159-167.

Flood, I. and Kartam, N. (1994). Neural networks in civil engineering. I: Principles and understanding. *Journal of Computing in Civil Engineering*, 8(2), 131-148.

Fontoura, A.P. and De Pauw, N. (1994). Microhabitat preference of stream macrobenthos and its significance in water quality assessment. *Verhandlungen Internationale Vereinigung für theoretische und angewandte Limnologie*, 25, 1936-1940.

Gabriels, W., Goethals, P.L.M. and De Pauw, N. (2002). Prediction of macroinvertebrate communities in sediments of Flemish watercourses based on artificial neural networks. *Verhandlungen Internationale Vereinigung für theoretische und angewandte Limnologie*, 28(2), 777-780.

Gabriels, W., Dedecker, A., Goethals, P.L.M., Lek, S. and De Pauw, N. (2004). Input variables selection of artificial neural networks predicting aquatic macrobenthos communities in Flanders (Belgium). *Aquatic Ecology*. (in preparation)

Garson, G.D. (1991). Interpreting neural-network connection weights. *Artificial Intelligence Expert*, 6, 47-51.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (1995). Bayesian data analysis. Chapman and Hall, London.

Gevrey, M., Dimopoulos, I. and Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160, 249-264.

Ghetti, P.F. (1997). Manuale di applicazione Indice Biotico Esteso (IBE). Provincia Autonoma Di Trento, pp. 222.

Ghetti, P.F. and Ravera, O. (1994). European perspective on biological monitoring. 46. In: Loeb, L. and Spacie, A. (eds.), *Biological monitoring of aquatic systems*. Lewis Publishers, Boca Raton, pp. 381.

Giller, P.S. and Malmqvist, B. (1998). The biology of streams and rivers, Oxford University Press, Oxford, pp. 296.

Gledhill, T., Sutcliffe, D.W. and Williams, W.D. (1976). A revised key to the British species of Crustacea: Malacostraca occurring in fresh water, with notes on their ecology and distribution. Freshwater Biological Association, Scientific Publication N°32, pp. 71.

Gledhill, T., Sutcliff, D.W. and Williams, W.D. (1993). British Freshwater Crustacea Malacostraca: A Key with Ecological Notes. Freshwater Biological Association, pp. 173.

Goedmaker, A. and Pinkster, S. (1981). Population dynamics of three Gammarid species (Crustacea, Amphipoda) in a French chalk stream. Part III. Migration. *Bijdragen tot de Dierkunde*, 51(2), 145-180.

Goethals, P.L.M., Bouma, J.J., François, D., D'heygere, T., Dedecker, A., Adriaenssens, V. and De Pauw, N. (2003). Coupling ecosystem valuation methods to the WAECO decision support system in the Zwalm Catchment (Belgium). Vol. 3, p. 971-976. In: Post, D.A., Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ), Proceedings 'Integrative Modelling of Biophysical, Social and Economic Systems for Resource Management Solutions MODSIM 2003', 14-17 July 2003, Townsville, Australia. 2066 p.

Goethals, P.L.M. and De Pauw, N. (2001). Development of a concept for integrated ecological river assessment in Flanders, Belgium. *Journal of Limnology*, 60 (1), 7-16.

Goethals, P.L.M., Dedecker, A.P., Bouma, J.J., François, D., Verstraete, A. and De Pauw, N. (2004). The Water Ecology decision support system (WAECO-DSS) for integrated cost-benefit analyses in river restoration management: case study of the Zwalm river basin (Belgium). *Journal of Environmental Management*. (submitted)

Goethals, P.L.M., Dedecker, A.P., Gabriels, W. and De Pauw, N. (2002). Development and Application of Predictive River Ecosystem Models Based on Classification Trees and Artificial neural Networks. In: Recknagel, F. (ed.), *Ecological Informatics: Understanding Ecology by Biologically-Inspired Computation*. Springer-Verlag, Berlin, pp. 432.

Goethals, P., Dedecker, A., Raes, N., Adriaenssens, V., Gabriels, W. and De Pauw, N. (2001). Development of river ecosystem models for Flemish watercourses: case studies in the Zwalm river basin. *Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen,* 66(1), 71-86.

Goh, A.T.C. (1995). Back-propagation neural networks for modelling complex systems. *Artificial Intelligence in Engineering*, 9, 143-151.

Goldberg, D.E. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Publishing Company, Reading, MA., pp. 412.

Gongrijp, A. (1981). Biologische beoordeling van slootkwaliteit. Een literatuurstudie en een voorstel voor onderzoek in Zuid-Holland. Hoogheemraadschap van Rijnland, technische dienst, Afd. Chemie en technologie.

Guégan, J.F., Lek, S. and Oberdorff, T. (1998). Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature*, 391, 382-384.

Guisan, A. and Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147-168.

Haas, T.C., Mowrer, H.T. and Shepperd, W.D. (1994). Modelling aspen stand growth with a temporal Bayes Network. *Artif. Insemen. Appl.*, 8, 15-28.

Hagan, M.T., Demuth, H.B. and Beale, M. (1996). Neural network design. PWS Publishing Company, Boston, USA., pp. 712.

Hanley, N., Schäpfer, F. and Spurgeon, J. (2003). Aggregating the benefits of environmental improvements: the distance-decay functions for use and non-use values. *Journal of Environmental Management*, 68, 297-304.

Hansell, R.I.C., Hansell, T.M. and Fenech, A. (2003). A new market instrument for sustainable economic and environmental development. *Environmental Monitoring and Assessment*, 86, 203-209.

Hare, L., Tessier, A. and Warren, L. (2001). Cadmium accumulation by invertebrates living at the sediment-water interface. *Environmental Toxicology and Chemistry*, 20, 880–889.

Hawkes, H.A. (1979). Invertebrates as indicators of river water quality. In: James, A. and Evinson, L. (Eds.), *Biological indicators of water quality*. John Wiley and Sons, Ltd., Chichester, UK.

Hawkes, H.A. and Davies, L.J. (1971). Some effects of organic enrichment on benthic invertebrate communities in stream riffles. p. 271-299. In: Duffey, E. and Watt, A. (Eds.). The scientific management of animal and plant communities for conservation. Blackwell, Oxford, UK.

Hayden, W. and Clifford, H.F. (1974). Seasonal movements of the mayfly Leptophlebia cupida (Say) in a brown-water stream of Alberta, Canada. *American Midland Naturalist*, 91(1), 90-102.

Haykin, S. (1999). Neural Networks: A Comprehensive Foundation. Second Edition. Prentice Hall, New Jersey, USA.

Hecht-Nielsen, R. (1987). Kolmogorov's mapping neural network existence theorem. p. 11–14. First IEEE International Conference on Neural Networks. San Diego, USA.

Helawell, J.M. (1986). Biological indicators of freshwater pollution and environmental management. Elsevier Applied Science Publishers, London, pp. 545.

Henry, C.P. and Amoros, C. (1995). Restoration ecology of riverine wetlands: I. A scientific base. *Environmental Management*, 19, 891-902.

Hering, D., Moog, O., Sandin, L. and Verdonschot, P. (2004). Overview and application of the AQEM assessment system. *Hydrobiologia*, 516, 1-20.

Hershey, A.E., Pastor, J., Peterson, B.J. and Kling, G.W. (1993). Stable isotope resolve the drift paradox for Baetis mayflies in an arctic river. *Ecology*, 74, 2315-2325.

Hilsenhoff, W.L. (1988). Rapid field assessment of organic pollution with a family-level biotic index. *Journal of the North American Benthological Society*, 7, 65-68.

Hirzel, A. and Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, 157, 331-341.

Hoang, H., Recknagel, F., Marshall, J. and Choy, S. (2001). Predictive modelling of macroinvertebrate assemblages for stream habitat assessments in Queensland (Australia). *Ecological Modelling*, 146, 195-206.

Hoang, H., Recknagel, F., Marshall, J. and Choy, S. (2002). Elucidation of hypothetical relationships between habitat conditions and macroinvertebrate assemblages in freshwater streams by Artificial Neural Networks. Ecological Informatics. p. 179-192. In: Recknagel, F. (ed.), *Ecological Informatics: Understanding Ecology by Biologically-Inspired Computation*. Springer-Verlag, Berlin, pp. 432.

Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.

Humphries, S. and Ruxton, G.D. (2002). Is there really a drift paradox? Journal of Animal *Ecology*, 71, 151-154.

Hung, M.S., Hu, M.Y., Shanker, M.S. and Patuwo, B.E. (1996). Estimating posterior probabilities in classification problems with neural networks. *International Journal of Computational Intelligence and Organizations*, 1(1), 49-60.

Hynes, H.B.N. (1971). The biology of polluted waters. University Toronto Press, Toronto.

Illies, J. (1961). Versuch einer allgemeinen biozönotische Gliedering der Fliesgewässer. *Int. Rev. ges. Hydrobiol.*, 46, 205-213.

IBN (1984). Norme Belge T 92-402. Biological water quality: determination of the biotic index based on aquatic macroinvertebrates. Institut Belge de Normalisation. (in Dutch and French).

Jaccard, P. (1908). Nouvelles reserches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.*, XLIV,223-269.

Janse, J.H. (1997). A model of nutrient dynamics in shallow lakes in relation to multiple stable states. *Hydrobiologia*, 342-343, 1-8.

Jeuken, M.H.J.L., Janse, J.H. and Aldenberg, T. (1999). Processes description of DUFLOW for Windows. 44 pp.

Johnson, R.K. (1998). Classification of Swedish lakes and rivers using benthic macroinvertbrates. In: Wiederholm, T. (ed.), *Bakgrundsrapport 2 till bedömningsgrunder för sjöar och vattendrag – biologiska parametrar*. Swedish Environmental Protection Agency Report 4921.

Jolma, A., De Marchi, C., Smith, M., Perera, B.J.C. and Somlyódy, L. (1997). StreamPlan: a support system for water quality managment on a river basin scale. *Environmental Modelling & Software*, 12, 275-284.

Jorde, K., Schneider, M. and Zoellner, F. (2000). Analysis of instream habitat quality – Preference functions and fuzzy models. In: Wang and Hu (eds.), *Stochastic Hydraulics*. Balkema, Rotterdam, pp. 671-680.

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

Kampichler, C., Barthel, J. and Wieland, R. (2000). Species density of foliage-dwelling spiders in field margins: a simple, fuzzy rule-based model. *Ecological Modelling*, 129, 87-99.

Kanellopoulos, I. and Wilkinson, G.G. (1997). Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, 18(4), 711-725.

Karaman, G.S and Pinkster, S. (1977). Freshwater Gammarus species from Europe, North Africa and adjacent regions of Asia (Crustacea-Amphipoda): part 1. Gammarus pulex group and related species. *Bijdrage tot de dierkunde*, 47(1).

Karr, J.R. (1981). Assessment of biotic integrity using fish communities. Fisheries, 6, 21-27. Karr, J.R. and Chu, E.W. (1997). Biological monitoring: Essential foundations for ecological risk assessment. *Human and Ecological Risk Assessment*, 3, 933-1004.

Karr, J.R. and Chu, E.W. (1999). Restoring life in running waters. Better biological monitoring. Island Press, Washington DC, pp. 206.

Karul, C., Soyupak, S.,Cilesiz, A.F., Akbay, N. and Germen, E. (2000). Case studies on the use of neural networks in eutrophication modeling. *Ecological Modelling*, 134, 145-152.

Knoben, R.A.E., Roos, C. and van Oirschot, M.C.M. (1995). Biological assessment methods for watercourses. UN/ECE Task Force on Monitoring and Assessment, Vol. 3. RIZA, P.O. Box 17, 8200 Lelystad, The Netherlands, pp. 86.

Knöpp, H. (1954). Ein neuer Weg zur Darstellung biologischer Vorfluteruntersuchungen, erläutert an einem Gütelangschnitt des Maines. *Wasserwirtschaft*, 45, 9-15.

Knowler, D.J., MacGregor, B.W., Bradford, M.J. and Peterman, R.M. (2003). Valuing freshwater salmon habitat on the west coast of Canada. *Journal of Environmental Management*, 69, 261-273.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.

Kompare, B., Bratko, I., Steinman, F. and Deroski, S. (1994). Using machine learning techniques in the construction of models. Part I: Introduction. *Ecological Modelling*, 75–76, 617–628.

Kolkwitz, R. and Marsson, M. (1902). Grundsatze für die biologische Beurteiung des Wassers nach zeiner Flora und Dauna. *Mitt. Prüfungsanst.Wasserversorg. Abwasserrein*, 1, 33-72.

Laë, R., Lek, S., and Moreau, J. (1999). Predicting fish yield of African lakes using neural networks. *Ecological Modelling*, 120, 325-335.

Lam, D., Leon, L., Hamilton, S. Crookshank, N., Bonin, D. and Swayne, D. (2004). Multi-model integration in a decision support system: a technical user interface approach for watershed and lake management scenarios. *Environmental Modelling and Software*, 19, 317-324.

Lam, D. and Swayne, D. (2001). Issues of EIS software design: some lessons learned in the past decade. *Environmental Modelling & Software*, 16, 419-425.

Larson, B.D. and Sengupta, R.R. (2004). A spatial decision support system to identify species-specific critical habitats based on size and accessibility using US GAP data. *Environmental Modelling & Software*, 19, 7-18.

Lau, L., Young, R.A., McKeon, G., Syktus, J., Duncalfe, F., Graham, N. and McGregor, J. (1999). Downscaling global information for regional benefit: coupling spatial models at varying space and time scales. *Environmental Modelling & Software*, 14, 519-529.

Lauryssen, F., Tack, F. and Verloo, M. (1994). Nitrogen transport in the Zwalm river basin. *Water*, 75, 46–49. (in Dutch)

Lawson, S.R., Manning, R.E., Valliere, W.A. and Wang, B. (2003). Proactive monitoring and adaptive management of social carrying capacity in Arches National Park: an application of computer simulation modeling. *Journal of Environmental Management*, 68, 305-313.

Lee, D.C. and Riemann, B.E. (1997). Population viability assessment of salmonids by using probabilistic networks. *North American Journal of Fisheries Management*, 17, 1144-1157.

Leeks, G.J.L., Neal, C., Jarvie, H.M. and Leach, D.V. (1997). The LOIS river monitoring network: strategy and implementation. *Science of the Total Environment*, 194-195, 101-109.

Legendre, P. and Legendre L. (1998). Numerical Ecology. Developments in Environmental Modelling, 20. Elsevier, Amsterdam, pp. 853.

Lek, S. and Guégan, J.F. (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, 120, 65-73.

Lek, S., Beland, A., Dimopoulos, I., Lauga, J. and Moreau, J. (1995). Improved estimation, using neural networks, of the food consumption of fish populations. *Marine and Freshwater Research*, 46, 1229-1236.

Lek, S., Belaud, A., Baran, P., Dimopoulos, I. and Delacoste, M . (1996a). Role of some environmental variables in trout abundance models using neural networks. *Aquatic Living Resources*, 9, 23-29.

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

Lenard, M.J., Alam, P. and Madey, G.R. (1995). The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decision Sciences*, 26(2), 209-227.

Liano, K. (1996). Robust error measure for supervised neural network learning with outliers. *IEEE Transactions on Neural Networks*, 7(1), 246-250.

Liebmann, H. (1962). Handbuch der Frischwasser- und Abwasserbiologie. Vol.1. 2nd ed.R. Oldenburg, München, pp. 588.

Linke, S., Bailey, R.C. and Schwindt, J. (1999). Temporal variability of stream bioassessments using benthic macroinvertebrates. *Freshwater Biology*, 42(3), 575-584.

Linnaeus, C. (1758). Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tenth edition, pp. 824.

Lorenz, A., Hering, D., Feld, C.K. and Rolauffs, P. (2004). A new method for assessing the impact of hydromorphological degradation on the macroinvertebrate fauna of five German stream types. *Hydrobiologia*, 516, 107-127.

Mackenthun, K.M. (1969). The practice of water pollution biology. FWPCA, Washington, pp. 281.

Mackinson, S. (2000). An adaptive fuzzy expert system for predicting structure, dynamics and distribution of herring shoals. *Ecological Modelling*, 126, 155-178.

MacNeil, C., Dick, J.T.A., Bigsby, E., Elwood, R.W., Montgomery, W.I., Gibbins, C.N. and Kelly, D.W. (2002). The validity of the Gammarus:Asellus ratio as an index of organic pollution: abiotic and biotic influences. *Water Research*, 36(2), 75-84.

Macrofauna-atlas of North Holland (1990). Distribution maps and responses to environmental factors of aquatic invertebrates. H.A. Steenbergen. Haarlem, pp. 650.

Maier, H.R. and Dandy, G.C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and Software*, 15, 101-124.

Manel, S., Dias, J.M., Buckton, S.T. and Ormerod, S.J. (1999). Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecolology*, 36, 734-747.

Manel, S., Williams, H.C. and Ormerod, S.J. (2001). Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38, 921-931.

Marcot, B.G., Holthausen, R.S., Raphael, M.G., Rowland, M. and Wisdom, M. (2001). Using Bayesian Belief Networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management*, 153, 29-42.

Marshall, J., Hoang, H., Choy, S. and Recknagel, F. (2002). Relationships between habitat properties and the occurrence of macroinvertebrates in Queensland streams (Australia) discovered by a sensitivity analysis with artificial neural networks. *Verhandlungen Internationale Vereinigung für theoretische und angewandte Limnologie*, 28, 1415-1419.

Masters, T. (1993). Practical Neural Network Recipes in C++. Academic Press, San Diego, USA.

Mastrorillo, S., Lek, S. and Dauba, F. (1997a). Predicting the abundance of minnow Phoxinus phoxinus (Cyprinidae) in the River Ariege (France) using artificial neural networks. *Aquatic Living Resources*, 10, 169-176.

Mastrorillo, S., Lek, S., Dauba, F. and Beland, A. (1997b). The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater biology*, 38, 237-246.

Matthews, R.A., Buikema, A.L., Cairns, J. and Rogers, J.H. (1982). Biological monitoring Part IIA – Receiving system functional methods, relationship and indices. *Water Research*, 6, 129-139.

McCoy, N.H. (2003). Behavioral externalities in natural resource production possibility frontiers: integrating biology and economics to model human-wildlife interactions. *Journal of Environmental Management*, 69, 105-113.

Meesters, E.H., Bak, R.P.M., Westmacott, S., Ridgley, M. and Dollar, S. (1997). A fuzzy logic model to predict coral reef development under nutrient and sediment stress. *Conservation Biology*, 12(5), 957-965.

Melloul, A.J. and Collin, M.L. (2003). Harmonizing water management and social needs: a necessary condition for sustainable development. The case of Israel’s coastal aquifer. *Journal of Environmental Management*, 67, 385-394.

Metcalfe, J.L. (1989). Biological water quality assessment of running water based on macroinvertebrate communities: history and present status in Europe. *Environmental Pollution*, 60, 101-139.

Metcalfe-Smith, J.L. (1994). Biological water quality assessment of rivers: use of macroinvertebrate communities. In: Calow, P. and Petts, G.E. (eds.), *The rivers handbook*. Vol.2. Blackwell Scientific Publications, Oxford, pp. 144-170.

Mohapl, J. (2000). Measurement diagnostics by analysis of last digits. *Environmental Monitoring & Assessment*, 61, 407-417.

Moog, O. (1995). Fauna Aquatica Austriaca – a comprehensive species inventory of Austrian aquatic organisms with ecological data. First edition, Wasserwirtschafskataster, Bundesministeriumf für Land- und Forstwirtschaft, Wien.

Mustajoli, J., Hämäläinen, R.P. and Marttunen, M. Participatory multicriteria decision analysis with Web-HIPRE: a case of lake regulation policy. *Environmental Modelling & Software*. (in press)

Myatt, L.B., Scrimshaw, M.D. and Lester, J.N. (2003). Public perceptions and attitudes towards an established managed realignment scheme: Orplands, Essex, UK. *Journal of Environmental Management*, 68, 173-181.

Nabhan, T.M. and Zomaya, A.Y. (1994). Toward generating neural network structures for function approximation. *Neural Networks*, 7(1), 89-99.

Nixon, S. (2003). An overview of the biological assessment of surface water quality in Europe. In: Symoens, J.J. and Wouters, K. (eds.), *Biological evaluation and monitoring of the quality of surface waters*. Nat.Com. of Biological Sciences and SCOPE Nat. Com., Brussels, pp. 9-15.

Norris, R.H. and Georges, A. (1993). Analysis and interpretation of benthic macroinvertebrate surveys. In: Rosenberg, D.M. and Resh, V.H. (eds.), *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman and Hall, New York.

Norris, R.H. and Thoms, M.C. (1999). What is river health? *Freshwater Biology*, 41, 197-209.

Obach, M., Wagner, R., Werner, H. and Schmidt, H.H. (2001). Modelling population dynamics of aquatic insects with artificial neural networks. *Ecological Modelling*, 146, 207-217.

Olden, J.D. and Jackson, D.A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154, 135-150.

Olden, J.D., Joy, M.K. and Death, R.G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178, 389-397.

Olson, R.L., Willers, J.L. and Wagner, T.L. (1990). A framework for modelling uncertain reasoning in ecosystem management II. Bayesian Belief Networks. *Artificial Intelligence Applications in Natural Resources Management*, 4(4), 11-24.

Olsson, T. and Söderström, O. (1978). Springtime migration and growth of Parameletus chelifer (Ephemeroptera) in a temporary stream in northern Sweden. *Oikos*, 31, 284-289.

Özesmi, S.L. and Özesmi, U. (1999). An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological Modelling*, 116(1), 15-31.

Paggio, R., Agre, G., Dichev, C., Umann, G., Rozman, T., Batachia, L. and Stocchero, M. (1999). A cost-effective programmable environment for developing environmental decision support systems. *Environmental Modelling & Software, 14, 367-382.*

Pantle, R. and Buck, H. (1955). Die biologische Uberwachung der Gewasser und die Darstellung der Ergebnisse. *Gas-u Wasserfach*, 96, 604.

Park, Y.S., Céréghino, R., Compin, A. and Lek, S. (2003a). Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. *Ecological Modelling*, 160, 265-280.

Park, R.A. and Clough, J.S. (2004). AQUATOX (Release 2). Modeling environmental fate and ecological effects in aquatic ecosystems. Volume 2: Technical documentation. US-EPA, Office of Water, Office of Science and Technology, pp. 210.

Park, Y.S., Kwak, I.S., Chon, T.S., Kim, J.K. and Jorgensen, S.E. (2001). Implementation of artificial neural networks in patterning and prediction of exergy in response to temporal dynamics of benthic macroinvertebrate communities in streams. *Ecological Modelling*, 146, 143-157.

Park, Y.S., Verdonschot, P.F.M., Chon, T.S. and Lek, S. (2003b). Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network. *Water Research*, 37, 1749-1758.

Patten, B.C. (1962). Species diversity in net phytoplankton of Raritan Bay. *Journal of Marine Research*, 20, 57-75.

Patuwo, E., Hu, M.Y. and Hung, M.S. (1993). Two-group classification using neural networks. *Decision Sciences*, 24(4), 825–845.

Pavlikakis, G.E. and Tsihrintzis, V.A. (2003). A quantitative method for accounting human opinion, preferences and perceptions in ecosystem management. *Journal of Environmental Management*, 68, 193-205.

Pavluk, T.I., Bij de Vaate, A. and Leslie, H.A. (2000). Development of an Index of Trophic Completeness for benthic macroinvertebrate communities in flowing waters. *Hydrobiologia*, 427, 135-141.

Pearce, D. (2001). Valuing Biological Diversity: Issues and Overview. In: *Valuation of Biodiversity Benefits*. Selected Studies, OECD, pp. 27-44.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems. Morgan Kauffman Publishers, San Francisco, CA, USA.

Peeters, E.T.H.M. (2001). Benthic macroinvertebrates and multiple stressors. Quantification of the effects of multiple stressors in field, laboratory and model settings. PhD thesis, Wageningen University, The Netherlands, pp. 168.

Persoone, G. and De Pauw, N. (1979). Systems of biological indicators for water quality assessment. In: Ravera O (ed.), *Biological aspects of freshwater pollution*, Pergamon Press, Oxford, pp. 39-75.

Piramuthu, S., Shaw, M. and Gentry, J. (1994). A classification approach using multi-layered neural networks. *Decision Support Systems*, 11(5), 509-525.

Poch, M., Comas, J., Rodríguez-Roda, I., Sànchez-Marrè, M. and Cortés, U. Designing and building real environmental decision support systems. *Environmental Modelling & Software*. (in press)

Pullar, D. and Springer, D. (2000). Towards integrating GIS and catchment models. *Environmental Modelling & Software*, 15, 451-459.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1), 145-151.

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.

Quinlan, J.R. (1993). C4.5: Programs for machine learning. Morgan Kaufmann, San Francisco, USA.

Rabeni, C.F. and Minshall, G.W. (1977). Factors affecting microdistribution of stream benthic insects. *Oikos*, 29, 33-43.

Reckhow, K.H. (2002). Bayesian Approaches in Ecological Analysis and Modeling. In: Canham, C.D., Cole, J.J. and Lauenroth, W.K. (eds.), *The Role of Models in Ecosystem Science*. Princeton University Press.

Recknagel, F. (2001). Applications of machine learning to ecological modeling. *Ecological Modelling*, 146, 303-310.

Recknagel, F. (2003). Ecological Informatics: Understanding Ecology by Biologically-Inspired Computation. Springer-Verlag, Berlin, Germany, pp. 432.

Reed, M., Cuddy, S.M. and Rizzoli, A.E. (1999). A framework for modeling multiple resource management issues – an open modeling approach. *Environmental Modelling & Software*, 14, 503-509.

Regan, H.M. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications*, 12, 618-628.

Reynoldson, T.B., Day, K.E. and Pascoe, T. (2000). The development of the BEAST: a predictive approach for assessing sediment quality in the North American Great Lakes. In: Wright, J.F., Sutcliffe, D.W. and Furse, M.T. (eds.), *Assessing the biological quality of fresh waters: RIVPACS and other techniques*. Freshwater Biological Association, UK, 165-180.

Richardson, R.E. (1928). The bottom faunaof the middle Illinois River 1913-1925: its distribution, abundance, valuation and index value in the study of stream pollution. *Bull. Illinois State Nat. Hist. Survey*, 17, 387-475.

Rizzoli, A.E. and Young, W.J. (1997). Delivering environmental decision support systems: software tools and techniques. *Environmental Modelling & Software*,12, 237-249.

Rodriguez Capitulo, A., Tangorra, M. and Ocon, C. (2001). Use of benthic macroinvertebrates to assess the biological status of Pampean streams in Argentina. *Aquatic Ecology*, 35, 109-119.

Roldan, G. (1992). Guia para el estudio de los macroinvertebrados acuaticos de Antioquia, Colombia.

Rogers, L.L. and Dowla, F.U. (1994). Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resources Research*, 30(2), 457-481.

Rosenberg, D.M. and Resh, V.H. (1993). Freshwater Monitoring and Benthic macroinvertebrates, Chapman and Hall, New York.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986). Learning representations by back-propagation errors. *Nature*, 323, 533-536.

Sandin, L., Sommerhäuser, M., Stubauer, I., Hering, D. and Johnson, R. (2000). Stream assessment methods, stream typology approaches and outlines of a European stream typology. AQEM-EU project EVK1-CT1999-00027, pp. 43.

Sandin, L. (2001). SWEPACS: a Swedish running water prediction and classification system using benthic macroinvertebrates. In: Bäck, S. and Karttunen, K. (eds.), *Classification of ecological status of lakes and rivers*. TemaNord Environment. pp. 44-46.

Scardi, M. and Harding, L.W. (1999). Developing an empirical model of phytoplankton primary production: A neural network case study. *Ecological Modelling*, 120(2-3), 213-223.

Schlegel, S. and Baumann, P. (1996). Requirements with respect to on-line analysers for N and P. *Water Science & Technology*, 33(1), 139-146.

Schleiter, I.M., Borchardt, D., Wagner, R., Dapper, T., Schmidt, K.D., Schmidt H.H. and Werner, H. (1999). Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks. *Ecological Modelling*, 120(2-3), 271-286.

Schleiter, I.M., Obach, M., Borchardt, D. and Werner, H. (2001). Bioindication of chemical and hydromorphological habitat characteristics with benthic macro-invertebrates based on artificial neural networks. *Aquatic Ecology*, 35, 147-158.

Scoccimarro, M., Walker, A., Dietrich, C., Schreider, S., Jakeman, T. and Ross, H. (1999). A framework for integrated catchment assessment in northern Thailand. *Environmental Modelling & Software*, 14, 567-577.

Seegert, G. (2000). The development, use, and misuse of biocriteria with an emphasis on the index of biotic integrity. *Environmental Science & Policy*, 3, 51-58.

Shannon CE and Weaver W (1949). The mathematical theory of communication. University of Illinois Press, Urbana.

Sheffer, M. (1990). Multiplicity of stable states in freshwater systems. *Hydrobiologia*, 200-201, 475–486.

Sharma, S. and Moog, O. (2001). Introducing the NEPBIOS method of surface water quality monitoring. Aquatic Ecology Centre, Kathmandu University, Nepal.

Shi, T. (2004). Ecological economics as a policy science: rhetoric or commitment towards an improved decision-making process on sustainability. *Ecological Economics*, 48, 23-36.

Sigua, G.C. and Tweedale, W.A. (2003). Watershed scale assessment of nitrogen and phosphorus loadings in the Indian River Lagoon basin, Florida. *Journal of Environmental Management*, 67, 363-372.

Silvert, W. (2000). Fuzzy indices of environmental conditions. *Ecological Modelling*, 130, 111-119.

Skriver, J., Friberg, N. and Kirkegaard, (2001). Biological assessment of watercourse quality in Denmark: Introduction of the Danish Stream Fauna Index (DSFI) as the official biomonitoring method. *Verhandlungen Internationale Vereinigung für theoretische und angewandte Limnologie*, 27, 1822-1830.

Sladecek V. (1973). The reality of three British biotic indices. *Water Research*, 7, 995-1002.

Smith, M.J., Kay, W.R., Edward, D.H.D., Papas, P.J., Richardson, K.S.J., Simpson, J.C., Pinder, A.M., Cale, D.J., Horwitz, P.H.J., Davis, J.A., Yung, F.H., Norris R.H. and Halse S.A. (1999). AusRivAS: using macroinvertebrates to assess ecological condition of rivers in Western Australia. *Freshwater Biology*, 41, 269-282.

Söderström, O. (1987). Upstream movements of invertebrates in running waters – a review. *Archiv fur Hydrobiologie*, 111, 197-208.

Sorensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analysis of the vegetation on Danish commons. Biologiske Skifter. *Det Konglige Danske Videnskabernes Selskab*, 5, 1-34.

Speirs, D.C. and Gurney, W.S.C. (2001). Population persistence in rivers and estuaries. *Ecology*, 82, 1219-1237.

STOWA (1992). Ecologische beoordeling en beheer van oppervlaktewater: Beoordelingssstseem voor stromende wateren op basis van macrofauna. STOWA Rapport, 92-8, 1-86.

Sudaryanti, S., Trihadiningrum, Y., Hart, B.T., Davies, P.E., Humphrey, C., Norris, R., Simpson, J. and Thurtell L. (2001). Assessment of the biological health of the Brantas River, East Java, Indonesia using the Australian River Assessment System (AUSRIVAS) methodology. *Aquatic Ecology*, 35, 135-146.

Sung, A.H. (1998). Ranking importance of input parameters of neural networks. *Expert systems with Applications*, 15, 405-411.

Sweeting, R., Lowson, D., Hale, P. and Wright, J. (1992). 1990 Biological assessment of rivers in the UK. In: Newman, P., Piavaux, A. and Sweeting, R. (eds.), *River water quality – assessment and control*. Commission of the European Communities. EUR 14606 EN-FR, 1992-III. Brussels, 319-326.

Tachet, H., Richoux, P., Bournaud, M. and Usseglio-Polatera, P. (2002). Invertébrés d'eau douce. Systématique, biologie, écologie.CNRS Editions, Paris, pp. 587.

Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control. IEEE Trans. *Systems, Mananagement and Cybernetics*, 15(1), 116-132.

Tattari, S., Schultz, T. and Kuussaari, M. (2003). Use of belief network modelling to assess the impact of buffer zones on water protection and biodiversity. *Agriculture Ecosystems & Environment*, 96, 119-132.

Terano, T., Asai, K. and Sugeno, M. (1994). Applied Fuzzy Systems. C.G. Aschmann III, trans. Academic Press, Boston, pp. 302.

Tolkamp, H.H. (1980). Organic-substrate relationships in lowland streams. PhD thesis, Wageningen, the Netherlands.

Tolkamp, H.H. (1982). Microdistribution of macroinvertebrates in lowland streams. *Hydrobiological Bulletin*, 16(2-3), 133-148.

Toran, F., Ramirez, D., Navarro, A.E., Casans, S., Pelegri, J. and Espi, J.M. (2001). Design of a virtual instrument for water quality monitoring across the Internet. *Sensors and Actuators B-Chemical*, 76, 281-285.

Trigg, D.J., Walley, W.J. and Ormerod, S.J. (2000). A prototype Bayesian belief network for the diagnosis of acidification in Welsh rivers. In: Brebbia C.A., Ibarra-Berastegui, G. and Zannetti, P. (eds.), *Development and Application of Computer Techniques to Environmental Studies*. ENVIROSOFT 2000, Bilbao, Spain.

Turner, R.K., Bateman, I.J. and Adger, W.N. (2001). Ecological Economics and Coastal Zone Ecosystems' values: an overview. In: Turner, R.K., Bateman, I.J. and Adger, W.N. (eds.), *Economics of Coastal and Water Resources: Valuing Environmental Functions*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 1-43.

Turner, R., Pearce, D. and Bateman, I. (1994). Environmental Economics, Harlow, UK.

US-EPA (1996). Biological criteria: Technical guidance for streams and small rivers. U.S. Environmental Protection Agency, Office of Water, Washington DC. EPA-822-B96-001.

Uzunov, Y., Penev, L., Kovachev, S. and Baev, P. (1998). Bulgarian Biotic Index (BGBI) – an express method for bioassessment of the quality of running waters. *Comptes rendus de l'Académie bulgare des Sciences*, 51(11-12), 117-120.

Vandenberghe, V., Goethals, P.L.M., van Griensven, A., Meirlaen, J., De Pauw, N., Vanrolleghem, P. and Bauwens, W. (2004). Applications of automated measurement stations for continuous water quality monitoring of the Dender river in Flanders, Belgium. *Environmental Monitoring and Assessment*. (accepted)

Vandenberghe, V., van Griensven, A. and Bauwens, W. (2001). Detection of the most optimal measuring points for water quality variables: application to the river water quality model of the river Dender in ESWAT'. In: Proceedings of the 2nd World Water Congress of the International Water Association, October 15-19, 2001, Berlin, Germany.

Vanden Bossche, J.P. and Josens, G. (2003). Macrozoobenthos biodiversity and biological quality monitoring of watercourses in Wallonia (Belgium). In: Symoens, J.J. and Wouters, K. (eds.), *Biological evaluation and monitoring of the quality of surface waters*. Nat.Com. of Biological Sciences and SCOPE Nat. Com., Brussels, pp. 77-79.

Van de Guchte, C. (1992). The sediment quality TRIAD: an integrated approach to assess contaminated sediments. p. 417-431. In: *River water quality. Ecological assessment and control*. Commission of the European Communities, Brussels, Belgium.

van der Molen, D. (1999). The role of eutrophication models in water management. PhD thesis, Landbouwuniversiteit Wageningen. pp. 167.

Vangheluwe, M., Vandenbroele, M. and Van Sprang, P. (2002). Exposure assessment of copper and zinc for the European sediment compartment. Euras, pp. 68.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell J.R. and Cushing, C.E. (1980). The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, 37, 130-137.

van Puijenbroek, P.J.T.M. and Knoop, J.M. (2002). Integrated modelling for nutrient loading of polder lakes. Integrated Assessment and Decision Support. In: Rizzoli, A.E. and Jakeman, A.J. (eds.), *Integrated Assessment and Decision Support*, Proceedings of the 1st biennial meeting of the International Environmental Modelling and Software Society, Lugano, Switzerland, 1, pp. 287–292.

van Puijenbroek, P.J.T.M., Janse, J.H. and Knoop, J.M. (2004). Integrated modelling for nutrient loading and ecology of lakes in The Netherlands. *Ecological Modelling*, 174, 127-141.

Vanrolleghem, P.A. and Lee, D.S. (2003). On-line measurement equipment for wastewater treatment processes: state of the art. *Water Science & Technology*, 47(2), 1-34.

Vanrolleghem, P.A., Schilling, W., Rauch, W., Krebs, P. and Aalderink, H. (1999). Setting up measuring campaigns for integrated wastewater modelling. *Water Science & Technology*, 39(4), 257-268.

Verdonschot, P.F.M. (1990). Ecological characterization of surface waters in the Province of Overijssel (the Netherlands). Research Institute for Nature Management, Leersum, pp. 255.

Verdonschot, P.F.M., Driessen, J.M.C., Mosterdijk, H.G., and Schot, J.A. (1998). The 5-S-Model, an integrated approach for stream rehabilitation. p. 36-44. In: Hansen, H.O. and Madsen, B.L. (eds.), *River Restoration '96*. European Centre for River Restoration, Denmark.

Verdonschot, P.F.M. and Nijboer, R.C. (2002). Towards a decision support system for stream restoration in the Netherlands: an overview of restoration projects and future needs. *Hydrobiologia*, 478, 131-148.

VMM (2000). Water quality - water discharges 1999. Flemish Environmental Agency, VMM, Erembodemgem. (in Dutch)

Voinov, A., Fitz, C., Boumans, R. and Costanza, R. (2004). Modular ecosystem modeling. *Environmental Modelling and Software*, 19, 285-304.

Wagner, R., Dapper, T. and Schmidt, H.H. (2000). The influence of environmental variables on the abundance of aquatic insects: a comparison of ordination and artificial neural networks. *Hydrobiologia*, 422-423, 143-152.

Walczak, S. and Cerpa, N. (1999). Heuristic principles for the design of artificial neural networks. *Information and Software Technology*, 41(2), 107–117.

Walczak, S. (1995). Developing neural nets currency trading. *Artificial Intelligence in Finance*, 2(1), 27-34.

Walley, W.J. and Dzeroski, S. (1995). Biological monitoring: a comparison between bayesian, neural and machine learning methods of water quality classification. In: Denzer, R., Schimak, G. and Russell, D. (eds.), *Environmental Software Systems*. Chapman & Hall, London, pp. 229-240.

Walley, W.J. and Fontama, V.N. (1998). Neural network predictors of average score per taxon and number of families at unpolluted river sites in Great Britain. *Water Research*, 32(3), 613-622.

Walley, W.J., O'Connor, M.A., Trigg, D.J. and Martin, R.W. (2002). Diagnosing and predicting river health from biological survey data using pattern recognition and plausible reasoning. R&D Technical Report E1-056/TR. Environment Agency, Swindon, UK.

Wang, F. (1994). The use of artificial neural networks in a geographical information system for agricultural land-suitability assessment. *Environment and Planning A*, 26(2), 265-284.

Wang, J. and Thongngamdee, S. (2003). On-line electrochemical monitoring of (TNT) 2,4,6-trinitrotoluene in natural waters. *Anal. Chim. Acta*, 485, 139-144.

Warren, L.A., Tessier, A. and Hare, L. (1998). Modelling cadmium accumulation by benthic invertebrates in situ: the relative contributions of sediment and overlying water reservoirs to organism cadmium concentrations. *Limnol. Oceanogr*., 43, 1442–1454.

Washington H.G. (1984). Diversity, biotic and similarity indices: a review with special relevance to aquatic ecosystems. *Water Research*,18(6), 653-694.

Waters, T.F. (1965). Interpretation of invertebrate drift in streams. *Ecology*, 46, 327-334.

Weigend, A.S., Huberman, B.A. and Rumelhart, D.E. (1990). Predicting the future: A connectionist approach. *International Journal of Neural Systems*, 1(3), 193-209.

Wesenberg-Lund, C. (1982). Biologie der süsswassertiere. Wirbellose tiere. Julius springer, Wien, Cramer, Braunschweig, Koeltz, Koenigstein.

Whitehead, P.G., Howard, A. and Arulmani, C. (1997). Modelling algal growth and transport in rivers. A comparison of time series anlaysis, dynamic mass balance and neural network techniques. *Hydrobiologia*, 349, 39-46.

Whitehurst, I.T. (1988). Factors affecting the Gammarus to Asellus ratio in unpolluted and polluted waters. PhD thesis, Brighton Polytechnic, Brighton, England.

Whitehurst, I.T. and Lindsey, B.I. (1990). Impact of organic enrichment on the benthic macroinvertebrate communities of a lowland river. *Water Research*, 24(5), 625-630.

Whittaker, R.H. (1952). A study of summer foliage insect communities in the Great Smoky Mountains. *Ecological Monographs*, 22,1-44.

Williams, W.T. (1971). Principles of clustering. *A. Rev. Ecol. Syst.*, 2, 303-326.

Wilhm, J.L. and Dorris (1966). Species diversity of benthic macroinvertebrates in a stream receiving domestic and oil refinery effluents. *American Midland Naturalist*, 76, 427-449.

Witten, I.H. and Frank, E. (2000). Data Mining: practical machine learning tools and techniques with Java implementations. Morgan Kaufmann Publishers, San Francisco. 369 pp.

Wolfson, L.J., Kadane, J.B. and Small, M.J. (1996). Bayesian environmental policy decisions: two case studies. *Ecological Applications*, 6(4), 1056-1066.

Woodiwiss, F.S. (1964). The biological system of stream classification used by the Trent River Board. *Chemical Industry*, 14, 443-447.

Woodiwiss, F.S. (1980). Biological monitoring of surface water quality. Summary Report. Commission of the European Communities. ENV/787/80-EN, pp. 45.

Wright, J.F., Furse, M.T. and Armitage, P.D. (1993). RIVPACS – a technique for evaluating the biological quality of rivers in the UK. *European Water Quality Control*, 3,15-25.

Wright, J.F., Sutcliffe, D.W., and Furse, M.T. (2000). Assessing the biological quality of fresh waters: RIVPACS and other techniques. FBA, Ambleside (UK), pp. 373.

Yao, J., Teng, N., Poh, H.L. and Tan, C.L. (1998). Forecasting and analysis of marketing data using neural networks. *Journal of Information Science and Engineering*, 14, 843-862.

Young, P.C. and Beck, M.B. (1974). The modeling and control of water quality in a river system. *Automatica*, 10, 455-468.

Young, W.J., Lam, D.C.L., Ressel, V. and Wong, I.W. (2000). Development of an environmental flows decision support system. *Environmental Modelling & Software*, 15, 257-265.

Yuan, W., James, P., Hodgson, K., Hutchinson, S.M. and Shi, C. (2003). Development of sustainability indicators by communities in China: a case study of Chongming County, Shanghai. *Journal of Environmental Management*, 68, 253-261.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8, 338-353.

Zelinka, M. and Marvan, P. (1961). Zur präzisierung der biologische Klassifikation der reinheit fliessender gewässer. *Archiv fur Hydrobiologie*, 57, 389-407.

Zimmerman, H.J. (1990). Fuzzy Sets Theory and Its Applications. Kluwer-Nijhoff.

Summary

The European Water Framework Directive (WFD) 2000/60/EC aims at a good ecological status for all water bodies in the member states of the European Community by 2015. A major part of these water bodies can be classified as running waters or rivers. According to the WFD, rivers are to be assessed by comparing the actual status to a reference status. To this end, reference communities must be described that represent a good ecological status. Additionally, for the development of a representative set of metrics for ecological river assessment, one needs to gain insight in the relation between the aquatic communities and the human activities affecting these water systems. Insights in these relations will also be valuable for detection of causes of particular river conditions (environmental impact assessment) as well as for decision-making in river restoration and protection management to meet and sustain the requirements set by the WFD.

Until now, ecological models have rarely been used to support river management and water policy. Models have however several interesting applications in this context. First of all, through these models a better interpretation of the river status can be possible, the causes of the status of a river can be detected and assessment methods can be optimised. Secondly, these models can allow for calculating the effect of future river restoration actions on aquatic ecosystems and supporting the selection of the most sustainable options. Thirdly, these models can help to find the major gaps in our knowledge of river systems and help to set-up cost effective monitoring programmes.

The present thesis aimed at determining the appropriate variables and ecosystem processes by using classification trees as well as artificial neural networks to predict biological communities present in rivers. The research focused on macroinvertebrates in brooks and small rivers in Flanders (Belgium). The applied modelling techniques in this research are all data driven approaches. In this manner, an *a priori* and often biased knowledge of ecological experts has not been used during the model development process. However, when discussing the results, the outcome of the data driven models has been compared to expert rules from literature. This approach allows for deriving rules that contribute to a better understanding of river ecosystems and support of their management.

The developed models have been applied to support decision-making in water management. In this way, a crucial validation step, often lacking in many model development and assessment studies has been made and this can probably also help to pursue river managers of the added value of such ecological models. These models can in this manner support the appropriate selection of sustainable management options and help to convince stakeholders to make the necessary investments and/or activity changes as desired by society.

Samenvatting

De Europese Kaderrichtlijn Water (KRW) 2000/60/EC stelt voor alle lidstaten van de Europese Unie een goede ecologische status van alle waterlichamen tegen 2015 voorop. Het overgrote deel van deze waterlichamen kan als stromende wateren of rivieren aanzien worden. Volgens de KRW moet de waterkwaliteit van de rivieren beoordeeld worden door de actuele condities te vergelijken met de referentiecondities. Daarom moeten eerst referentiecondities, die een goede ecologische status voorstellen, beschreven worden. Bijkomend moet de relatie tussen aquatische gemeenschappen en de menselijke activiteiten die deze watersystemen aantasten beter begrepen worden om zo een representatieve set van indices voor ecologisch rivierbeoordeling te kunnen ontwikkelen. Kennis over deze relaties kan eveneens nuttig zijn bij zowel het opsporen van oorzaken van bepaalde riviercondities (milieueffectrapportering) als bij beslissingsondersteuning inzake rivierherstel en beheer om zo aan de eisen van de KRW te voldoen.

Tot nu toe werden ecologische modellen zelden gebruikt bij de ondersteuning van rivierbeheer en waterbeleid. Modellen kennen in deze context nochtans verscheidene interessante toepassingen. Ten eerste kunnen deze modellen bijdragen tot een betere interpretatie van de huidige riviercondities, de oorzaken van bepaalde riviercondities kunnen achterhaald worden en beoordelingsmethoden kunnen geoptimaliseerd worden. Ten tweede kunnen deze modellen het effect van toekomstige rivierherstelmaatregelen op aquatische ecosystemen doorrekenen en de selectie van de beste herstelopties ondersteunen. Ten derde kunnen deze modellen de belangrijkste hiaten in onze kennis over riviersystemen helpen opvullen en helpen bij het opzetten van kostefficiënte monitoringsprogramma’s.

Deze thesis beoogt het bepalen van geschikte variabelen en ecosysteemprocessen door beslissingsbomen en artificiële neurale netwerken toe te passen bij het voorspellen van biologische gemeenschappen in rivieren. Het onderzoek richt zich vooral op macro-invertebraten in beken en smalle rivieren in Vlaanderen (België). De gebruikte modelleringstechnieken tijdens dit onderzoek zijn allen gegevensgebaseerd. Op deze manier werd tijdens het ontwikkelen van de modellen geen gebruik gemaakt van *a priori* en vaak vooringenomen kennis van ecologische experts. Bij de discussie werden de resultaten van de gegevensgebaseerde modellen echter wel vergeleken met expertregels uit de literatuur. Deze benadering laat toe om regels of te leiden die een beter inzicht geven in rivierecosystemen en de ondersteuning van hun beheer.

De ontwikkelde modellen werden praktisch toegepast om zo beslissingen te ondersteunen in het waterbeheer. Op deze manier werd een cruciale validatie toegevoegd, die bij het ontwikkelen van modellen en beoordelingsstudies vaak ontbreekt. Dit kan een belangrijke toegevoegde waarde betekenen voor rivierbeheerders. Deze modellen kunnen zo bijdragen tot het selecteren van geschikte beheersopties en kunnen helpen de beheerders te overtuigen de nodige investeringen en/of wijzigingen in activiteiten door te voeren zoals gewenst door de samenleving.

Curriculum vitae

Curriculum vitae

Name:	Peter Leon Maria Goethals
Date and place of birth:	31 May 1972, Sleidinge (Belgium)
Address (private):	Normaalschoolstraat 12, B-9000 Gent
Address (work)	Department Applied Ecology and Environmental Biology Ghent University J. Plateaustraat 22, B-9000 Gent E-mail: Peter.Goethals@UGent.be Tel.: 0032 (0)9 264 37 76 Fax.: 0032 (0)9 264 41 99

1 Education

1.1 Diplomas

1996: **Bio-Engineer, option Biotechnology**, Ghent University, Faculty of Agricultural and Applied Biological Sciences, Coupure Links 653, B-9000 Gent (5 years full time academic training, 300 credits). Script 'Genetic and eco-physiological aspects of aggregation in activated sludge', Promoter: Prof. Dr. ir. W. Verstraete.

1999: **Master of Science in Industrial Management** (Bedrijfskundig Ingenieur), Ghent University, Faculty of Applied Sciences, Jozef Plateaustraat 22, B-9000 Gent in co-operation with the Vlerick Leuven-Gent Management School, Reep 1, B-9000 Gent (2 years half time academic training, 60 credits). Script 'Development and application of automated measurement stations for surface water monitoring', Promoters: Prof. Dr. ir. R. Van Landeghem, Prof. Dr. N. De Pauw and Prof. Dr. ir. P. Vanrolleghem.

2001: **Master of Science in Knowledge Technology**, Ghent University, Faculty of Arts and Philosophy, Blandijnberg 2, B-9000 Gent (2 years full time academic training, 120 credits). Script 'Knowledge and information management in integrated ecological water research', Promoters: Prof. Dr. D. Vervenne and Prof. Dr. F. Vandamme.

2003: **Academic Teachers' Training Program, option Applied Biological Sciences**, Ghent University, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, B-9000 Gent (1 year half time academic training, 34 credits).

1.2 Certificates based on exams

1998: '**Expert systems**', Prof. Dr. ir. D. Matthys, Ghent University, Faculty Applied Sciences, Gent (15 h, 3 credits).

1998: ‘**Integrated water management**’ (Leerstoel Integraal Waterbeheer), Prof. Dr. P. Meire, Antwerp University (UIA), Antwerpen (15 h). Paper ‘Concepts for a knowledge centre for integrated water management’, Promotor: Prof. Dr. P. Meire.

2001: ‘**Practical English, level 4**’, Ghent University, Talencentrum, Gent (32 h).

1.3 Certificates based on attendance

1998: ‘**Modelling of rivers with AQUASIM**’, Dr. G. Goudsmit, Dr. P. Reichert and Dr. O. Wanner, EAWAG, Dübendorf, Suisse (12 h).

1999: ‘**Emotional intelligence**’, Prof. Dr. H. Van den Broeck, Vlerick Leuven-Gent Management School, Les Carroz D’Araches, France (15 h).

2000: ‘**GIS – Basic concepts, water management and geostatistics**’ Prof. Dr. ir. M. Van Meirvenne, Ghent University, IRI, Gent, Belgium (18 h).

2000: ‘**Analysis of environmental data with machine learning methods**’, Prof. Dr. S. Dzeroski, Jozef Stefan Institute, Ljubljana, Slovenia (30 h).

2001: ‘**Intelligent data analysis**’, Dr. R. Silipo, European School on Intelligent Data Analysis, University of Calcolo, Palermo, Italy (30 h).

1.4 Other courses I followed

1996: ‘**Meeting management**’ and ‘**Project management**’, Royal Society of Flemish Engineers (KVIV), ‘Communication days’ at the Ghent University campus, Gent, Belgium (6 h).

1999: **‘Environmental law’**, Prof. Dr. K. Deketelaere, Catholic University of Leuven (KULeuven), HIVA-training centre, Leuven, Belgium (90 h).

2001: ‘**Career planning**’, Royal Society of Flemish Engineers (KVIV) and Nicholson International, Antwerp, Belgium (6 h).

2002: ‘**From idea to spin-off: how to set up a spin-off company from academic research at the university**’, Dr. J. Bil, Prof. Dr. ir. B. Clarysse and Prof. Dr. S. Manigart, Ghent University, Gent, Belgium (15 h).

2 Work experience

1 November 1996 - 31 December 1997: Scientific assistant in the Laboratory of Microbial Ecology and Technology (promoter: Prof. W. Verstraete), Ghent University.

Project research related to bio-remediation and ecological risk assessment of soils contaminated with mineral oil (project supported by OVAM) and to the use of compost in the remediation of contaminated soils (project supported by VLACO).

1 Ferbruary 1998 – present: Scientific assistant in the Laboratory of Ecotoxicology and Aquatic Ecology (promoter: Prof. N. De Pauw), Ghent University.

Project research related to the use of knowledge technology for the optimization of monitoring, assessment and modeling of ecosystems to support water management and policy making.

3 Scientific activities and results

3.1 Publications

3.1.1 International journal publications with peer-review

Bossier, P., Goethals, P. & Rodrigues-Pousada, C. (1997). Constitutive flocculation in *Saccharomyces cerevisiae* through overexpression of GTS1 gene, coding a 'Glo'-type Zn-finger-containing protein. Yeast 13: 717-725.

Goethals, P. & De Pauw, N. (2001). Development of a concept for integrated river assessment in Flanders, Belgium. Journal of Limnology 60: 7-16.

Dedecker, A.P., Goethals, P.L.M. & De Pauw, N. (2002). Comparison of artificial neural network (ANN) model development methods for prediction of macroinvertebrate communities in the Zwalm river basin in Flanders, Belgium. The Scientific World Journal 2: 96-104.

Adriaenssens, V., Goethals, P. & De Pauw, N. (2002). Assessment of land-use impact on macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium) using multivariate analysis and Geographical Information Systems. The Scientific World Journal 2: 546-557.

D'heygere, T., Goethals, P. & De Pauw, N. (2002). Optimisation of the monitoring strategy of macroinvertebrate communities in the river Dender, in relation to the EU Water Framework Directive. The Scientific World Journal 2: 607-617.

Gabriels, W., Goethals, P.L.M. & De Pauw, N. (2002). Prediction of macroinvertebrate communities in sediments of Flemish watercourses based on artificial neural networks. Verhandlungen Internationale Vereinigung für theoretische und angewandte Limnologie 28: 777-780.

D'heygere, T., Goethals, P.L.M. & De Pauw N. (2003). Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates. Ecological Modelling 160: 291-300.

Adriaenssens, V., De Baets, B., Goethals, P.L.M. & De Pauw, N. (2004). Fuzzy rule-based models for decision support in ecosystem management. The Science of the Total Environment 319: 1-12.

Adriaenssens, V., Simons, F., Nguyen, L.T.H., Goddeeris, B., Goethals, P.L M. and De Pauw, N. (2004). Potential of bio-indication of Chironomid communities for assessment of running water quality in Flanders (Belgium). Belgian Journal of Zoology 134: 15-24.

Dedecker, A.P., Goethals, P.L.M., Gabriels, W. & De Pauw, N. (2004). Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). Ecological Modelling 174 (1-2): 161-173.

Breine, J.J., Simoens, I., Goethals, P., Quataert, P., Ercken, D., Van Liefferinghe, C. & Belpaire, C. (2004). A fish-based index of biotic integrity for upstream brooks in Flanders (Belgium). Hydrobiologia 522 (1-3): 133-148.

Adriaenssens, V., Goethals, P.L.M., Trigg, D. & De Pauw, N. Application of Bayesian belief networks to develop transparent ecological models to support river restoration management. Annales de Limnologie – International Journal of Limnology 40(3): 181-191.

D'heygere, T., Goethals, P.L.M. & De Pauw, N. (in press). Genetic algorithms for optimisation of predictive ecosystem models based on decision trees and neural networks. Ecological Modelling.

Adriaenssens, V., Goethals, P.L.M. and De Pauw, N. (in press). Fuzzy knowledge-based models for prediction of macroinvertebrate taxa in watercourses in Flanders, Belgium. Ecological Modelling.

Gabriels, W., Goethals, P. & De Pauw, N. (in press). Implications of taxonomic identification levels on the Belgian Biotic Index assessment method. Hydrobiologia.

Vandenberghe, V., Goethals, P.L.M., van Griensven, A., Meirlaen, J., De Pauw, N., Vanrolleghem, P. & Bauwens, W. (in press). Application of automated measurement stations for continuous water quality monitoring of the Dender river in Flanders, Belgium. Environmental Monitoring and Assessment.

Dedecker, A.P., Goethals, P.L.M., D’heygere, T., Gevrey, M., Lek, S. & De Pauw, N. (in press). Application of Artificial Neural Network models to analyse the relationships between *Gammarus pulex* L. (Crustacea, Amphipoda) and river characteristics. Environmental Monitoring and Assessment.

Gabriels, W., Goethals, P.L.M. & De Pauw, N. (accepted). Development of a multimetric assessment system based on macroinvertebrates for rivers in Flanders (Belgium) according to the European Water Framework Directive. Verhandlungen Internationale Vereinigung für theoretische und angewandte Limnologie.

Goethals, P.L.M., Dedecker, A.P., Bouma, J.J., François, D., Verstraete, A. & De Pauw, N. (submitted). The Water Ecology decision support system (WAECO-DSS) for integrated cost-benefit analyses in river restoration management: case study of the Zwalm river basin (Belgium). Journal of Environmental Management.

Dedecker, A.P., Goethals, P.L.M., D’heygere, T. & De Pauw, N. (submitted). Development and validation of an in-stream migration model for *Gammarus pulex* based on expert rules and GIS-layers. Aquatic Ecology.

Dedecker, A.P., Goethals, P.L.M., D’heygere, T., Gevrey, M., Lek, S. & De Pauw, N. (submitted). Habitat preference study of *Asellus* (Crustacea, Isopoda) by applying input variable contribution methods to Artificial Neural Networks. Environmental Modelling and Assessment.

Adriaenssens, V., Verdonschot, P.F.M., Goethals, P.L.M. & De Pauw, N. (submitted). Application of clustering techniques for the characterisation of macroinvertebrate communities to support river restoration management. Aquatic Ecology.

Dakou, E., Lazaridou-Dimitriadou, M., D’heygere, T., Dedecker, A, Goethals, P.L.M. & De Pauw, N. (submitted). Development of models predicting macroinvertebrate communities in Greek rivers using rule induction techniques. Aquatic Ecology.

Dakou, E., Lazaridou-Dimitriadou, M., D’heygere, T., Dedecker, A., Goethals, P.L.M. & De Pauw, N. (submitted). Development of supervised artificial neural network models predicting macroinvertebrate communities in Greek rivers. Annales de Limnologie – International Journal of Limnology.

3.1.2 Publications in international proceedings with peer-review

Goethals, P.L.M., Nicolletto, L., Kersters, I., Bossier, P. & Verstraete, W. (1997). Effect of reactive oxygen intermediates on microbial cell and communities: a model for the beneficial effect of Nutrifloc 50 S on activated sludge, p. 241-244. In: Verachtert, H. & Verstraete, W. (Eds.), Royal Flemish Society of Engineers (KVIV), Proceedings ‘International Symposium Environmental Biotechnology’, 21-23 April 1997, Oostende, Belgium.

van Griensven, A., Vandenberghe, V., Bols, J., De Pauw, N., Goethals, P., Meirlaen, J., Vanrolleghem, P.A., Van Vooren, L. & Bauwens, W. (2000). Experience and organisation of automated measuring stations for river water quality monitoring. In: International Water Association (IWA), Proceedings ‘1st World Congress of the International Water Association’, 3-7 July 2000, Paris, France. CD-Rom publication.

Gabriels, W., Goethals, P.L.M., Heylen, S., De Cooman, W. & De Pauw, N. (2000). Modelling benthic macro-invertebrate communities in Flanders using artificial neural networks, p. 1.143-1.146. In: International Water Association (IWA), Proceedings ‘5th International Symposium System Analysis and Computing in Water Quality Management (Watermatex 2000)’, 18-20 September 2000, Gent, Belgium.

D'heygere, T., Goethals, P.L.M. & De Pauw, N. (2002). Use of genetic algorithms to select input variables in artificial neural network models for the prediction of benthic

macroinvertebrates, Vol. 2, p. 136-141. In: Rizzoli, A.E. & Jakeman, A.J. (Eds.), The International Environmental Modelling and Software Society (iEMSs), Proceedings ‘Integrated Assessment and Decision Support’, 24-27 June 2002, Lugano, Switzerland. 618 p.

Dedecker, A., Goethals, P.L.M., Gabriels, W. & De Pauw, N. (2002). Optimisation of Artificial Neural Network (ANN) model design for prediction of macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium), Vol. 2, p. 142-147. In: Rizzoli, A.E. & Jakeman, A.J. (Eds.), The International Environmental Modelling and Software Society (iEMSs), Proceedings ‘Integrated Assessment and Decision Support’, 24-27 June 2002, Lugano, Switzerland. 618 p.

D'heygere, T., Goethals, P.L.M., Dedecker, A. & De Pauw, N. (2003). Development of a monitoring network to model the habitat suitability of macroinvertebrates in the Zwalm river basin (Flanders, Belgium), Vol. 2, p. 753-758. In: Post, D.A., Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ), Proceedings ‘Integrative Modelling of Biophysical, Social and Economic Systems for Resource Management Solutions (MODSIM 2003)’, 14-17 July 2003, Townsville, Australia. 2066 p.

Goethals, P.L.M., Bouma, J.J., François, D., D'heygere, T., Dedecker, A., Adriaenssens, V. & De Pauw, N. (2003). Coupling ecosystem valuation methods to the WAECO decision support system in the Zwalm Catchment (Belgium). Vol. 3, p. 971-976. In: Post, D.A., Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ), Proceedings ‘Integrative Modelling of Biophysical, Social and Economic Systems for Resource Management Solutions (MODSIM 2003)’, 14-17 July 2003, Townsville, Australia. 2066 p.

Adriaenssens, V., Dedecker, A., D’heygere, T., Goethals, P.L.M. & De Pauw, N. (2003). Relations between structural characteristics and macroinvertebrate communities in the Zwalm river basin at different spatial scales, p. 81-86. In: Symoens, J.-J. & Wouters, K. (Eds.), National Committee of Biological Sciences, Proceedings ‘Biological Evaluation and Monitoring of the Quality of Surface Waters’, 10 October 2002, Brussels, Belgium. 129 p.

Adriaenssens, V., Simons, F., Goethals, P.L.M., Goddeeris, B. & De Pauw, N. (2003). Ecological analysis of the Chironomid communities in the Zwalm river basin, p. 87-93. In: Symoens, J.-J. & Wouters, K. (Eds.), National Committee of Biological Sciences, Proceedings ‘Biological Evaluation and Monitoring of the Quality of Surface Waters’, 10 October 2002, Brussels, Belgium. 129 p.

Dedecker, A.P., Van Melckebeke, K., D'heygere, T., Goethals, P.L.M. & De Pauw, N. (2004). Studying the impact of weirs for water quantity control on aquatic macroinvertebrates in rivers by means of neural networks and migration models. p. 1157-1161. In: García de Jalón, D. (Ed.), International Association of Hydraulic Research (IAHR), Proceedings ‘The 5th International Symposium on Ecohydraulics: Aquatic Habitats: Analysis & Restoration’, 12-17 September 2004, Madrid, Spain. 1453 p.

3.1.3 National journal publications with peer-review

Adriaenssens, V., Goethals, P.L.M., Breine, J.J., Maes, J., Simoens, I., Ercken, D., Belpaire, C., Ollevier, F. & De Pauw, N. (2002). Importance of references in the development of an estuarine fish index in Flanders (in Dutch). Landschap 19: 58-61.

3.1.4 Other journal publications

Goethals, P.L.M., Nicolletto, L., Bossier, P. & Verstraete, W. (1996). A model for the beneficial effects of Nutrifloc 50 S on the settlement of activated sludge. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen: 2035-2042.

Goethals, P.L.M., Devliegher, W., Van Acker, E. & Verstraete, W. (1997). Applications of compost in soil sanitation: bioremediation of soils contaminated with mineral oil. VLARIO - December.

Goethals, P. (1997). Genetic and eco-physiological aspects of aggregation in activated sludge (in Dutch). Energie & Milieu 5: 235-237.

Bols, J., Goethals, P.L.M., Meirlaen, J., van Griensven, A., Vandenberghe, V., Van Vooren, L., De Pauw, N., Vanrolleghem, P. & Bauwens, W. (1999). Automated measurement stations for river water quality monitoring. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 64(5a): 107-110.

Rousseau, D., Goethals, P., Meirlaen, J., Verboven, J., Vanrolleghem, P. & De Pauw, N. (1999). Monitoring and modeling of free-water-surface constructed wetlands. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 64(5a): 191-196.

Adriaenssens, V., Goethals, P. & De Pauw, N. (2000). Development of an integrated system for ecological river quality assessment. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 65(4): 215-218.

Gabriels, W., Goethals, P.L.M. & De Pauw, N. (2000). Modelling of river ecosystems based on the use of artificial neural networks. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 65(4): 211-214.

Vandenberghe, V., van Griensven, A., Bauwens, W., Goethals, P., De Pauw, N., Meirlaen, J., Van Vooren, L. & Vanrolleghem, P. (2000). The importance of on-line water quality measurements, application on the river Dender (in Dutch). @WEL 5: 1-6.

Goethals, P.L.M., Wieme, U., Bols, J., Rousseau, D., De Pauw, N., Meirlaen, J., Van Vooren, L., Vanrolleghem, P.A., Vandenberghe, V., van Griensven, A. & Bauwens, W. (2000). Quality assurance of automated measurement stations for the 'on-line' quality monitoring of surface waters (in Dutch). @WEL 6: 1-9.

Adriaenssens, V., Goethals, P., Luypaert, P. & De Pauw, N. (2001). Impact of land-use on macroinvertebrate communities in the Zwalm river basin. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66(1): 51-61.

Gabriels, W., Goethals, P.L.M., Hermans, P. & De Pauw, N. (2001). Development of short and long-term management options for Bergelenput to avoid fish kills caused by algal blooms. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66(1): 63-70.

Goethals, P., Dedecker, A., Raes, N., Adriaenssens, V., Gabriels, W. & De Pauw, N. (2001). Development of river ecosystem models for Flemish watercourses: case studies in the Zwalm river basin. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66(1): 71-86.

Goethals, P. (2001). Resources guide: water and ecology. Waterlines 20: 32.

D'heygere, T., Goethals, P., van Griensven, A., Vandenberghe, V., Bauwens, W., Vanrolleghem, P. & De Pauw, N. (2001). Optimisation of the discrete conductivity and dissolved oxygen monitoring using continuous data series obtained with automated measurement stations. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66(3): 149-153.

Adriaenssens, V., Goethals, P. & De Pauw, N. (2001). Development of a fuzzy expert system for the prediction of macroinvertebrate taxa. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66(4): 225-228.

Dedecker, A., Goethals, P., Gabriels, W. & De Pauw, N. (2001). River management applications of ecosystem models predicting aquatic macroinvertebrate communities based on artificial neural networks (ANNs). Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66(4): 207-211.

D'heygere, T., Goethals, P. & De Pauw, N. (2001). Application of genetic algorithms for input variables selection of decision tree models predicting mollusca in unnavigable Flemish watercourses. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66(4): 219-223.

Goethals, P., Gasparyan, K. & De Pauw, N. (2001). River restoration simulations by ecosystem models predicting aquatic macroinvertebrate communities based on J48 classification trees. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66(4): 213-217.

Raes, N., Goethals, P., Adriaenssens, V. & De Pauw, N. (2001). Predicting Gammaridae (Crustaceae, Isopoda) in the Zwalm river basin (Flanders, Belgium) by means of fuzzy logic models. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66(4): 229-232.

Dedecker, A., Goethals, P.L.M., D'heygere, T. & De Pauw N. (2002). Use of Artificial Neural Networks (ANNs) and Geographical Information Systems (GIS) to simulate the migration of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 67: 105-108.

D'heygere, T., Adriaenssens, V., Dedecker, A., Gabriels, W., Goethals, P.L.M. & De Pauw, N. (2002). Development of a decision support system for integrated water management in the Zwalm river basin, Belgium. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 67: 159-162.

Gabriels, W., Adriaenssens, V., Goethals, P.L.M. & De Pauw, N. (2002). Monitoring of macroinvertebrate communities for the ecological evaluation of valuable upstream brooks in

Flanders, Belgium. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 67: 145-147.

Adriaenssens, V., D’heygere, T., Dedecker, A., Goethals, P.L.M. & De Pauw, N. (2002). Relations between structural characteristics and macroinvertebrate communities in the Zwalm river basin at different spatial scales. Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 67: 129-132.

Adriaenssens, V., Goethals, P.L.M., De Pauw, N., Breine, J.J., Simoens, I., Maes, J., Ercken, D., Belpaire, C. & Ollevier, F. (2002). Development of an estuarine fish index in Flanders (in Dutch). Water 2: 1-13.

Dedecker, A.P., Goethals, P.L.M. & De Pauw, N. (2003). Overview and quantification of the factors affecting the upstream and downstream movements of *Gammarus pulex* (Amphipoda). Communications in Applied Biological Sciences 68: 25-31.

Goethals, P. (2003). Ecological informatics applications in water management. Newsletter IBW-IN January 2003: 3.

Charles, J., Goethals, P.L.M., Adriaenssens, V. & De Pauw, N. (2003). Bayesian belief networks for the prediction of macroinvertebrate taxa in the Zwalm river basin. Communications in Applied Biological Sciences 68: 79-82.

De Pauw, N., Rousseau, D., Goethals, P. & Van Minh, P. (2003). Wastewater treatment by natural systems in Ho Chi Minh City (Vietnam). Center for Environmental Sanitation Newsletter 1: 20-21.

Stuer, V., De Ridder, K., Verbist, B., Adriaenssens, V., Goethals, P. & De Pauw, N. (2004). Development of a biological water quality assessment system for the Sumberjaya watershed in West-Lampung, Sumatra (Indonesia). Center for Environmental Sanitation Newsletter 2: 40-41.

Ambelu Bayih, A., Goethals, P., Legesse, W. & De Pauw, N. (2004). Development of decision support techniques for monitoring, assessment and management of rivers in Ethiopia. Center for Environmental Sanitation Newsletter 2: 38-39.

Dominguez, L., Matamoros, D., Goethals, P. & De Pauw, N. (2004). Development of a monitoring network to assess the impact of banana farming on the Chaguana river in Ecuador. Center for Environmental Sanitation Newsletter 2: 46-47.

Van Minh, P., Goethals, P., Rousseau, D., Cahn, D. & De Pauw, N. (2004). Wastewater-fed aquaculture in the South of Vietnam. Center for Environmental Sanitation Newsletter 2: 43.

3.1.5 Books as editor

Goethals, P.L.M. & De Pauw, N. (Eds., accepted). Applications of Ecological Informatics in Water Management. Special issue of Aquatic Ecology, Kluwer, Dordrecht, The Netherlands.

3.1.6 Chapters in books

Goethals, P. & De Pauw, N. (2001). Ecological informatics applied to decision support in river management: case studies for education purposes. p. 164-172. In: Hens, L., Boon, E.K., Sinitsyn, M., Rusanov, A. & Solntsev, V. (Eds.), Ecological field training in the boreal forest zone. A manual for students and teachers, Russian Academy of Sciences, Moscow, Russia.

Goethals, P.L.M. (2002). 2.3.9. Ecological informatics in river management. p. 203-213. In: Meij, J.M. (Ed.), Dealing with the data flood. Mining data, text and multimedia. Netherlands Study Centre for Technology Trends (STT), The Hague, The Netherlands, 896 p.

Goethals, P., Dedecker, A., Gabriels, W. & De Pauw, N. (2002). Development and application of predictive river ecosystem models based on classification trees and artificial neural networks, p. 91-107. In: Recknagel, F. (Ed.), Ecological Informatics: Understanding ecology by biologically-inspired computation, Springer-Verlag, Berlin Heidelberg, Germany. 398 p.

Dedecker, A. P., Goethals, P.L.M. De Pauw, N. (in press). Sensitivity and robustness of predictive neural network ecosystem models for simulation of different management scenarios. In: Modelling community structure in freshwater ecosystems, Scardi, M. (Ed.), Springer-Verlag, Berlin Heidelberg, Germany.

Goethals, P.L.M. & De Pauw, N. (in press). Integrated water management: headword definition. In: Deconinck, S., Dictionary of Sustainable Development, Center of Sustainable Development (CDO), Ghent University, Gent, Belgium. CD-Rom publication.

Goethals, P.L.M. & De Pauw, N. (in press). Drought impacts: headword definition. In: Deconinck, S., Dictionary of Sustainable Development, Center of Sustainable Development (CDO), Ghent University, Gent, Belgium. CD-Rom publication.

De Pauw, N., Gabriels, W. & Goethals, P.L.M. (accepted). Monitoring and assessment of macroinvertebrates in freshwaters. In: Biomonitoring. Wiley, New York, USA.

3.1.7 Books of abstracts as editor

Goethals, P. & De Pauw, N., Eds. (1998). Workshop: 'Ecological assessment of surface waters in The Netherlands and Flanders', 8 October 1998: book of abstracts (in Dutch). Dutch Aquatic Ecology Society, Antwerp, Belgium. 43 p.

Goethals, P. & De Pauw, N., Eds. (1999). Workshop: 'Natural systems for treatment of (waste)water in The Netherlands and Flanders', October 21, 1999: book of abstracts (in Dutch). Dutch-Flemish Ecology Society, Antwerp, Belgium. 67 p.

Goethals, P. & De Pauw, N. (2002). Ecological informatics applications in water management conference, 6-7 November 2002: book of abstracts, Dutch-Flemish Ecology Society, Ghent, Belgium. 44 p.

Goethals, P., Duel, H., Dunbar, M., Harby, A. & De Pauw, N., Eds. (2002). Scaling Subgroup Meeting. European Aquatic Modelling Network COST626. 11-14 December 2002. Book of abstracts. Ghent University, Ghent, Belgium. 21 p.

3.1.8 Research reports

Goethals, P., Devliegher, W. & Verstraete, W. (1997). Biological monitoring and ecological risk assessment of the bioremediation project BP-Gent (in Dutch). Ghent University, Laboratory for Microbial Ecology and Technology, Ghent, Belgium. 30 p.

Goethals, P., De Boever, P., Nollet, L. & Verstraete, W. (1997). Risk assessment on PAH bioavailability in soils introduced in the humane digestive system (in Dutch). Ghent University, Laboratory for Microbial Ecology and Technology, Ghent, Belgium. 17 p.

Goethals, P., Devliegher, W. & Verstraete, W. (1997). Research on the use of compost in soil bioremediation (in Dutch). Ghent University, Laboratory for Microbial Ecology and Technology, Ghent, Belgium. 63 p.

De Pauw, N., Bauwens, W. & Goethals, P. (1999). Canal Tan Hoa - Lo Gom sanitation and urban upgrading project. Expert mission on assessment of wastewater, sediment treatment, embankment and industrial pre-treatment action plans. University of Ghent, Laboratory of Environmental Toxicology and Aquatic Ecology; Free University of Brussels, Department of Hydrology, Ghent, Belgium. 43 p.

Heylen, S., Goethals, P.L.M. & De Pauw, N. (1999). Audit on the water quality determination of the major watercourses with the Belgian Biotic Index and proposal for an improved measurement strategy (in Dutch). Group for Applied Ecology and Ghent University, Department of Applied Ecology and Environmental biology, Laboratory for Environmental Toxicology and Aquatic Ecology, Ghent, Belgium. 58 p.

Goethals, P.L.M., Rousseau, D. & De Pauw, N. (2000). Canal Tan Hoa - Lo Gom sanitation and urban upgrading project. Expert mission on assessment of wastewater, sediment treatment, embankment and industrial pre-treatment action plans. University of Ghent, Department of Applied Ecology and Environmental Biology, Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent, Belgium. 55 p.

Vasel, J.-L., De Pauw, N., Jupsin, H., Rousseau, D. & Goethals, P. (2000). Canal Tan Hoa – Lo Gom sanitation and urban upgrading project. Pre-feasibility study of a waste stabilisation pond system for the treatment of polluted canal water. Final report. Fondation Universitaire Luxembourgeoise, Département Eau et Environnement & Ghent University, Department of Applied Ecology and Environmental Biology. 18 p.

Gabriels, W., Goethals, P. & De Pauw, N. (2001). Hydrobiological research of Bergelenput at Wevelgem. Study by order of the Provincial Commission for Fisheries of Western Flanders (in Dutch). Ghent University, Laboratory for Environmental Toxicology and Aquatic Ecology, Research Unit Aquatic Ecology, Ghent, Belgium. 30 p.

Detemmerman, L., Goethals, P. & De Pauw, N. (2001). Analysis of the sampling and processing procedures of the Biotic Sediment Index (in Dutch). Ghent University, Laboratory of Environmental Toxicology and Aquatic Ecology by order of the Flemish Environmental Agency, Gent, België. 34 p. + appendices.

Detemmerman, L., Goethals, P. & De Pauw, N. (2001). Comparative study of the Biotic Sediment Index and the Belgian Biotic Index (in Dutch). Ghent University, Department Applied Ecology and Environmental Biology, Research Unit Aquatic Ecology by order of the Flemish Environmental Agency, Ghent, Belgium. 83 p. + appendices.

Detemmerman, L., Heylen, S., Goethals, P. & De Pauw, N. (2001). Manual for the determination of mentum deformities in Chironomus-larvae for the assessment of water sediments in Flanders (in Dutch). Ghent University and Flemish Environment Agency, Ghent, Belgium. 30 p + appendices.

Breine, J.J., Goethals, P., Simoens, I., Ercken, D., Van Liefferinge, C., Verhaegen, G., Belpaire, C., De Pauw, N., Meire, P. & Ollevier, F. (2001). The Fish Index as a tool for measuring the biotic integrity of the Flemish inland waters. Final report of project VLINA 9901, study carried out by order of the Flemish Community within the framework of the Flemish Nature Impulse Program. D/2001/3241/261 (in Dutch). Institute for Forestry and Game Management, Groenendaal, Belgium. 174 p. + appendices.

De Pauw, N. & Goethals, P. (2001). Canal Tan Hoa - Lo Gom sanitation and urban upgrading project. Expert mission on assessment of wastewater, sediment treatment, embankment and industrial pre-treatment action plans: course on 'sustainable water management'. Ghent University, Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent, Belgium. 54 p.

Triest, L., D'heere, E., De Pauw, N., Goethals, P., Adriaenssens, V., Gabriels, W., Belpaire, C., Breine, J.J. & Simoens, I. (2001). Comparison of bio-indicators for the ecological evaluation of valuable upstream brook stretches. Final report of project VLINA 0008, study carried out by order of the Flemish Community within the framework of the Flemish Nature Impulse Program (in Dutch). Free University Brussels; Ghent University; Institute for Forestry and Game Management, Brussels, Belgium. 137 p.

Gabriels, W., Goethals, P.L.M., Adriaenssens, V., Heylen, S. & De Pauw, N. (2003). Development of a score system for the assessment of macroinvertebrates for the implementation of the Water Framework Directive in Flanders. Ghent University, Gent, Belgium. 72 p.

Goethals, P.L.M., Rousseau, D., Vasel, J.-L. & De Pauw, N. (2003). Audit PMU415 lagoon project: appreciation report for phase 2 of the detailed design study: comments on the final design and research possibilities. Ghent University, Gent, Belgium. 96 p.

3.1.9 Proceedings without peer-review

Goethals, P., Vanrolleghem, P.A., Bauwens, W. & De Pauw, N. (1999). Modelling, indices of biotic integrity and (meta-)databases as a bridge between monitoring and management of aquatic ecosystems: abstract for oral presentation (in Dutch). In: NecoV Wintermeeting 1999, Antwerp, Belgium, 8-9 December. p. 3-6.

Goethals, P., Vanrolleghem, P.A., Bauwens, W. & De Pauw, N. (1999). Modelling, indices of biotic integrity and (meta-)databases as a bridge between monitoring and management of aquatic ecosystems: abstract for oral presentation (in Dutch). In: NecoV Wintermeeting 1999, Antwerp, Belgium, 8-9 December. p. 3-6.

Goethals, P. & De Pauw, N. (2001). Predictive ecosystem models as basis for a more sustainable water management: abstract for oral presentation (in Dutch). In: NecoV Wintermeeting 2001, Antwerp, Belgium, 12-13 December. p. 18-21.

Bauwens, W. & Goethals, P.L.M. (2002). Reflections on the feasibility and implications of the EU Water Framework Directive. In: XXVII General Assembly of the European Geophysical Society, Nice, France, 21-26 April 2002.

Dedecker, A., Goethals, P., Gabriels, W. & De Pauw, N. (in press). Prediction of macroinvertebrate communities in the Zwalm catchment by means of artificial neural networks. In: Scheldt catchment (in Dutch). In: K. Buis, Ed. KVAB, Antwerp, Belgium.

D'heygere, T., Goethals, P. & De Pauw, N. (in press). Development of a monitoring strategy for river quality assessment of the river Dender by means of macroinvertebrates with respect to the European Water Framework Directive. In: Scheldt catchment (in Dutch). In: K. Buis, Ed. KVAB, Antwerp, Belgium.

Gabriels, W., Goethals, P., Heylen, S. & De Pauw, N. (in press). Analysis and prediction of macroinvertebrate communities in the Scheldt catchment. In: Scheldt catchment (in Dutch). In: K. Buis, Ed. KVAB, Antwerp, Belgium.

Raes, N., Goethals, P. & De Pauw, N. (in press). Prediction of macroinvertebrate communities in the Zwalm catchment by means of fuzzy logic models. In: Scheldt catchment (in Dutch). In: K. Buis, Ed. KVAB, Antwerp, Belgium.

3.2 Platform presentations

3.2.1 Platform presentations as presenter

Goethals, P. Genetic and eco-physiological aspects of aggregation in activated sludge. Tenth Anniversary of WEL, Brussels. Belgium, 19 June 1997.

Goethals, P., van Griensven, A., Bols, J., De Pauw, N., Vanrolleghem, P., Van Vooren, L. & Bauwens, W. Automated measurement stations and water quality modelling: abstract for oral presentation. Ninth European Congress on Biotechnology, Brussels, Belgium, 11-15 July 1999.

Goethals, P., Vanrolleghem, P.A., Bauwens, W. & De Pauw, N. Modelling, indices of biotic integrity and (meta-)databases as a bridge between monitoring and management of aquatic ecosystems. NecoV Wintermeeting 1999, Antwerp, Belgium, 8-9 December 1999.

Goethals, P.L.M., Vanrolleghem, P.A. & De Pauw, N. Water quality models and indices as tools for an integrated ecological assessment and management. Annual Scientific Meeting of the Freshwater Biological Association, Birmingham, UK, 13-15 September 2000.

Goethals, P., Dedecker, A., Gabriels, W. & De Pauw, N. Development and application of predictive river ecosystem models based on classification trees and artificial neural networks. International Conference on Applications of Machine Learning to Ecological Modelling, Adelaide, Australia, 27 November – 1 December 2000.

Goethals, P. & De Pauw, N. Development of educational case studies on integrated ecological assessment systems for nature conservation and restoration in river basins. Workshop on sustainable development in natural resources management. Moscow, Russia, 1-2 February 2001.

Goethals, P., Gabriels, W., Adriaenssens, V. & De Pauw, N. Monitoring, modelling, assessment and management of rivers in Flanders (Belgium). Workshop of the European Aquatic Modelling Network (COST626). Stuttgart, Germany, 8-9 March 2001.

Goethals, P. & De Pauw, N. Ecological informatics applied to decision support in river management: case studies for educational purposes. Workshop on ecological field training in the boreal forest zone. Monturova, Russia, 2-7 June 2001.

Goethals, P., Dzeroski, S., Dedecker, A., Raes, N., Adriaenssens, V., Gabriels, W. & De Pauw, N. Comparing classification trees, artificial neural networks and fuzzy logic models to predict macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium). Second Symposium for European Freshwater Sciences (SEFS2). Toulouse, France, 8-12 July 2001.

Goethals, P., Adriaenssens, V., Breine, J.J., Simoens, I., Ercken, D., Van Liefferinge, C., Belpaire, C. & De Pauw, N. Development and application of fish community models for biotic integrity assessment of the Scheldt Estuary (Belgium). Third European Ecological Modelling Conference, Dubrovnik, Croatia, 8-12 September 2001.

Goethals, P., Adriaenssens, V., Gabriels, W. & De Pauw, N. Applying computational techniques in decision support systems for integrated river basin management: case study of the Zwalm river basin (Flanders, Belgium). Annual Scientific Meeting of the Freshwater Biological Association, London, UK, 12-14 September 2001.

Goethals, P., Breine, J.J., Simoens, I., Ercken, D., Van Liefferinge, C., Belpaire, C. & De Pauw, N. Development and application of fish community models for biotic integrity assessment of lotic streams in Flanders (Belgium). Workshop on parameter selection in modelling aquatic community structure, Namur, Belgium, 15-16 September 2001.

Goethals, P., Dzeroski, S., Vanrolleghem, P. & De Pauw, N. Prediction of benthic macroinvertebrate taxa (Asellidae and Tubificidae) in watercourses of Flanders by means of classification trees. Second World Water Congress of the International Water Association, Berlin, Germany, 15-19 October 2001.

Goethals, P., D'heygere, T., Gabriels, W., Dedecker, A., Adriaenssens, V. & De Pauw, N. Ecological informatics applications in water management. B-IWA Happy Hour, Brussels, Belgium, 22 October 2001.

Goethals, P. & De Pauw, N. Predictive ecosystem models as basis for a more sustainable water management. NecoV Wintermeeting, Antwerp, Belgium, 12-13 December 2001.

Goethals, P. & De Pauw, N. Ecological models for decision support in river management: state-of-the-art. Workshop on cost-benefit analyses in water management. Neeltje-Jans, The Netherlands, 7 December 2001.

Goethals, P.L.M., François, D., Bouma, J.J., De Pauw, N. & De Clercq, M. Environmental Management Accounting applications of the WAECO decision support system in the Zwalm Catchment (Belgium). Fifth Annual Conference of the Environmental Management Accounting Network - Europe, Gloucestershire, UK, 11-12 February 2002.

Goethals, P., Adriaenssens, V., De Baets, B. & De Pauw, N. Development and assessment of fuzzy logic models predicting aquatic macroinvertebrate taxa in the Zwalm catchment. Third Conference of the International Society for Ecological Informatics, Rome, Italy, 26-30 August 2002.

Gabriels, W., Goethals, P.L.M.* & De Pauw, N. Input variables selection of artificial neural networks predicting aquatic macrobenthos communities in Flanders (Belgium). Third Conference of the International Society for Ecological Informatics, Rome, Italy, 26-30 August 2002. (*presenter)

Adriaenssens, V., Goethals, P.L.M.* & De Pauw, N. River quality assessment based on fuzzy logic. Third Conference of the International Society for Ecological Informatics, Rome, Italy, 26-30 August 2002. (*presenter)

Goethals, P., Gabriels, W. & De Pauw, N. Water quality assessment of rivers and lakes based on benthic macroinvertebrates in relation to the Water Framework Directive. Symposium on biological evaluation and monitoring of the quality of surface waters (SCOPE), Brussels, Belgium, 10 October 2002.

Goethals, P. & Niels De Pauw, N. Ecological informatics applications in water management: state-of-the-art. Ecological informatics applications in water management conference, Gent, Belgium, 6-7 November 2002.

Goethals ,P. Ecological informatics applications in water management: conclusions of the symposium and future prospects. Ecological informatics applications in water management conference, Gent, Belgium, 6-7 November 2002.

Goethals, P., Breine, J., Simoens, I. & Belpaire, C. Development of a multimetric fishindex for the implementation of the European Water Framework Directive in Flanders. Workshop of the Flemish Integrated Water Management Committee. Brussels, Belgium, 28 November 2002.

Goethals, P.L.M., Bouma, J.J., François, D., D'heygere, T., Dedecker, A., Adriaenssens, V. & De Pauw, N. Ecohydraulic modelling and assessment at multiple spatial scales: case studies in the Zwalm and Scheldt river basin. Scaling Subgroup Meeting of the European Aquatic Modelling Network COST626. Gent, Belgium, 11-14 December 2002.

Goethals, P.L.M., Bouma, J.J., François, D., D’heygere, T., Dedecker, A., Adriaenssens, V. & De Pauw, N. Coupling ecosystem valuation methods to the WAECO decision support system in the Zwalm catchment (Belgium). Integrative Modelling of Biophysical, Social and

Economic Systems for Resource Management Solutions Conference (MODSIM 2003), Townsville, Australia, 14-17 July 2003.

Goethals, P., D'heygere, T. & De Pauw, N. The effects of flow variability on macrobenthos communities in the Zwalm river basin. Workshop of the European Aquatic Modelling Network (COST626), Aix-en-Provence, France, 30-31 October 2003.

Goethals, P., Breine, J., Simoens, I. & Belpaire, C. Evaluation of multimetric fish community indices for the implementation of the European Water Framework Directive in Flanders. SCALDIT Workshop of the Fresh Surface Waters Subgroup (P09). Brussels, Belgium, 19 December 2003.

Goethals, P. Modelling for the European Water Framework Directive. Workshop of the European Aquatic Modelling Network (COST626), Salzburg, Austria, 24-26 March 2004.

Goethals, P., Breine, J., Simoens, I. & Belpaire, C. Description of reference communities of fish for the implementation of the European Water Framework Directive in Flanders: practical problems and pragmatic solutions. SCALDIT Workshop of the Fresh Surface Waters Subgroup (P09). Brussels, Belgium, 22 June 2004.

3.2.2 Platform presentations as co-author (presenter is underlined)

Goethals, P.L.M., Nicolletto, L., Bossier, P. & Verstraete, W. A model for the beneficial effects of Nutrifloc 50 S on the settlement of activated sludge. Tenth Forum for Applied Biotechnology, Gent, Belgium, September 1996.

Goethals, P.L.M., Nicolletto, L., Kersters, I., Bossier, P. & Verstraete, W. Effect of reactive oxygen intermediates on microbial cell and communities: a model for the beneficial effect of Nutrifloc 50 S on activated sludge. International Symposium Environmental Biotechnology, Oostende, Belgium, 1997.

Goethals, P.L.M. & De Pauw, N. Development of an integrated ecological assessment system for the implementation of the EU Water Framework Directive: case of the Zwalm river basin. Workshop Biological Monitoring, Verbania Pallanza, Italy, 4-5 September 2000.

De Pauw, N., Goethals, P. & Bauwens, W. Integrated ecological monitoring and assessment of the river Dender. Workshop on Ecological research in the Scheldt catchment, Brussels, Belgium, 29-30 March 2001.

Gabriels, W., Goethals, P., Heylen, S. & De Pauw, N. Analysis and prediction of benthic macroinvertebrate communities in the Scheldt catchment. Workshop on Ecological research in the Scheldt catchment, Brussels, Belgium, 29-30 March 2001.

Adriaenssens, V., Luypaert, P., Goethals, P. & De Pauw, N. (2001). Impact of land use on macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium). Second Symposium for European Freshwater Sciences (SEFS2). Toulouse, France, 8-12 July 2001.

Simoens, I., Van Liefferinge, C., Breine, J.J., Goethals, P., Ercken, D., Verhaegen, G., De Pauw, N., Ollevier, F., Meire, P. & Belpaire, C. The influence of seasonal changes in the fish populations on the index of biotic integrity in brooks of the Schelde- and Maas-basin. Second Symposium for European Freshwater Sciences (SEFS2). Toulouse, France, 8-12 July 2001.

Breine, J.J., Simoens, I., Goethals, P., Quataert, P., Ercken, D., Van Liefferinge, C. & Belpaire, C. Development of an Index of Biotic Integrity for lotic streams in Flanders (Belgium). Tenth Congress of Ichthyology ECI-X, Prague, Czech Republic, 3-7 September 2001.

Adriaenssens, V., Goethals, P. & De Pauw, N. Effect of input variable selection on the performance of macroinvertebrate taxa prediction in the Zwalm River basin by means of fuzzy logic. Workshop on parameter selection in modelling aquatic community structure, Namur, Belgium, 15-16 September 2001.

D'heygere, T., Goethals, P. & De Pauw, N. Application of evolutionary algorithms for input variable selection of classification tree models predicting benthic macroinvertebrate communities in watercourses of Flanders (Belgium). Workshop on parameter selection in modelling aquatic community structure, Namur, Belgium, 15-16 September 2001.

Gabriels, W., Goethals, P. & De Pauw, N. (2001). Calibration of models to predict macrobenthos communities representing unimpacted river sediments in Flanders, Belgium. Workshop on parameter selection in modelling aquatic community structure, Namur, Belgium, 15-16 September 2001.

Bauwens, W. & Goethals, P.L.M. Reflections on the feasibility and implications of the EU Water Framework Directive. XXVII General Assembly of the European Geophysical Society, Nice, France, 21-26 April 2002.

Dedecker, A., Goethals, P., Gabriels, W. & De Pauw, N. Artificial Neural Network (ANN) model design methodologies for applications in river basin management. International Conference on Integrated Assessment and Decision Support of the International Environmental Modelling and Software Society (iEMSs), Lugano, Switzerland, 24-27 June 2002.

D'heygere, T., Goethals, P. & De Pauw, N. Use of genetic algorithms to select input variables in neural network models for the prediction of benthic macroinvertebrates. International Conference on Integrated Assessment and Decision Support of the International Environmental Modelling and Software Society (iEMSs), Lugano, Switzerland, 24-27 June 2002.

Dedecker, A., Goethals, P.L.M. & De Pauw, N. Sensitivity and robustness of predictive neural network ecosystem models for simulations of 'extreme' management scenarios. Third Conference of the International Society for Ecological Informatics, Rome, Italy, 26-30 August 2002.

D'heygere, T., Goethals, P. & De Pauw, N. Optimisation of predictive decision tree and neural network ecosystem models with genetic algorithms. Third Conference of the International Society for Ecological Informatics, Rome, Italy, 26-30 August 2002.

Adriaenssens, V., Goethals, P., Dedecker, A., D'heygere, T. & De Pauw, N. Data collection in the Zwalm river basin for the prediction of the effect of restoration actions on macroinvertebrate communities. Ecological informatics applications in water management conference, Gent, Belgium, 6-7 November 2002.

D'heygere, T., Goethals, P. & De Pauw, N. Input variables selection in neural network ecosystem models: comparison of senso-nets and genetic algorithms. Ecological informatics applications in water management conference, Gent, Belgium, 6-7 November 2002.

Gabriels, W., Goethals, P. & De Pauw, N. Prediction of benthic macroinvertebrate abundance in Flemish watercourses using artificial neural networks and multiple regression. Ecological informatics applications in water management conference, Gent, Belgium, 6-7 November 2002.

Verslycke, T., Goethals, P., Vandenbergh, G., Callebaut, K. & Janssen, C.R. Application of rule induction techniques for detecting the possible impact of endocrine disruptors on the North Sea ecosystem. Colours of Ocean Data: International symposium on oceanographic data and information management with special attention on biological data. Brussels, Belgium, 25-27 November 2002.

De Pauw, N., Rousseau, D., Goethals, P. & Van Minh, P. Natural systems for wastewater treatment: training and research opportunities in Vietnam. Workshop on Lagoon Technologies in Southern Vietnam, Ho Chi Minh City, Vietnam, 4 December 2002.

D'heygere, T., Goethals, P.L.M., Dedecker, A. & De Pauw, N. (2003). Development of a monitoring network to model the habitat suitability of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). Integrative Modelling of Biophysical, Social and Economic Systems for Resource Management Solutions Conference (MODSIM 2003), Townsville, Australia, 14-17 July 2003.

Gabriels, W., Goethals, P. & De Pauw, N. Development of a multimetric macrobenthos index for rivers for the implementation of the European Water Framework Directive in Flanders. Workshop of the Flemish Integrated Water Management Committee. Brussels, Belgium, 28 November 2002.

Dedecker, A., Goethals, P. & De Pauw, N. Use of Artificial Neural Network (ANN) models and Geographical Information Systems (GIS) to simulate the migration of Gammarus pulex in the Zwalm river basin (Belgium). Symposium for European Freshwater Sciences (SEFS3). Edinburgh, UK, 13-18 July 2003.

Adriaenssens, V., Goethals, P., Charles, J. & De Pauw, N. Using knowledge and data in ecological modeling of freshwater biological communities. Symposium for European Freshwater Sciences (SEFS3). Edinburgh, UK, 13-18 July 2003.

Goethals, P.L.M., Adriaenssens, V., Breine, J., Simoens, I., Van Liefferinghe, C., Ercken, D., Maes, J., De Pauw, N. & Belpaire, C. Developing an index of biotic integrity to assess fish communities of the Scheldt estuary in Flanders (Belgium). Estuaries on the edge conference, Seattle, USA, 14-18 September 2003.

Gabriels, W., Goethals, P. & De Pauw, N. Development and evaluation of multimetric macrobenthos indices for rivers and lakes for the implementation of the European Water

Framework Directive in Flanders. SCALDIT Workshop of the Fresh Surface Waters Subgroup (P09). Brussels, Belgium, 19 December 2003.

3.3 Poster presentations

Goethals, P.L.M. & Verstraete, W. Bioavailability of organic pollutants in soil. Presented at:
- Third PhD Symposium FLTBW, Ghent University. Gent, Belgium, 1997.

Bols, J., Goethals, P.L.M., Meirlaen, J., van Griensven, A., Vandenberghe, V., Van Vooren, L., De Pauw, N., Vanrolleghem, P. & Bauwens, W. Automated measurement stations for river water quality monitoring. Presented at:
- Thirteenth Forum for Applied Biotechnology. Gent, Belgium, 22-23 September 1999.

Rousseau, D., Goethals, P., Meirlaen, J., Verboven, J., Vanrolleghem, P. & De Pauw, N. Monitoring and modelling of free-water-surface constructed wetlands. Presented at:
- Thirteenth Forum for Applied Biotechnology. Gent, Belgium, 22-23 September 1999.

van Griensven, A., Vandenberghe, V., Bols, J., De Pauw, N., Goethals, P., Meirlaen, J., Vanrolleghem, P.A., Van Vooren, L. & Bauwens, W. Experience and organisation of automated measuring stations for river water quality monitoring. Presented at:
- First World Congress of the International Water Association. Paris, France, 3-7 July 2000.
- B-IWA Happy Hour. Brussels, Belgium, 21 May 2001.

Gabriels, W., Goethals, P.L.M., Heylen, S., De Cooman, W. & De Pauw, N. Modelling benthic macro-invertebrate communities in Flanders using artificial neural networks. Presented at:
- Sixth PhD Symposium FLTBW, Ghent University. Gent, Belgium, 2000.
- International Symposium System Analysis and Computing in Water Quality Management (Watermatex 2000). Ghent, Belgium, 18-20 September 2000.
- XXVIII SIL Congress. Melbourne, Australia, 4-10 February 2001.
- NecoV Wintermeeting. Antwerp, Belgium, 12-13 December 2001.

Adriaenssens, V., Goethals, P. & De Pauw, N. Development of an integrated system for ecological river quality assessment. Presented at:
- Sixth PhD Symposium FLTBW, Ghent University. Gent, Belgium, 2000.

Adriaenssens, V., Goethals, P.L.M., Breine, J.J., Maes, J., Simoens, I., Ercken, D., Belpaire, C., Ollevier, F. & De Pauw, N. Development of an Index of Biotic Integrity for Estuaries in Flanders based on fish communities. Presented at:
- NecoV Wintermeeting. Wageningen, The Netherlands,13-14 December 2000.
- Workshop on ecological research in the Scheldt catchment. Brussels, Belgium, 29-30 March 2001.
- The Aquatic Biodiversity Symposium, Beveren, Belgium, 11-13 August 2003.
- Biomonitoring of the Environment. Gent, Belgium, 3 December 2003.

Breine, J.J., Simoens, I., Belpaire, C., D'heere, E., Triest, L., Goethals, P., Gabriels, W., Adriaenssens, V. & De Pauw, N. Comparison of bio-indicators for the ecological evaluation of valuable upstream brook trajects in the Scheldt catchment. Presented at:
- Workshop on ecological research in the Scheldt catchment. Brussels, Belgium, 29-30 March 2001.

Dedecker, A., Goethals, P., Gabriels, W. & De Pauw, N. Prediction of macro-invertebrate communities in the Zwalm catchment with the use of artificial neural networks. Presented at:
- Workshop on ecological research in the Scheldt catchment. Brussels, Belgium, 29-30 March 2001.
- Second Symposium for European Freshwater Sciences (SEFS2). Toulouse, France, 8-12 July 2001.

Goethals, P., D'heygere, T., van Griensven, A., Vanrolleghem, P., Bauwens, W. & De Pauw, N. On-line water quality monitoring with automated measurement stations on the river Dender. Presented at:
- Workshop on ecological research in the Scheldt catchment. Brussels, Belgium, 29-30 March 2001.
- Second Symposium for European Freshwater Sciences (SEFS2). Toulouse, France, 8-12 July 2001.
- Biomonitoring of the Environment. Gent, Belgium, 3 December 2003.

Goethals, P. & De Pauw, N. Development of an integrated system for the ecological management of the Zwalm catchment. Presented at:
- B-IWA Happy Hour. Brussels, Belgium, 19 February 2001.
- Workshop on ecological research in the Scheldt catchment. Brussels, Belgium, 29-30 March 2001.
- Second Symposium for European Freshwater Sciences (SEFS2). Toulouse, France, 8-12 July 2001.

Raes, N., Goethals, P., Adriaenssens, V., De Baets, B. & De Pauw, N. Prediction of macroinvertebrate communities in the Zwalm catchment based on fuzzy logic. Presented at:
- Workshop on ecological research in the Scheldt catchment. Brussels, Belgium, 29-30 March 2001.
- Second Symposium for European Freshwater Sciences (SEFS2). Toulouse, France, 8-12 July 2001.

Simoens, I., Van Liefferinge, C., Breine, J.J., Goethals, P., Ercken, D., Verhaegen, G., De Pauw, N., Ollevier, F., Meire, P. & Belpaire, C. Seasonal variations in the fish populations of some watercourses in the Scheldt catchment. Presented at:
- Workshop on ecological research in the Scheldt catchment. Brussels, Belgium, 29-30 March 2001.

Gabriels, W., Goethals, P. & De Pauw, N. Self Organising Maps for analysing and classifying benthic macroinvertebrate communities in Flanders. Presented at:
- Second Symposium for European Freshwater Sciences (SEFS2). Toulouse, France, 8-12 July 2001.

D'heygere, T., Goethals, P., van Griensven, A., Vandenberghe, V., Bauwens, W., Vanrolleghem, P. & De Pauw, N. Optimisation of the discrete conductivity and dissolved oxygen monitoring using continuous data series obtained with automated measurement stations. Presented at:
- Fifteenth Forum Applied Biotechnology. 24-25 September 2001.

Adriaenssens, V., Goethals, P. & De Pauw, N. Development of a fuzzy expert system for the prediction of macroinvertebrate taxa. Presented at:
- Seventh PhD Symposium FLTBW. Ghent, Belgium, 2001.

Dedecker, A., Goethals, P., Gabriels, W. & De Pauw, N. River management applications of ecosystem models predicting aquatic macroinvertebrate communities based on artificial neural networks (ANNs). Presented at:
- Seventh PhD Symposium FLTBW. Ghent, Belgium, 2001.

D'heygere, T., Goethals, P. & De Pauw, N. Application of genetic algorithms for input variables selection of decision tree models predicting mollusca in unnavigable Flemish watercourses. Presented at:
- Seventh PhD Symposium FLTBW. Ghent, Belgium, 2001.

Goethals, P., Gasparyan, K. & De Pauw, N. River restoration simulations by ecosystem models predicting aquatic macroinvertebrate communities based on J48 classification trees. Presented at:
- Seventh PhD Symposium FLTBW. Ghent, Belgium, 2001.

Raes, N., Goethals, P., Adriaenssens, V. & De Pauw, N. Predicting Gammaridae (Crustaceae, Isopoda) in the Zwalm river basin (Flanders, Belgium) by means of fuzzy logic models. Presented at:
- Seventh PhD Symposium FLTBW. Ghent, Belgium, 2001.

Adriaenssens, V., Goethals, P. & De Pauw, N. Development of a fuzzy expert system for the prediction of macroinvertebrate taxa. Presented at:
- NecoV Wintermeeting. Antwerp, Belgium, 12-13 December 2001.

Dedecker, A., Goethals, P., Gabriels, W. & De Pauw, N. (2001). Development and application of river ecosystem models based on Artificial Neural Networks (ANNs). Presented at:
- NecoV Wintermeeting. Antwerp, Belgium, 12-13 December 2001.

D'heygere, T., Goethals, P., van Griensven, A., Vandenberghe, V., Bauwens, W., Vanrolleghem, P. & De Pauw, N. (2001). Optimization of the discrete monitoring of conductivity and dissolved oxygen by means of continuous measurement series obtained with automated measurement stations. Presented at:
- NecoV Wintermeeting. Antwerp, Belgium, 12-13 December 2001.

Dakou, E., D’heygere, T., Goethals, P., De Pauw, N. & Lazaridou-Dimitriadou, M. Comparison of the performance of decision tree and ANN-models predicting benthic macroinvertebrates in the Axios River (Greece). Presented at:
- Ecological informatics applications in water management conference, Gent, Belgium, 6-7 November 2002.

D'heygere, T., Adriaenssens, V., Dedecker, A., Gabriels, W., Goethals, P. & De Pauw, N. Development of a Decision Support System for integrated water management in the Zwalm river basin, Belgium. Presented at:
- B-IWA Happy Hour. Brussels, Belgium, 10 June 2002.
- Eighth PhD Symposium FLTBW. Ghent, Belgium, 2002.
- Ecological informatics applications in water management conference, Gent, Belgium, 6-7 November 2002.
- Third Conference of the International Society for Ecological Informatics, Rome, Italy, 26-30 August 2002.
- Biomonitoring of the Environment. Gent, Belgium, 3 December 2003.

Gabriels, W., Adriaenssens, V., Goethals, P.L.M. & De Pauw, N. Monitoring of macroinvertebrate communities for the ecological evaluation of valuable upstream brooks in Flanders, Belgium. Presented at:
- Eighth PhD Symposium FLTBW. Ghent, Belgium, 2002.

Adriaenssens, V., D'heygere, T., Dedecker, A., Goethals, P.L.M. & De Pauw, N. (2002). Relations between structural characteristics and macroinvertebrate communities in the Zwalm river basin at different spatial scales. Presented at:
- Eighth PhD Symposium FLTBW. Ghent, Belgium, 2002.
- Symposium on biological evaluation and monitoring of the quality of surface waters (SCOPE), Brussels, Belgium, 10 October 2002.

Adriaenssens, V., Simons, F., Goethals, P.L.M., Goddeeris, B. & De Pauw, N. Ecological analysis of the Chironomid communities in the Zwalm river basin. Presented at:
- Symposium on biological evaluation and monitoring of the quality of surface waters (SCOPE), Brussels, Belgium, 10 October 2002.

Dedecker, A., Goethals, P. & De Pauw, N. Use of Artificial Neural Network (ANN) models and Geographical Information Systems (GIS) to simulate the migration of Gammarus pulex in the Zwalm river basin (Flanders, Belgium). Presented at:
- Eighth PhD Symposium FLTBW. Ghent, Belgium, 2002.
- Ecological informatics applications in water management conference, Gent, Belgium, 6-7 November 2002.

Gabriels, W., Goethals, P. & De Pauw, N. Implications of taxonomic identification levels on the Belgian Biotic Index assessment method. Presented at:
- The Aquatic Biodiversity Symposium, Beveren, Belgium, 11-13 August 2003.

Charles, J., Goethals, P.L.M., Adriaenssens, V. & De Pauw, N. (2003). Bayesian belief networks for the prediction of macroinvertebrate taxa in the Zwalm river basin. Presented at:
- Ninth PhD Symposium FLTBW. Leuven, Belgium, 2003.

Gabriels, W., Goethals, P., Adriaenssens, V., Heylen, S. & De Pauw, N. Development of an assessment system for aquatic macroinvertebrates in Flanders
according to the European Water Framework Directive. Presented at:
- Biomonitoring of the Environment. Gent, Belgium, 3 December 2003.

Mouton, A., Depestele, J., D’heygere, T., Goethals, P. & De Pauw, N. Development of ecosystem models to predict the effects of river restoration projects at different spatial scales. Presented at:
- B-IWA Happy Hour. , Belgium, 27 October 2003.
- Biomonitoring of the Environment. Gent, Belgium, 3 December 2003.
- NecoV Wintermeeting. Gent, Belgium, 14-15 January 2004.

Depestele, J., Mouton, A., D’heygere, T., Goethals, P. & De Pauw, N. Development of Predictive Models for the Management of Fish Communities in the Zwalm River Basin, Belgium. Presented at:
- B-IWA Happy Hour. Brussels, Belgium, 22 March 2004.

Goethals, P., Dedecker, A, D’heygere, T., Adriaenssens, V., Gabriels, W., Depestele, J., Mouton, A., Dominguez, L., Zarkami, R., Ambelu, A. & De Pauw, N. Ecotechnological solutions for river management. Presented at:
- B-IWA Happy Hour. Brussels, Belgium, 7 June 2004.

3.4 Organization of scientific meetings and conferences

Workshop: 'Ecological assessment of surface waters in The Netherlands and Flanders', 8 October 1998. Dutch Aquatic Ecology Society, Antwerp, Belgium. Organizers: De Pauw, N. & Goethals, P.

Workshop: 'Natural systems for treatment of (waste)water in The Netherlands and Flanders', 21 October 1999. Dutch-Flemish Ecology Society, Antwerp, Belgium. Organizers: De Pauw, N. & Goethals, P.

Ecological informatics applications in water management conference, 6-7 November 2002. Dutch-Flemish Ecology Society (NecoV), Ghent, Belgium. Organizers: Goethals, P. & De Pauw, N.

Scaling Subgroup Meeting. European Aquatic Modelling Network COST626. 11-14 December 2002. Ghent, Belgium. Organizers: Goethals, P., Duel, H., Dunbar, M., Harby, A. & De Pauw, N.

Fifth International Symposium on Ecohydraulics: ‘Restoration of aquatic habitats’. 12-17 September 2004. Madrid, Spain. Chairman of the session B: Environmental Flows for Fluvial Maintenance and Conservation.

Tenth PhD Symposium FLTBW. Gent, Belgium, 29 September 2004. Member of the organizing committee; chairman of the session: Environmental Science and Technology.

Workshop: ‘Typologies and Reference Conditions in the Rhine, Meuse and Schelde Basins’, 30 November 2004, Luxembourg. Moderator of the round table discussion on ‘Typologies for the European Water Framework Directive’.

3.5 Startup and supervision of PhD studies

Adriaenssens Veronique (2000-2004): Knowledge-based macroinvertebrate habitat suitability models for use in ecological river management.

Van Minh Phan (since 1999): Wastewater-fed aquaculture in the South of Vietnam.

Gabriels Wim (since 2000): Analysis, modelling and assessment of macroinvertebrates in Flemish surface waters.

D'heygere Tom (since 2001): Application of evolutionary algorithms for the optimisation of predictive ecological models based on decision trees and neural networks.

Dedecker Andy (since 2001): Development of neural network models to predict macroinvertebrate communities for application in water management.

Zarkami Rahmat (since 2003): Monitoring, modeling, assessment and management of pike (*Esox Lucius*).

Dominguez Luis Elvin (since 2004): Development of monitoring and assessment methods based on macroinvertebrates to support decision making in river management in Ecuador.

Mouton Ans (since 2004): Linking and integrating ecological models in decision support systems for biological community management in rivers.

4 Teaching and educational activities

4.1 Practical exercises and other academic training at the Ghent University

1996-1997: Microbial Processes and Technologies for Environmental Sanitation (Prof. Dr. ir. W. Verstraete): practical exercises for Bio-engineers (Option Environmental Technology). (15h)

1997-1998: Microbial Ecological Processes (Prof. Dr. ir. W. Verstraete): practical exercises for Bio-engineers (Option Land and Forest Management) and Civil Engineers (selected as optional course). (15h)

1997-1998/1998-1999/1999-2000: Algoculture (Prof. Dr. N. De Pauw): practical exercises for MSc. Aquaculture. (15h)

1997-1998/1998-1999/1999-2000: Biological Monitoring and Water Quality Assessment (Prof. Dr. N. De Pauw): practical exercises for MSc. Environmental Sanitation. (15h)

1997-1998/1998-1999/1999-2000: Biological Monitoring and Water Quality Assessment (Prof. Dr. N. De Pauw): practical exercises for Bio-engineers (Option Environmental Technology). (15h)

1997-1998/1998-1999/1999-2000: Biological Monitoring and Water Quality Assessment (Prof. Dr. N. De Pauw): practical exercises for Master in Environmental Sanitation and Technology (GAS, Dutch program). (15h)

1997-1998/1998-1999/1999-2000: Aquatic Ecology (Prof. Dr. N. De Pauw): practical exercises for Bio-engineers (Option Environmental Technology). (15h)

1998-1999/1999-2000: Aquatic Ecology (Prof. Dr. N. De Pauw): practical exercises for MSc. Environmental Sanitation and MSc. Aquaculture. (15h)

2000-2001/2001-2002/2002-2003/2003-2004: Scientific information retrieval on the internet. Hands-on training for MSc. Environmental Sanitation. (3h)

2003-2004: Project for Bio-engineers (Option Environmental Technology). (one project of four students).

4.2 International training courses

Biological monitoring and assessment of rivers. 5-9 November 1999, Klitorea, Greece (BISEL-project). Co-ordination of the field training, laboratory analyses and data interpretation (20h). Program organizers: Lic. D. Vanderveken, Lic. W. Aerts and Prof. Dr. N. De Pauw. (40h)

Biological monitoring and assessment of rivers. 11-13 April 2000, Gent, Belgium (BISEL-project). Co-ordination of the field training, laboratory analyses and data interpretation (20h). Program organizers: Prof. Dr. N. De Pauw and ir. P. Goethals. (30h)

Sustainable Water Management, 3-10 October 2001, Ho Chi Minh City, Vietnam (BTC-project). Co-ordination of field excursions (6h), courses on ‘Advanced and industrial wastewater treatment’ (2h) and ‘Integrated water management’ (2h). Program organizers: Prof. Dr. N. De Pauw, ir. P. Goethals and ir. J. Van Lint. (28h)

Biological monitoring and assessment, 29 July – 6 August 2003, Guayaquil, Ecuador (VLIR-project). Courses on ‘Ecological Informatics’ (6h) and co-ordination of the field training, laboratory analyses and data interpretation (20h). Program organizers: Prof. Dr. N. De Pauw and ir. P. Goethals. (40h)

Advances in Wastewater treatment, 18-19 May 2004, Zagreb, Croatia (EU-Tempus-project). Courses on ‘Natural systems for wastewater treatment’ (1h), ‘The European Water Framework Directive and integrated water management’ (1h), ‘Ecological Informatics applications in water management’ (1h) and ‘Industrial wastewater treatment’ (1.5h). Program organizers: Dr. S. Novak and Prof. Dr. W. Verstraete. (9h)

Advances in Wastewater treatment, 25 June 2005, Skopje, Macedonia (EU-Tempus-project). Courses on ‘Natural systems for wastewater treatment’ (1h), ‘The European Water Framework Directive and integrated water management’ (0.5h). Program organizer: Prof. Dr. O. Cukaliev. (6h)

Up to date Water Treatment and Reuse, 23-27 August 2004, Gent, Belgium (VLIR-project): course on ‘Decision support systems for water management’ (0.75h). Program organizers: Prof. Dr. M. Van den Heede and Prof. Dr. ir. W. Verstraete. (40h)

4.3 Supervision of scripts and stages

Boucault Christophe (1996-1997). Stage. ‘Ecotoxicity of mineral oils and effects on plant production‘. Promotor: Prof. Dr. ir. W. Verstraete. Tutors: ir. P. Goethals & ir. A. De Sloovere.

Bekaert Inge (1997-1998). Stage. ‘Kwantificering van metabolische activiteiten van algen‘. Hogeschool Gent, Gegradueerde in Chemie Optie Milieuzorg. Promotor: Prof. Dr. N. De Pauw. Tutors: ir. P. Goethals & ir. C. Vlerick. 102 p. + appendices.

Bols Jan (1998-1999). Script. ‘On-line meetsystemen voor de opvolging van rivierwaterkwaliteit’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotor: Prof. Dr. N. De Pauw. Co-promotor: Prof. Dr. ir. P. Vanrolleghem. Tutor: ir. P. Goethals. 134 p. + appendices.

Rousseau Diederik (1998-1999). Script. ‘Monitoring en modellering van vloeirietvelden’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotor: Prof. Dr. N. De Pauw. Co-promotor: Prof. Dr. ir. P. Vanrolleghem. Tutors: ir. P. Goethals & ir. J. Meirlaen. 124 p. + appendices.

Vandevelde Dieter (1998-1999). Script. ‘Modellering van percollatierietvelden’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotor: Prof. Dr. N. De Pauw. Co-promotor: ir. P. Goethals. 105 p.

De Schauvre Els (1998-1999). Stage. ‘Monitoring en inventarisatie van aquatische ecosystemen in rivieren’. Hogeschool Gent, Gegradueerde in Chemie Optie Milieuzorg. Promotor: Prof. Dr. N. De Pauw. Tutors: ir. P. Goethals & ir. C. Vlerick. 72 p. + appendices.

Campens Astrid (1998-1999). Stage. ‘Monitoring van rivierkwaliteit met on-line meetsystemen’. Hogeschool Gent, Gegradueerde in Chemie Optie Milieuzorg. Promotor: Prof. Dr. N. De Pauw. Tutors: ir. P. Goethals & ir. C. Vlerick. 131 p. + appendices.

De Kuyper Ann (1998-1999). Stage. ‘Fysisch-chemische en biologische monitoring van zes vloeirietvelden te Wontergem’. Katholieke Hogeschool Sint-Lieven. Graduaat Chemie Optie Milieuzorg. Promotor: Prof. Dr. N. De Pauw. Tutors: ir. P. Goethals & G. Van der Maelen. 71 p. + appendices.

Gabriels Wim (1999-2000). Script. ‘Modelleren van macro-invertebratengemeenschappen in de Dender’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotor: Prof. Dr. N. De Pauw. Co-promotor: ir. P. Goethals. 73 p.

Luyckx Liesbeth (1999-2000). Script. ‘Ecosysteemmonitoring van de Dender’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotor: Prof. Dr. N. De Pauw. Co-promotor: ir. P. Goethals. 91 p. + appendices.

Wieme Ulrik (1999-2000). Script. ‘Optimalisatie van on-line meetstations voor de monitoring van oppervlaktewaterkwaliteit’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotor: Prof. Dr. N. De Pauw. Co-promotor: Prof. Dr. ir. P. Vanrolleghem. Tutor: ir. P. Goethals. 138 p. + appendices.

Adriaenssens Veronique (1999-2000). Script. ‘Ontwikkeling van een Visindex voor brakke wateren’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotor: Prof. Dr. N. De Pauw. Co-promotor: ir. P. Goethals. 106 p. + appendices.

Nguyen Ngoc Thi (1999-2000). Script. ‘Study of industrial wastewater discharges in the Binh Hung Hoa Canal (Ho Chi Minh City – Vietnam). Univertsiteit Gent, Master of Science in Environmental Sanitation. Promotor: Prof. Dr. N. De Pauw. Co-promotor: ir. P. Goethals. 81 p.

Vandergunst Iselinde (1999-2000). Stage. ‘Onderzoek van aquatische ecosystemen in het stroombekken van de Zwalm’. Katholieke Hogeschool Sint-Lieven. Graduaat Chemie Optie Milieuzorg. Promotor: Prof. Dr. N. De Pauw. Tutors: ir. P. Goethals & ir. M. Dujardin. 65 p. + appendices.

Huyghe Jeroen (1999-2000). Stage. ‘Ecosysteemverstoring van de waterlopen binnen het Zwalmbekken’. Hogeschool Gent, Gegradueerde in Chemie Optie Milieuzorg. Promotor: Prof. Dr. N. De Pauw. Tutors: ir. P. Goethals & ir. C. Vlerick. 98 p. + appendices.

Raes Nico (2000-2001). Script. ‘Ecosysteemmodellering en –monitoring van het Zwalmbekken met behulp van vaaglogica’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: Lic. V. Adriaenssens. 125 p. + appendices.

Dedecker Andy (2000-2001). Script. ‘Ecosysteemmonitoring en –modellering van de Zwalm aan de hand van artificiële neurale netwerken’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. W. Gabriels. 104 p. + appendices.

D’heygere Tom (2000-2001). Script. ‘Ecosysteemmonitoring van de Dender tussen Geraardsbergen en Denderleeuw’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. 117 p. + appendices.

Hermans Peter (2000-2001). Script. ‘Hydrobiologisch onderzoek en beheersanalyse van Bergelenput’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotor: Prof. Dr. N. De Pauw. Co-promotor: ir. P. Goethals. Tutor: ir. W. Gabriels. 65 p. + appendices.

Hendrickx Jan (2000-2001). Stage. ‘Canal Tan Hoa-Lo Gom sanitation and urban upgrading project’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: Ing. Y. Dervaux. 29 p. + appendices.

Michiels Ben (2000-2001). Script. ‘Ontwikkeling van Indexen voor Biotische Integriteit voor de beoordeling van visgemeenschappen in de vlagzalm-, forel- en brakwaterzone’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. 90 p. + appendices.

Hendrickx Jan (2000-2001). Script. ‘Uitwerking van beheersalternatieven voor het ecologisch herstel van het Zwalm stroombekken’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. 113 p. + appendices.

Verween Annick (2000-2001). Script. ‘Uitwerking van beheersalternatieven voor het herstel van de biologie van het Denderecosysteem’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotor: Prof. Dr. N. De Pauw. Co-promotor: ir. P. Goethals. 85 p. + appendices.

Billiaert Bruno (2000-2001). Script. ‘Uitwerken van beheersalternatieven voor het herstel van de fysisch-chemische en structurele verstoring van het Denderecosysteem’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotor: Prof. Dr. N. De Pauw. Co-promotor: ir. P. Goethals. 75 p. + appendices.

De Clercq Katrien (2001-2002). Script. ‘Relaties tussen structurele eigenschappen en macro-invertebratengemeenschappen in de waterlopen van het zwalmbekken’. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: Lic. V. Adriaenssens. 79 p. + appendices.

Houzet Hannelore en Decraene Evelien (2001-2002). Script. ‘Voorspelling van de habitatpreferenties van macro-invertebraten in de Zwalm met behulp van inductief logisch programmeren’. Katholieke Universiteit Leuven, Licentiaat Informatica. Promotor: Prof. Dr. ir. M. Bruynooghe. Co-promotor: ir. P. Goethals. Tutor: Prof. Dr. ir. H. Blockeel. 160 p. + appendices.

Abuaku Ebenezer (2001-2002). Script. ‘Groei van eendekroos op anaerobe digestor effluent’. Universiteit Gent & Vrije Universiteit Brussel, GGS Physical Land Resources. Promotors: Prof. Dr. ir. W. Verstraete & ir. P. Goethals. 46 p.

Simons Frank (2001-2002). Script. ‘Ecologische analyse van Chironomidengemeenschappen in het Zwalmbekken’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotor: Prof. Dr. N. De Pauw. Co-promotor: ir. P. Goethals. 113 p. + appendices.

Dakou Eleni (2001-2002). Script. ‘Development of ecological models for prediction of macroinverterbrates in Greek rivers’. Universiteit Gent, Erasmus student van de Aristotle University of Thessaloniki. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutors: ir. T. D’heygere & ir. W. Gabriels. 65 p.

Charles Joke (2002-2003). Script. ‘Ontwikkeling van een modeldatabank voor de voorspelling van macro-invertebratengemeenschappen in de waterlopen van het zwalmbekken’. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: Lic. V. Adriaenssens. 100 p. + appendices.

Van Haverbeke Emmanuel (2002-2003). Script. ‘Uitwerking van herstelmaatregelen voor snoek (*Esox Lucius*) in rivieren’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. T. D’heygere. 61 p. + appendices.

De Ridder Kathelijne (2002-2003). Script. ‘Indices voor de beoordeling macroinvertebratengemeenschappen van stromende wateren in Vlaanderen’. Universiteit

Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. N. De Pauw & ir. W. Gabriels. Tutor: ir. P. Goethals. 55 p. + appendices.

Stuer Veerle (2002-2003). Script. 'Indices voor de beoordeling macroinvertebratengemeenschappen van stilstaande wateren in Vlaanderen'. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. W. Gabriels. 41 p. + appendices.

Vandromme Jan (2002-2003). Script. 'Uitwerking van herstelmaatregelen voor snoek (*Esox Lucius*) in stilstaande wateren'. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. T. D'heygere. 74 p. + appendices.

Kakuli Viana (2002-2003). Script. 'Assessment and sustainable management of water systems in Sudan'. Univertsiteit Gent, Master of Science in Environmental Sanitation. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. 78 p.

Goessens Xenia (2002-2003). Script. 'Ecosysteemwaarderingstechnieken voor de evaluatie en selectie van beheersalternatieven inzake structuurherstel van de waterlopen in het Zwalmbekken'. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. J.J. Bouma & ir. P. Goethals. Tutor: ir. A. Dedecker & Lic. D. François. 76 p. + appendices.

Verstraete Ann (2002-2003). Script. 'Ecosysteemwaarderingstechnieken voor de evaluatie en selectie van beheersalternatieven inzake waterkwaliteitsherstel van de waterlopen in het Zwalmbekken'. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. J.J. Bouma & ir. P. Goethals. Tutor: ir. A. Dedecker & Lic. D. François. 79 p.

Jacquemin Djordi (2002-2003). Script. 'Ontwikkeling van indexen voor biotische integriteit voor de beoordeling van stilstaande wateren volgens de Kaderrichtlijn Water'. KULeuven, Licenciaat Biologie. Promotor: Prof. Dr. F. Ollevier. Tutor: ir. P. Goethals.

Depestele Jochen (2003-2004). Script. 'Ontwikkeling van een beslissingsondersteunend systeem voor het beheer van visgemeenschappen in het Zwalmbekken'. Universiteit Gent, Bio-ingenieur Optie Land- en bosbeheer. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. T. D'heygere. 123 p. + appendices.

Mouton Ans (2003-2004). Script. 'Gebruik van predictieve modellen voor het uitwerken en selecteren van herstelplannen voor macro-invertebratengemeenschappen in het Zwalmbekken'. Universiteit Gent, Bio-ingenieur Optie Milieutechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. T. D'heygere. 133 p. + appendices.

Van De Walle An (2003-2004). Script. 'Uitwerking en implementatie van een internationale methode voor het beoordelen van macro-invertebratengemeenschappen in Vlaanderen'. Universiteit Gent, Bio-ingenieur Optie Cel- en Genbiotechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. A. Dedecker. 110 p. + appendices.

Deltour Judith (2003-2004). Script. 'Uitwerken van herstelplannen voor de snoek in het Zwalmbekken'. Universiteit Gent, Bio-ingenieur Optie Land- en bosbeheer. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. T. D'heygere.

Van Melckebeke Koen (2003-2004). Script. ‘Ontwikkeling en gebruik van neurale netwerkmodellen en migratiemodellen voor het voorspellen van macro-inverterbratengemeenschappen in het Zwalmbekken‘. Universiteit Gent, Bio-ingenieur Optie Milieutechnologie. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. A. Dedecker.

Janssen Katrien (2003-2004). Stage. ‘Ecologische monitoring en beoordeling van rivieren’. Hogeschool Limburg. Promotor: Lic. B. Cornelis. Tutors: ir. A. Dedecker & ir. P. Goethals. 84 p. + appendices.

Adjei Augustina (2003-2004). Script. ‘Development of the Pentad methodology for river monitoring and assessment in Ghana’. Univertsiteit Gent, Master of Science in Environmental Sanitation. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals.

Van Hees Stijn (2003-2004). Script. ‘Analyse en beheer van snoek in het Demerbekken’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. T. D’heygere. 71 p. + appendices.

Meert Carolien (2003-2004). Script. ‘Analyse en beheer van snoek in het Netebekken’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. Tutor: ir. T. D’heygere. 72 p. + appendices.

Van De Velde Benjamin (2003-2004). Script. ‘Uitwerking van een digitaal databanksysteem voor analyse en beheer van macro-invertebraten in oppervlaktewateren’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. 58 p. + appendices.

Milotic Tanja (2003-2004). Script. ‘Analyse van de recrutering van vispopulaties voor de optimalisatie van Visindexen voor stromende wateren’. Universiteit Gent, GAS Milieuwetenschappen en –technologieën. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals. 115 p.

Janssen Katrien (2003-2004). Project (stage). ‘Onzekerheidsanalyse bij de ecologische monitoring van rivieren’. Hogeschool Limburg. Promotor: Lic. B. Cornelis.Tutors: ir. A. Dedecker & ir. P. Goethals. 72 p. + appendices.

Lukaw Yatta Samuel Laku (2003-2004). Script. ‘Development of a biological monitoring network for river assessment in Sudan’. Univertsiteit Gent, Master of Science in Environmental Sanitation. Promotors: Prof. Dr. N. De Pauw & ir. P. Goethals.

4.4 Member in the jury of scripts and stages

Universiteit Gent

Vandevelde Dieter (1998-1999); Gabriels Wim (1999-2000); Luyckx Liesbeth (1999-2000); Wieme Ulrik (1999-2000); Adriaenssens Veronique (1999-2000); Nguyen Ngoc Thi (1999-2000); Raes Nico (2000-2001); Dedecker Andy (2000-2001); D’heygere Tom (2000-2001); Hermans Peter (2000-2001); Hendrickx Jan (2000-2001); Michiels Ben (2000-2001); Verween Annick (2000-2001); Billiaert Bruno (2000-2001); De Clercq Katrien (2001-2002);

Abuaku Ebenezer (2001-2002); Simons Frank (2001-2002); Dakou Eleni (2001-2002); Charles Joke (2002-2003); Lisette Talan (2002-2003); Van Haverbeke Emmanuel (2002-2003); De Ridder Kathelijne (2002-2003); Stuer Veerle (2002-2003); Vandromme Jan (2002-2003); Kakuli Viana (2002-2003); K Nzila Charles (2002-2003); Goessens Xenia (2002-2003); Verstraete Ann (2002-2003); Depestele Jochen (2003-2004); Mouton Ans (2003-2004); Van De Walle An (2003-2004); Deltour Judith (2003-2004); Van Melckebeke Koen (2003-2004); Adjei Augustina (2003-2004); Van Hees Stijn (2003-2004); Meert Carolien (2003-2004); Van De Velde Benjamin (2003-2004); Milotic Tanja (2003-2004), Lukaw Yatta Samuel Laku (2003-2004).

Hogeschool Gent

Bekaert Inge (1997-1998); De Schauvre Els (1998-1999); Campens Astrid (1998-1999); Huyghe Jeroen (1999-2000).

KULeuven

Houzet Hannelore en Decraene Evelien (2001-2002).

Hogeschool Limburg

Janssen Katrien (2003-2004); Dreezen Caroline (2003-2004); Spapen Bianca (2003-2004).

4.5 Member in PhD-juries

Adriaenssens Veronique (Ghent University, 15 October 2004).

5 Social, scientific and academic support activities

5.1 Water management and policy support as scientific expert in scientific steering committees of projects

2002-2003: Flemish Environment Agency (VMM). Development of ecological indices for estuaries to implement the European Water Framework Directive in Flanders.

2000-present: European Aquatic Modelling Network (COST 626). National representative in the Management Committee.

5.2 Review activities as scientific expert

Editorial boards
European Water Management

Editor of special issues of international journals
Aquatic Ecology: 1 issue

Reviewed papers of International journals
Water Science and Technology: 1 paper
Aquatic Ecology: 4 papers
Ecological Modelling: 2 papers
European Water Management: 4 papers
Hydrobiologia: 1 paper
Journal of Biogeography: 1 paper
Oecologia: 1 paper

Review of book chapters
Scardi, M. (Ed., 2004). Modelling community structure in freshwater ecosystems. Springer Verlag, Heidelberg, Germany: 5 chapters
Environment and Nature Report of Flanders (Mira-T) 2001. Garant, Leuven, Belgium: 1 chapter
Environment and Nature Report of Flanders (Mira-T) 2002. Garant, Leuven, Belgium: 2 chapters
Environment and Nature Report of Flanders (Mira-T) 2003 Garant, Leuven, Belgium: 3 chapters

Conference papers (peer-reviewed proceedings):
International Symposium System Analysis and Computing in Water Quality Management (Watermatex 2000). Ghent, Belgium, 18-20 September 2000: 1 paper
The Second International Nitrogen Conference (N2001). Potomac, Maryland, USA, 14-18 October 2001: 1 paper
Second World Water Conference of the International Water Association (IWA). Berlin, Germany, 15-19 October 2001: 14 papers
International Conference on Politics and Information Systems: Technologies and Applications (PISTA ’04). Orlando, Florida, USA, 21-24 July 2004: 10 papers

Reviewed project proposals:
Belgian Technical Co-operation (BTC): 7 proposals
Witec: 1 proposal

5.3 Committees in the Ghent University

1998-present: Management Board of the Department of Applied Ecology and Environmental Biology (Chair: Prof. Dr. R. Lemeur), Faculty Applied Biological Sciences.
2002: VAO Postgraduate Scholarships Committee (Chair: Prof. Dr. ir. O. Van Cleemput), Faculty Applied Biological Sciences.
2002-2003: Bachelor-Masters Programs Working Group (Chair: Prof. Dr. ir. E. Vandamme), Faculty Applied Biological Sciences.
2002-2003: Program Board of the Academic Teachers’ Training Program (Chair: Prof. Dr. A. Aelterman), Faculty of Psychology and Educational Sciences.
2002-2003: Program Reforming Workgroup of the Academic Teachers’ Training Program (Chair: Prof. Dr. A. Aelterman), Faculty of Psychology and Educational Sciences.
2003: Lecturer in Soil Management Selection Committee (Chair: Prof. Dr. ir. H. Van Langenhove), Faculty Applied Biological Sciences.

2003-present: Faculty Policy Plan Committee (Chair: Prof. Dr. ir. H. Van Langenhove), Faculty Applied Biological Sciences.
2003-present: Organizing Committee of the PhD Symposium (Chair: Prof. Dr. ir. J. Dewulf). Faculty Applied Biological Sciences.
2004: Lecturer in Hydrology Selection Committee (Chair: Prof. Dr. ir. H. Van Langenhove), Faculty Applied Biological Sciences.
2004: Associate Professors Evaluation Committee (Chair: Prof. Dr. ir. H. Van Langenhove), Faculty Applied Biological Sciences.

5.4 Memberships and activities in scientific organizations

1999-present: Vlerick Alumni
1996-present: Alumni RUG
1996-present: Koninklijke Vlaamse Ingenieursvereniging (KVIV)
1996-present: Verbond FLTBW
1998-present: Nederlands-Vlaamse Vereniging voor Ecologie (NecoV): chairman of the Ecological Informatics Workgroup
1999-present: International Water Association (IWA)
1999-present: Freshwater Biological Association (FBA)
2000-present: International Society of Ecological Informatics (ISEI): founding member
2000-present: Natuurpunt Vlaanderen
2001-present: Koninklijk Natuurwetenschappelijk Genootschap Dodonaea
2002-present: De Akademische Club (FLTBW-UGent)
2002-present: International Society for Ecological Engineering (ISEE)
2003-present: American Association for the Advancement of Science (AAAS)

6 Scientific awards and scholarships

6.1 Scientific awards

Water-Energy-Environment Society (WEL), ‘Environment’ student award 1996, on the basis of the script ‘Genetic and eco-physiological aspects of aggregation in activated sludge’, **P. Goethals**, Promoter: Prof. Dr. ir. W. Verstraete.

Dutch-Flemish Society for Ecology (NecoV), poster award, Wintermeeting 13-14 December 2000, Wageningen, The Netherlands, on the basis of the poster ‘Development of an estuarine fish index of biotic integrity for Flanders’, V. Adriaenssens, **P. Goethals**, J. Breine, J. Maes, C. Belpaire, F. Ollevier & N. De Pauw.

International Water Association – Belgium (B-IWA), poster award, B-IWA Happy Hour, 19 February 2001, Brussels, Belgium, on the basis of the poster ‘Development of an integrated ecological river management system: case study of the Zwalm river basin’, **P. Goethals** & N. De Pauw.

International Water Association – Belgium (B-IWA), poster award, B-IWA Happy Hour, 21 May 2001, Brussels, Belgium, on the basis of the poster ‘Experience and organization of automated measurement stations for river quality monitoring’, A. van Griensven, V.

Vandenberghe, J. Bols, N. De Pauw, **P. Goethals**, J. Meirlaen, P.Vanrolleghem, L. Van Vooren & W. Bauwens.

International Water Association – Belgium (B-IWA), poster award B-IWA Happy Hour, 10 June 2002, Brussels, Belgium, on the basis of the poster ‘Development of a Decision Support System for integrated water management in the Zwalm river basin, Belgium’, T. D’heygere, V. Adriaenssens, A. Dedecker, W. Gabriels, **P.L.M. Goethals** & N. De Pauw.

Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ), Student award for an excellent student paper and presentation at the ‘Integrative Modelling of Biophysical, Social and Economic Systems for Resource Management Solutions (MODSIM 2003)’, 14-17 July 2003, Townsville, Australia, on the basis of the paper and platform presentation ‘Coupling ecosystem valuation methods to the WAECO decision support system in the Zwalm catchment (Belgium)’, **P.L.M. Goethals**, J.J. Bouma, D. François, T. D’heygere, A. Dedecker, V. Adriaenssens & N. De Pauw.

Dutch-Flemish Society for Ecology (NecoV), poster award, Wintermeeting 14-15 January 2004, Gent, Belgium, on the basis of the poster ‘Development of ecosystem models to predict the effects of river restoration projects at different spatial scales’, A. Mouton, J. Depestele, T. D’heygere, **P.L.M. Goethals** & N. De Pauw.

6.2 Travelling scholarships

European Commission, European School on Intelligent Data Analysis, travelling and training scholarship. It allowed following the training: ‘Intelligent data analysis’, 26-30 March 2001, University of Calcolo, Palermo, Italy.

European Commission, Environmental Management Accounting Network – Europe, travelling scholarship. It allowed taking part in the conference ‘Environmental management accounting and government policy’, 11-12 February 2002, Cheltenham, United Kingdom.

Fund for Scientific Research – Flanders (FWO-Flanders), travelling scholarship. It allowed taking part in ‘The 3rd Conference of the International Society for Ecological Informatics (ISEI)’, 26-30 August 2002, Grottaferrata (Rome), Italy.

Appendices

Appendix 1: Classification tree *Gammarus*, Sediments Flanders, Subset 1, PCF=0.5

```
Clay <= 11
|  Conductivity <= 730.000019
|  |  Width <= 9
|  |  |  Day <= 23: 1 (5.0)
|  |  |  Day > 23
|  |  |  |  Flowvelocity = 0: 0 (2.0/1.0)
|  |  |  |  Flowvelocity = 1
|  |  |  |  |  Sand <= 96: 0 (11.0)
|  |  |  |  |  Sand > 96: 1 (2.0)
|  |  |  |  Flowvelocity = 2
|  |  |  |  |  Day <= 335
|  |  |  |  |  |  OM <= 2.7: 0 (41.0/3.0)
|  |  |  |  |  |  OM > 2.7
|  |  |  |  |  |  |  OM <= 7.2
|  |  |  |  |  |  |  |  T <= 11.8: 1 (5.0)
|  |  |  |  |  |  |  |  T > 11.8: 0 (6.0/1.0)
|  |  |  |  |  |  |  OM > 7.2: 0 (3.0)
|  |  |  |  |  Day > 335: 1 (3.0)
|  |  |  |  Flowvelocity = 3
|  |  |  |  |  Loam <= 1: 0 (5.0)
|  |  |  |  |  Loam > 1
|  |  |  |  |  |  T <= 11.8: 0 (5.0/1.0)
|  |  |  |  |  |  T > 11.8: 1 (4.0)
|  |  |  |  Flowvelocity = 4: 1 (3.0/1.0)
|  |  Width > 9: 1 (5.0)
|  Conductivity > 730.000019: 0 (54.0/2.0)
Clay > 11: 0 (74.0)

Number of Leaves  :   16
Size of the tree :        28

=== Evaluation on test set ===

Correctly Classified Instances          93               81.5789 %
Incorrectly Classified Instances        21               18.4211 %
Kappa statistic                          0.2192
Mean absolute error                      0.2149
Root mean squared error                  0.421
Relative absolute error                 82.0239 %
Root relative squared error            115.4308 %

=== Confusion Matrix ===

  a  b   <-- classified as
 88  8 |  a = 0
 13  5 |  b = 1
```

Appendix 2: Classification tree *Gammarus*, Sediments Flanders, Subset 1, PCF=0.25

```
Clay <= 11
|  Conductivity <= 730.000019
|  |  Width <= 9
|  |  |  Day <= 23: 1 (5.0)
|  |  |  Day > 23
|  |  |  |  Flowvelocity = 0: 0 (2.0/1.0)
|  |  |  |  Flowvelocity = 1
|  |  |  |  |  Sand <= 96: 0 (11.0)
|  |  |  |  |  Sand > 96: 1 (2.0)
|  |  |  |  Flowvelocity = 2
|  |  |  |  |  Day <= 335
|  |  |  |  |  |  OM <= 2.7: 0 (41.0/3.0)
|  |  |  |  |  |  OM > 2.7
|  |  |  |  |  |  |  OM <= 7.2
|  |  |  |  |  |  |  |  T <= 11.8: 1 (5.0)
|  |  |  |  |  |  |  |  T > 11.8: 0 (6.0/1.0)
|  |  |  |  |  |  |  OM > 7.2: 0 (3.0)
|  |  |  |  |  Day > 335: 1 (3.0)
|  |  |  |  Flowvelocity = 3
|  |  |  |  |  Loam <= 1: 0 (5.0)
|  |  |  |  |  Loam > 1
|  |  |  |  |  |  T <= 11.8: 0 (5.0/1.0)
|  |  |  |  |  |  T > 11.8: 1 (4.0)
|  |  |  |  Flowvelocity = 4: 1 (3.0/1.0)
|  |  Width > 9: 1 (5.0)
|  Conductivity > 730.000019: 0 (54.0/2.0)
Clay > 11: 0 (74.0)

Number of Leaves  :   16
Size of the tree :        28

=== Evaluation on test set ===

Correctly Classified Instances          93               81.5789 %
Incorrectly Classified Instances        21               18.4211 %
Kappa statistic                          0.2192
Mean absolute error                      0.2149
Root mean squared error                  0.421
Relative absolute error                 82.0239 %
Root relative squared error            115.4308 %

=== Confusion Matrix ===

  a  b   <-- classified as
 88  8 |  a = 0
 13  5 |  b = 1
```

Appendix 3: Classification tree *Gammarus*, Sediments Flanders, Subset 1, PCF=0.1

```
Clay <= 11
|  Conductivity <= 730.000019
|  |  Width <= 9
|  |  |  Day <= 23: 1 (5.0)
|  |  |  Day > 23
|  |  |  |  Day <= 335: 0 (85.0/18.0)
|  |  |  |  Day > 335: 1 (5.0/1.0)
|  |  Width > 9: 1 (5.0)
|  Conductivity > 730.000019: 0 (54.0/2.0)
Clay > 11: 0 (74.0)

Number of Leaves  :   6
Size of the tree :      11

=== Evaluation on test set ===

Correctly Classified Instances          95               83.3333 %
Incorrectly Classified Instances        19               16.6667 %
Kappa statistic                          0.2135
Mean absolute error                      0.2112
Root mean squared error                  0.3631
Relative absolute error                 80.5959 %
Root relative squared error             99.575  %

=== Confusion Matrix ===

  a  b   <-- classified as
 91  5 |  a = 0
 14  4 |  b = 1
```

Appendix 4: Classification tree *Gammarus*, Sediments Flanders, Subset 1, PCF=0.01

```
Clay <= 11
|   Width <= 9
|   |   Day <= 23: 1 (5.0)
|   |   Day > 23: 0 (143.0/24.0)
|   Width > 9: 1 (6.0/1.0)
Clay > 11: 0 (74.0)

Number of Leaves  :   4
Size of the tree :        7

=== Evaluation on test set ===

Correctly Classified Instances          98               85.9649 %
Incorrectly Classified Instances        16               14.0351 %
Kappa statistic                          0.3184
Mean absolute error                      0.2066
Root mean squared error                  0.3351
Relative absolute error                 78.8588 %
Root relative squared error             91.8957 %

=== Confusion Matrix ===

  a  b   <-- classified as
 93  3 |  a = 0
 13  5 |  b = 1
```

Appendix 5: Classification tree *Gammarus*, Sediments Flanders, Subset 2, PCF=0.5

```
Pb <= 10
|  Day <= 42: 1 (6.0)
|  Day > 42
|  |  Depth <= 0.55
|  |  |  TOXT = 0
|  |  |  |  Pb <= 0
|  |  |  |  |  Totalphosphorus <= 734
|  |  |  |  |  |  DO <= 6.4: 1 (7.0/1.0)
|  |  |  |  |  |  DO > 6.4: 0 (9.0/1.0)
|  |  |  |  |  Totalphosphorus > 734: 0 (9.0)
|  |  |  |  Pb > 0: 0 (4.0)
|  |  |  TOXT = 1: 0 (4.0)
|  |  Depth > 0.55
|  |  |  Pb <= 0: 1 (10.0/2.0)
|  |  |  Pb > 0: 0 (3.0/1.0)
Pb > 10
|  Clay <= 11
|  |  As <= 4.2: 0 (45.0)
|  |  As > 4.2
|  |  |  Flowvelocity = 0: 0 (1.0)
|  |  |  Flowvelocity = 1: 0 (17.0/1.0)
|  |  |  Flowvelocity = 2
|  |  |  |  T <= 4.6: 1 (3.0)
|  |  |  |  T > 4.6
|  |  |  |  |  OM <= 0.8: 1 (2.0)
|  |  |  |  |  OM > 0.8: 0 (31.0/3.0)
|  |  |  Flowvelocity = 3
|  |  |  |  Pb <= 31: 0 (6.0)
|  |  |  |  Pb > 31: 1 (3.0)
|  |  |  Flowvelocity = 4: 0 (2.0/1.0)
|  Clay > 11: 0 (66.0)

Number of Leaves  :   18
Size of the tree :        32

=== Evaluation on test set ===

Correctly Classified Instances          86               75.4386 %
Incorrectly Classified Instances        28               24.5614 %
Kappa statistic                          0.1551
Mean absolute error                      0.2483
Root mean squared error                  0.4576
Relative absolute error                 95.8982 %
Root relative squared error            128.4415 %

=== Confusion Matrix ===

  a  b   <-- classified as
 80 17 |  a = 0
 11  6 |  b = 1
```

Appendix 6: Classification tree *Gammarus*, Sediments Flanders, Subset 2, PCF=0.25

```
Pb <= 10
|   Day <= 42: 1 (6.0)
|   Day > 42
|   |   Depth <= 0.55
|   |   |   TOXT = 0
|   |   |   |   Totalphosphorus <= 734
|   |   |   |   |   DO <= 6.4: 1 (8.0/2.0)
|   |   |   |   |   DO > 6.4: 0 (9.0/1.0)
|   |   |   |   Totalphosphorus > 734: 0 (12.0)
|   |   |   TOXT = 1: 0 (4.0)
|   |   Depth > 0.55
|   |   |   Pb <= 0: 1 (10.0/2.0)
|   |   |   Pb > 0: 0 (3.0/1.0)
Pb > 10: 0 (176.0/13.0)

Number of Leaves  :   8
Size of the tree :    15

=== Evaluation on test set ===

Correctly Classified Instances          87               76.3158 %
Incorrectly Classified Instances        27               23.6842 %
Kappa statistic                          0.13
Mean absolute error                      0.2673
Root mean squared error                  0.4371
Relative absolute error                103.2235 %
Root relative squared error            122.6701 %

=== Confusion Matrix ===

  a  b   <-- classified as
 82 15 |  a = 0
 12  5 |  b = 1
```

Appendix 7: Classification tree *Gammarus*, Sediments Flanders, Subset 2, PCF=0.1

```
Pb <= 10
|   Day <= 42: 1 (6.0)
|   Day > 42
|   |   Depth <= 0.55: 0 (33.0/7.0)
|   |   Depth > 0.55: 1 (13.0/4.0)
Pb > 10: 0 (176.0/13.0)

Number of Leaves  :   4
Size of the tree :        7

=== Evaluation on test set ===

Correctly Classified Instances          94               82.4561 %
Incorrectly Classified Instances        20               17.5439 %
Kappa statistic                          0.1909
Mean absolute error                      0.2426
Root mean squared error                  0.3948
Relative absolute error                 93.6743 %
Root relative squared error            110.8017 %

=== Confusion Matrix ===

  a  b   <-- classified as
 90  7 |  a = 0
 13  4 |  b = 1
```

Appendix 8: Classification tree *Gammarus*, Sediments Flanders, Subset 2, PCF=0.01

```
Pb <= 10
|   Day <= 42: 1 (6.0)
|   Day > 42
|   |   Depth <= 0.55: 0 (33.0/7.0)
|   |   Depth > 0.55: 1 (13.0/4.0)
Pb > 10: 0 (176.0/13.0)

Number of Leaves  :   4
Size of the tree :        7

=== Evaluation on test set ===

Correctly Classified Instances          94               82.4561 %
Incorrectly Classified Instances        20               17.5439 %
Kappa statistic                          0.1909
Mean absolute error                      0.2426
Root mean squared error                  0.3948
Relative absolute error                 93.6743 %
Root relative squared error            110.8017 %

=== Confusion Matrix ===

  a  b   <-- classified as
 90  7 |  a = 0
 13  4 |  b = 1
```

Appendix 9: Classification tree *Gammarus*, Sediments Flanders, Subset 3, PCF=0.5

```
Clay <= 11
|  DO <= 6
|  |  Loam <= 0
|  |  |  Ni <= 0: 0 (3.0)
|  |  |  Ni > 0: 1 (4.0/1.0)
|  |  Loam > 0: 0 (75.0/5.0)
|  DO > 6
|  |  Width <= 0.75: 1 (4.0)
|  |  Width > 0.75
|  |  |  Day <= 135: 0 (29.0/2.0)
|  |  |  Day > 135
|  |  |  |  Flowvelocity = 0: 0 (2.0/1.0)
|  |  |  |  Flowvelocity = 1: 0 (6.0/1.0)
|  |  |  |  Flowvelocity = 2
|  |  |  |  |  Pb <= 0: 1 (9.0/1.0)
|  |  |  |  |  Pb > 0
|  |  |  |  |  |  T <= 5.1: 1 (3.0)
|  |  |  |  |  |  T > 5.1: 0 (10.0)
|  |  |  |  Flowvelocity = 3
|  |  |  |  |  pH <= 7.13: 0 (3.0)
|  |  |  |  |  pH > 7.13: 1 (5.0/1.0)
|  |  |  |  Flowvelocity = 4: 1 (4.0)
Clay > 11: 0 (71.0)

Number of Leaves  :   14
Size of the tree :    24

=== Evaluation on test set ===

Correctly Classified Instances        102               89.4737 %
Incorrectly Classified Instances       12               10.5263 %
Kappa statistic                         0.6219
Mean absolute error                     0.1306
Root mean squared error                 0.2931
Relative absolute error                50.4408 %
Root relative squared error            82.2562 %

=== Confusion Matrix ===

  a  b   <-- classified as
 89  8 |  a = 0
  4 13 |  b = 1
```

Appendix 10: Classification tree *Gammarus*, Sediments Flanders, Subset 3, PCF=0.25

```
Clay <= 11
|   DO <= 6: 0 (82.0/8.0)
|   DO > 6
|   |   Width <= 0.75: 1 (4.0)
|   |   Width > 0.75
|   |   |   Day <= 135: 0 (29.0/2.0)
|   |   |   Day > 135
|   |   |   |   Flowvelocity = 0: 0 (2.0/1.0)
|   |   |   |   Flowvelocity = 1: 0 (6.0/1.0)
|   |   |   |   Flowvelocity = 2
|   |   |   |   |   Pb <= 0: 1 (9.0/1.0)
|   |   |   |   |   Pb > 0
|   |   |   |   |   |   T <= 5.1: 1 (3.0)
|   |   |   |   |   |   T > 5.1: 0 (10.0)
|   |   |   |   Flowvelocity = 3
|   |   |   |   |   pH <= 7.13: 0 (3.0)
|   |   |   |   |   pH > 7.13: 1 (5.0/1.0)
|   |   |   |   Flowvelocity = 4: 1 (4.0)
Clay > 11: 0 (71.0)

Number of Leaves  :   12
Size of the tree :    20

=== Evaluation on test set ===

Correctly Classified Instances         103               90.3509 %
Incorrectly Classified Instances        11                9.6491 %
Kappa statistic                          0.6288
Mean absolute error                      0.1342
Root mean squared error                  0.2879
Relative absolute error                 51.8054 %
Root relative squared error             80.8061 %

=== Confusion Matrix ===

  a  b   <-- classified as
 91  6 |  a = 0
  5 12 |  b = 1
```

Appendix 11: Classification tree *Gammarus*, Sediments Flanders, Subset 3, PCF=0.1

```
Clay <= 11
|   DO <= 6: 0 (82.0/8.0)
|   DO > 6
|   |   Width <= 0.75: 1 (4.0)
|   |   Width > 0.75
|   |   |   Day <= 135: 0 (29.0/2.0)
|   |   |   Day > 135
|   |   |   |   Flowvelocity = 0: 0 (2.0/1.0)
|   |   |   |   Flowvelocity = 1: 0 (6.0/1.0)
|   |   |   |   Flowvelocity = 2
|   |   |   |   |   Pb <= 0: 1 (9.0/1.0)
|   |   |   |   |   Pb > 0
|   |   |   |   |   |   T <= 5.1: 1 (3.0)
|   |   |   |   |   |   T > 5.1: 0 (10.0)
|   |   |   |   Flowvelocity = 3
|   |   |   |   |   pH <= 7.13: 0 (3.0)
|   |   |   |   |   pH > 7.13: 1 (5.0/1.0)
|   |   |   |   Flowvelocity = 4: 1 (4.0)
Clay > 11: 0 (71.0)

Number of Leaves  :   12
Size of the tree :    20

=== Evaluation on test set ===

Correctly Classified Instances         103               90.3509 %
Incorrectly Classified Instances        11                9.6491 %
Kappa statistic                          0.6288
Mean absolute error                      0.1342
Root mean squared error                  0.2879
Relative absolute error                 51.8054 %
Root relative squared error             80.8061 %

=== Confusion Matrix ===

  a  b   <-- classified as
 91  6 |  a = 0
  5 12 |  b = 1
```

Appendix 12: Classification tree *Gammarus*, Sediments Flanders, Subset 3, PCF=0.01

```
: 0 (228.0/35.0)

Number of Leaves  :   1
Size of the tree :         1

=== Evaluation on test set ===

Correctly Classified Instances          97               85.0877 %
Incorrectly Classified Instances        17               14.9123 %
Kappa statistic                          0
Mean absolute error                      0.2568
Root mean squared error                  0.3562
Relative absolute error                 99.1835 %
Root relative squared error             99.986  %

=== Confusion Matrix ===

  a  b   <-- classified as
 97  0 |  a = 0
 17  0 |  b = 1
```

Appendix 13: Classification tree *Asellus*, Sediments Flanders, Subset 1, PCF=0.5

```
DO <= 3.1: 0 (37.0/1.0)
DO > 3.1
|   TOXR = 0
|   |   Day <= 70
|   |   |   Cd <= 0.3: 0 (18.0)
|   |   |   Cd > 0.3
|   |   |   |   DO <= 7.3: 0 (11.0)
|   |   |   |   DO > 7.3: 1 (4.0/1.0)
|   |   Day > 70
|   |   |   Width <= 3.5
|   |   |   |   Flowvelocity = 0: 1 (1.0)
|   |   |   |   Flowvelocity = 1
|   |   |   |   |   As <= 6.3: 0 (9.0)
|   |   |   |   |   As > 6.3
|   |   |   |   |   |   Conductivity <= 790.000022: 1 (8.0)
|   |   |   |   |   |   Conductivity > 790.000022: 0 (2.0)
|   |   |   |   Flowvelocity = 2
|   |   |   |   |   DO <= 5.1: 0 (10.0)
|   |   |   |   |   DO > 5.1
|   |   |   |   |   |   Day <= 329
|   |   |   |   |   |   |   Day <= 79: 1 (6.0/1.0)
|   |   |   |   |   |   |   Day > 79
|   |   |   |   |   |   |   |   Depth <= 0.45
|   |   |   |   |   |   |   |   |   Depth <= 0.35
|   |   |   |   |   |   |   |   |   |   Day <= 127: 1 (3.0/1.0)
|   |   |   |   |   |   |   |   |   |   Day > 127: 0 (8.0)
|   |   |   |   |   |   |   |   |   Depth > 0.35: 1 (4.0/1.0)
|   |   |   |   |   |   |   |   Depth > 0.45: 0 (8.0)
|   |   |   |   |   |   Day > 329: 1 (5.0)
|   |   |   |   Flowvelocity = 3: 0 (17.0/1.0)
|   |   |   |   Flowvelocity = 4: 1 (3.0/1.0)
|   |   |   Width > 3.5
|   |   |   |   Clay <= 12
|   |   |   |   |   Kjeldahlnitrogen <= 2520
|   |   |   |   |   |   Kjeldahlnitrogen <= 660
|   |   |   |   |   |   |   Loam <= 0: 0 (4.0)
|   |   |   |   |   |   |   Loam > 0
|   |   |   |   |   |   |   |   Kjeldahlnitrogen <= 490: 1 (8.0)
|   |   |   |   |   |   |   |   Kjeldahlnitrogen > 490: 0 (4.0)
|   |   |   |   |   |   Kjeldahlnitrogen > 660
|   |   |   |   |   |   |   Cd <= 0.3
|   |   |   |   |   |   |   |   Ni <= 8: 1 (6.0)
|   |   |   |   |   |   |   |   Ni > 8: 0 (4.0/1.0)
|   |   |   |   |   |   |   Cd > 0.3: 1 (19.0)
|   |   |   |   |   Kjeldahlnitrogen > 2520: 0 (5.0)
|   |   |   |   Clay > 12
|   |   |   |   |   Cd <= 0.5: 0 (10.0)
|   |   |   |   |   Cd > 0.5: 1 (2.0)
|   TOXR = 1: 0 (12.0/1.0)

Number of Leaves  :   27
Size of the tree :    50
```

```
=== Evaluation on test set ===

Correctly Classified Instances          75               65.7895 %
Incorrectly Classified Instances        39               34.2105 %
Kappa statistic                          0.1033
Mean absolute error                      0.3338
Root mean squared error                  0.5495
Relative absolute error                 78.9473 %
Root relative squared error            119.11   %

=== Confusion Matrix ===

  a  b   <-- classified as
 66 13 |  a = 0
 26  9 |  b = 1
```

Appendix 14: Classification tree *Asellus*, Sediments Flanders, Subset 1, PCF=0.25

```
DO <= 3.1: 0 (37.0/1.0)
DO > 3.1
|  Day <= 70: 0 (36.0/3.0)
|  Day > 70
|  |  Width <= 3.5
|  |  |  Flowvelocity = 0: 1 (1.0)
|  |  |  Flowvelocity = 1
|  |  |  |  As <= 6.3: 0 (10.0)
|  |  |  |  As > 6.3
|  |  |  |  |  Conductivity <= 790.000022: 1 (9.0/1.0)
|  |  |  |  |  Conductivity > 790.000022: 0 (3.0)
|  |  |  Flowvelocity = 2
|  |  |  |  DO <= 5.1: 0 (11.0)
|  |  |  |  DO > 5.1
|  |  |  |  |  Day <= 329
|  |  |  |  |  |  Day <= 79: 1 (7.0/1.0)
|  |  |  |  |  |  Day > 79: 0 (24.0/5.0)
|  |  |  |  |  Day > 329: 1 (5.0)
|  |  |  Flowvelocity = 3: 0 (18.0/1.0)
|  |  |  Flowvelocity = 4: 1 (3.0/1.0)
|  |  Width > 3.5
|  |  |  Clay <= 12
|  |  |  |  Kjeldahlnitrogen <= 2520
|  |  |  |  |  Kjeldahlnitrogen <= 660
|  |  |  |  |  |  Loam <= 0: 0 (4.0)
|  |  |  |  |  |  Loam > 0
|  |  |  |  |  |  |  Kjeldahlnitrogen <= 490: 1 (8.0)
|  |  |  |  |  |  |  Kjeldahlnitrogen > 490: 0 (4.0)
|  |  |  |  |  Kjeldahlnitrogen > 660
|  |  |  |  |  |  Cd <= 0.3
|  |  |  |  |  |  |  Ni <= 8: 1 (6.0)
|  |  |  |  |  |  |  Ni > 8: 0 (4.0/1.0)
|  |  |  |  |  |  Cd > 0.3: 1 (19.0)
|  |  |  |  Kjeldahlnitrogen > 2520: 0 (5.0)
|  |  |  Clay > 12
|  |  |  |  Cd <= 0.5: 0 (12.0)
|  |  |  |  Cd > 0.5: 1 (2.0)

Number of Leaves  :   21
Size of the tree :       38

=== Evaluation on test set ===

Correctly Classified Instances          77               67.5439 %
Incorrectly Classified Instances        37               32.4561 %
Kappa statistic                          0.1179
Mean absolute error                      0.3329
Root mean squared error                  0.5311
Relative absolute error                 78.7253 %
Root relative squared error            115.1184 %
```

```
=== Confusion Matrix ===

  a  b   <-- classified as
 69 10 |  a = 0
 27   8 |  b = 1
```

Appendix 15: Classification tree *Asellus*, Sediments Flanders, Subset 1, PCF=0.1

```
DO <= 3.1: 0 (37.0/1.0)
DO > 3.1
|  Day <= 70: 0 (36.0/3.0)
|  Day > 70
|  |  Width <= 3.5
|  |  |  Flowvelocity = 0: 1 (1.0)
|  |  |  Flowvelocity = 1
|  |  |  |  As <= 6.3: 0 (10.0)
|  |  |  |  As > 6.3
|  |  |  |  |  Conductivity <= 790.000022: 1 (9.0/1.0)
|  |  |  |  |  Conductivity > 790.000022: 0 (3.0)
|  |  |  Flowvelocity = 2
|  |  |  |  DO <= 5.1: 0 (11.0)
|  |  |  |  DO > 5.1
|  |  |  |  |  Day <= 329
|  |  |  |  |  |  Day <= 79: 1 (7.0/1.0)
|  |  |  |  |  |  Day > 79: 0 (24.0/5.0)
|  |  |  |  |  Day > 329: 1 (5.0)
|  |  |  Flowvelocity = 3: 0 (18.0/1.0)
|  |  |  Flowvelocity = 4: 1 (3.0/1.0)
|  |  Width > 3.5
|  |  |  Clay <= 12
|  |  |  |  Kjeldahlnitrogen <= 2520
|  |  |  |  |  Kjeldahlnitrogen <= 660
|  |  |  |  |  |  Loam <= 0: 0 (4.0)
|  |  |  |  |  |  Loam > 0
|  |  |  |  |  |  |  Kjeldahlnitrogen <= 490: 1 (8.0)
|  |  |  |  |  |  |  Kjeldahlnitrogen > 490: 0 (4.0)
|  |  |  |  |  Kjeldahlnitrogen > 660: 1 (29.0/3.0)
|  |  |  |  Kjeldahlnitrogen > 2520: 0 (5.0)
|  |  |  Clay > 12
|  |  |  |  Cd <= 0.5: 0 (12.0)
|  |  |  |  Cd > 0.5: 1 (2.0)

Number of Leaves  :   19
Size of the tree :        34

=== Evaluation on test set ===

Correctly Classified Instances          80               70.1754 %
Incorrectly Classified Instances        34               29.8246 %
Kappa statistic                          0.2112
Mean absolute error                      0.3167
Root mean squared error                  0.5126
Relative absolute error                 74.9157 %
Root relative squared error            111.1077 %
```

```
=== Confusion Matrix ===

  a  b   <-- classified as
 69 10 |  a = 0
 24 11 |  b = 1
```

Appendix 16: Classification tree *Asellus*, Sediments Flanders, Subset 1, PCF=0.01

```
DO <= 3.1: 0 (37.0/1.0)
DO > 3.1
|  Day <= 70: 0 (36.0/3.0)
|  Day > 70
|  |  Width <= 3.5: 0 (91.0/28.0)
|  |  Width > 3.5
|  |  |  Clay <= 12
|  |  |  |  Kjeldahlnitrogen <= 2520
|  |  |  |  |  Kjeldahlnitrogen <= 660
|  |  |  |  |  |  Loam <= 0: 0 (4.0)
|  |  |  |  |  |  Loam > 0
|  |  |  |  |  |  |  Kjeldahlnitrogen <= 490: 1 (8.0)
|  |  |  |  |  |  |  Kjeldahlnitrogen > 490: 0 (4.0)
|  |  |  |  |  Kjeldahlnitrogen > 660: 1 (29.0/3.0)
|  |  |  |  Kjeldahlnitrogen > 2520: 0 (5.0)
|  |  |  Clay > 12
|  |  |  |  Cd <= 0.5: 0 (12.0)
|  |  |  |  Cd > 0.5: 1 (2.0)

Number of Leaves  :   10
Size of the tree :    19

=== Evaluation on test set ===

Correctly Classified Instances          82               71.9298 %
Incorrectly Classified Instances        32               28.0702 %
Kappa statistic                          0.1846
Mean absolute error                      0.3452
Root mean squared error                  0.4881
Relative absolute error                 81.6473 %
Root relative squared error            105.8043 %

=== Confusion Matrix ===

  a  b   <-- classified as
 75  4 |  a = 0
 28  7 |  b = 1
```

Appendix 17: Classification tree *Asellus*, Sediments Flanders, Subset 2, PCF=0.5

```
Conductivity <= 910.000026
|  Width <= 3.5
|  |  T <= 13.7
|  |  |  Day <= 69: 0 (18.0/1.0)
|  |  |  Day > 69
|  |  |  |  Width <= 2
|  |  |  |  |  Zn <= 82
|  |  |  |  |  |  Zn <= 37: 1 (5.0/1.0)
|  |  |  |  |  |  Zn > 37: 0 (9.0)
|  |  |  |  |  Zn > 82
|  |  |  |  |  |  DO <= 5.7
|  |  |  |  |  |  |  Depth <= 0.35: 1 (4.0/1.0)
|  |  |  |  |  |  |  Depth > 0.35: 0 (4.0/1.0)
|  |  |  |  |  |  DO > 5.7: 1 (13.0)
|  |  |  |  Width > 2
|  |  |  |  |  Clay <= 1: 1 (3.0)
|  |  |  |  |  Clay > 1: 0 (23.0/3.0)
|  |  T > 13.7: 0 (24.0)
|  Width > 3.5
|  |  Kjeldahlnitrogen <= 2040
|  |  |  Flowvelocity = 0: 1 (1.0)
|  |  |  Flowvelocity = 1
|  |  |  |  Clay <= 12: 1 (3.0)
|  |  |  |  Clay > 12: 0 (2.0)
|  |  |  Flowvelocity = 2
|  |  |  |  As <= 7.1
|  |  |  |  |  T <= 19.6: 0 (8.0/1.0)
|  |  |  |  |  T > 19.6: 1 (4.0)
|  |  |  |  As > 7.1
|  |  |  |  |  Sand <= 97: 1 (16.0)
|  |  |  |  |  Sand > 97: 0 (3.0/1.0)
|  |  |  Flowvelocity = 3: 1 (8.0)
|  |  |  Flowvelocity = 4: 1 (0.0)
|  |  Kjeldahlnitrogen > 2040: 0 (17.0/2.0)
Conductivity > 910.000026: 0 (63.0/5.0)

Number of Leaves  :   20
Size of the tree :    36

=== Evaluation on test set ===

Correctly Classified Instances          79               69.2982 %
Incorrectly Classified Instances        35               30.7018 %
Kappa statistic                          0.176
Mean absolute error                      0.3264
Root mean squared error                  0.5129
Relative absolute error                 77.5245 %
Root relative squared error            112.0928 %
```

```
=== Confusion Matrix ===

  a  b   <-- classified as
 69 11 |  a = 0
 24 10 |  b = 1
```

Appendix 18: Classification tree *Asellus*, Sediments Flanders, Subset 2, PCF=0.25

```
Conductivity <= 910.000026
|  Width <= 3.5
|  |  T <= 13.7
|  |  |  Day <= 69: 0 (18.0/1.0)
|  |  |  Day > 69
|  |  |  |  Width <= 2
|  |  |  |  |  Zn <= 82
|  |  |  |  |  |  Zn <= 37: 1 (5.0/1.0)
|  |  |  |  |  |  Zn > 37: 0 (9.0)
|  |  |  |  |  Zn > 82
|  |  |  |  |  |  DO <= 5.7
|  |  |  |  |  |  |  Depth <= 0.35: 1 (4.0/1.0)
|  |  |  |  |  |  |  Depth > 0.35: 0 (4.0/1.0)
|  |  |  |  |  |  DO > 5.7: 1 (13.0)
|  |  |  |  Width > 2
|  |  |  |  |  Clay <= 1: 1 (3.0)
|  |  |  |  |  Clay > 1: 0 (23.0/3.0)
|  |  T > 13.7: 0 (24.0)
|  Width > 3.5
|  |  Kjeldahlnitrogen <= 2040
|  |  |  Flowvelocity = 0: 1 (1.0)
|  |  |  Flowvelocity = 1
|  |  |  |  Clay <= 12: 1 (3.0)
|  |  |  |  Clay > 12: 0 (2.0)
|  |  |  Flowvelocity = 2
|  |  |  |  As <= 7.1
|  |  |  |  |  T <= 19.6: 0 (8.0/1.0)
|  |  |  |  |  T > 19.6: 1 (4.0)
|  |  |  |  As > 7.1
|  |  |  |  |  Sand <= 97: 1 (16.0)
|  |  |  |  |  Sand > 97: 0 (3.0/1.0)
|  |  |  Flowvelocity = 3: 1 (8.0)
|  |  |  Flowvelocity = 4: 1 (0.0)
|  |  Kjeldahlnitrogen > 2040: 0 (17.0/2.0)
Conductivity > 910.000026: 0 (63.0/5.0)

Number of Leaves  :   20
Size of the tree :    36

=== Evaluation on test set ===

Correctly Classified Instances          79               69.2982 %
Incorrectly Classified Instances        35               30.7018 %
Kappa statistic                          0.176
Mean absolute error                      0.3264
Root mean squared error                  0.5129
Relative absolute error                 77.5245 %
Root relative squared error            112.0928 %
```

```
=== Confusion Matrix ===

  a  b   <-- classified as
 69 11 |  a = 0
 24 10 |  b = 1
```

Appendix 19: Classification tree *Asellus*, Sediments Flanders, Subset 2, PCF=0.1

```
Conductivity <= 910.000026
|   Width <= 3.5
|   |   T <= 13.7
|   |   |   Day <= 69: 0 (18.0/1.0)
|   |   |   Day > 69
|   |   |   |   Width <= 2
|   |   |   |   |   Zn <= 82
|   |   |   |   |   |   Zn <= 37: 1 (5.0/1.0)
|   |   |   |   |   |   Zn > 37: 0 (9.0)
|   |   |   |   |   Zn > 82: 1 (21.0/4.0)
|   |   |   |   Width > 2
|   |   |   |   |   Clay <= 1: 1 (3.0)
|   |   |   |   |   Clay > 1: 0 (23.0/3.0)
|   |   T > 13.7: 0 (24.0)
|   Width > 3.5
|   |   Kjeldahlnitrogen <= 2040
|   |   |   Flowvelocity = 0: 1 (1.0)
|   |   |   Flowvelocity = 1
|   |   |   |   Clay <= 12: 1 (3.0)
|   |   |   |   Clay > 12: 0 (2.0)
|   |   |   Flowvelocity = 2
|   |   |   |   As <= 7.1
|   |   |   |   |   T <= 19.6: 0 (8.0/1.0)
|   |   |   |   |   T > 19.6: 1 (4.0)
|   |   |   |   As > 7.1
|   |   |   |   |   Sand <= 97: 1 (16.0)
|   |   |   |   |   Sand > 97: 0 (3.0/1.0)
|   |   |   Flowvelocity = 3: 1 (8.0)
|   |   |   Flowvelocity = 4: 1 (0.0)
|   |   Kjeldahlnitrogen > 2040: 0 (17.0/2.0)
Conductivity > 910.000026: 0 (63.0/5.0)

Number of Leaves  :   18
Size of the tree :    32

=== Evaluation on test set ===

Correctly Classified Instances          76               66.6667 %
Incorrectly Classified Instances        38               33.3333 %
Kappa statistic                          0.1301
Mean absolute error                      0.3445
Root mean squared error                  0.5239
Relative absolute error                 81.8152 %
Root relative squared error            114.4986 %

=== Confusion Matrix ===

  a  b   <-- classified as
 66 14 |  a = 0
 24 10 |  b = 1
```

Appendix 20: Classification tree *Asellus*, Sediments Flanders, Subset 2, PCF=0.01

```
Conductivity <= 910.000026
|   Width <= 3.5
|   |   T <= 13.7
|   |   |   Day <= 69: 0 (18.0/1.0)
|   |   |   Day > 69
|   |   |   |   Width <= 2
|   |   |   |   |   Zn <= 82
|   |   |   |   |   |   Zn <= 37: 1 (5.0/1.0)
|   |   |   |   |   |   Zn > 37: 0 (9.0)
|   |   |   |   |   Zn > 82: 1 (21.0/4.0)
|   |   |   |   Width > 2
|   |   |   |   |   Clay <= 1: 1 (3.0)
|   |   |   |   |   Clay > 1: 0 (23.0/3.0)
|   |   T > 13.7: 0 (24.0)
|   Width > 3.5
|   |   Kjeldahlnitrogen <= 2040: 1 (45.0/11.0)
|   |   Kjeldahlnitrogen > 2040: 0 (17.0/2.0)
Conductivity > 910.000026: 0 (63.0/5.0)

Number of Leaves  :    10
Size of the tree :     19

=== Evaluation on test set ===

Correctly Classified Instances          74               64.9123 %
Incorrectly Classified Instances        40               35.0877 %
Kappa statistic                          0.1618
Mean absolute error                      0.3563
Root mean squared error                  0.5117
Relative absolute error                 84.6277 %
Root relative squared error            111.8488 %

=== Confusion Matrix ===

  a  b   <-- classified as
 60 20 |  a = 0
 20 14 |  b = 1
```

Appendix 21: Classification tree *Asellus*, Sediments Flanders, Subset 3, PCF=0.5

```
DO <= 2.7: 0 (35.0)
DO > 2.7
|   TOXR = 0
|   |   As <= 16.7
|   |   |   Day <= 273
|   |   |   |   Depth <= 0.2: 0 (17.0)
|   |   |   |   Depth > 0.2
|   |   |   |   |   Flowvelocity = 0: 0 (2.0)
|   |   |   |   |   Flowvelocity = 1
|   |   |   |   |   |   Totalphosphorus <= 1920
|   |   |   |   |   |   |   Pb <= 18
|   |   |   |   |   |   |   |   Width <= 9: 1 (4.0/1.0)
|   |   |   |   |   |   |   |   Width > 9: 0 (2.0)
|   |   |   |   |   |   |   Pb > 18: 0 (11.0)
|   |   |   |   |   |   Totalphosphorus > 1920: 1 (5.0/1.0)
|   |   |   |   |   Flowvelocity = 2
|   |   |   |   |   |   pH <= 7.97
|   |   |   |   |   |   |   TOXT = 0
|   |   |   |   |   |   |   |   Hg <= 0.243
|   |   |   |   |   |   |   |   |   Zn <= 149
|   |   |   |   |   |   |   |   |   |   As <= 4.4: 1 (6.0/1.0)
|   |   |   |   |   |   |   |   |   |   As > 4.4
|   |   |   |   |   |   |   |   |   |   |   Clay <= 10: 0 (9.0)
|   |   |   |   |   |   |   |   |   |   |   Clay > 10
|   |   |   |   |   |   |   |   |   |   |   |   Depth <= 0.75: 1 (4.0)
|   |   |   |   |   |   |   |   |   |   |   |   Depth > 0.75: 0 (3.0)
|   |   |   |   |   |   |   |   |   Zn > 149: 0 (15.0)
|   |   |   |   |   |   |   |   Hg > 0.243: 1 (3.0)
|   |   |   |   |   |   |   TOXT = 1: 0 (3.0)
|   |   |   |   |   |   pH > 7.97: 1 (5.0)
|   |   |   |   |   Flowvelocity = 3: 0 (23.0/3.0)
|   |   |   |   |   Flowvelocity = 4: 0 (5.0)
|   |   |   Day > 273
|   |   |   |   Clay <= 12
|   |   |   |   |   Width <= 3
|   |   |   |   |   |   Depth <= 0.2: 1 (3.0)
|   |   |   |   |   |   Depth > 0.2
|   |   |   |   |   |   |   Day <= 280: 1 (6.0/1.0)
|   |   |   |   |   |   |   Day > 280: 0 (14.0/2.0)
|   |   |   |   |   Width > 3: 1 (11.0)
|   |   |   |   Clay > 12: 0 (5.0)
|   |   As > 16.7
|   |   |   Conductivity <= 939.999998
|   |   |   |   Conductivity <= 250: 0 (3.0)
|   |   |   |   Conductivity > 250: 1 (23.0/2.0)
|   |   |   Conductivity > 939.999998: 0 (5.0)
|   TOXR = 1: 0 (6.0)

Number of Leaves  :   26
Size of the tree :    48
```

```
=== Evaluation on test set ===

Correctly Classified Instances          68               59.6491 %
Incorrectly Classified Instances        46               40.3509 %
Kappa statistic                          0.1509
Mean absolute error                      0.391
Root mean squared error                  0.5816
Relative absolute error                 92.8539 %
Root relative squared error            127.1211 %

=== Confusion Matrix ===

  a  b   <-- classified as
 49 31 |  a = 0
 15 19 |  b = 1
```

Appendix 22: Classification tree *Asellus*, Sediments Flanders, Subset 3, PCF=0.25

```
DO <= 2.7: 0 (35.0)
DO > 2.7
|   TOXR = 0
|   |   As <= 16.7
|   |   |   Day <= 273
|   |   |   |   Depth <= 0.2: 0 (17.0)
|   |   |   |   Depth > 0.2
|   |   |   |   |   Flowvelocity = 0: 0 (2.0)
|   |   |   |   |   Flowvelocity = 1
|   |   |   |   |   |   Totalphosphorus <= 1920
|   |   |   |   |   |   |   Pb <= 18
|   |   |   |   |   |   |   |   Width <= 9: 1 (4.0/1.0)
|   |   |   |   |   |   |   |   Width > 9: 0 (2.0)
|   |   |   |   |   |   |   Pb > 18: 0 (11.0)
|   |   |   |   |   |   Totalphosphorus > 1920: 1 (5.0/1.0)
|   |   |   |   |   Flowvelocity = 2
|   |   |   |   |   |   pH <= 7.97
|   |   |   |   |   |   |   Hg <= 0.243
|   |   |   |   |   |   |   |   Zn <= 149
|   |   |   |   |   |   |   |   |   As <= 4.4: 1 (7.0/2.0)
|   |   |   |   |   |   |   |   |   As > 4.4
|   |   |   |   |   |   |   |   |   |   Clay <= 10: 0 (10.0)
|   |   |   |   |   |   |   |   |   |   Clay > 10
|   |   |   |   |   |   |   |   |   |   |   Depth <= 0.75: 1 (4.0)
|   |   |   |   |   |   |   |   |   |   |   Depth > 0.75: 0 (3.0)
|   |   |   |   |   |   |   |   Zn > 149: 0 (16.0)
|   |   |   |   |   |   |   Hg > 0.243: 1 (3.0)
|   |   |   |   |   |   pH > 7.97: 1 (5.0)
|   |   |   |   |   Flowvelocity = 3: 0 (23.0/3.0)
|   |   |   |   |   Flowvelocity = 4: 0 (5.0)
|   |   |   Day > 273
|   |   |   |   Clay <= 12
|   |   |   |   |   Width <= 3
|   |   |   |   |   |   Depth <= 0.2: 1 (3.0)
|   |   |   |   |   |   Depth > 0.2
|   |   |   |   |   |   |   Day <= 280: 1 (6.0/1.0)
|   |   |   |   |   |   |   Day > 280: 0 (14.0/2.0)
|   |   |   |   |   Width > 3: 1 (11.0)
|   |   |   |   Clay > 12: 0 (5.0)
|   |   As > 16.7
|   |   |   Conductivity <= 939.999998
|   |   |   |   Conductivity <= 250: 0 (3.0)
|   |   |   |   Conductivity > 250: 1 (23.0/2.0)
|   |   |   Conductivity > 939.999998: 0 (5.0)
|   TOXR = 1: 0 (6.0)

Number of Leaves  :  25
Size of the tree :   46
```

```
=== Evaluation on test set ===

Correctly Classified Instances          67               58.7719 %
Incorrectly Classified Instances        47               41.2281 %
Kappa statistic                          0.1389
Mean absolute error                      0.3976
Root mean squared error                  0.5834
Relative absolute error                 94.4412 %
Root relative squared error            127.5166 %

=== Confusion Matrix ===

  a  b   <-- classified as
 48 32 |  a = 0
 15 19 |  b = 1
```

Appendix 23: Classification tree *Asellus*, Sediments Flanders, Subset 3, PCF=0.1

```
DO <= 2.7: 0 (35.0)
DO > 2.7
|   As <= 16.7
|   |   Day <= 273: 0 (122.0/27.0)
|   |   Day > 273
|   |   |   Clay <= 12
|   |   |   |   Width <= 3
|   |   |   |   |   Depth <= 0.2: 1 (3.0)
|   |   |   |   |   Depth > 0.2
|   |   |   |   |   |   Day <= 280: 1 (6.0/1.0)
|   |   |   |   |   |   Day > 280: 0 (15.0/2.0)
|   |   |   |   Width > 3: 1 (11.0)
|   |   |   Clay > 12: 0 (5.0)
|   As > 16.7
|   |   Conductivity <= 939.999998
|   |   |   Conductivity <= 250: 0 (3.0)
|   |   |   Conductivity > 250: 1 (23.0/2.0)
|   |   Conductivity > 939.999998: 0 (5.0)

Number of Leaves  :   10
Size of the tree :    19

=== Evaluation on test set ===

Correctly Classified Instances          82               71.9298 %
Incorrectly Classified Instances        32               28.0702 %
Kappa statistic                          0.2808
Mean absolute error                      0.3491
Root mean squared error                  0.4813
Relative absolute error                 82.9114 %
Root relative squared error            105.2033 %

=== Confusion Matrix ===

  a  b   <-- classified as
 68 12 |  a = 0
 20 14 |  b = 1
```

Appendix 24: Classification tree *Asellus*, Sediments Flanders, Subset 3, PCF=0.01

```
DO <= 2.7: 0 (35.0)
DO > 2.7
|   As <= 16.7
|   |   Day <= 273: 0 (122.0/27.0)
|   |   Day > 273
|   |   |   Clay <= 12
|   |   |   |   Width <= 3
|   |   |   |   |   Depth <= 0.2: 1 (3.0)
|   |   |   |   |   Depth > 0.2
|   |   |   |   |   |   Day <= 280: 1 (6.0/1.0)
|   |   |   |   |   |   Day > 280: 0 (15.0/2.0)
|   |   |   |   Width > 3: 1 (11.0)
|   |   |   Clay > 12: 0 (5.0)
|   As > 16.7
|   |   Conductivity <= 939.999998
|   |   |   Conductivity <= 250: 0 (3.0)
|   |   |   Conductivity > 250: 1 (23.0/2.0)
|   |   Conductivity > 939.999998: 0 (5.0)

Number of Leaves  :   10
Size of the tree :    19

=== Evaluation on test set ===

Correctly Classified Instances          82               71.9298 %
Incorrectly Classified Instances        32               28.0702 %
Kappa statistic                          0.2808
Mean absolute error                      0.3491
Root mean squared error                  0.4813
Relative absolute error                 82.9114 %
Root relative squared error            105.2033 %

=== Confusion Matrix ===

  a  b   <-- classified as
 68 12 |  a = 0
 20 14 |  b = 1
```

Appendix 25: Classification tree *Gammarus*, Zwalm river basin, Subset 1, PCF=0.5

```
Totalphosphorus <= 0.37
|   Depth <= 19: 1 (41.0)
|   Depth > 19
|   |   Flowvelocity <= 0.32: 1 (17.0)
|   |   Flowvelocity > 0.32
|   |   |   pH <= 7.68: 0 (4.0)
|   |   |   pH > 7.68: 1 (14.0/3.0)
Totalphosphorus > 0.37
|   Distmouth <= 7396.07
|   |   T <= 16.5
|   |   |   Totalphosphorus <= 0.575: 0 (14.0)
|   |   |   Totalphosphorus > 0.575: 1 (3.0/1.0)
|   |   T > 16.5: 1 (2.0)
|   Distmouth > 7396.07
|   |   Totalphosphorus <= 0.76
|   |   |   Loamclay <= 0: 0 (4.0/1.0)
|   |   |   Loamclay > 0: 1 (17.0/1.0)
|   |   Totalphosphorus > 0.76: 0 (3.0)

Number of Leaves  :   10
Size of the tree :    19

=== Evaluation on test set ===

Correctly Classified Instances          42               70      %
Incorrectly Classified Instances        18               30      %
Kappa statistic                          0.1615
Mean absolute error                      0.3132
Root mean squared error                  0.5271
Relative absolute error                 85.6821 %
Root relative squared error            124.5527 %

=== Confusion Matrix ===

  a  b   <-- classified as
  5  9 |  a = 0
  9 37 |  b = 1
```

Appendix 26: Classification tree *Gammarus*, Zwalm river basin, Subset 1, PCF=0.25

```
Totalphosphorus <= 0.37
|   Depth <= 19: 1 (41.0)
|   Depth > 19
|   |   Flowvelocity <= 0.32: 1 (17.0)
|   |   Flowvelocity > 0.32
|   |   |   pH <= 7.68: 0 (4.0)
|   |   |   pH > 7.68: 1 (14.0/3.0)
Totalphosphorus > 0.37
|   Distmouth <= 7396.07
|   |   T <= 16.5
|   |   |   Totalphosphorus <= 0.575: 0 (14.0)
|   |   |   Totalphosphorus > 0.575: 1 (3.0/1.0)
|   |   T > 16.5: 1 (2.0)
|   Distmouth > 7396.07
|   |   Totalphosphorus <= 0.76
|   |   |   Loamclay <= 0: 0 (4.0/1.0)
|   |   |   Loamclay > 0: 1 (17.0/1.0)
|   |   Totalphosphorus > 0.76: 0 (3.0)

Number of Leaves  :   10
Size of the tree :    19

=== Evaluation on test set ===

Correctly Classified Instances          42               70      %
Incorrectly Classified Instances        18               30      %
Kappa statistic                          0.1615
Mean absolute error                      0.3132
Root mean squared error                  0.5271
Relative absolute error                 85.6821 %
Root relative squared error            124.5527 %

=== Confusion Matrix ===

  a  b   <-- classified as
  5  9 |  a = 0
  9 37 |  b = 1
```

Appendix 27: Classification tree *Gammarus*, Zwalm river basin, Subset 1, PCF=0.1

```
Totalphosphorus <= 0.37: 1 (76.0/7.0)
Totalphosphorus > 0.37
|   Distmouth <= 7396.07
|   |   T <= 16.5
|   |   |   Totalphosphorus <= 0.575: 0 (14.0)
|   |   |   Totalphosphorus > 0.575: 1 (3.0/1.0)
|   |   T > 16.5: 1 (2.0)
|   Distmouth > 7396.07
|   |   Totalphosphorus <= 0.76
|   |   |   Loamclay <= 0: 0 (4.0/1.0)
|   |   |   Loamclay > 0: 1 (17.0/1.0)
|   |   Totalphosphorus > 0.76: 0 (3.0)

Number of Leaves  :   7
Size of the tree :    13

=== Evaluation on test set ===

Correctly Classified Instances          44               73.3333 %
Incorrectly Classified Instances        16               26.6667 %
Kappa statistic                          0.1724
Mean absolute error                      0.2994
Root mean squared error                  0.4744
Relative absolute error                 81.8943 %
Root relative squared error            112.0963 %

=== Confusion Matrix ===

  a  b   <-- classified as
  4 10 |  a = 0
  6 40 |  b = 1
```

Appendix 28: Classification tree *Gammarus*, Zwalm river basin, Subset 1, PCF=0.01

```
Totalphosphorus <= 0.37: 1 (76.0/7.0)
Totalphosphorus > 0.37
|   Distmouth <= 7396.07
|   |   T <= 16.5
|   |   |   Totalphosphorus <= 0.575: 0 (14.0)
|   |   |   Totalphosphorus > 0.575: 1 (3.0/1.0)
|   |   T > 16.5: 1 (2.0)
|   Distmouth > 7396.07
|   |   Totalphosphorus <= 0.76
|   |   |   Loamclay <= 0: 0 (4.0/1.0)
|   |   |   Loamclay > 0: 1 (17.0/1.0)
|   |   Totalphosphorus > 0.76: 0 (3.0)

Number of Leaves  :   7
Size of the tree :        13

=== Evaluation on test set ===

Correctly Classified Instances          44               73.3333 %
Incorrectly Classified Instances        16               26.6667 %
Kappa statistic                          0.1724
Mean absolute error                      0.2994
Root mean squared error                  0.4744
Relative absolute error                 81.8943 %
Root relative squared error            112.0963 %

=== Confusion Matrix ===

  a  b   <-- classified as
  4 10 |  a = 0
  6 40 |  b = 1
```

Appendix 29: Classification tree *Gammarus*, Zwalm river basin, Subset 2, PCF=0.5

```
Width <= 700
|  Ammonium <= 0.39
|  |  T <= 12.1
|  |  |  T <= 11.8: 1 (8.0)
|  |  |  T > 11.8: 0 (4.0/1.0)
|  |  T > 12.1: 1 (40.0)
|  Ammonium > 0.39
|  |  Poolriffle = 1: 1 (7.0/1.0)
|  |  Poolriffle = 2: 1 (0.0)
|  |  Poolriffle = 3
|  |  |  Flowvelocity <= 0.48: 1 (7.0)
|  |  |  Flowvelocity > 0.48: 0 (2.0)
|  |  Poolriffle = 4
|  |  |  Hollowbanks = 1: 1 (0.0)
|  |  |  Hollowbanks = 2: 1 (2.0)
|  |  |  Hollowbanks = 3: 1 (4.0)
|  |  |  Hollowbanks = 4
|  |  |  |  Banks = 0: 0 (5.0/1.0)
|  |  |  |  Banks = 1: 1 (2.0)
|  |  |  |  Banks = 2: 0 (0.0)
|  |  |  Hollowbanks = 5
|  |  |  |  pH <= 7.7
|  |  |  |  |  Boulders <= 3.3: 1 (4.0)
|  |  |  |  |  Boulders > 3.3: 0 (3.0/1.0)
|  |  |  |  pH > 7.7: 0 (4.0)
|  |  |  Hollowbanks = 6
|  |  |  |  pH <= 7.71: 1 (2.0)
|  |  |  |  pH > 7.71: 0 (2.0)
|  |  Poolriffle = 5
|  |  |  Conductivity <= 714: 1 (5.0)
|  |  |  Conductivity > 714: 0 (4.0/1.0)
|  |  Poolriffle = 6
|  |  |  Suspendedsolids <= 14: 1 (5.0)
|  |  |  Suspendedsolids > 14: 0 (2.0)
Width > 700: 0 (7.0/1.0)

Number of Leaves  :   23
Size of the tree :    36

=== Evaluation on test set ===

Correctly Classified Instances          43               71.6667 %
Incorrectly Classified Instances        17               28.3333 %
Kappa statistic                          0.2956
Mean absolute error                      0.2714
Root mean squared error                  0.4793
Relative absolute error                 74.2491 %
Root relative squared error            113.2644 %
```

```
=== Confusion Matrix ===

  a  b   <-- classified as
  8  6 |  a = 0
 11 35 |  b = 1
```

Appendix 30: Classification tree *Gammarus*, Zwalm river basin, Subset 2, PCF=0.25

```
Width <= 700: 1 (112.0/23.0)
Width > 700: 0 (7.0/1.0)

Number of Leaves  :   2
Size of the tree :        3

=== Evaluation on test set ===

Correctly Classified Instances          46               76.6667 %
Incorrectly Classified Instances        14               23.3333 %
Kappa statistic                         0.1322
Mean absolute error                       0.3429
Root mean squared error                   0.4271
Relative absolute error                 93.7884 %
Root relative squared error            100.913  %

=== Confusion Matrix ===

  a  b   <-- classified as
  2 12 |  a = 0
  2 44 |  b = 1
```

Appendix 31: Classification tree *Gammarus*, Zwalm river basin, Subset 2, PCF=0.1

```
Width <= 700: 1 (112.0/23.0)
Width > 700: 0 (7.0/1.0)

Number of Leaves  :   2
Size of the tree :        3

=== Evaluation on test set ===

Correctly Classified Instances          46               76.6667 %
Incorrectly Classified Instances        14               23.3333 %
Kappa statistic                          0.1322
Mean absolute error                      0.3429
Root mean squared error                  0.4271
Relative absolute error                 93.7884 %
Root relative squared error            100.913  %

=== Confusion Matrix ===

  a  b   <-- classified as
  2 12 |  a = 0
  2 44 |  b = 1
```

Appendix 32: Classification tree *Gammarus*, Zwalm river basin, Subset 2, PCF=0.01

```
Width <= 700: 1 (112.0/23.0)
Width > 700: 0 (7.0/1.0)

Number of Leaves  :   2
Size of the tree :        3

=== Evaluation on test set ===

Correctly Classified Instances          46               76.6667 %
Incorrectly Classified Instances        14               23.3333 %
Kappa statistic                          0.1322
Mean absolute error                      0.3429
Root mean squared error                  0.4271
Relative absolute error                 93.7884 %
Root relative squared error            100.913  %

=== Confusion Matrix ===

  a  b   <-- classified as
  2 12 |  a = 0
  2 44 |  b = 1
```

Appendix 33: Classification tree *Gammarus*, Zwalm river basin, Subset 3, PCF=0.5

```
Loamclay <= 49.1
|   Hollowbanks = 1: 1 (3.0)
|   Hollowbanks = 2: 1 (8.0)
|   Hollowbanks = 3: 1 (16.0)
|   Hollowbanks = 4: 1 (14.0/1.0)
|   Hollowbanks = 5
|   |   Totalnitrogen <= 6.46: 0 (4.0)
|   |   Totalnitrogen > 6.46
|   |   |   Meandering = 1: 1 (0.0)
|   |   |   Meandering = 2: 1 (2.0)
|   |   |   Meandering = 3
|   |   |   |   Totalnitrogen <= 22.5: 1 (2.0)
|   |   |   |   Totalnitrogen > 22.5: 0 (2.0)
|   |   |   Meandering = 4: 1 (10.0/1.0)
|   |   |   Meandering = 5
|   |   |   |   Depth <= 42: 1 (13.0/1.0)
|   |   |   |   Depth > 42: 0 (4.0/1.0)
|   |   |   Meandering = 6: 1 (0.0)
|   Hollowbanks = 6: 1 (15.0/1.0)
Loamclay > 49.1
|   Orthophosphate <= 0.235: 1 (5.0)
|   Orthophosphate > 0.235
|   |   Ammonium <= 0.6: 0 (9.0)
|   |   Ammonium > 0.6
|   |   |   COD <= 16: 1 (3.0)
|   |   |   COD > 16
|   |   |   |   Conductivity <= 583: 1 (3.0)
|   |   |   |   Conductivity > 583: 0 (6.0)

Number of Leaves  :   19
Size of the tree :    29

=== Evaluation on test set ===

Correctly Classified Instances          44               73.3333 %
Incorrectly Classified Instances        16               26.6667 %
Kappa statistic                          0.2195
Mean absolute error                      0.2835
Root mean squared error                  0.4937
Relative absolute error                 76.6562 %
Root relative squared error            113.9914 %

=== Confusion Matrix ===

  a  b   <-- classified as
  5 10 |  a = 0
  6 39 |  b = 1
```

Appendix 34: Classification tree *Gammarus*, Zwalm river basin, Subset 3, PCF=0.25

```
Loamclay <= 49.1: 1 (93.0/13.0)
Loamclay > 49.1
|   Orthophosphate <= 0.235: 1 (5.0)
|   Orthophosphate > 0.235
|   |   Ammonium <= 0.6: 0 (9.0)
|   |   Ammonium > 0.6
|   |   |   COD <= 16: 1 (3.0)
|   |   |   COD > 16
|   |   |   |   Conductivity <= 583: 1 (3.0)
|   |   |   |   Conductivity > 583: 0 (6.0)

Number of Leaves  :   6
Size of the tree :        11

=== Evaluation on test set ===

Correctly Classified Instances          48              80      %
Incorrectly Classified Instances        12              20      %
Kappa statistic                          0.3514
Mean absolute error                      0.2629
Root mean squared error                  0.4156
Relative absolute error                 71.0867 %
Root relative squared error             95.9503 %

=== Confusion Matrix ===

  a  b   <-- classified as
  5 10 |  a = 0
  2 43 |  b = 1
```

Appendix 35: Classification tree *Gammarus*, Zwalm river basin, Subset 3, PCF=0.1

```
Loamclay <= 49.1: 1 (93.0/13.0)
Loamclay > 49.1
|   Orthophosphate <= 0.235: 1 (5.0)
|   Orthophosphate > 0.235
|   |   Ammonium <= 0.6: 0 (9.0)
|   |   Ammonium > 0.6
|   |   |   COD <= 16: 1 (3.0)
|   |   |   COD > 16
|   |   |   |   Conductivity <= 583: 1 (3.0)
|   |   |   |   Conductivity > 583: 0 (6.0)

Number of Leaves  :   6
Size of the tree :        11

=== Evaluation on test set ===

Correctly Classified Instances          48               80      %
Incorrectly Classified Instances        12               20      %
Kappa statistic                          0.3514
Mean absolute error                      0.2629
Root mean squared error                  0.4156
Relative absolute error                 71.0867 %
Root relative squared error             95.9503 %

=== Confusion Matrix ===

  a  b   <-- classified as
  5 10 |  a = 0
  2 43 |  b = 1
```

Appendix 36: Classification tree *Gammarus*, Zwalm river basin, Subset 3, PCF=0.01

```
Loamclay <= 49.1: 1 (93.0/13.0)
Loamclay > 49.1
|   Orthophosphate <= 0.235: 1 (5.0)
|   Orthophosphate > 0.235
|   |   Ammonium <= 0.6: 0 (9.0)
|   |   Ammonium > 0.6
|   |   |   COD <= 16: 1 (3.0)
|   |   |   COD > 16
|   |   |   |   Conductivity <= 583: 1 (3.0)
|   |   |   |   Conductivity > 583: 0 (6.0)

Number of Leaves  :   6
Size of the tree :        11

=== Evaluation on test set ===

Correctly Classified Instances          48               80      %
Incorrectly Classified Instances        12               20      %
Kappa statistic                          0.3514
Mean absolute error                      0.2629
Root mean squared error                  0.4156
Relative absolute error                 71.0867 %
Root relative squared error             95.9503 %

=== Confusion Matrix ===

  a  b   <-- classified as
  5 10 |  a = 0
  2 43 |  b = 1
```

Appendix 37: Classification tree *Asellus*, Zwalm river basin, Subset 1, PCF=0.5

```
Width <= 114
|  Banks = 0: 0 (47.0/4.0)
|  Banks = 1: 0 (6.0/1.0)
|  Banks = 2: 1 (2.0)
Width > 114
|  Strorder = 1
|  |  Poolriffle = 1: 1 (0.0)
|  |  Poolriffle = 2: 1 (0.0)
|  |  Poolriffle = 3: 1 (0.0)
|  |  Poolriffle = 4: 1 (4.0)
|  |  Poolriffle = 5: 0 (2.0)
|  |  Poolriffle = 6: 1 (0.0)
|  Strorder = 2
|  |  pH <= 7.47: 1 (3.0)
|  |  pH > 7.47: 0 (5.0)
|  Strorder = 3
|  |  Orthophosphate <= 0.23
|  |  |  Poolriffle = 1: 0 (0.0)
|  |  |  Poolriffle = 2: 0 (2.0/1.0)
|  |  |  Poolriffle = 3: 1 (4.0/1.0)
|  |  |  Poolriffle = 4: 0 (3.0)
|  |  |  Poolriffle = 5: 1 (1.0)
|  |  |  Poolriffle = 6: 0 (1.0)
|  |  Orthophosphate > 0.23: 1 (7.0)
|  Strorder = 4: 1 (32.0/1.0)

Number of Leaves  :   19
Size of the tree :       26

=== Evaluation on test set ===

Correctly Classified Instances          43               71.6667 %
Incorrectly Classified Instances        17               28.3333 %
Kappa statistic                          0.4308
Mean absolute error                      0.3183
Root mean squared error                  0.5187
Relative absolute error                 63.741  %
Root relative squared error            103.79   %

=== Confusion Matrix ===

  a  b   <-- classified as
 24  7  |  a = 0
 10 19 |  b = 1
```

Appendix 38: Classification tree *Asellus*, Zwalm river basin, Subset 1, PCF=0.25

```
Width <= 114: 0 (55.0/7.0)
Width > 114
|   Strorder = 1
|   |   Poolriffle = 1: 1 (0.0)
|   |   Poolriffle = 2: 1 (0.0)
|   |   Poolriffle = 3: 1 (0.0)
|   |   Poolriffle = 4: 1 (4.0)
|   |   Poolriffle = 5: 0 (2.0)
|   |   Poolriffle = 6: 1 (0.0)
|   Strorder = 2
|   |   pH <= 7.47: 1 (3.0)
|   |   pH > 7.47: 0 (5.0)
|   Strorder = 3: 1 (18.0/6.0)
|   Strorder = 4: 1 (32.0/1.0)

Number of Leaves  :   11
Size of the tree :    15

=== Evaluation on test set ===

Correctly Classified Instances          45               75      %
Incorrectly Classified Instances        15               25      %
Kappa statistic                          0.4978
Mean absolute error                      0.2873
Root mean squared error                  0.439
Relative absolute error                 57.5344 %
Root relative squared error             87.8385 %

=== Confusion Matrix ===

  a  b   <-- classified as
 25  6 |  a = 0
  9 20 |  b = 1
```

Appendix 39: Classification tree *Asellus*, Zwalm river basin, Subset 1, PCF=0.1

```
Width <= 114: 0 (55.0/7.0)
Width > 114: 1 (64.0/14.0)

Number of Leaves  :   2
Size of the tree :        3

=== Evaluation on test set ===

Correctly Classified Instances          46               76.6667 %
Incorrectly Classified Instances        14               23.3333 %
Kappa statistic                          0.5339
Mean absolute error                      0.3241
Root mean squared error                  0.4267
Relative absolute error                 64.9057 %
Root relative squared error             85.3906 %

=== Confusion Matrix ===

  a  b   <-- classified as
 23  8 |  a = 0
  6 23 |  b = 1
```

Appendix 40: Classification tree *Asellus*, Zwalm river basin, Subset 1, PCF=0.01

```
Width <= 114: 0 (55.0/7.0)
Width > 114: 1 (64.0/14.0)

Number of Leaves  :   2
Size of the tree :        3

=== Evaluation on test set ===

Correctly Classified Instances          46               76.6667 %
Incorrectly Classified Instances        14               23.3333 %
Kappa statistic                          0.5339
Mean absolute error                      0.3241
Root mean squared error                  0.4267
Relative absolute error                 64.9057 %
Root relative squared error             85.3906 %

=== Confusion Matrix ===

  a  b   <-- classified as
 23  8 |  a = 0
  6 23 |  b = 1
```

Appendix 41: Classification tree *Asellus*, Zwalm river basin, Subset 2, PCF=0.5

```
Width <= 267
|  Hollowbanks = 1: 0 (0.0)
|  Hollowbanks = 2
|  |  Flowvelocity <= 0.43: 0 (2.0)
|  |  Flowvelocity > 0.43: 1 (2.0)
|  Hollowbanks = 3: 0 (11.0/1.0)
|  Hollowbanks = 4
|  |  Depth <= 22: 0 (15.0)
|  |  Depth > 22: 1 (2.0)
|  Hollowbanks = 5
|  |  Strorder = 1
|  |  |  Totalphosphorus <= 0.13: 1 (2.0)
|  |  |  Totalphosphorus > 0.13: 0 (20.0/1.0)
|  |  Strorder = 2: 1 (5.0/1.0)
|  |  Strorder = 3
|  |  |  Flowvelocity <= 0.71: 0 (4.0)
|  |  |  Flowvelocity > 0.71: 1 (2.0)
|  |  Strorder = 4: 0 (1.0)
|  Hollowbanks = 6
|  |  Gravel <= 16.3
|  |  |  Suspendedsolids <= 23
|  |  |  |  Nitrate <= 2.19: 1 (3.0)
|  |  |  |  Nitrate > 2.19
|  |  |  |  |  T <= 14.7: 0 (3.0)
|  |  |  |  |  T > 14.7: 1 (3.0/1.0)
|  |  |  Suspendedsolids > 23: 0 (5.0)
|  |  Gravel > 16.3: 1 (5.0)
Width > 267: 1 (34.0/1.0)

Number of Leaves  :   18
Size of the tree :       29

=== Evaluation on test set ===

Correctly Classified Instances          44               73.3333 %
Incorrectly Classified Instances        16               26.6667 %
Kappa statistic                          0.4649
Mean absolute error                      0.2868
Root mean squared error                  0.4895
Relative absolute error                 57.4347 %
Root relative squared error             97.9583 %

=== Confusion Matrix ===

  a  b   <-- classified as
 24  7 |  a = 0
  9 20 |  b = 1
```

Appendix 42: Classification tree *Asellus*, Zwalm river basin, Subset 2, PCF=0.25

```
Width <= 267
|  Hollowbanks = 1: 0 (0.0)
|  Hollowbanks = 2
|  |  Flowvelocity <= 0.43: 0 (2.0)
|  |  Flowvelocity > 0.43: 1 (2.0)
|  Hollowbanks = 3: 0 (11.0/1.0)
|  Hollowbanks = 4
|  |  Depth <= 22: 0 (15.0)
|  |  Depth > 22: 1 (2.0)
|  Hollowbanks = 5
|  |  Strorder = 1
|  |  |  Totalphosphorus <= 0.13: 1 (2.0)
|  |  |  Totalphosphorus > 0.13: 0 (20.0/1.0)
|  |  Strorder = 2: 1 (5.0/1.0)
|  |  Strorder = 3
|  |  |  Flowvelocity <= 0.71: 0 (4.0)
|  |  |  Flowvelocity > 0.71: 1 (2.0)
|  |  Strorder = 4: 0 (1.0)
|  Hollowbanks = 6
|  |  Gravel <= 16.3
|  |  |  Suspendedsolids <= 23
|  |  |  |  Nitrate <= 2.19: 1 (3.0)
|  |  |  |  Nitrate > 2.19
|  |  |  |  |  T <= 14.7: 0 (3.0)
|  |  |  |  |  T > 14.7: 1 (3.0/1.0)
|  |  |  Suspendedsolids > 23: 0 (5.0)
|  |  Gravel > 16.3: 1 (5.0)
Width > 267: 1 (34.0/1.0)

Number of Leaves  :   18
Size of the tree :       29

=== Evaluation on test set ===

Correctly Classified Instances          44               73.3333 %
Incorrectly Classified Instances        16               26.6667 %
Kappa statistic                          0.4649
Mean absolute error                      0.2868
Root mean squared error                  0.4895
Relative absolute error                 57.4347 %
Root relative squared error             97.9583 %

=== Confusion Matrix ===

  a  b   <-- classified as
 24  7 |  a = 0
  9 20 |  b = 1
```

Appendix 43: Classification tree *Asellus*, Zwalm river basin, Subset 2, PCF=0.1

```
Width <= 267: 0 (85.0/24.0)
Width > 267: 1 (34.0/1.0)

Number of Leaves  :   2
Size of the tree :         3

=== Evaluation on test set ===

Correctly Classified Instances          45               75      %
Incorrectly Classified Instances        15               25      %
Kappa statistic                          0.4921
Mean absolute error                      0.3322
Root mean squared error                  0.4195
Relative absolute error                 66.523  %
Root relative squared error             83.9366 %

=== Confusion Matrix ===

  a  b   <-- classified as
 30  1  |  a = 0
 14 15 |  b = 1
```

Appendix 44: Classification tree *Asellus*, Zwalm river basin, Subset 2, PCF=0.01

```
Width <= 267: 0 (85.0/24.0)
Width > 267: 1 (34.0/1.0)

Number of Leaves  :   2
Size of the tree :        3

=== Evaluation on test set ===

Correctly Classified Instances          45              75      %
Incorrectly Classified Instances        15              25      %
Kappa statistic                          0.4921
Mean absolute error                      0.3322
Root mean squared error                  0.4195
Relative absolute error                 66.523  %
Root relative squared error             83.9366 %

=== Confusion Matrix ===

  a  b   <-- classified as
 30   1 |  a = 0
 14 15 |  b = 1
```

Appendix 45: Classification tree *Asellus*, Zwalm river basin, Subset 3, PCF=0.5

```
Width <= 250
|   Width <= 123
|   |   Banks = 0: 0 (44.0/5.0)
|   |   Banks = 1
|   |   |   T <= 15.1: 0 (8.0)
|   |   |   T > 15.1: 1 (4.0)
|   |   Banks = 2
|   |   |   Conductivity <= 738: 1 (2.0)
|   |   |   Conductivity > 738: 0 (3.0)
|   Width > 123
|   |   Distmouth <= 15778.284
|   |   |   Width <= 144: 1 (9.0)
|   |   |   Width > 144
|   |   |   |   Ammonium <= 0.23: 1 (3.0)
|   |   |   |   Ammonium > 0.23: 0 (7.0/1.0)
|   |   Distmouth > 15778.284: 0 (4.0)
Width > 250: 1 (35.0/2.0)

Number of Leaves  :   10
Size of the tree :    18

=== Evaluation on test set ===

Correctly Classified Instances          52               86.6667 %
Incorrectly Classified Instances         8               13.3333 %
Kappa statistic                          0.733
Mean absolute error                      0.1908
Root mean squared error                  0.3532
Relative absolute error                 38.2041 %
Root relative squared error             70.6811 %

=== Confusion Matrix ===

  a  b   <-- classified as
 27  4 |  a = 0
  4 25 |  b = 1
```

Appendix 46: Classification tree *Asellus*, Zwalm river basin, Subset 3, PCF=0.25

```
Width <= 250
|  Width <= 123
|  |  Banks = 0: 0 (44.0/5.0)
|  |  Banks = 1
|  |  |  T <= 15.1: 0 (8.0)
|  |  |  T > 15.1: 1 (4.0)
|  |  Banks = 2
|  |  |  Conductivity <= 738: 1 (2.0)
|  |  |  Conductivity > 738: 0 (3.0)
|  Width > 123
|  |  Distmouth <= 15778.284
|  |  |  Width <= 144: 1 (9.0)
|  |  |  Width > 144
|  |  |  |  Ammonium <= 0.23: 1 (3.0)
|  |  |  |  Ammonium > 0.23: 0 (7.0/1.0)
|  |  Distmouth > 15778.284: 0 (4.0)
Width > 250: 1 (35.0/2.0)

Number of Leaves  :   10
Size of the tree :       18

=== Evaluation on test set ===

Correctly Classified Instances          52               86.6667 %
Incorrectly Classified Instances         8               13.3333 %
Kappa statistic                          0.733
Mean absolute error                      0.1908
Root mean squared error                  0.3532
Relative absolute error                 38.2041 %
Root relative squared error             70.6811 %

=== Confusion Matrix ===

  a  b   <-- classified as
 27  4 |  a = 0
  4 25 |  b = 1
```

Appendix 47: Classification tree *Asellus*, Zwalm river basin, Subset 3, PCF=0.1

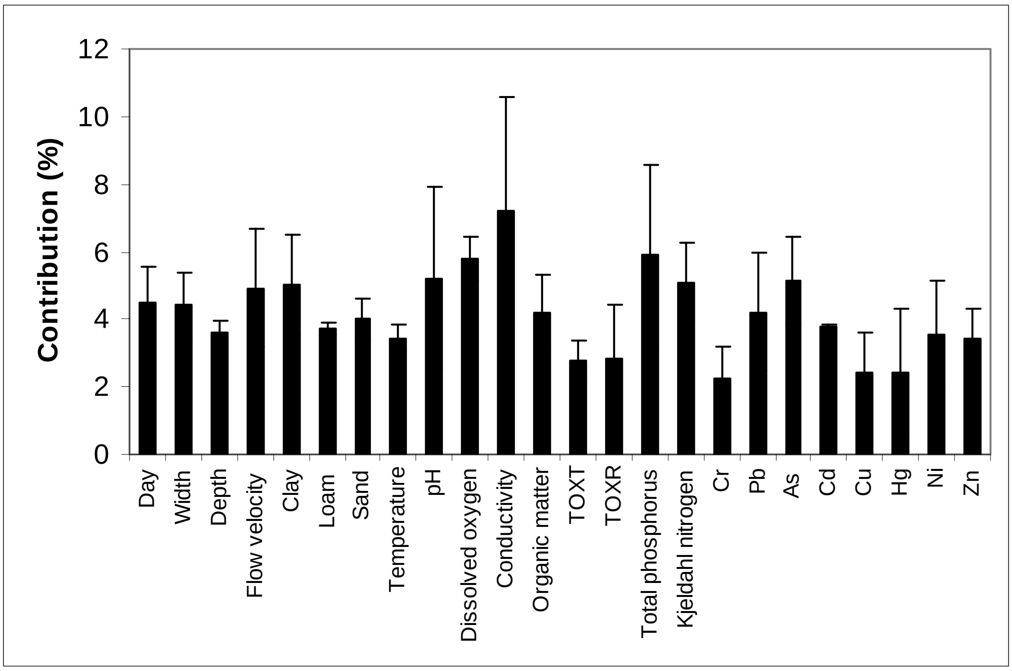
```
Width <= 250
|  Width <= 123
|  |  Banks = 0: 0 (44.0/5.0)
|  |  Banks = 1
|  |  |  T <= 15.1: 0 (8.0)
|  |  |  T > 15.1: 1 (4.0)
|  |  Banks = 2
|  |  |  Conductivity <= 738: 1 (2.0)
|  |  |  Conductivity > 738: 0 (3.0)
|  Width > 123
|  |  Distmouth <= 15778.284
|  |  |  Width <= 144: 1 (9.0)
|  |  |  Width > 144
|  |  |  |  Ammonium <= 0.23: 1 (3.0)
|  |  |  |  Ammonium > 0.23: 0 (7.0/1.0)
|  |  Distmouth > 15778.284: 0 (4.0)
Width > 250: 1 (35.0/2.0)

Number of Leaves  :   10
Size of the tree :       18

=== Evaluation on test set ===

Correctly Classified Instances          52               86.6667 %
Incorrectly Classified Instances         8               13.3333 %
Kappa statistic                          0.733
Mean absolute error                      0.1908
Root mean squared error                  0.3532
Relative absolute error                 38.2041 %
Root relative squared error             70.6811 %

=== Confusion Matrix ===

  a  b   <-- classified as
 27  4 |  a = 0
  4 25 |  b = 1
```

Appendix 48: Classification tree *Asellus*, Zwalm river basin, Subset 3, PCF=0.01

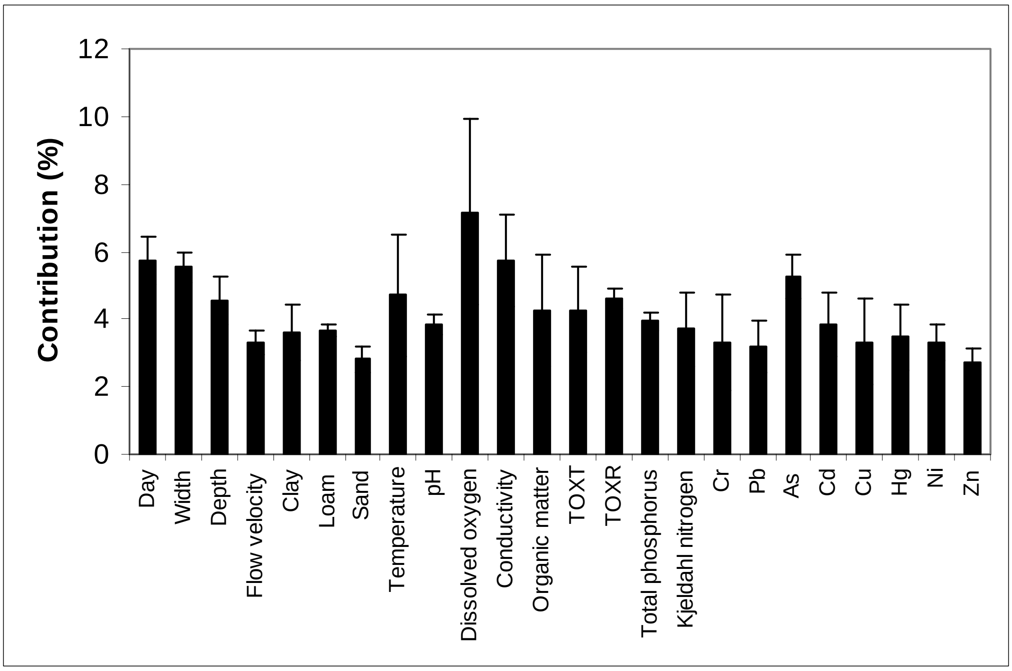
```
Width <= 250
|   Width <= 123: 0 (61.0/11.0)
|   Width > 123
|   |   Distmouth <= 15778.284
|   |   |   Width <= 144: 1 (9.0)
|   |   |   Width > 144
|   |   |   |   Ammonium <= 0.23: 1 (3.0)
|   |   |   |   Ammonium > 0.23: 0 (7.0/1.0)
|   |   Distmouth > 15778.284: 0 (4.0)
Width > 250: 1 (35.0/2.0)

Number of Leaves  :   6
Size of the tree :       11

=== Evaluation on test set ===

Correctly Classified Instances          52               86.6667 %
Incorrectly Classified Instances         8               13.3333 %
Kappa statistic                          0.7324
Mean absolute error                      0.2168
Root mean squared error                  0.3517
Relative absolute error                 43.415  %
Root relative squared error             70.3686 %

=== Confusion Matrix ===

  a  b   <-- classified as
 28  3 |  a = 0
  5 24 |  b = 1
```

Appendix 49: ANN, P/A, *Gammarus* and *Asellus*, Sediments Flanders, Weights

Gammarus

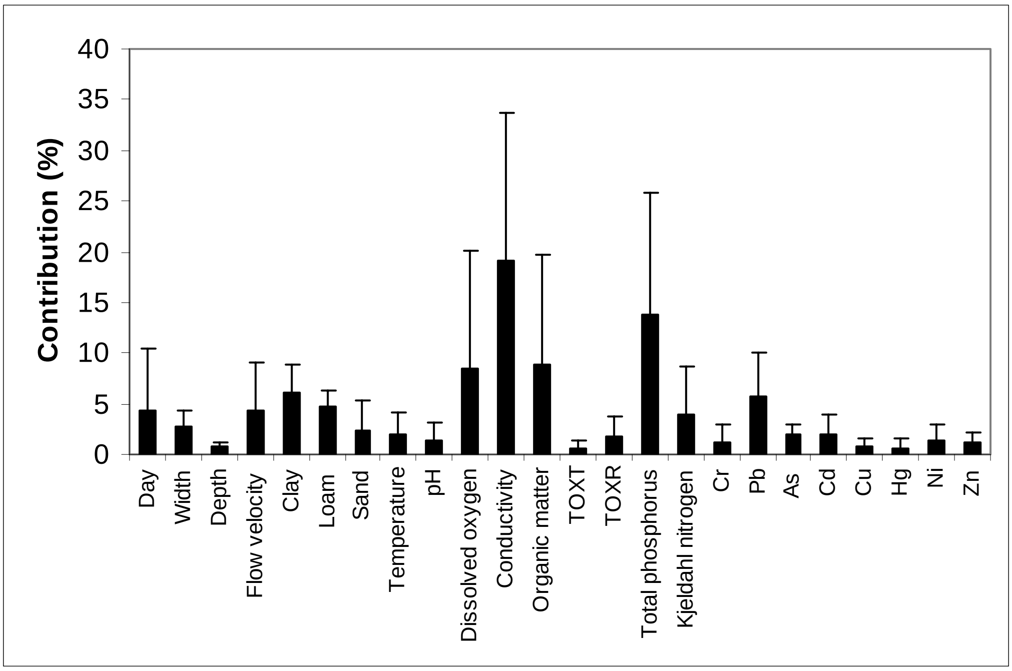


Asellus

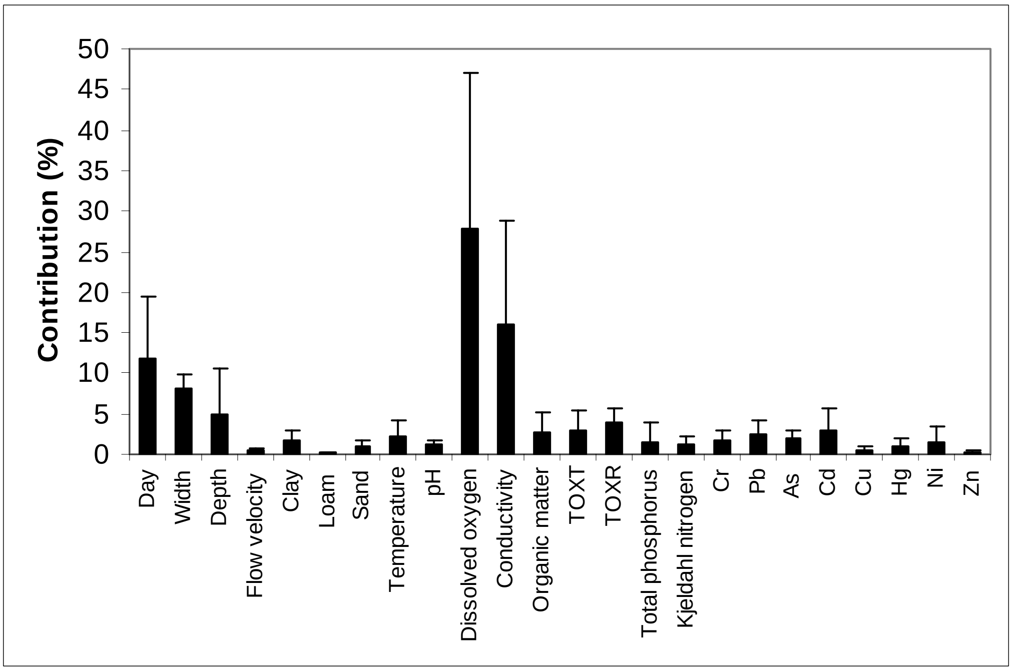


Appendix 50: ANN, P/A, *Gammarus* and *Asellus*, Sediments Flanders, PaD

Gammarus

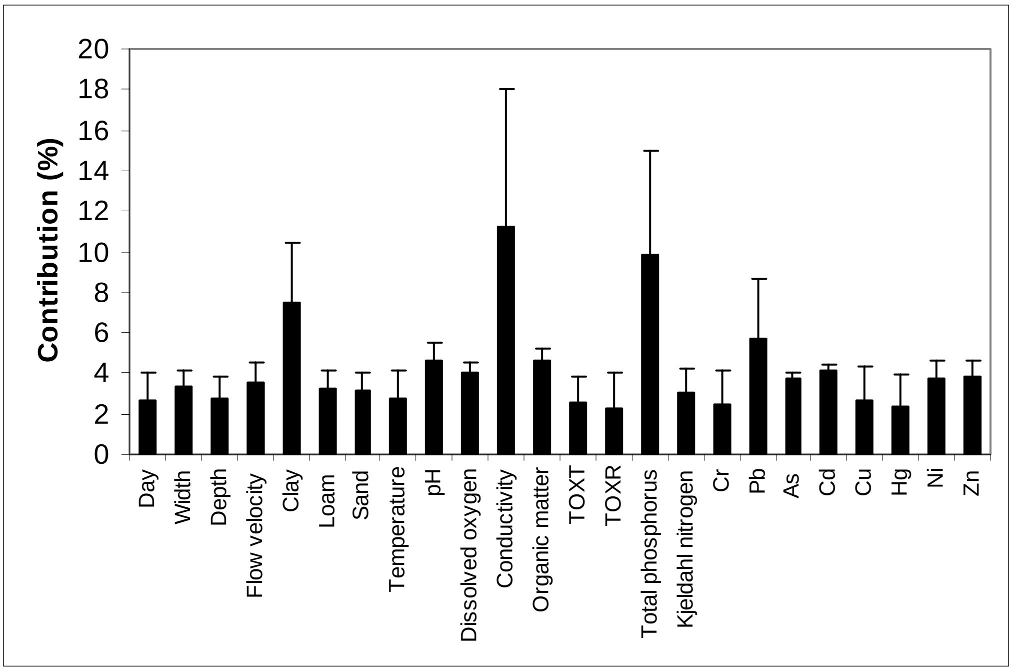


Asellus

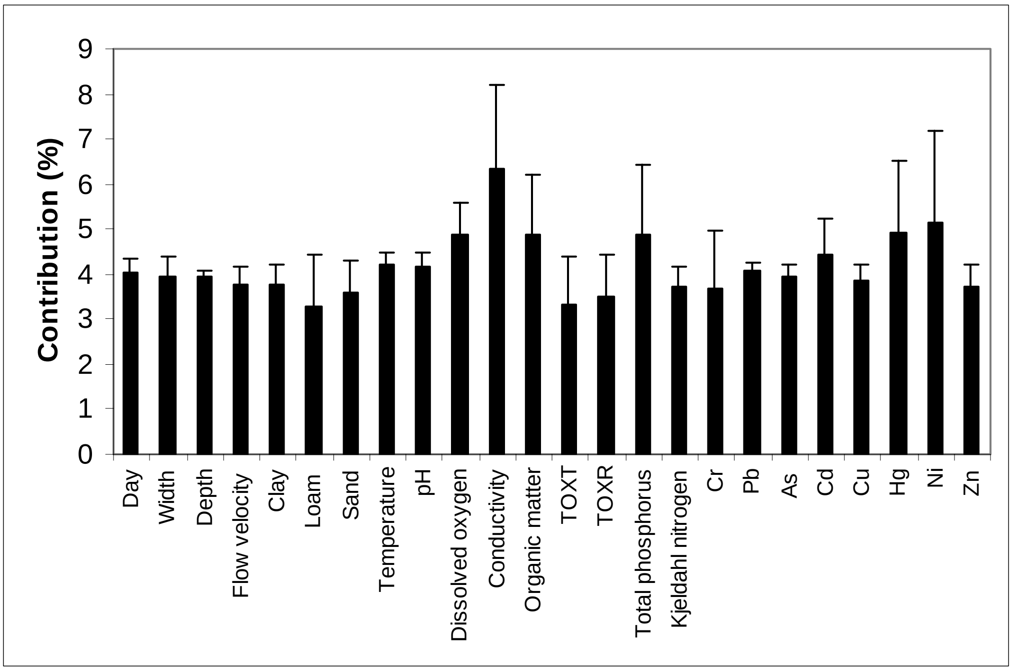


Appendix 51: ANN, P/A, *Gammarus* and *Asellus*, Sediments Flanders, Perturb

Gammarus

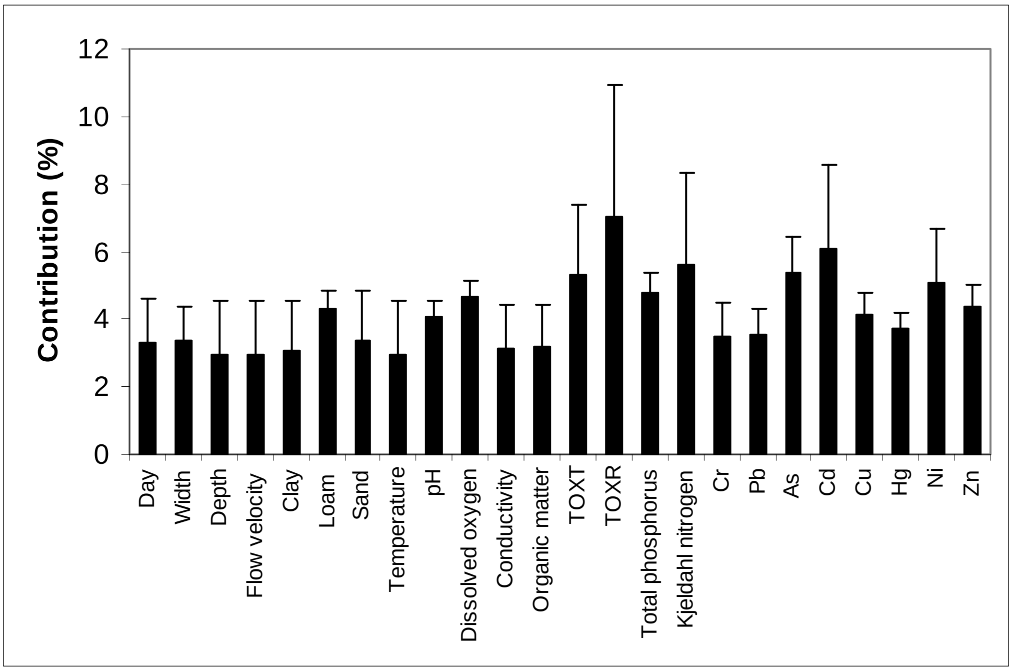


Asellus

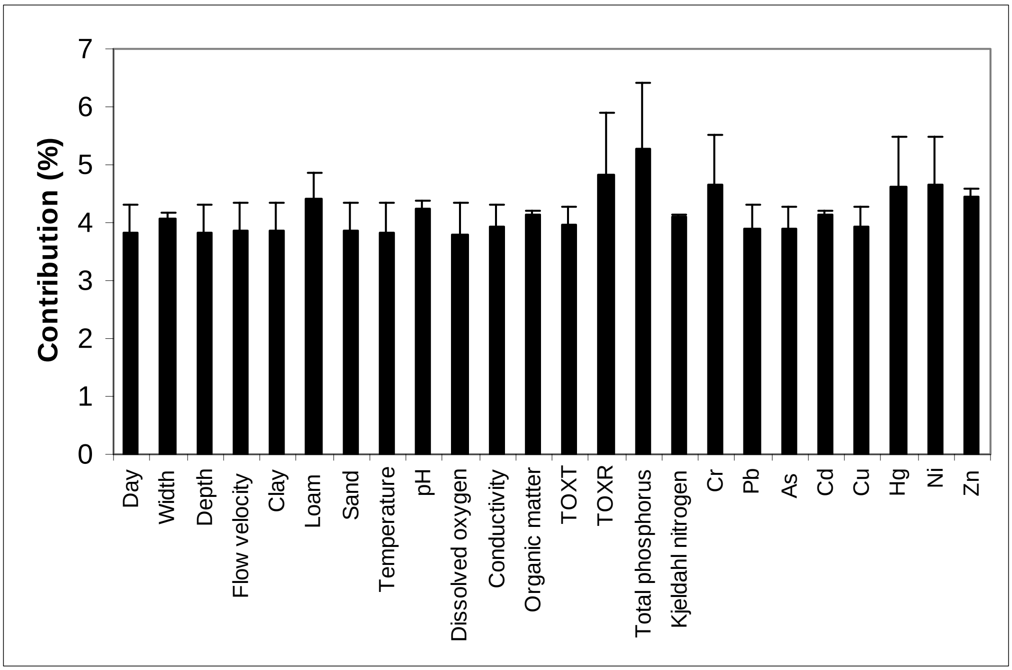


Appendix 52: ANN, P/A, *Gammarus* and *Asellus*, Sediments Flanders, Stepwise Reg

Gammarus

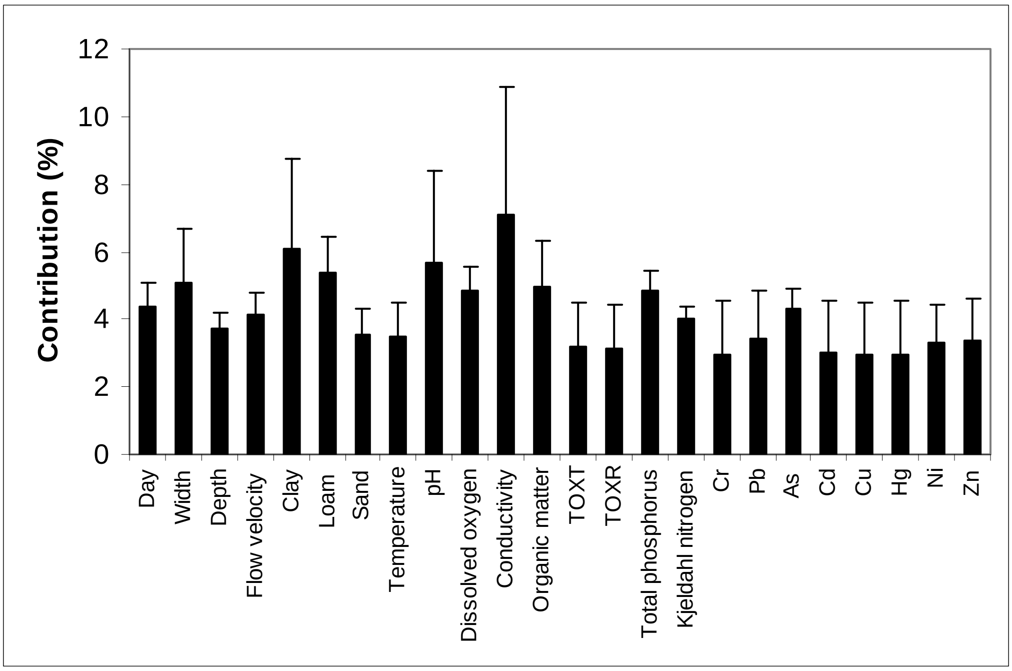


Asellus

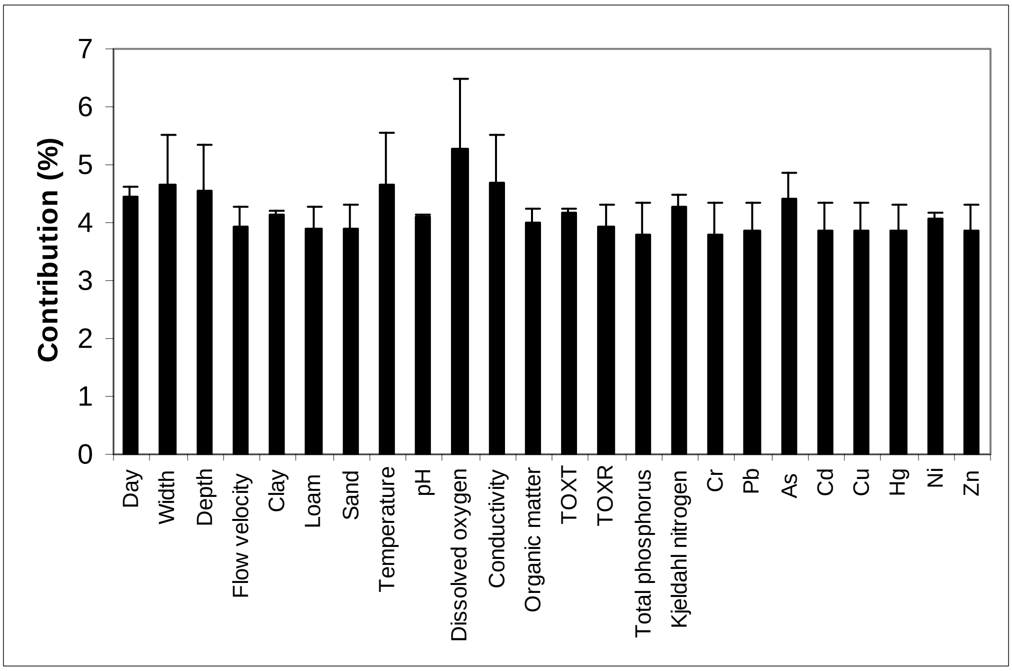


Appendix 53: ANN, P/A, *Gammarus* and *Asellus*, Sediments Flanders, Stepwise Imp

Gammarus

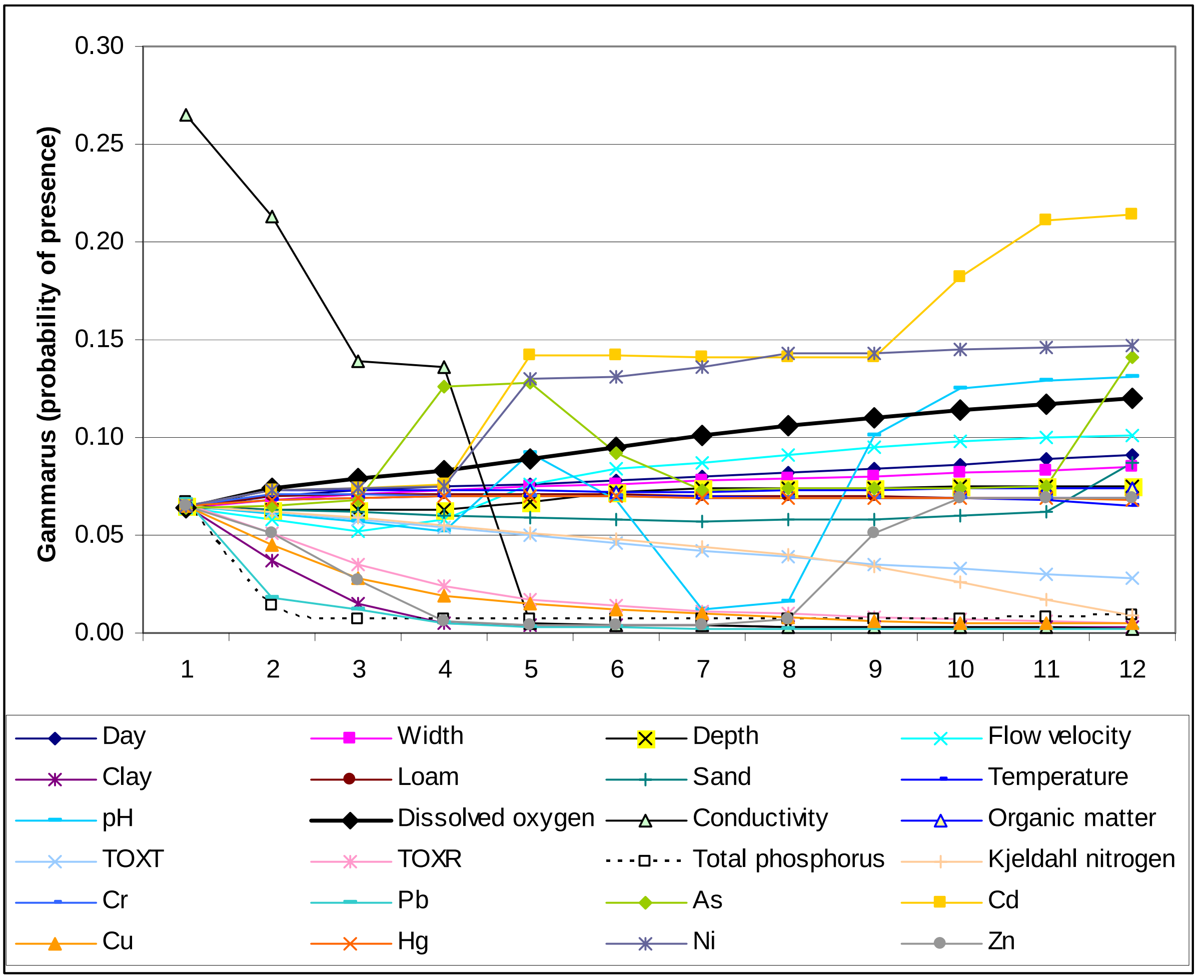


Asellus

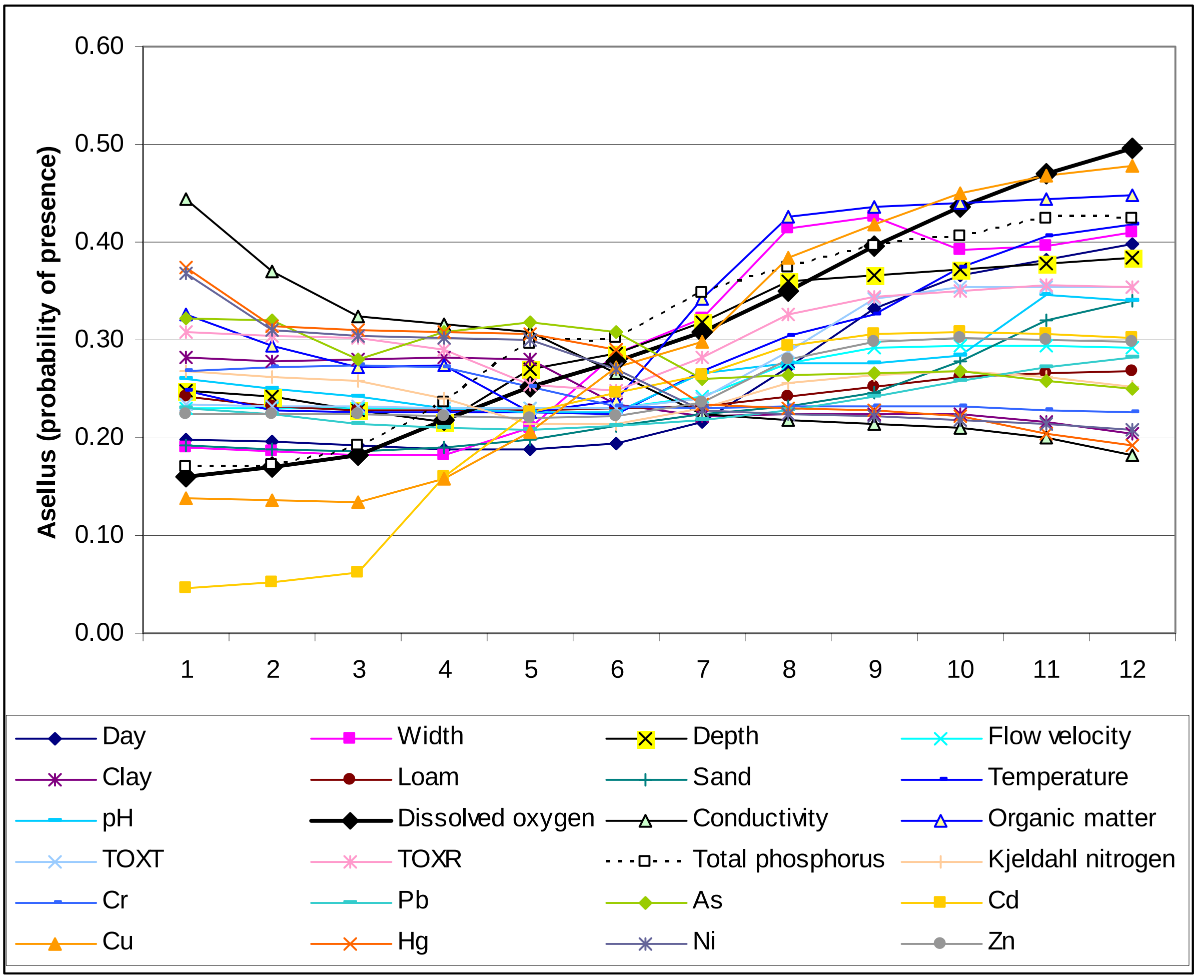


Appendix 54: ANN, P/A, *Gammarus* and *Asellus*, Sediments Flanders, Profile

Gammarus

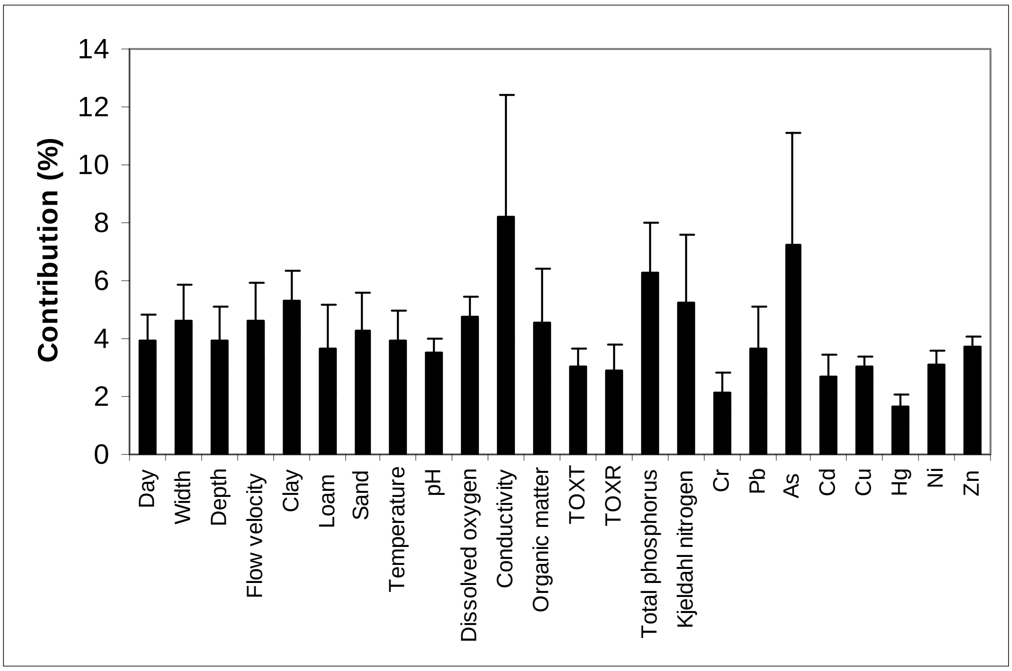


Asellus

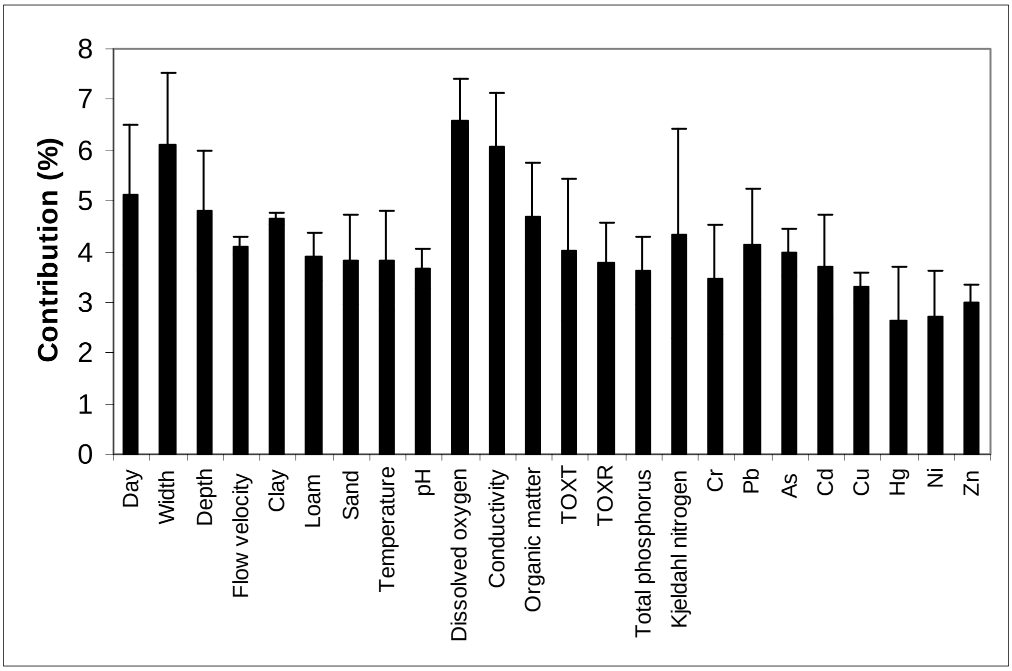

Asellus (probability of presence)
0.00
0.10
0.20
0.30
0.40
0.50
0.60
1
2
3
4
5
6
7
8
9
10
11
12
Day
Width
Depth
Flow velocity
Clay
Loam
Sand
Temperature
pH
Dissolved oxygen
Conductivity
Organic matter
TOXT
TOXR
Total phosphorus
Kjeldahl nitrogen
Cr
Pb
As
Cd
Cu
Hg
Ni
Zn

Appendix 55: ANN, ABUN, *Gammarus* and *Asellus*, Sediments Flanders, Weights

Gammarus

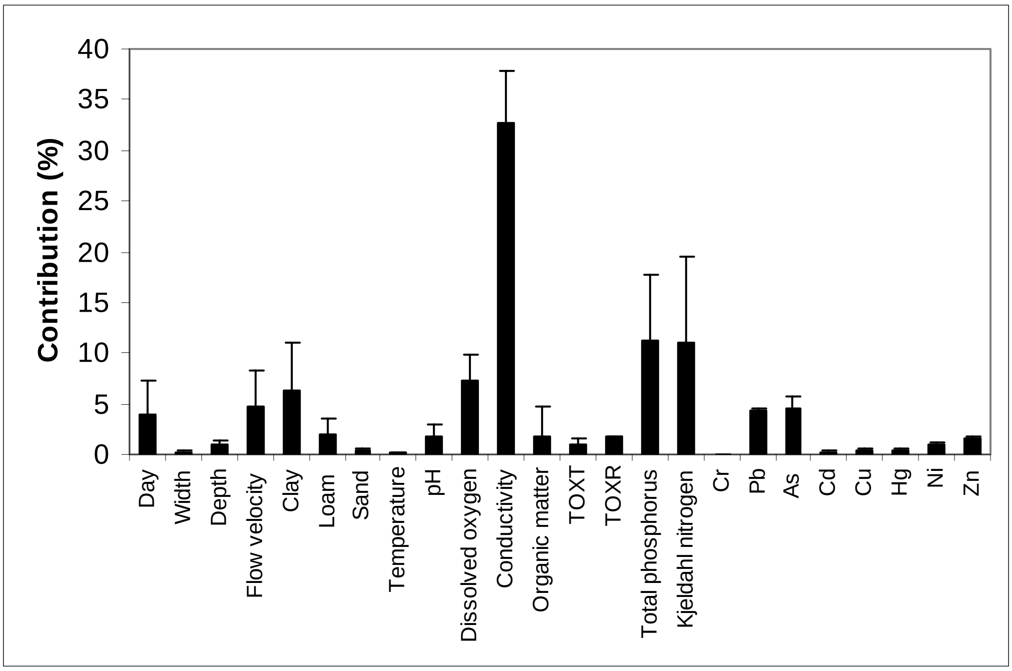


Asellus

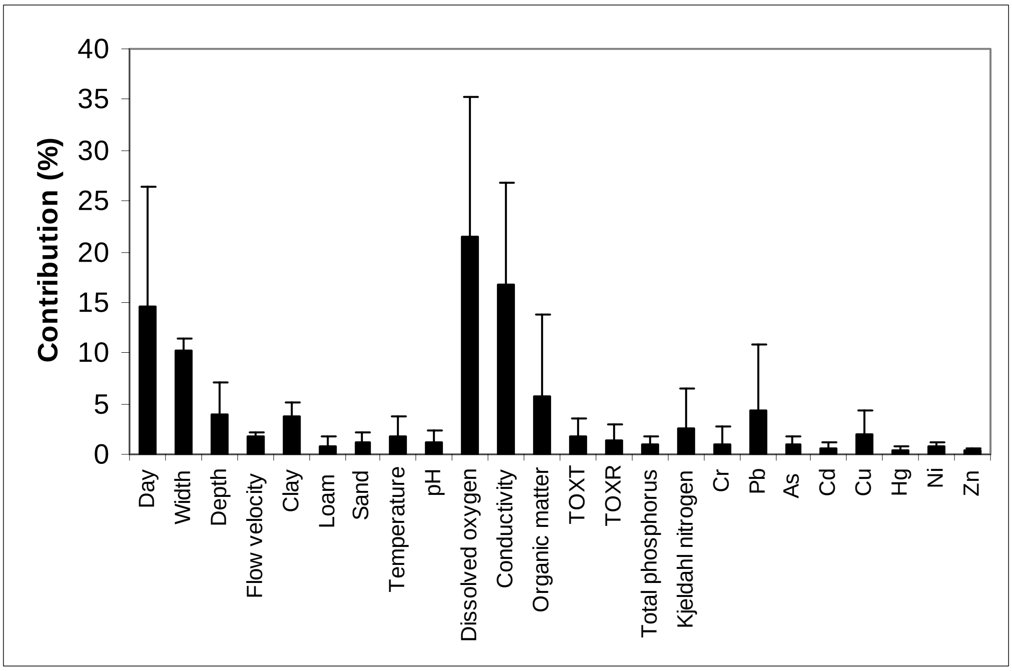


Appendix 56: ANN, ABUN, *Gammarus* and *Asellus*, Sediments Flanders, PaD

Gammarus

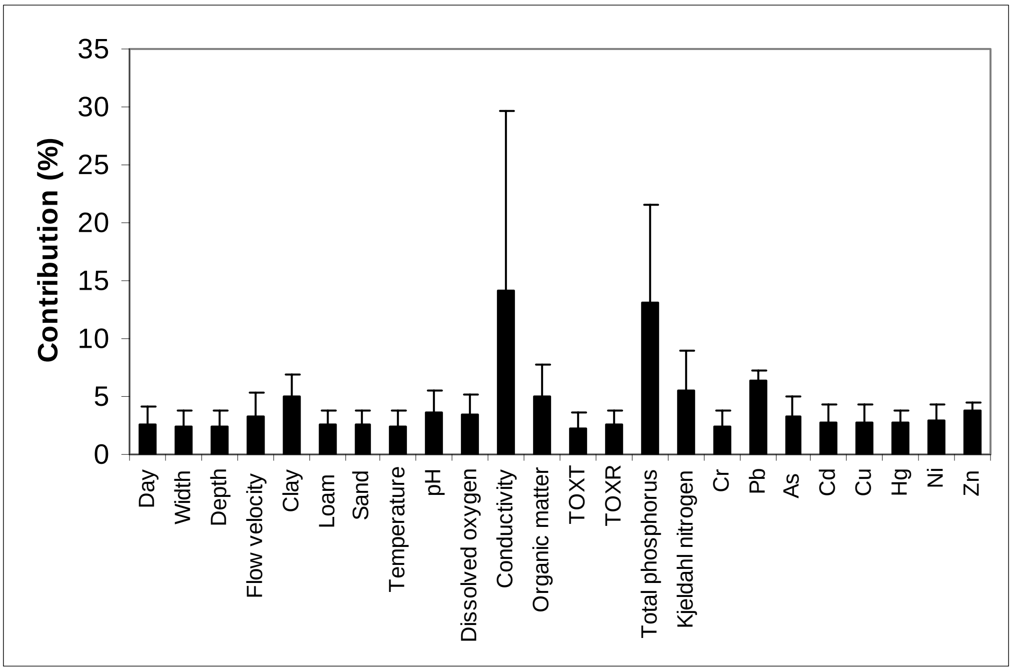


Asellus

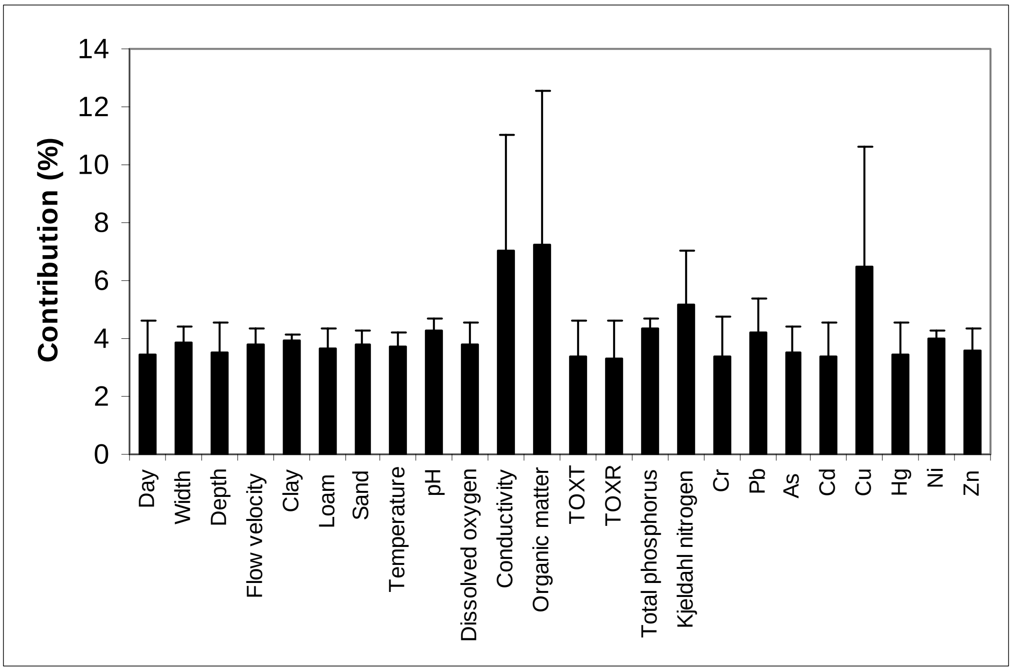


Appendix 57: ANN, ABUN, *Gammarus* and *Asellus*, Sediments Flanders, Perturb

Gammarus

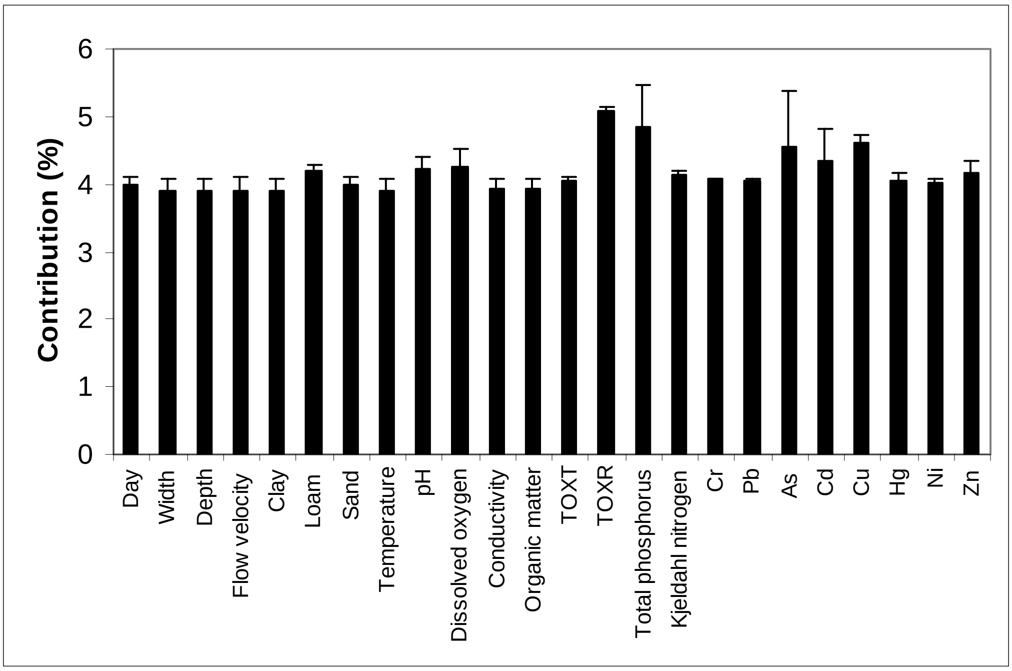


Asellus

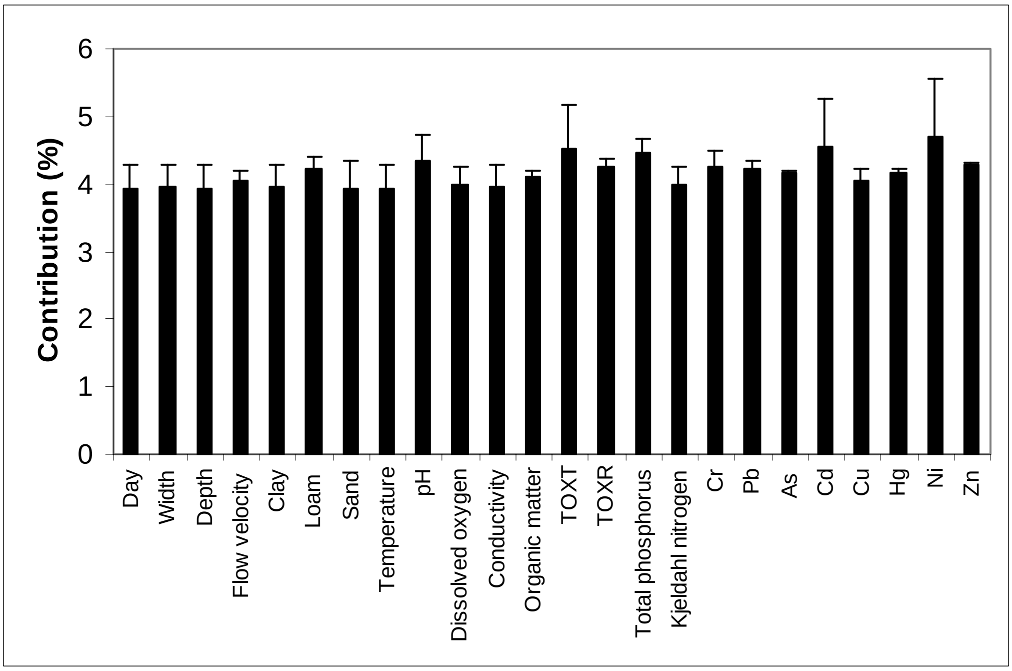


Appendix 58: ANN, ABUN, *Gammarus* and *Asellus*, Sediments Flanders, Stepwise Reg

Gammarus

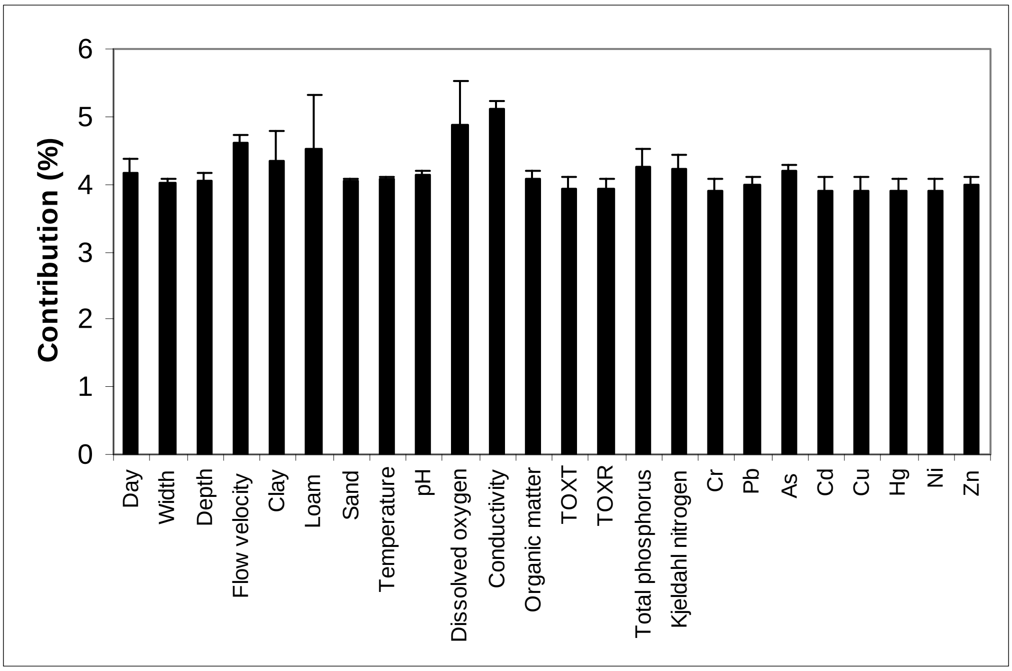


Asellus

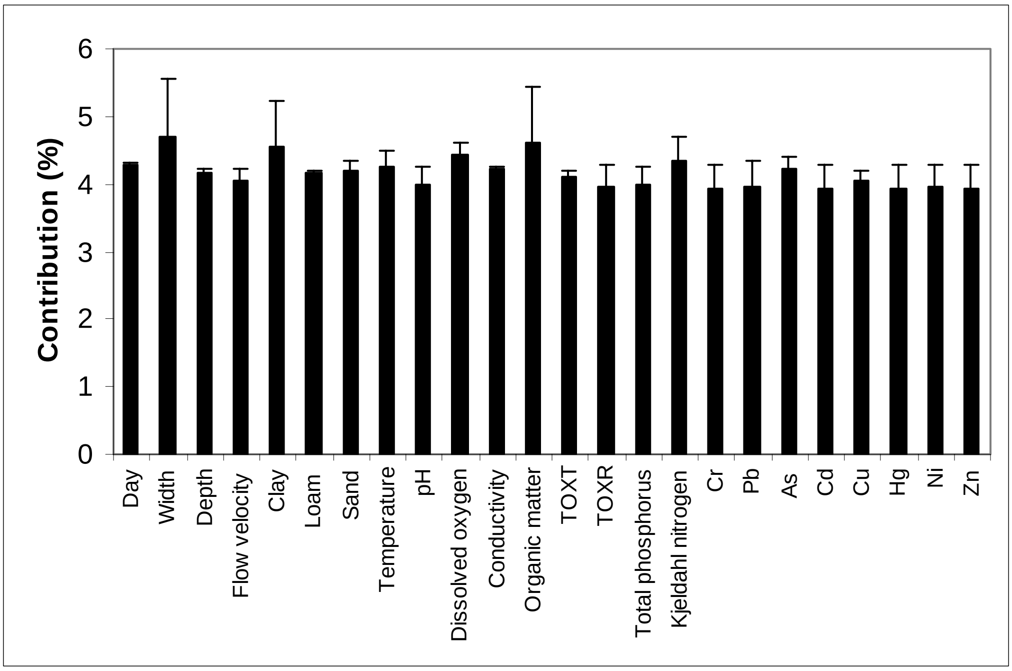


Appendix 59: ANN, ABUN, *Gammarus* and *Asellus*, Sediments Flanders, Stepwise Imp

Gammarus

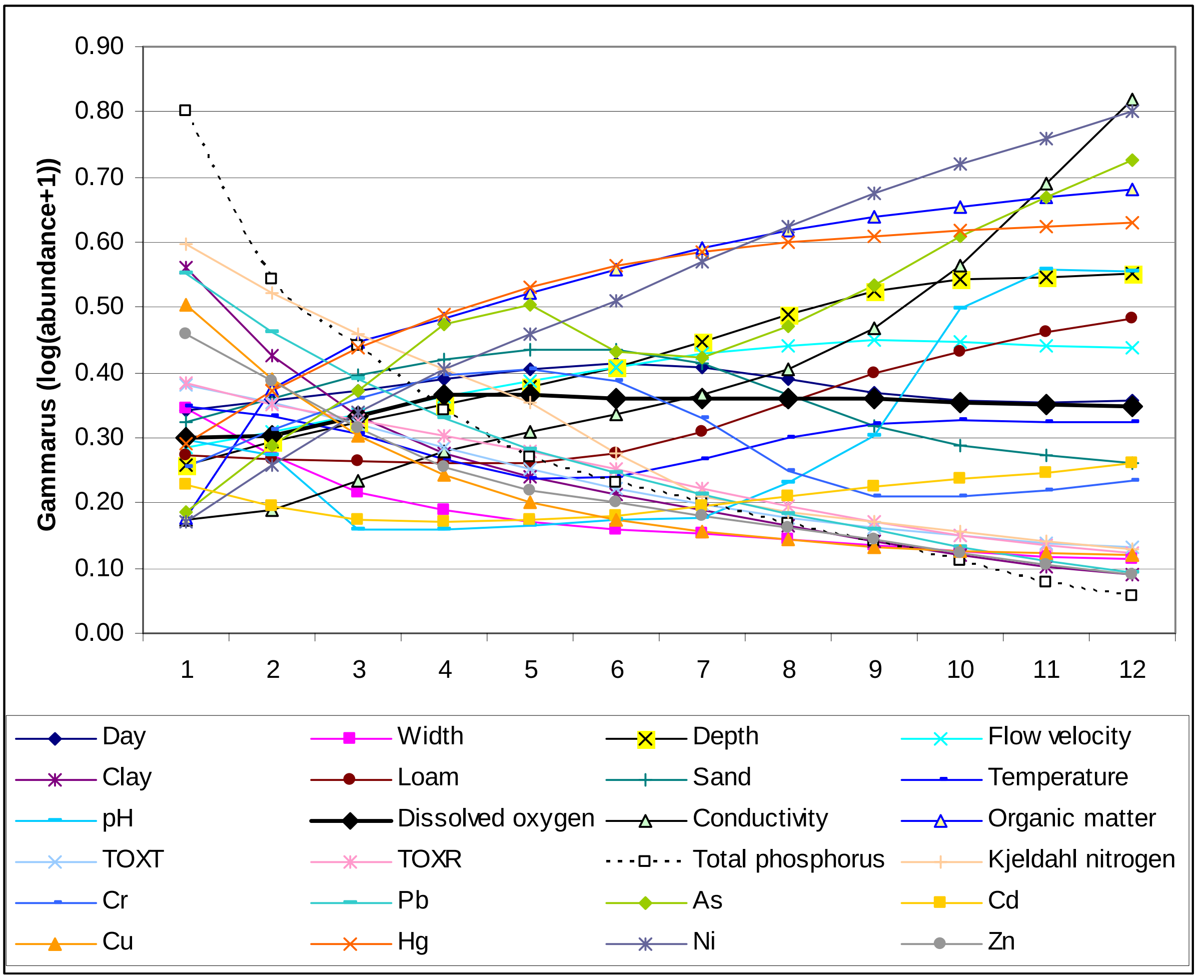


Asellus

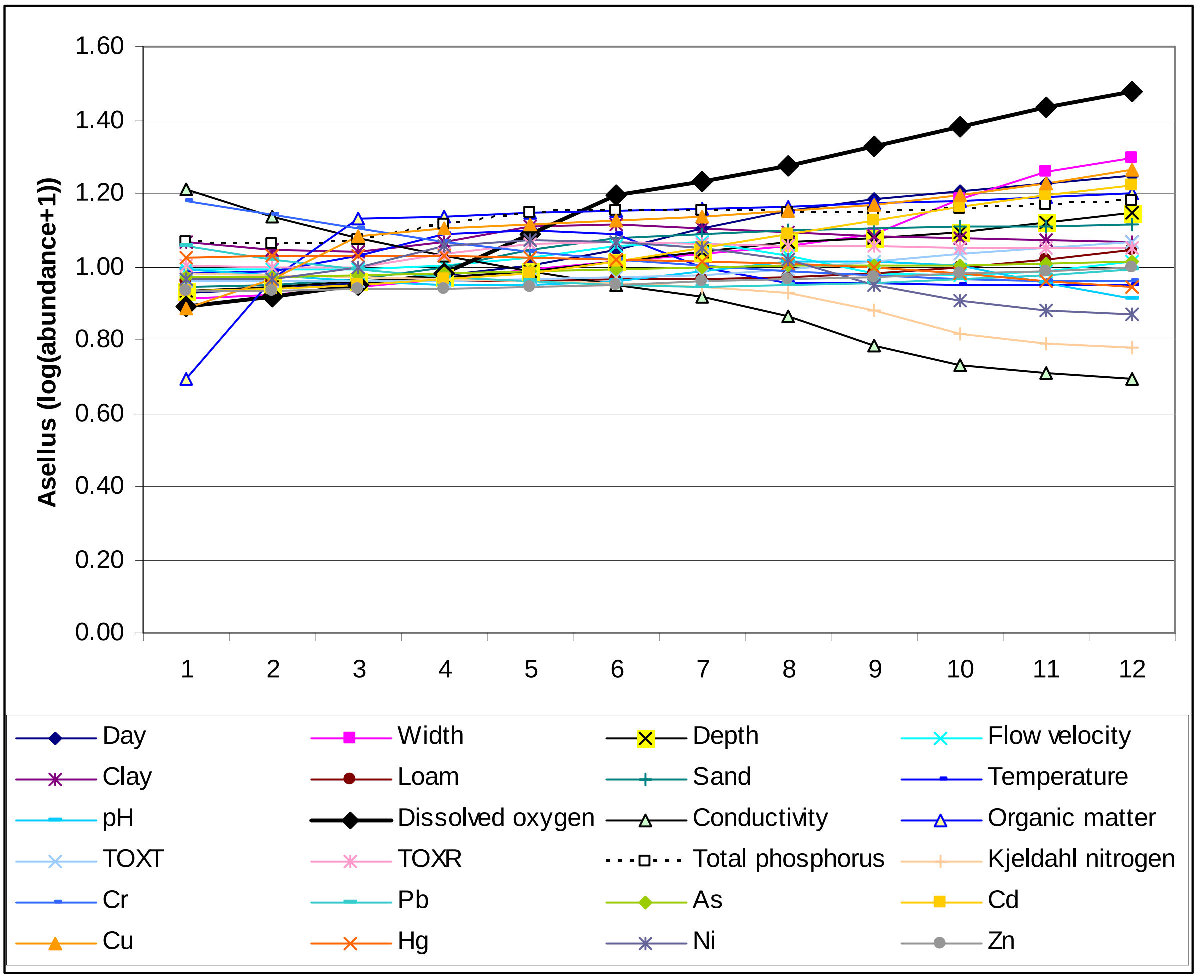


Appendix 60: ANN, ABUN, *Gammarus* and *Asellus*, Sediments Flanders, Profile

Gammarus

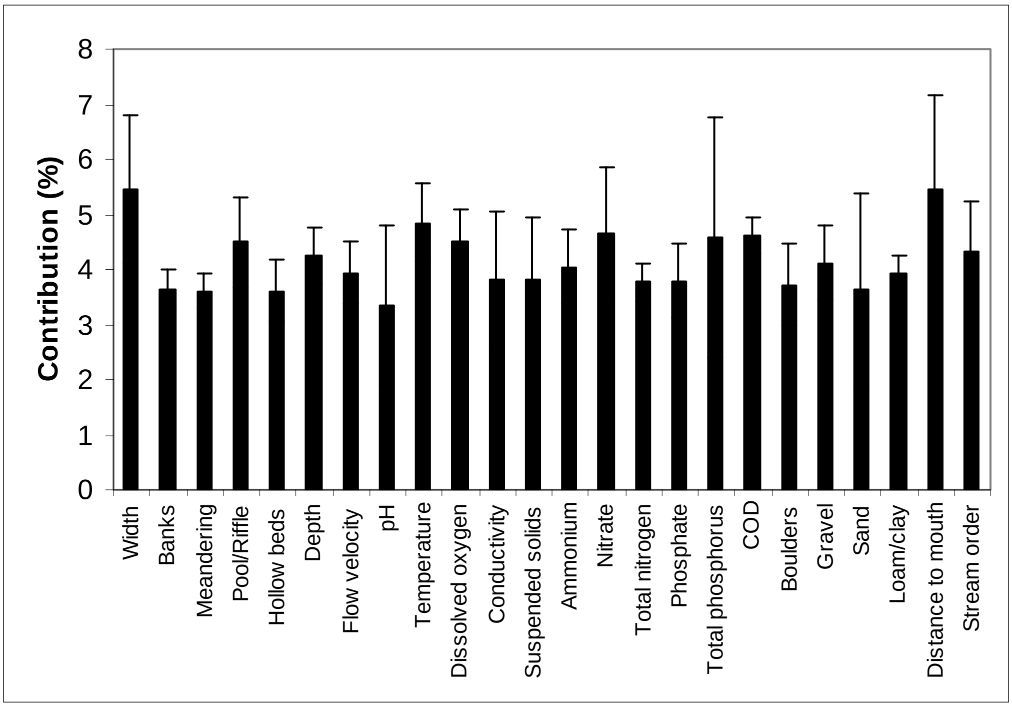


Asellus

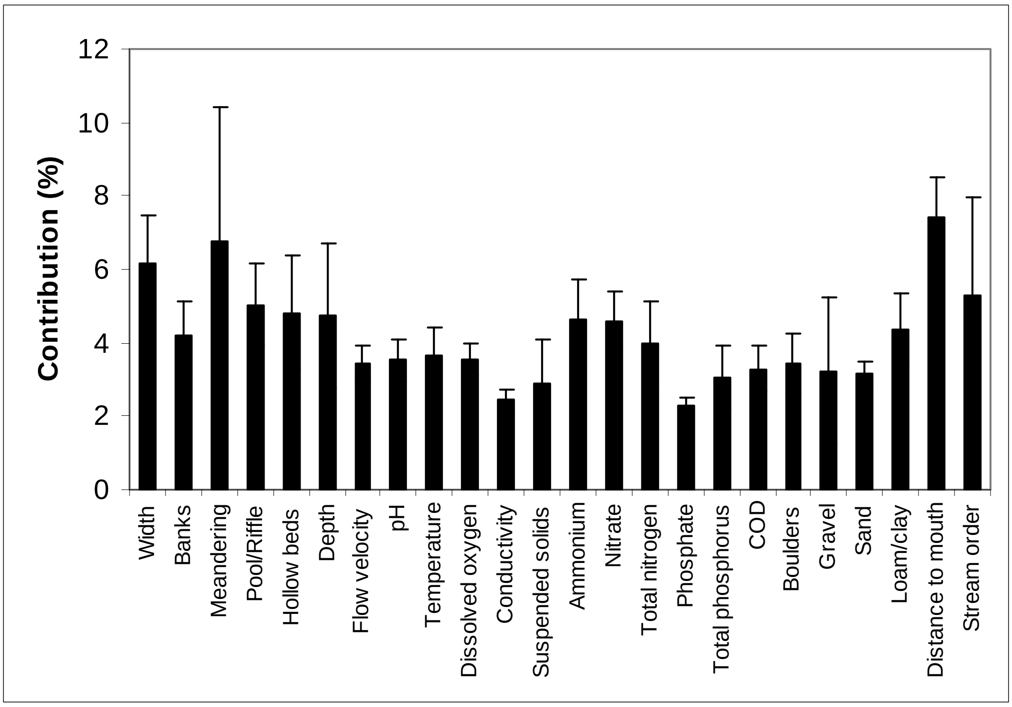

Asellus (log(abundance+1))
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
1
2
3
4
5
6
7
8
9
10
11
12
Day
Width
Depth
Flow velocity
Clay
Loam
Sand
Temperature
pH
Dissolved oxygen
Conductivity
Organic matter
TOXT
TOXR
Total phosphorus
Kjeldahl nitrogen
Cr
Pb
As
Cd
Cu
Hg
Ni
Zn

Appendix 61: ANN, P/A, *Gammarus* and *Asellus*, Zwalm river basin, Weights

Gammarus

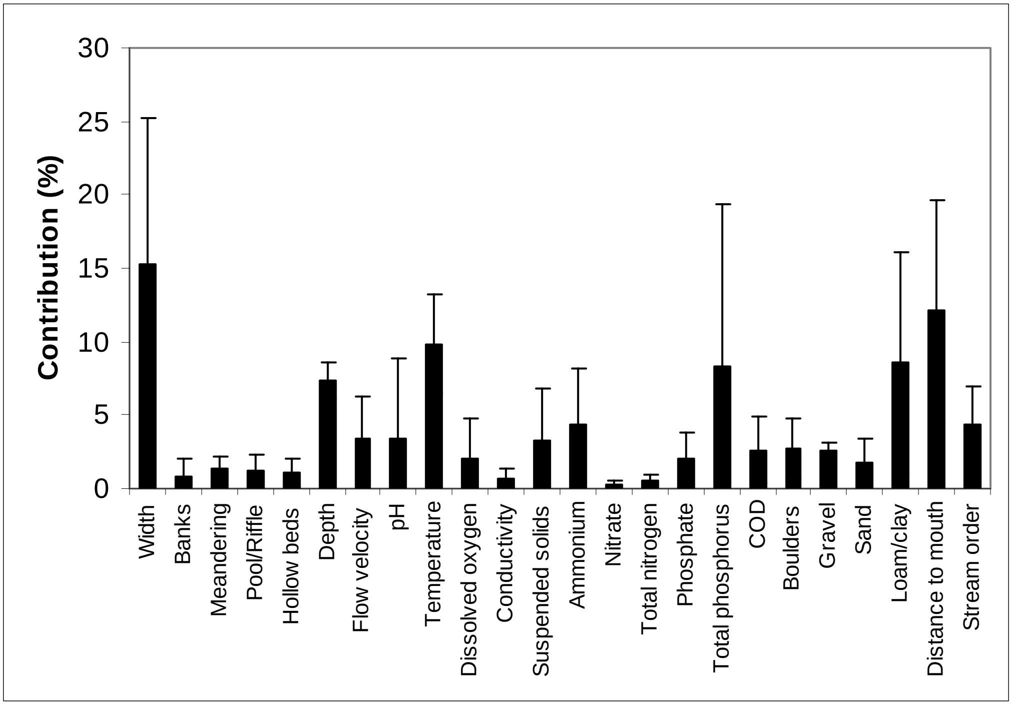


Asellus

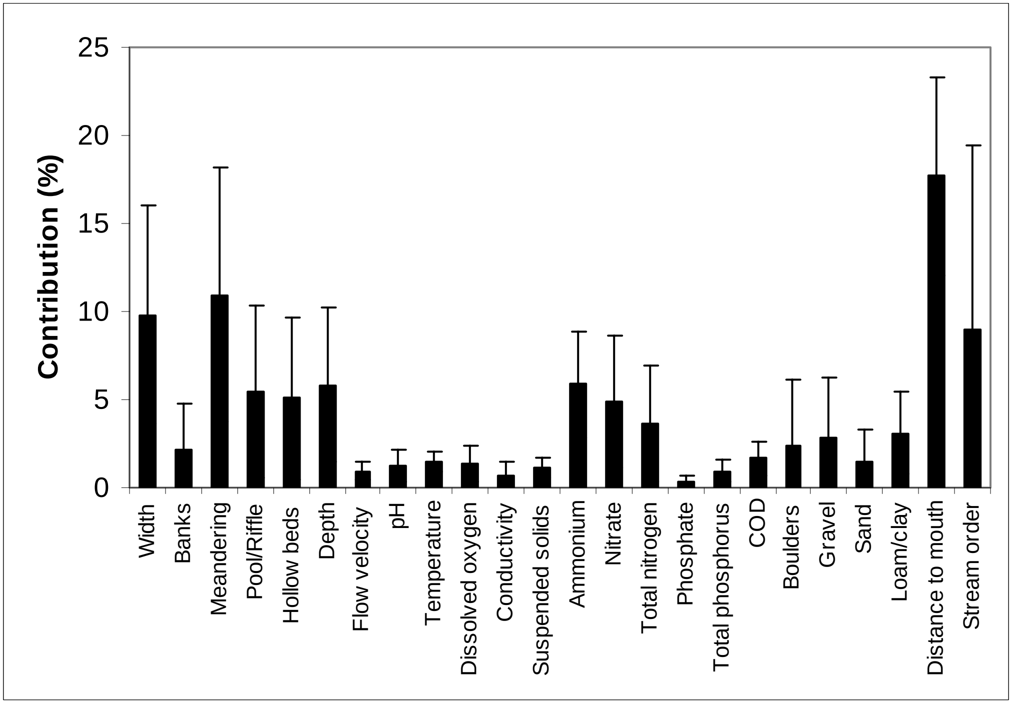


Appendix 62: ANN, P/A, *Gammarus* and *Asellus*, Zwalm river basin, PaD

Gammarus

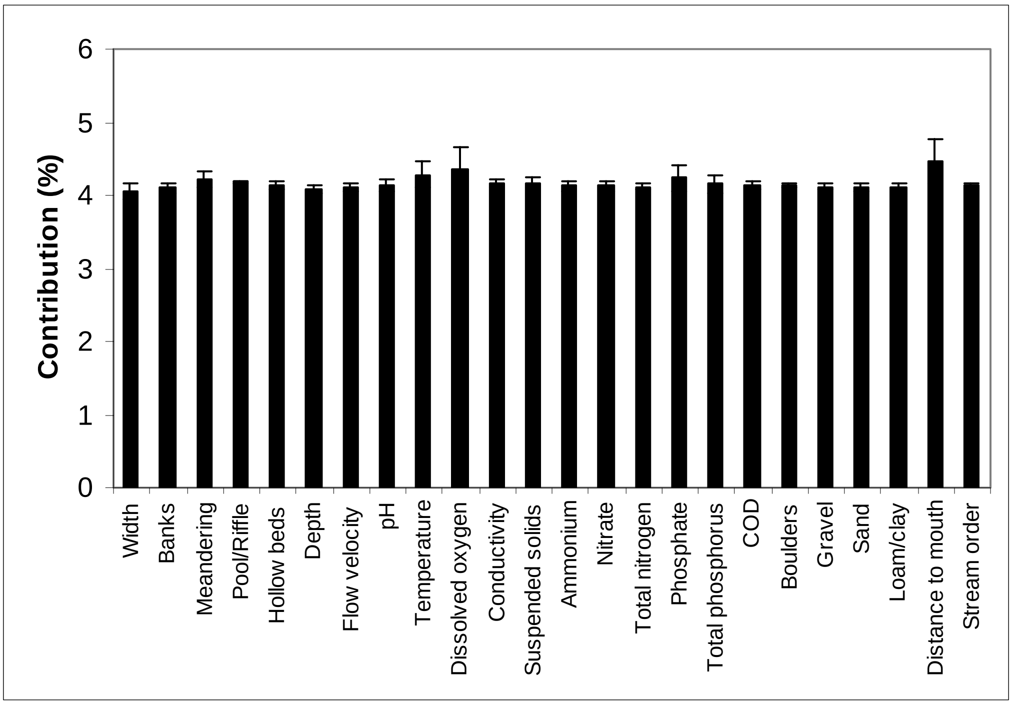


Asellus

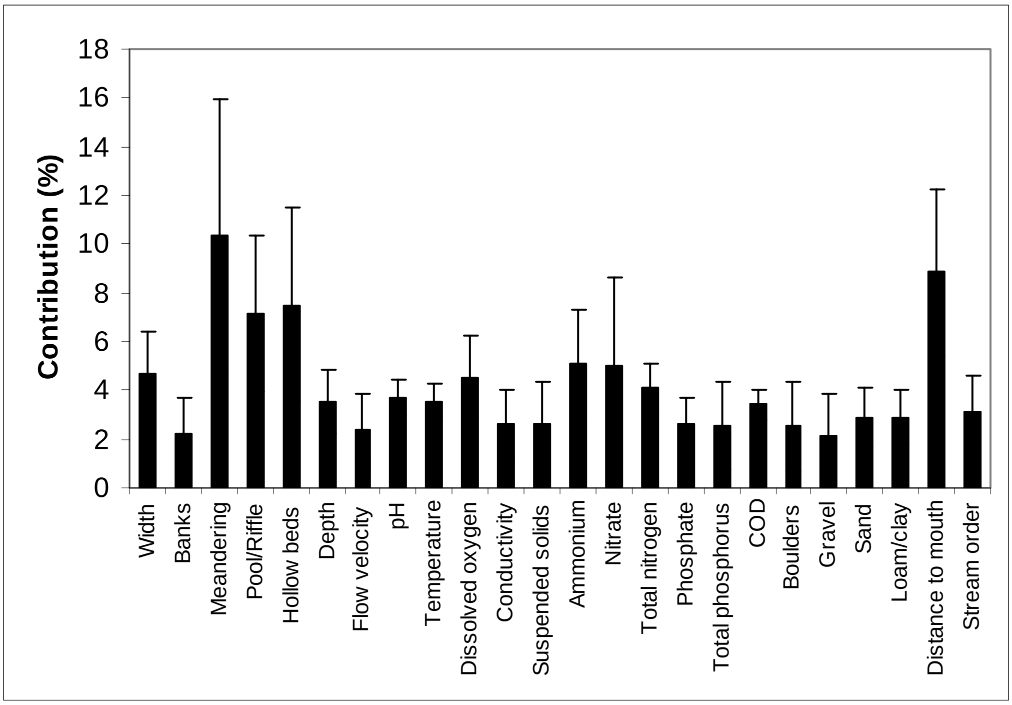


Appendix 63: ANN, P/A, *Gammarus* and *Asellus*, Zwalm river basin, Perturb

Gammarus

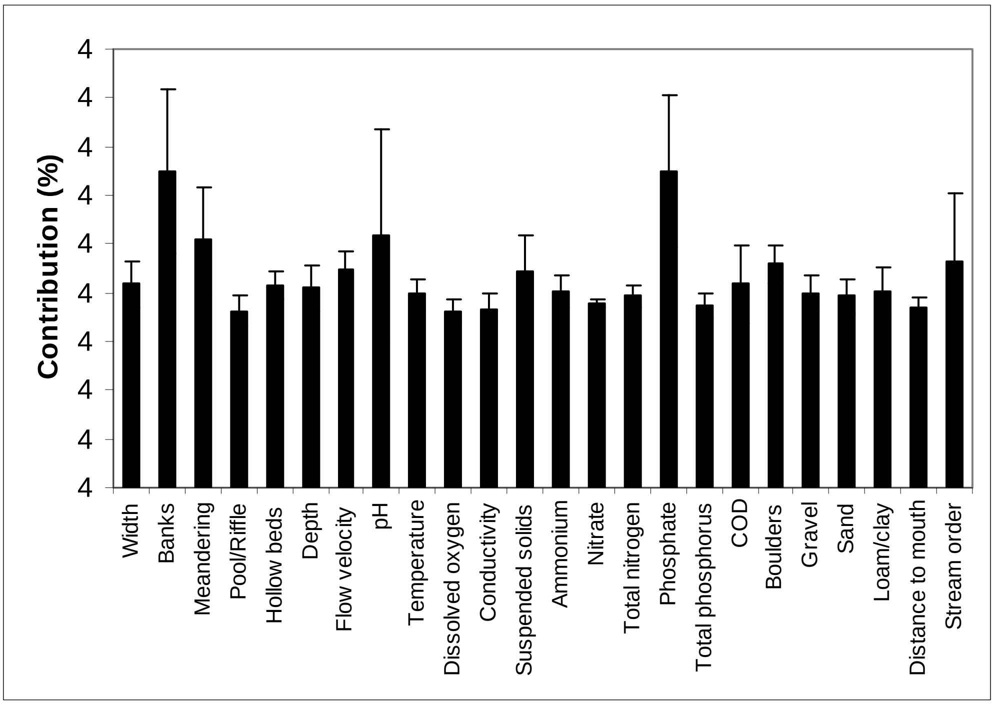


Asellus

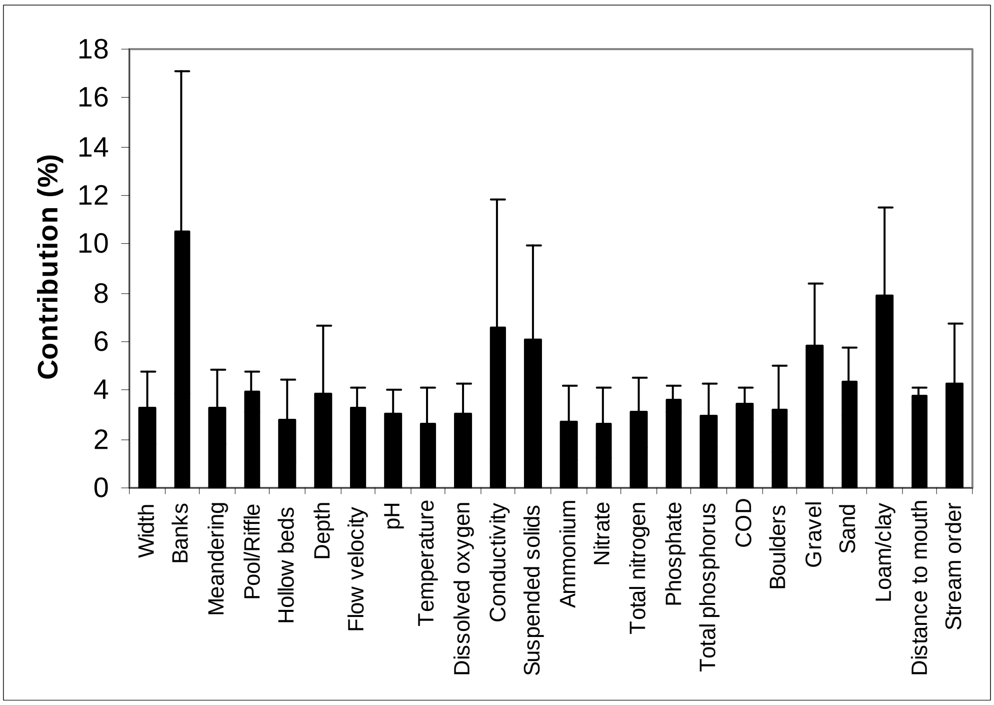


Appendix 64: ANN, P/A, *Gammarus* and *Asellus*, Zwalm river basin, Stepwise Reg

Gammarus

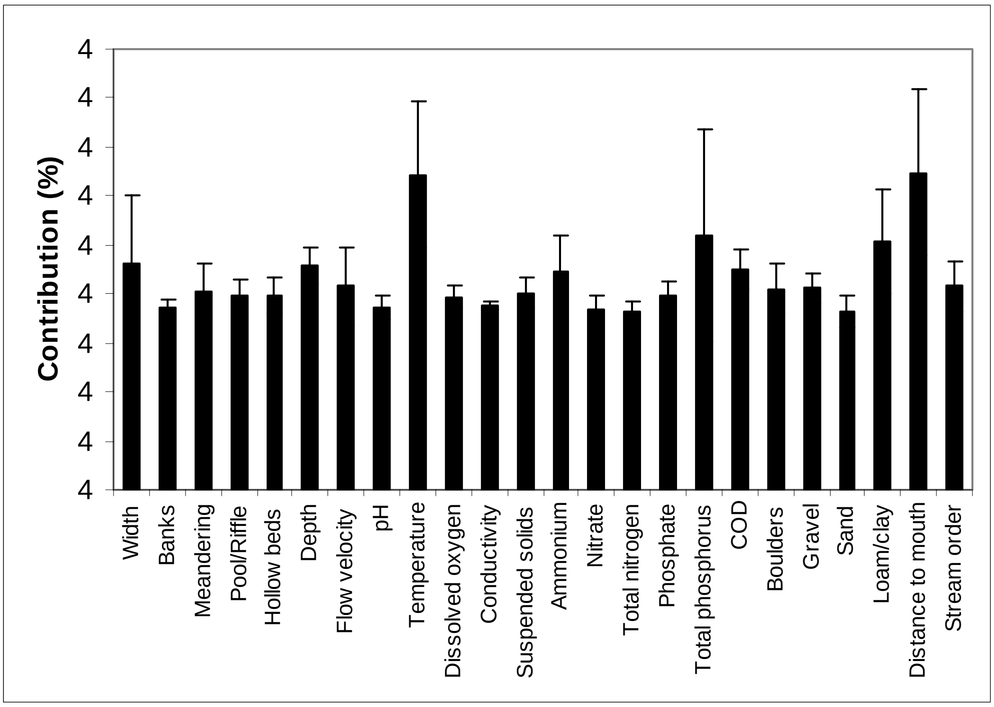


Asellus

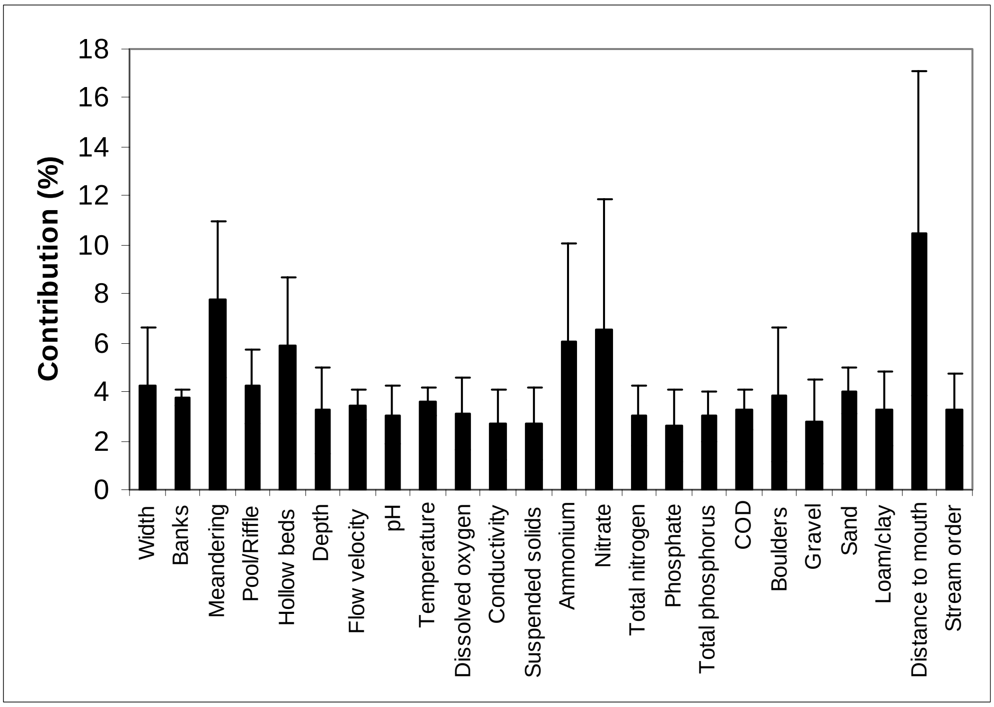


Appendix 65: ANN, P/A, *Gammarus* and *Asellus*, Zwalm river basin, Stepwise Imp

Gammarus

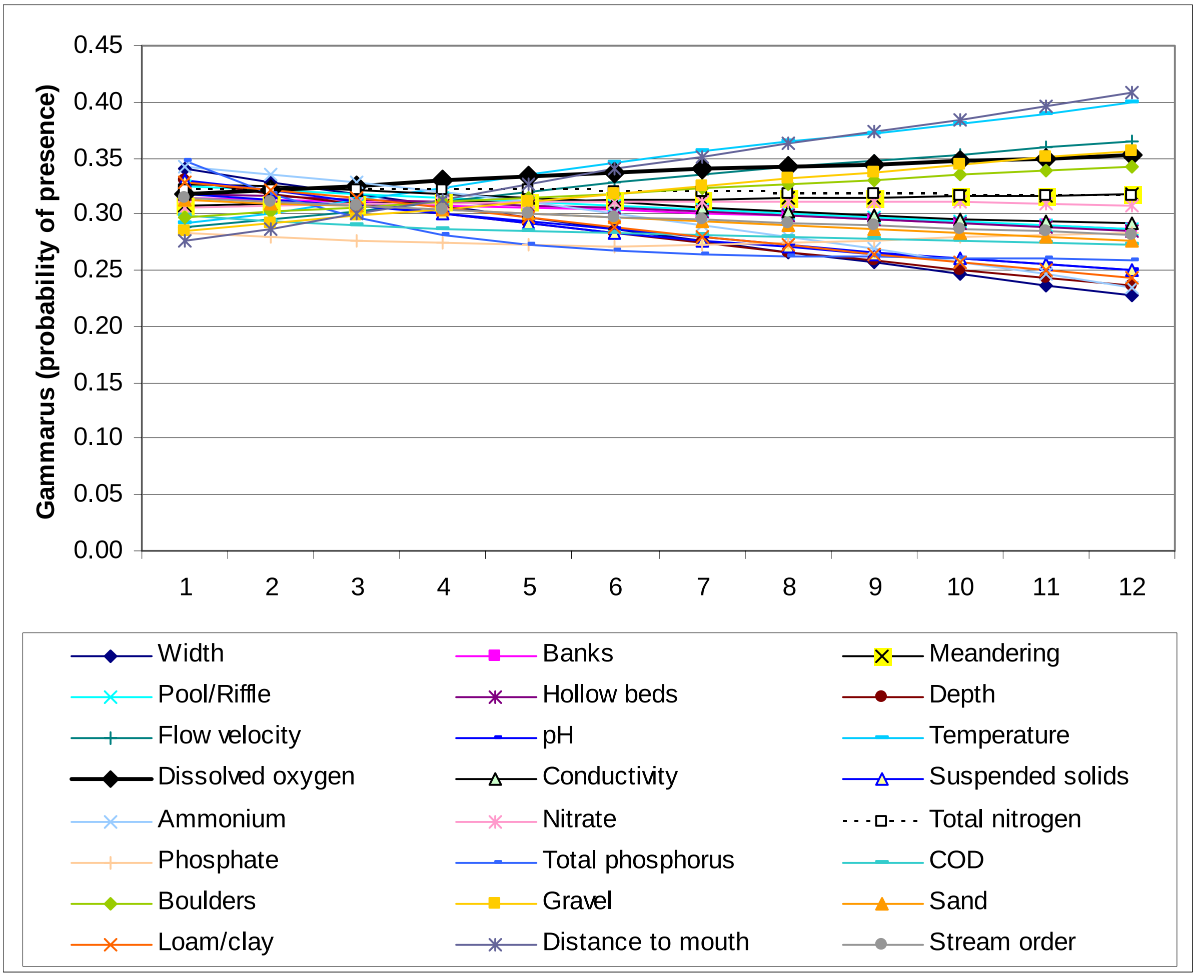


Asellus

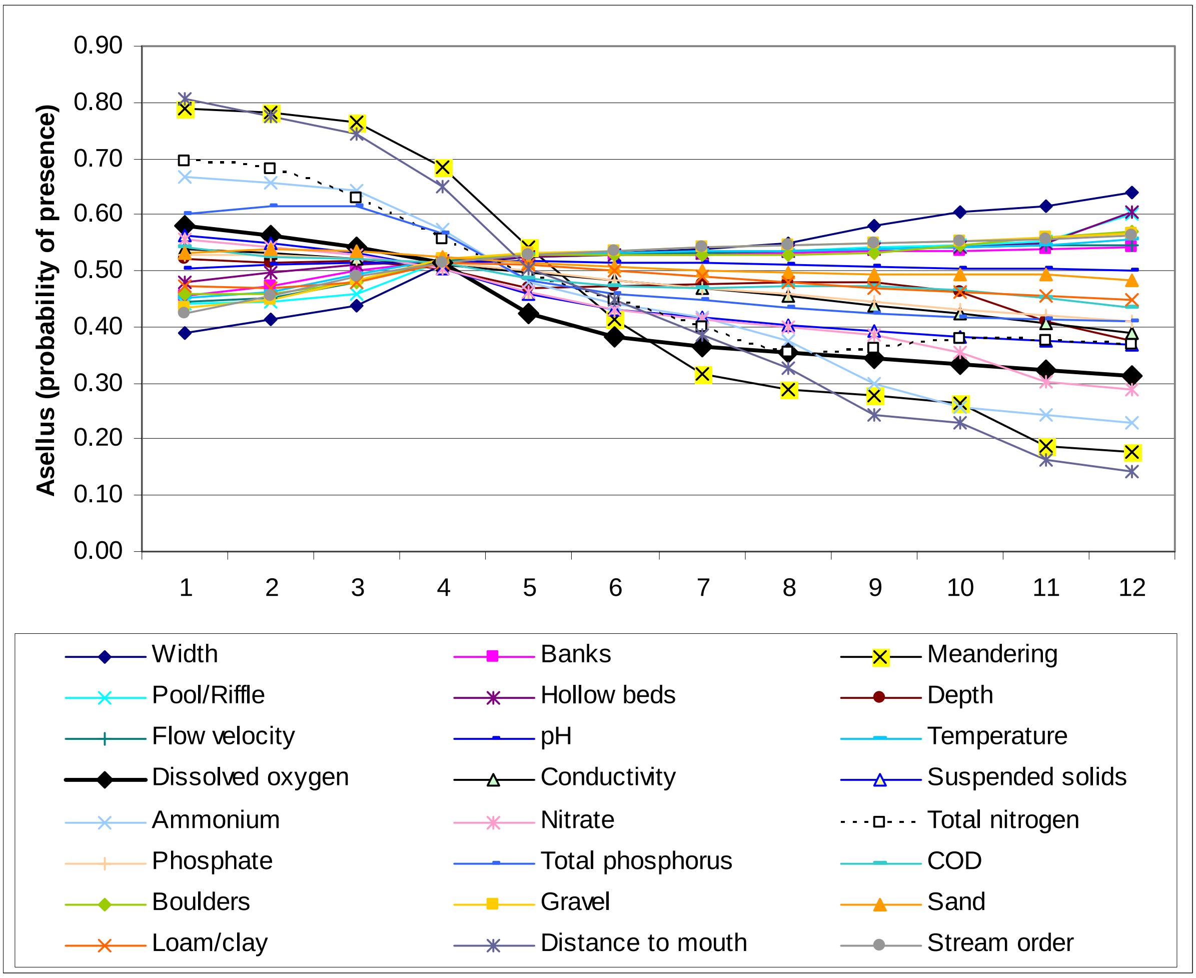


Appendix 66: ANN, P/A, *Gammarus* and *Asellus*, Zwalm river basin, Profile

Gammarus

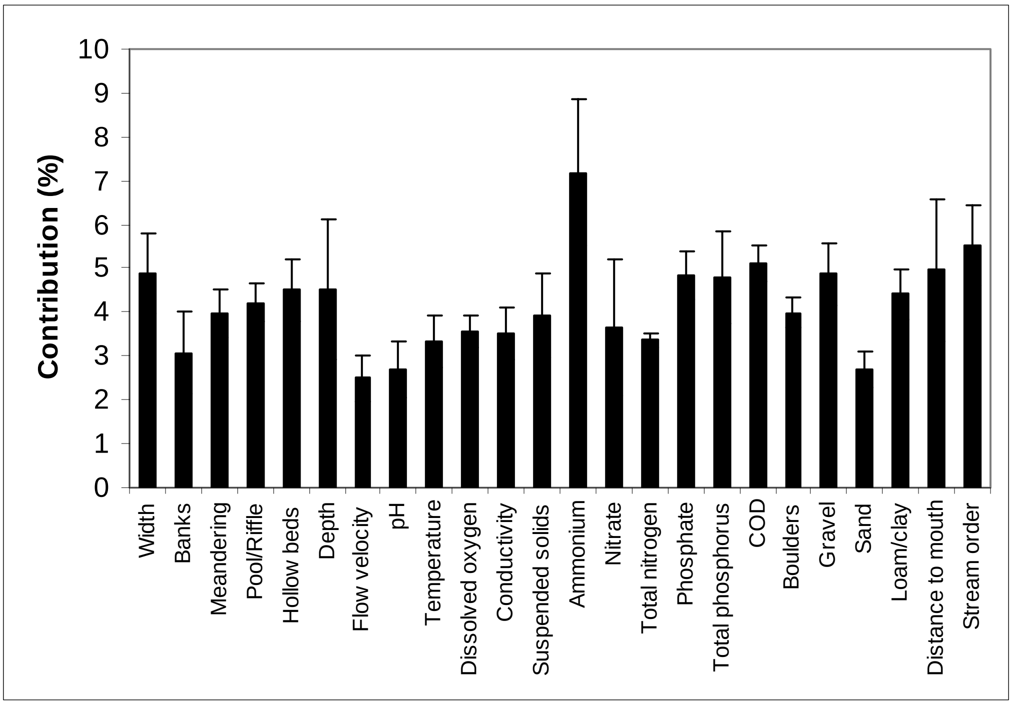


Asellus

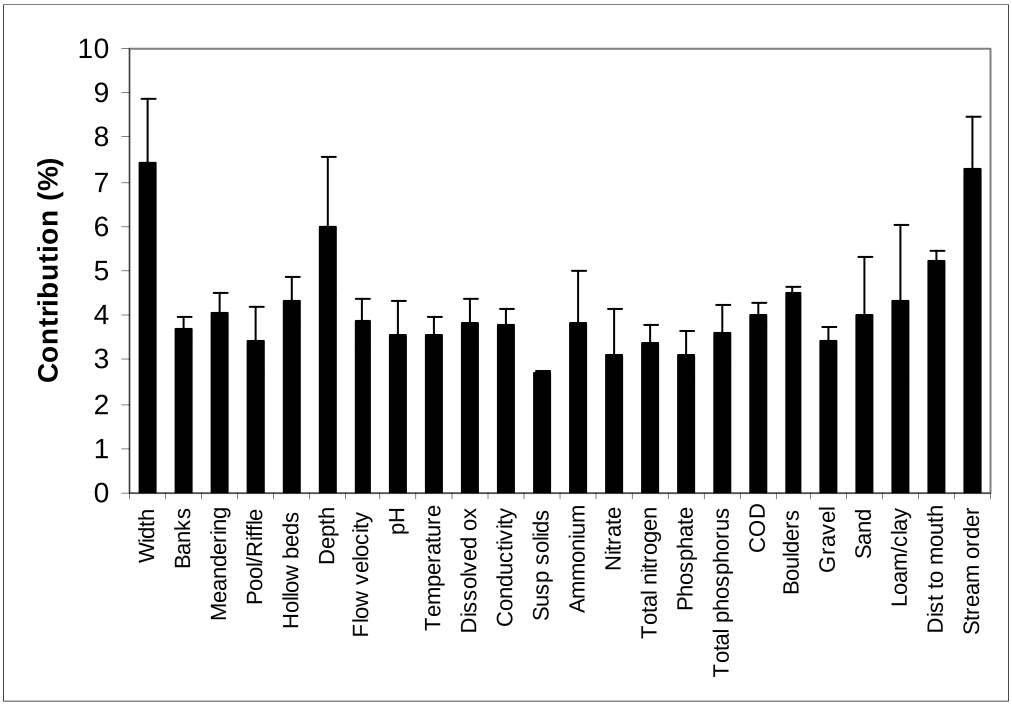

Asellus (probability of presence)
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
2
3
4
5
6
7
8
9
10
11
12
Width
Banks
Meandering
Pool/Riffle
Hollow beds
Depth
Flow velocity
pH
Temperature
Dissolved oxygen
Conductivity
Suspended solids
Ammonium
Nitrate
Total nitrogen
Phosphate
Total phosphorus
COD
Boulders
Gravel
Sand
Loam/clay
Distance to mouth
Stream order

Appendix 67: ANN, ABUN, *Gammarus* and *Asellus*, Zwalm river basin, Weights

Gammarus

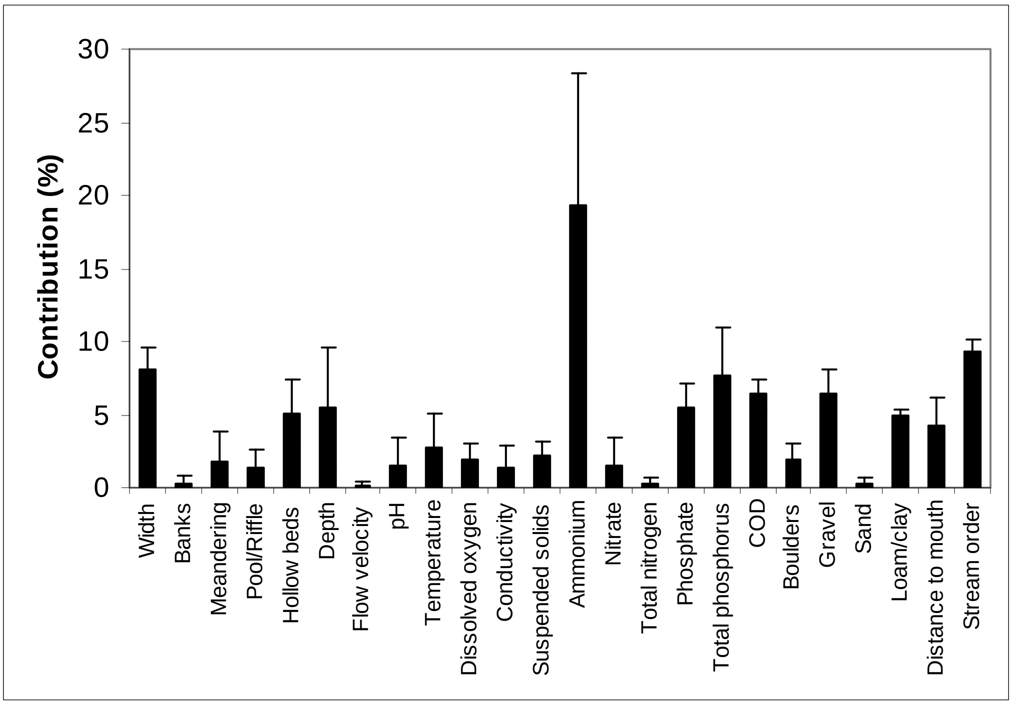


Asellus

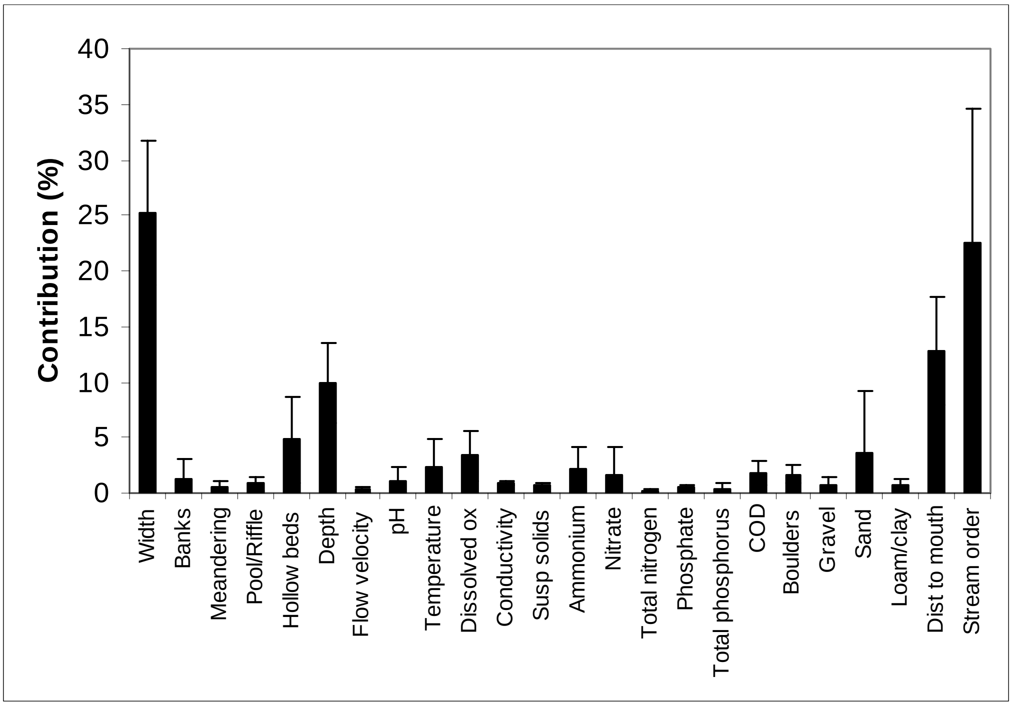


Appendix 68: ANN, ABUN, *Gammarus* and *Asellus*, Zwalm river basin, PaD

Gammarus

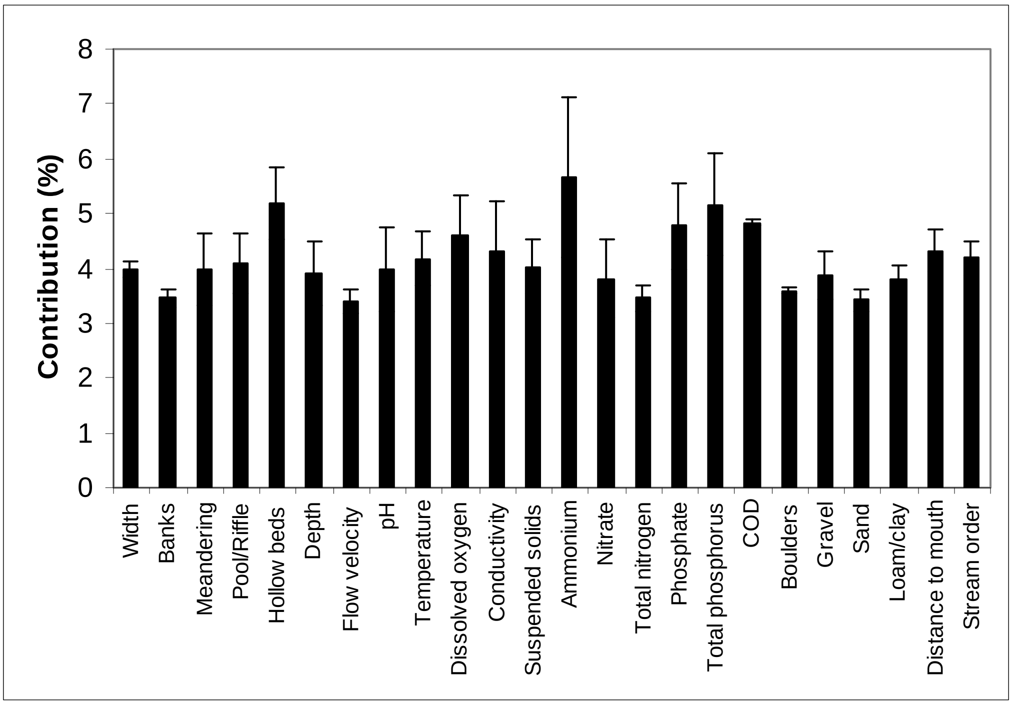


Asellus

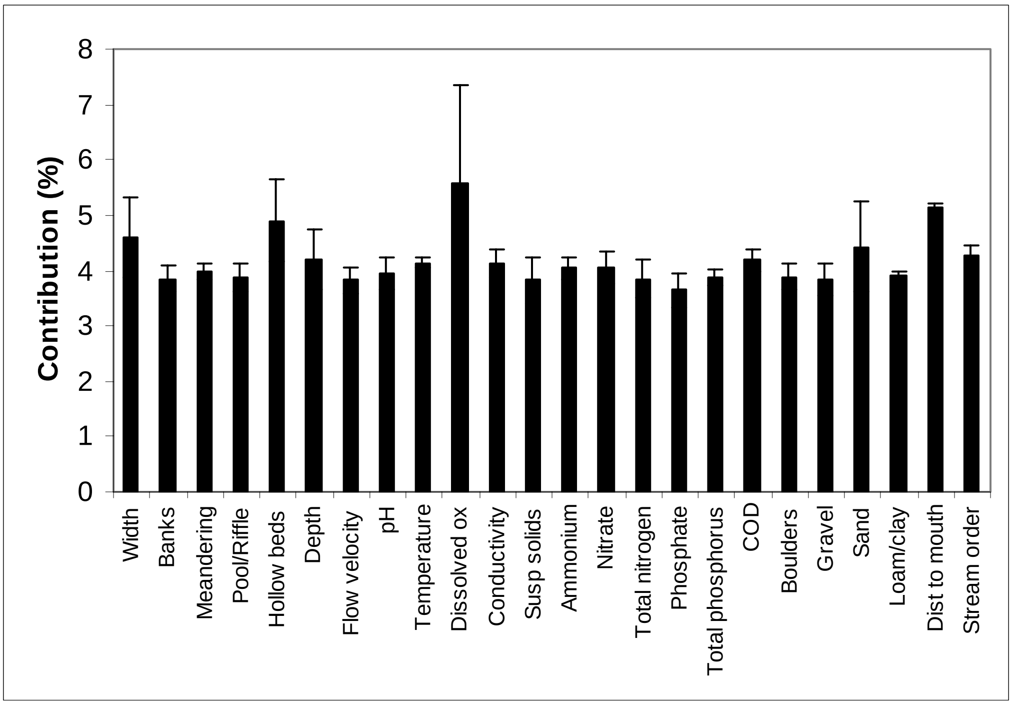


Appendix 69: ANN, ABUN, *Gammarus* and *Asellus*, Zwalm river basin, Perturb

Gammarus

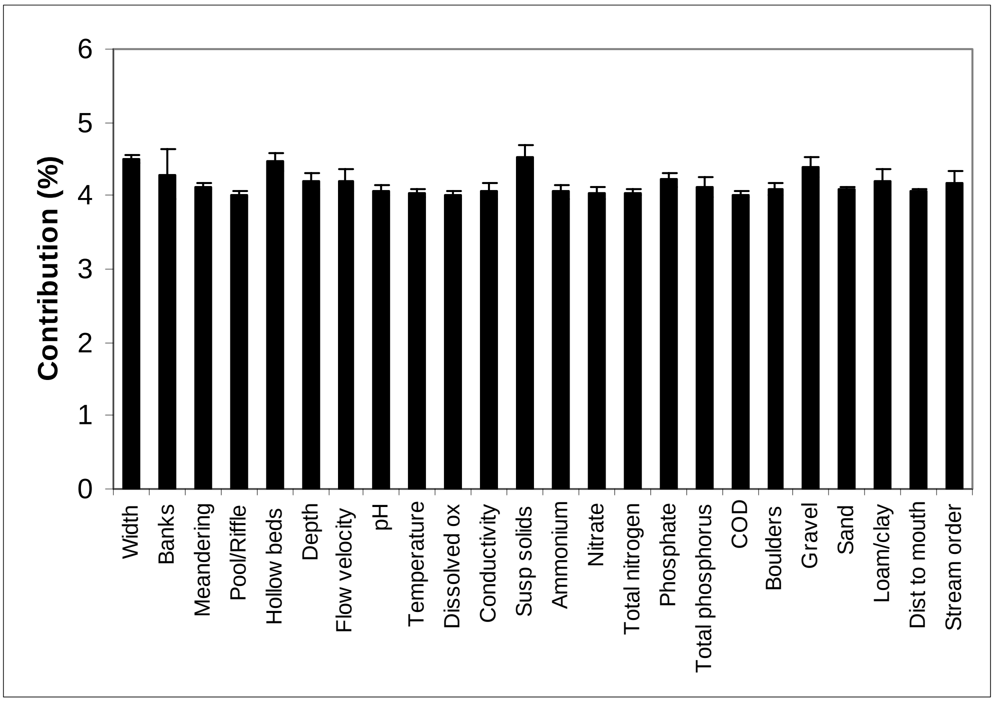


Asellus

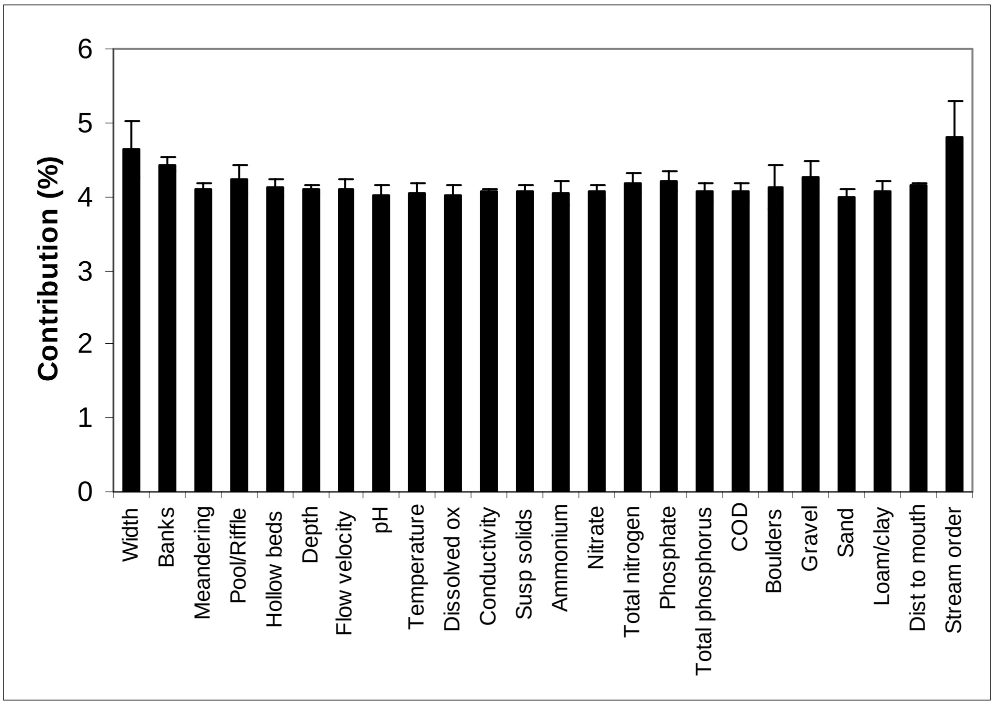


Appendix 70: ANN, ABUN, *Gammarus* and *Asellus*, Zwalm river basin, Stepwise Reg

Gammarus

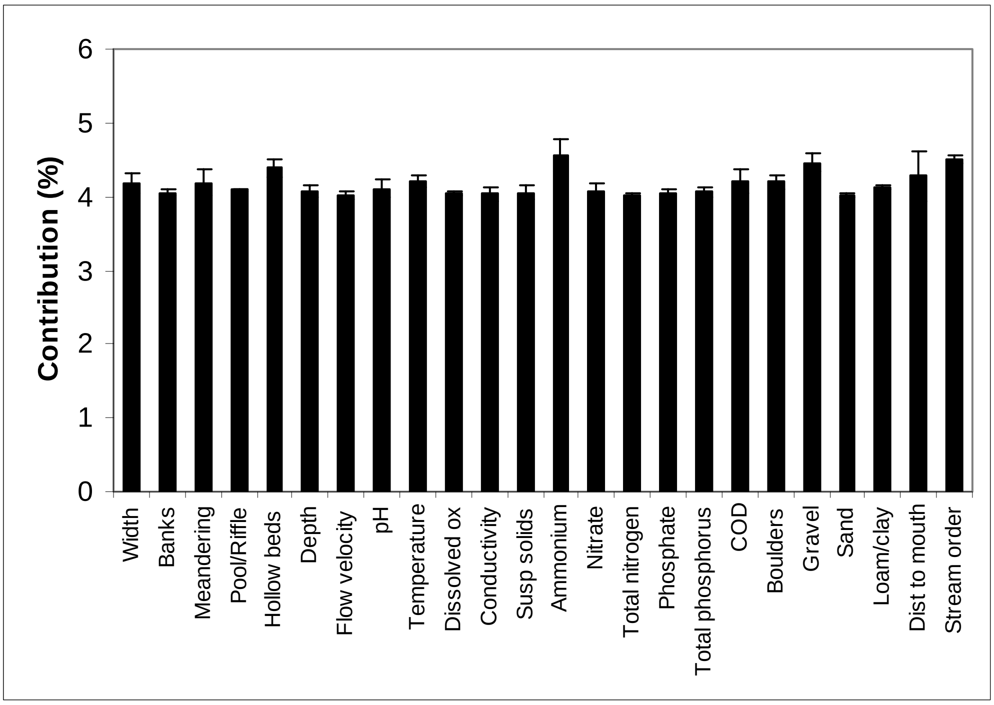


Asellus

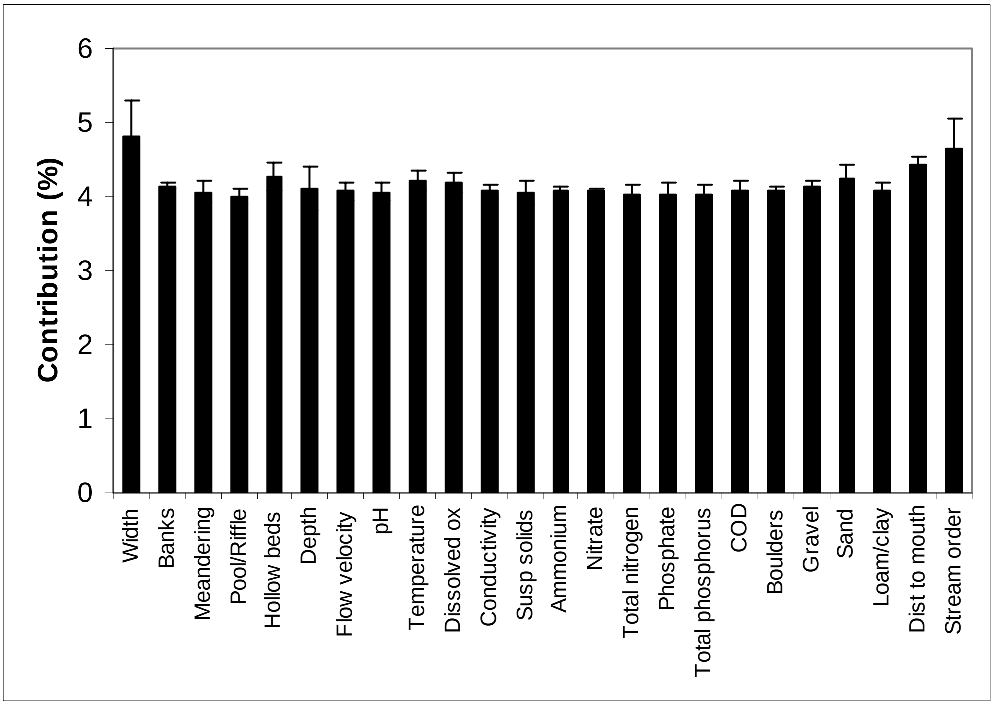


Appendix 71: ANN, ABUN, *Gammarus* and *Asellus*, Zwalm river basin, Stepwise Imp

Gammarus

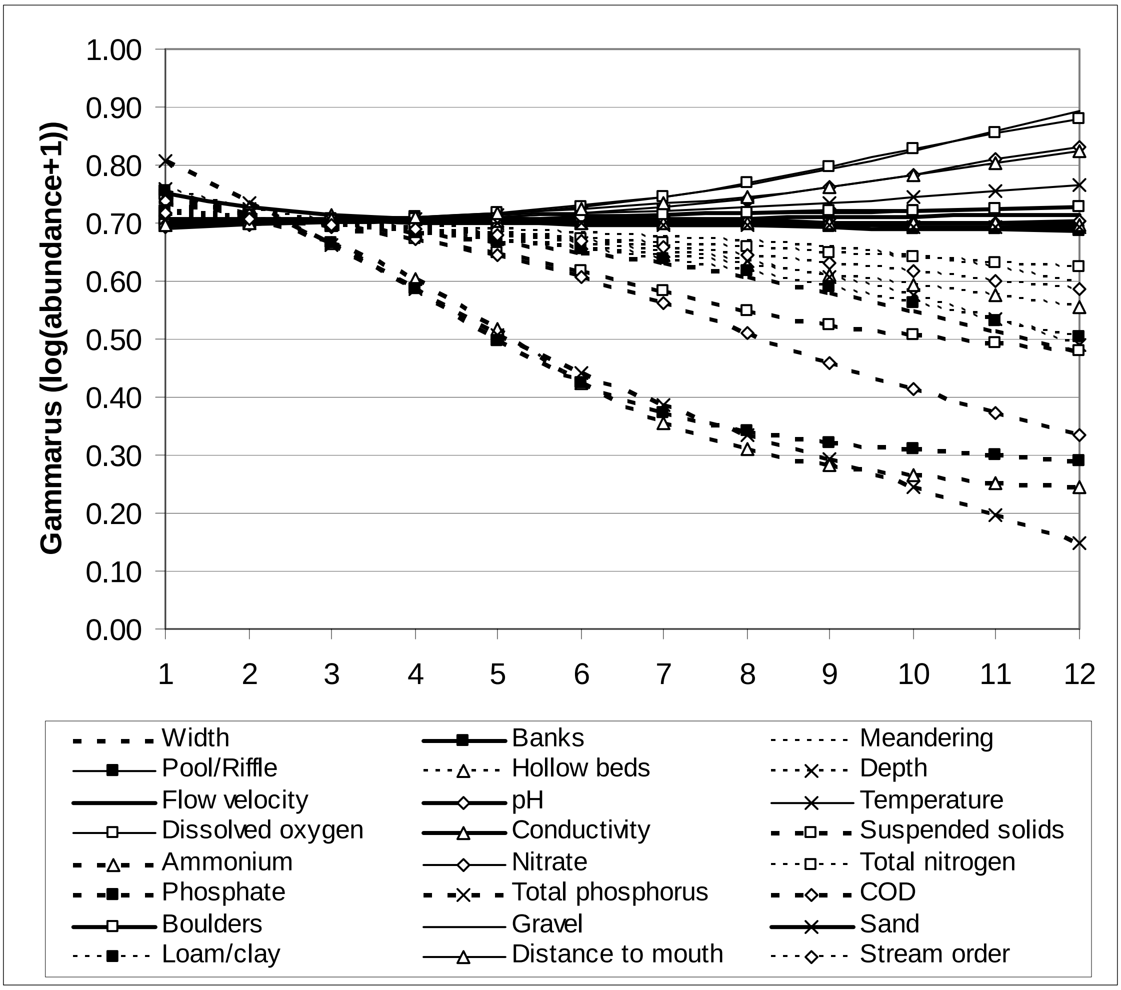


Asellus

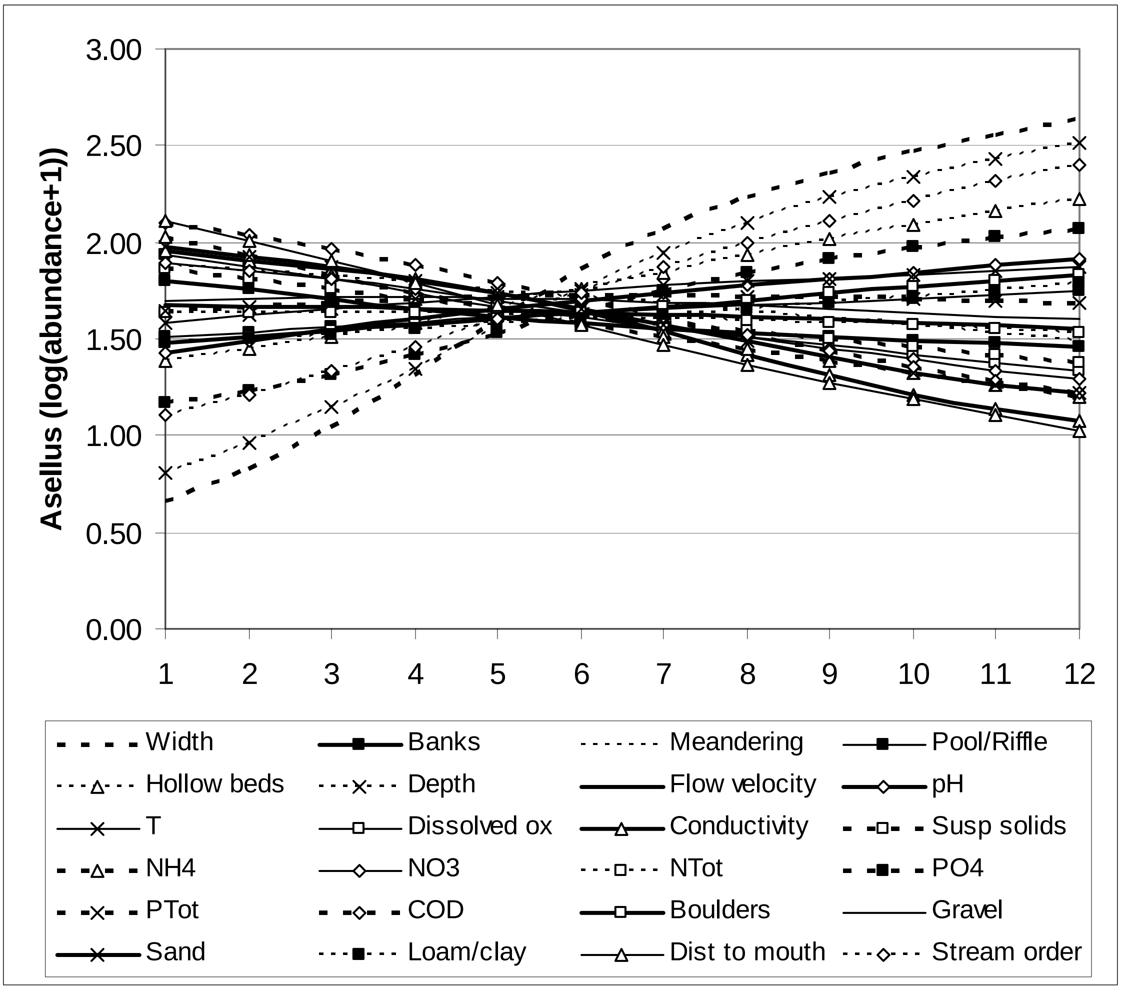


Appendix 72: ANN, ABUN, *Gammarus* and *Asellus*, Zwalm river basin, Profile

Gammarus



Asellus


Asellus (log(abundance+1))
3.00
2.50
2.00
1.50
1.00
0.50
0.00
1
2
3
4
5
6
7
8
9
10
11
12
Width
Banks
Meandering
Pool/Riffle
Hollow beds
Depth
Flow velocity
pH
T
Dissolved ox
Conductivity
Susp solids
NH4
NO3
NTot
PO4
PTot
COD
Boulders
Gravel
Sand
Loam/clay
Dist to mouth
Stream order

Appendix 73: numerical performance indicator values of the ANN models applied for the simulation of the practical restoration options in Chapter 6

Asellus	**CCI**	***K***
Subset 1	0.78	0.57
Subset 2	0.77	0.53
Subset 3	0.78	0.57
Gammarus		
Subset 1	0.77	0.35
Subset 2	0.73	0.22
Subset 3	0.82	0.41
Erpobdella		
Subset 1	0.83	0.65
Subset 2	0.77	0.48
Subset 3	0.78	0.55
Baetis		
Subset 1	0.75	0.20
Subset 2	0.68	0.14
Subset 3	0.67	0.12

ISBN 90-5989-046-9