Faculteit
Bio-ingenieurswetenschappen

Academiejaar 2004-2005

DEVELOPMENT OF HABITAT AND MIGRATION MODELS FOR THE PREDICTION OF MACROINVERTEBRATES IN RIVERS

ONTWIKKELING VAN HABITAT- EN MIGRATIEMODELLEN VOOR DE VOORSPELLING VAN MACRO-INVERTEBRATEN IN RIVIEREN

door

ir. Andy DEDECKER

Thesis submitted in fulfilment of the requirements for the degree
of Doctor (Ph.D.) 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
dr. ir. P.L.M. GOETHALS

Faculteit
Bio-ingenieurswetenschappen

Academiejaar 2004-2005

DEVELOPMENT OF HABITAT AND MIGRATION MODELS FOR THE PREDICTION OF MACROINVERTEBRATES IN RIVERS

ONTWIKKELING VAN HABITAT- EN MIGRATIEMODELLEN VOOR DE VOORSPELLING VAN MACRO-INVERTEBRATEN IN RIVIEREN

door

ir. Andy DEDECKER

Thesis submitted in fulfilment of the requirements for the degree
of Doctor (Ph.D.) 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
dr. ir. P.L.M. GOETHALS

ISBN 90-5989-068-X

This research was supported by a fellowship of the Flemish Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), Bischoffsheimlaan 25, B-1000 Brussels, Belgium.

This research was conducted at the Ghent University, Faculty of Bio-engineering Sciences, Department of Applied Ecology and Environmental Biology, Laboratory of Environmental Toxicology and Aquatic Ecology, Jozef Plateaustraat 22, B-9000 Gent, Belgium (contact: Prof. dr. N. De Pauw, email: Niels.DePauw2UGent.be, tel: +32 (0)9 264 37 68, fax: +32 (0)9 264 41 99)

List of abbreviations

AMINAL	Administratie Milieu-, Natuur-, Land- en Waterbeheer (Administration Environment, Nature, Land and Water Management)
ANN	Artificial Neural Network
AQEM	The Development and Testing of an Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates
AQUAFIN	Afvalwaterzuivering Vlaanderen (Flanders Waste Water Company)
AROHM	Administratie Ruimtelijke Ordening, Huisvesting, Monumenten en Landschappen (Administration Town and Country Planning, Housing, Monuments and Landscapes)
ASPT	Average Score Per Taxon
AUSRIVAS	Australian River Assessment System
AWP	Algemeen Waterkwaliteitsplan (General Water Quality Plan)
AWZ	Administratie Waterwegen en Zeewezen (Administration Watercourses and Sea Issues)
BBI	Belgian Biotic Index
BEAST	Benthic Assessment of Sediment
BSI	Belgian Sediment Index
CCI	Percentage of Correctly Classified Instances
CIS	Common Implementation Strategy
CITYFISH	Modelling the ecological quality of urban rivers: ecotoxicological factors limiting restoration of fish populations
CIW	Coördinatiecommissie Integraal Waterbeleid (Coordination Committee for Integrated Water Management)
CK	Cohen's kappa
COD	Chemical Oxygen Demand
CVE	Cross Validation Error
DTM	Digitaal Terrein Model (Digital Terrain Model)
EQR	Ecological Quality Ratio
ESRI	Environmental Systems Research Institute
EU	European Union
EUROLAKES	Integrated water resource management for important deep European lakes and their catchment areas
FAME	Development, evaluation and implementation of a standardised Fish-based Assessment Method for the Ecological status of european rivers
GIS	Geografisch Informatie Systeem (Geographical Information System)
IBI	Index of Biotic Integrity
IMC	Internationale Maascommissie (International Meuse Commission)
IN	Instituut voor Natuurbehoud (Institute of Nature Conservation)
ISC	Internationale Scheldecommissie – International Scheldt Commission
LMLS	Least Mean Log Squares Error
MANTRA east	Integrated strategies for the management of transboundary waters on the eastern european fringe – The pilot study of lake Peipsi and its drainage basin
MATLAB®	Matrix Laboratory®
MIRA-T	Milieu- en Natuurrapport Vlaanderen, Thema's (Report on the Environment and Nature, Themes)
MSE	Mean Squared Error
NICHE	Nature Impact Assessment of Changes in Hydro-Ecological Systems

NID	Neural Interpretation Diagram
PAEQANN	Predicting Aquatic Ecosystem Quality using Artificial Neural Networks: impact of environmental characteristics on the structure of aquatic communities (Algae, Benthic and Fish Fauna)
PDM	Probability Distributed Moisture
QA/QC	Quality Assurance/Quality Control
r	Correlation coefficient
R^2	Determination coefficient
REFCOND	Development of a protocol for identification of reference conditions, and boundaries between high, good and moderate status in lakes and watercourses
RIVPACS	River Invertebrate Prediction And Classification System
RMSE	Root Mean Squared Error
SENTWA	System for the Evaluation of Nutrient Transport to Water
SEPTWA	System for the Evaluation of Pesticides Transport to Water
SOM	Kohonen Self-Organizing Maps, i.e. Kohonen networks
SSD	Sum of Square Derivatives
SSE	Sum Squared Error
STAR	Standardisation of River Classification
VHA	Vlaamse Hydrografische Atlas (Flemish Hydrographical Atlas)
VLAREM	Vlaams Reglement voor Milieuvergunning (Flemish Regulation for Environmental Licenses)
VLM	Vlaamse Landmaatschappij (Flemish Land Agency)
VMM	Vlaamse Milieumaatschappij (Flemish Environment Agency)
VVP	Vereniging van Vlaamse Provincies (Association of Flemish Provinces)
VVPW	Vereniging van Vlaamse Polders en Wateringen (Association of Flemish Polders and Waters)
VVSG	Vereniging van Vlaamse Steden en Gemeenten (Association of Flemish Cities and Municipalities)
WEKA	Waikata Environment for Knowledge Analysis
WFD	Water Framework Directive
WWTP	Waste-Water Treatment Plant

Introduction

Problem definition

The quality and availability of freshwater is one, if not the most essential determinants for the health of ecosystems and human societies world-wide. Like all living beings, humans are seriously dependent on freshwater and exploit its resources heavily in the natural environment. By doing so, human activities have severely deteriorated freshwater systems, and many functions such as drinking water supply, fishing, … are threatened. Hydrological changes, physical disturbances, point and non-point sources of pollution, from both rural and urban activities, are all examples of processes responsible for the large-scale deterioration of freshwater systems and running waters in particular. This degradation affects the well-being of our natural environment and its biodiversity and may also affect freshwater services for human consumption.

Due to the pressures of increasing population and developing economy all over the world, the present situation of water quality and river management is far from satisfactory (Biswas, 1991; Haimes, 1992; Plate, 1993; Simonovic, 1996a, b; Falkenmark, 1997; Kundzewicz, 1997; Xia and Taceuchi, 1999). Because the restoration of river and water systems in general entails drastic social and economic consequences, and the management of water resources systems is often driven by multiple objectives (Huang and Xia, 2001), the decisions should be taken with enough forethought. To enhance sustainable management of water and river quality, in-depth research of the related barriers and the relevant mitigation approaches is needed.

In order to address the complicated interrelationships between human activities and the state of freshwater health, legislation initiatives have been taken and freshwater directives have been supported at different levels (e.g. EU Water Framework Directive (WFD) at the European level and the Decree of Integrated Water Management at the local Flemish level). In contrast to former European guidelines (e.g. EU Nitrate Directive), the present objectives of the WFD concerning water quality (EU, 2000) are no longer exclusively focussed on human activities. Also the biological communities (fishes, macroinvertebrates, macrophytes, phytobenthos and phytoplankton) and their ecological demands have to be taken into account. The major aim of the WFD is to reach 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 that represent a good ecological status must be described. Additionally, for the development of a representative set of metrics for ecological river assessment, one needs to gain insight into the relation between the aquatic communities and the human activities affecting these water systems. Insights into 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.

Recently, technologies are becoming more and more important in water quality and river restoration management, due to the rapid development of computational problem-solving tools and the enhancement of scientific approaches for information support (Huang and Xia, 2001). However, at the watershed level, the related physical, chemical and biological processes are numerous with complicated interactions. To retrieve these interaction as requested by the WFD, significant efforts are required to analyze the relevant information, simulate the related processes, evaluate the resulting impacts, and generate sound decision alternatives. Mathematical models are useful tools for water quality simulation and prediction. They can project consequences of alternative management, planning, or policy level activities, such that effective management schemes can be identified. In the last decades, study of water quality modelling has been an active aspect in environmental management. In Flanders for example, several water system models are already applied by the different governmental administrations and institutions (Cauwenberghs, 2003). They include e.g. sewer system models (e.g. Hydroworks, InfoWorks CS), nutrient (e.g. SENTWA) and pesticide (e.g. SEPTWA) transport models, water quality models (e.g. SIMCAT, Pegase), hydrological (e.g. PDM) and hydraulic (e.g. Mike 11, ISIS) models. In spite of the ecological objectives of the WFD, ecological models have rarely been used so far to support river management and water policy. Ecological models have however several interesting applications in this context. In particular, 'habitat suitability models' that can predict the habitat requirements of organisms based on environmental river characteristics could be very useful. This type of models has only very recently been recognized as a significant component of conservation planning (e.g. Guisan and Zimmerman, 2000; Austin, 2002; Scott et al., 2002). They are however almost exclusively used in ecological research, and to improve the models themselves. Exceptions are RIVPACS (River Invertebrate Prediction And Classification System) (Armitage et al., 1983; Wright et al., 1984; Wright et al., 1993; Wright et al., 2000), AUSRIVAS (Australian

River Assessment System) (Norris and Norris, 1995; Smith et al., 1999; Davies, 2000) and BEAST (Benthic Assessment of Sediment) (Reynoldson et al., 1995) which are being used to support river management and water policy in respectively the UK, Australia and North America. To this end, it can be a challenge to develop habitat suitability based models to support river management and water policy in Flanders.

A summary of the potential value of (ecological) models in river management is presented by Goethals and De Pauw (2001) (Fig. 1). 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 optimized. 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 (see also Vanrolleghem *et al.*, 1999).

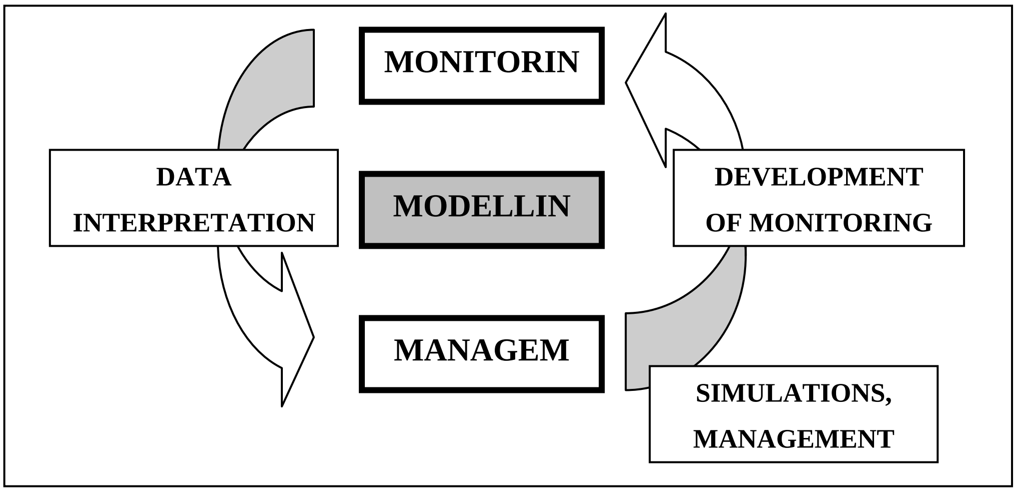


Fig. 1. Potential applications of models for information and decision support in river management (Goethals and De Pauw, 2001).

Before such ecological models can be effectively applied in river management, however, several research challenges still need to be tackled. The following two major actual problems are mentioned by Goethals (2005):

1. The availability of **reliable datasets**, which are needed to develop, train and validate ecological models. During the last decades, a lot of ecological data have indeed being collected, but these efforts are often too fragmented, resulting in databases that are not

compatible, lacking essential variables, ... Often these problems are (were?) related to the organizational structure of the water management boards and focussing at too specific goals to allow for an integrated water management;

2. 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 an essential 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 analyzed once the effect of management decisions has taken place.

Based on the above mentioned challenges and gaps, the scope and objectives of this thesis has been formulated as follows.

Scope and objectives

The overall aim of the present thesis is to determine the appropriate variables and ecosystem processes by using a modelling technique, based on Artificial Neural Networks (ANNs), to predict the habitat suitability of 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 focussed on macroinvertebrates in brooks and small rivers of the Zwalm river basin, a sub-basin of the Upper-Scheldt river basin (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 Artificial Neural Network models in this research are based on a data driven approach. 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.

In general, the assumption is made that ANN models are ‘perfectly mixed’, which means that all sites are freely and equally accessible. However, when accessibility is restricted (e.g. presence of migration barriers, unbridgeable distance between potential habitats and source populations, …), it is necessary to control for the effects of it before conclusions about preference and habitat suitability can be drawn. For example, the habitat of a restored river section can be predicted as suitable again (after river restoration actions took place) based on the habitat suitability models. It is possible however that this restored river section is inaccessible for the modelled organism due to the presence of migration barriers or an unbridgeable distance between the new potential habitat and the existing source populations. Based on habitat suitability models alone, this problem cannot be solved. Dedecker *et al.* (2005d) illustrated this problem. After water quality improvement, Limnephilidae (an indicator for good water quality) was predicted present at the restored river sections based on the habitat suitability model. However, existing source populations were located more than five kilometres away from these potential new habitats. In this way, recolonization of the restored sites was very unlikely. Nevertheless, this could not be retrieved from the habitat suitability models. To this end, migration models can be very useful as an extension of the habitat suitability models to determine the possibility of migration and recolonization.

The developed ANN and migration 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. Both models can thus become essential tools to convince stakeholders to make the necessary investments and/or activity changes as desired by society.

More specifically the research objectives of this thesis were fivefold (see also Fig. 2):

- establishment of monitoring networks and ecological databases to develop models predicting aquatic macroinvertebrates in rivers;
- development, optimization and validation of habitat suitability models based on ANNs;
- ecological interpretation of ANNs by application of input variable contribution methods;

- development and validation (based on sensitivity analysis) of migration models for macroinvertebrates as extension of the habitat suitability models;
- prediction and practical evaluation of river restoration scenarios.

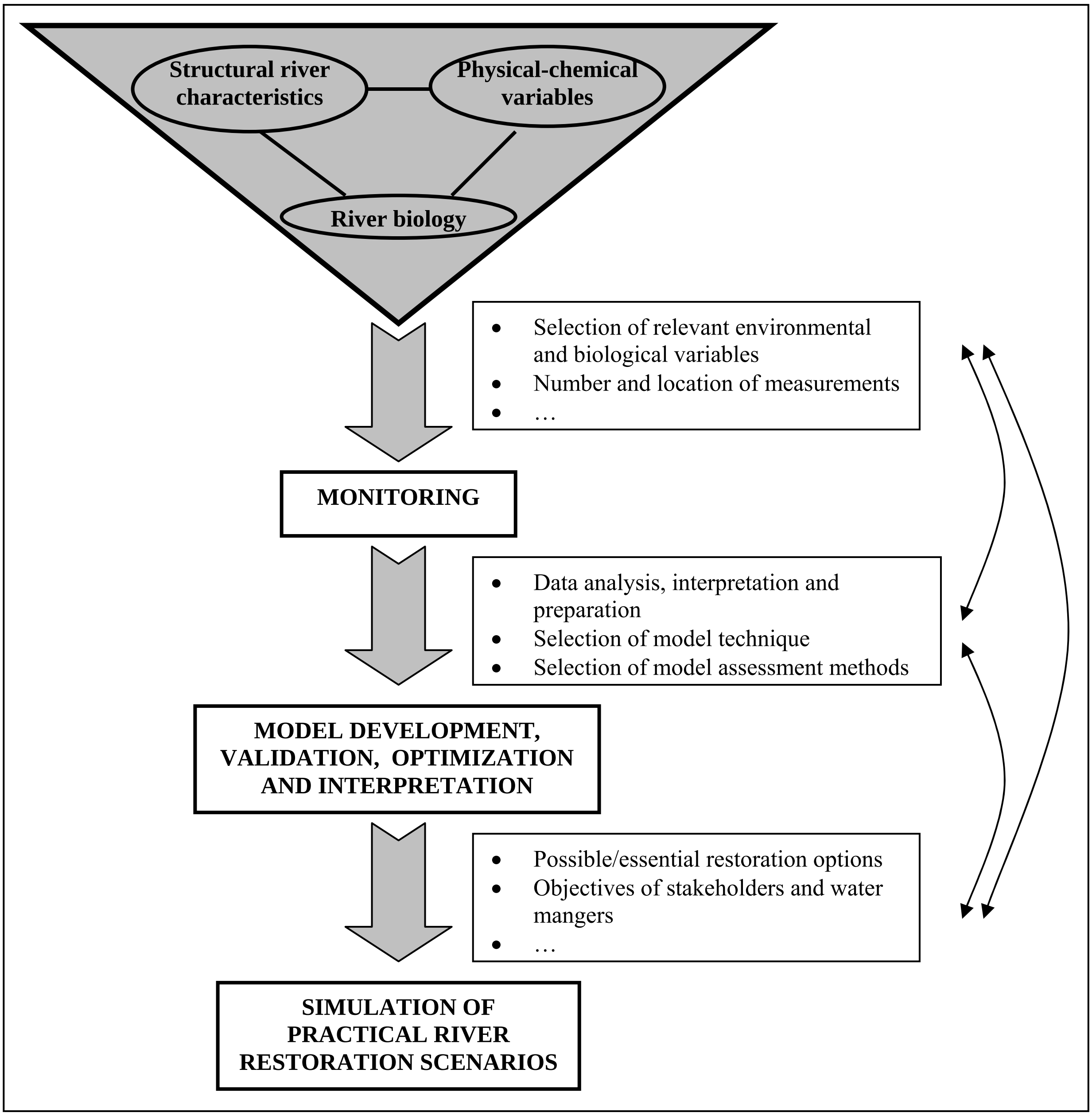


Fig. 2. The major steps related to the development and application of ecological models for decision support in water management that have been taken into consideration in this thesis.

The individual chapters are grouped in the above five parts and briefly described in the next paragraphs which further elucidate the specific goals of the research. The thesis consists of ten chapters and a general discussion.

Chapter 1 is an introduction into river assessment and river management in Flanders. First, the basis for the integrated water management in Flanders, the European Water Framework Directive (WFD), is briefly discussed and an overview of the major goals and projects related to the WFD is given. Although the Decree of Integrated Water Management (B.S. 14/11/03) entered into force in Flanders, the main tasks relating to water and river assessment and management are still scattered over several administrations. Therefore, a summary of the major responsibilities and examples of specific tasks of the most important administrations related to water and river management in Flanders is presented. Although several river models are already applied by the different governmental administrations and institutions in Flanders, the need of ecological habitat suitability and migration models can be illustrated. In particular insights are needed between river characteristics and biological communities for the optimization of ecological indices as well as for the prediction of the ecological effects of river restoration management.

Chapter 2 gives an overview of the state of the art of Artificial Neural Networks (ANNs) to predict macroinvertebrates in rivers. A general introduction on ANN models is given, illustrating the structure and theoretical basis on which the model is relying on. Major attention is focused on data analysis and processing, input variable selection, model architecture, model validation, optimization and interpretation. In addition a review is given of recent research discussing case studies on the prediction of macroinvertebrates by means of ANNs.

Chapter 3 is an introduction on the migration behaviour of macroinvertebrates is given. In addition, an overview of migration models is provided.

Chapter 4 is dealing with the collection of data and ecological information and contains a description of the study area, the monitoring networks and methods, the river site selection criteria, the database set-up and expert knowledge found in literature on the habitat preferences and factors affecting the migration behaviour of the taxa Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae. In comparison with Goethals (2005) who used only two taxa (Asellidae and Gammaridae), five taxa were selected in the present thesis ranging from very tolerant to very sensitive and from very common to very rare. A study, discussing the variability in the collection of macroinvertebrate data is included. Two databases are presented in this chapter. The first database consists of measurements conducted in 60 different sites in the entire Zwalm river basin during the period 2000-2003, while Goethals (2005) included only three years of data. This database was specifically constructed for the development of habitat suitability models. The second additional database was set up during the period 2002-2003 and contains also 60 sites. For this 'short distance' monitoring network, a part of the Zwalm river basin of about 12 km was selected, which consisted of the brooks Verrebeek, Dorenbosbeek and the upstream part of the Zwalm river. This database was constructed for the development of macroinvertebrate migration models. For both databases, macroinvertebrates were collected using hand nets and artificial substrates. In addition, water quality and physical habitat variables were used to describe the river characteristics of each site.

Chapter 5 describes the analysis of the collected data. 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. Therefore, a first and basic step before model development and application is getting insight into the range of inputs and outputs, what determines also the maximum application range of the ANN models. In addition, the mutual correlation between input variables and between input and output variables is calculated to help identify 'noise' variables. A visual relation analysis between input and output variables is conducted to get insight into outliers, the data clusters and distributions, missing or scarce variable combinations in certain ranges, … Also the geographical distribution of the variables over the study areas is presented based on a Geographical Information System to identify the priority sites for river restoration.

Chapter 6 describes the development, analysis and optimization of habitat suitability models based on Artificial Neural Networks (ANNs) for the prediction of macro-invertebrates. Based on the available datasets, ANN models were therefore applied. While Goethals (2005) applied a standard network architecture including 10 neurons in the hidden layer and 3-fold cross-validation (two third of the dataset used for training, one third for testing) as the standard validation method, the specific aim of the present study was to discuss different neural network models to obtain the best model configuration for the prediction of the selected macroinvertebrate taxa (Tubificidae, Asellidae, Gammaridae, *Baetis*, Limnephilidae). Based on this, the following actions were considered in this chapter:

- Since the size of the training and test set influences respectively the generalization capability of the model and accuracy of the model performance, the optimal size of training and test set was discussed. Therefore, nine cross-validation methods (2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation) were tested;
- the optimal ANN architecture for the five selected macroinvertebrates in both databases is searched for using the optimal training and test size concerning two questions:
    - how many hidden layers should exist in the ANN architecture?
    - how many neurons should be present in the hidden layer(s)?

  Additionally, the optimal ANN architecture was analyzed, examining the effect of annual testing. ANN models for the entire Zwalm river basin were trained with measured input and output data of 3 years, while data of the remaining independent year was used for testing. Similarly, the ANN models for the 'short distance' monitoring network were trained with data of one year and tested with data of the other year.

To retrieve the best ANN models, two evaluation criteria were applied. In this way, this chapter can be seen as the mathematical validation step of the habitat suitability models.

Chapter 7 describes the application of several methods, in addition and combined with the ANN models, to analyze the contribution of environmental variables to predict macroinvertebrates in a reliable manner and to detect the major river characteristics to describe the habitat suitability of the different taxa. In this manner, insight into the effect of river conditions on the presence/absence of macroinvertebrates is obtained and the outcomes of the models can be compared with ecological expert knowledge from literature. Based on this, Chapter 7 can perform as the ecological validation step of the habitat suitability models.

Ecosystem models can act as interesting tools to support decision-making in river restoration management. In particular, models such as ANNs, that can predict the habitat requirements of organisms, can be very useful. In general however, habitat suitability models do not include spatial and temporal relationships. Migration dynamics of the predicted organisms and migration barriers along the river may therefore deliver important additional information on the effectiveness of the restoration plans as illustrated before. In this context, migration models for Gammaridae (Crustacea, Amphipoda), *Baetis* (Insecta, Ephemeroptera) and Limnephilidae (Insecta, Trichoptera) were developed for the Zwalm river basin (**Chapter 8**).

Chapter 9 describes the application of sensitivity analysis to the migration models developed in Chapter 8. Therefore, the impact of the resistance values on the calculated migration cost of Gammaridae, *Baetis* and Limnephilidae is studied. To this end, resistance values ascribed to the different environmental characteristics determining the migration were modified. This sensitivity analysis can be seen as a theoretical validation of the developed migration models in Chapter 8.

Chapter 10 combines the predictive habitat suitability models (ANNs) and the migration models to allow for a more transparent and rational evaluation of the effects of specific management options. A crucial validation step is made by analyzing the practicability of using this type of data driven models in combination with migration 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 and can as such be seen as the practical validation step of the developed habitat suitability and migration models.

At the end, the thesis is closed with a general discussion with recommendations for further research.

Chapter 1
River assessment and management in Flanders and the need of habitat suitability models

1.1. River management

1.1.1. The Water Framework Directive as a guideline for integrated river basin management in Europe

On 23 October 2000, the ‘Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for the Community action in the field of water policy’ or short the EU Water Framework Directive (WFD) was finally adopted. The Directive that was published in the Official Journal (OJ L 327) on 22 December 2000, entered into force the same day (EU, 2000) (http://europa.eu.int/comm/environment/water/waterframe-work/index_en.html). An important shift compared to other European Directives related to water policy (e.g. Drinking Water Directive, Urban Waste Water Directive) is that an integrated approach was taken. The objectives of water management are not only defined with regard to human needs but attention is also paid to the ecological demands of the aquatic ecosystem (Elbersen *et al.*, 2003). With the implementation of the WFD, the different Member States are committed to invest in river management to improve river health. To guarantee that the WFD is implemented in a common way in all the Member States, a ‘Common Implementation Strategy’ (CIS) was adopted (EU, 2001).

In general, the directive distinguishes two major goals for rivers. First, the development of a management system has to be based on natural river basin districts which are the main units for river management (Blöch, 1999). Further, river basins and sub-basin are to be distinguished. The second goal of the directive is the development of river basin management plans and programmes of measures to achieve at least a ‘good surface water status’ in 2015 for al European rivers. In terms of environmental objectives, a ‘good surface water status’ means the status achieved by a surface water body is at least ‘good’ for both its ecological status, or its ecological potential for a heavily modified or an artificial water body, and its chemical status. The ecological status of a water body includes a combined effect of biological (phytoplankton, phytobenthos, macrophytes, macroinvertebrates and fishes), morphological and physical-chemical characteristics. A good chemical status implies that the environmental quality standards are respected for particular pollutants. The directive sets a number of monitoring obligations (surveillance, operational and investigative monitoring programmes). The results of these monitoring programmes are to be used for the implementation of a programme of measures in order to achieve good status for each type of water body, a status representative of the conditions at the site in the absence of interference by man (Chave, 2001). In addition, for each type of water body, reference conditions have to be established for the biological, hydromorphological and physical-chemical quality elements. Based on these type-specific reference conditions, an overall Ecological Quality Ratio (EQR = observed value/reference value) has to be defined. This EQR has a value between 0 (= minimal, bad ecological status) and 1 (= maximal, high ecological status), and distinguishes 5 classes going from a very good to a very bad ecological status (Wallin *et al.*, 2003).

Until now, several European Union-related research projects have been supported by the European Commission and are contributing to the implementation of the WFD. Examples are:

- **AQEM** (The Development and Testing of an Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates) (EVK1-CT-1999-00027) (http://www.aqem.de/);
- **STAR** (Standardisation of River Classification) (EVK1-CT-2001-00089) (http://www.eu-star.at/);
- **FAME** (Development, Evaluation and Implementation of a Standardised Fishbased Assessment Method for the Ecological Status of European Rivers) (EVK1-CT-2001-00094) (http://fame.boku.ac.at/);
- **REFCOND** (Development of a protocol for identification of reference conditions, and boundaries between high, good and moderate status in lakes and watercourses) (http://www-nrciws.slu.se/REFCOND/);
- **PAEQANN** (Predicting Aquatic Ecosystem Quality using Artificial Neural Networks: impact of environmental characteristics on the structure of aquatic communities (Algae, Benthic and Fish Fauna)) (EVK1-CT1999-00026) (http://aquaeco.ups-tlse.fr/);
- **CITYFISH** (Modelling the Ecological Quality of Urban Rivers: Ecotoxicological Factors Limiting Restoration of Fish Populations) (EVK1-1999-00041) (http://www.biosciences.bham.ac.uk/labs/taylor/CITYFISH.htm);
- **EUROLAKES** (Integrated water resource management for important deep European lakes and their catchment areas) (EVK1-CT-1999-00004) (http://pcs0.hydromod.de/Eurolakes/index.html)

- **MANTRA east** (Integrated Strategies for the Management of Transboundary Waters on the Eastern European Fringe – The Pilot Study of Lake Peipsi and its Drainage Basin) (EVK1- CT-2000-00076) (http://www.mantraeast.org/).

1.1.2. Integrated water management in Flanders

In Belgium, different water policies are being developed for the Flemish, Brussels and Walloon regions. Because parts of the major river basins (the Scheldt and the Meuse river basins, Fig. 1.1) are situated in these three regions, the water policies are often conflicting and 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 systems). 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, responsibilities relating to water and river assessment and management are scattered over several administrations in Belgium and Flanders, often resulting in specific and conflicting targets for the responsible managers.

In Flanders, the main tasks concerning river management have been allocated to the Flemish Environment Agency (VMM), the Flanders Waste Water Company (AQUAFIN), the Administration Watercourses and Sea Issues (AWZ), the Administration Environment, Nature, Land and Water Management (AMINAL), the Association of Flemish Provinces (VVP), the Association of Flemish Cities and Municipalities (VVSG), the Association of Flemish Polders and Waters (VVPW) and the Flemish Land Agency (VLM). A summary of their major responsibilities and examples of specific tasks concerning river management is given in Table 1.1.

Table 1.1. Summary of the major responsibilities and examples of specific tasks of key-role players related to water and river management in Flanders

Authority	**Major responsibilities**	**Examples of specific tasks**
VMM	Planning, reporting, collecting, informing, measuring and advising about surface water and sediment quality of watercourses	• Monitoring of the surface water and sediment quality (physical-chemical and biological) • Long-term planning for water treatment (the General Water Quality Plans, AWPs) • Giving advice in respect of the environmental permits issued within the VLAREM framework • Determining and collecting the levy on water pollution
AQUAFIN	Responsible for waste water treatment	• Financing, building and operating sewage processing systems, collectors and waste water treatment plants • Advising on the type of sewerage and treatment works
AWZ	River management of navigable watercourses	• Flood control • Increase the mobility of people and goods through a suitable and save shipping traffic • Guarantee an economical use of water • Stimulate recreation along the watercourses
AMINAL	River management of non-navigable watercourses 1st category	• Combat flooding • Ecological design of watercourses 1st category (re-meandering, ecological river restoration, remove fish migration barriers, …) • Rat control
VVP	River management of non-navigable watercourses 2nd category	• Maintenance of watercourses 2nd category (e.g. mowing, sludge removal and bank reinforcement) • Investment works (e.g. construction of controlled flood areas, weirs, fish passage and pumping-engines)
VVSG	River management of non-navigable watercourses 3rd category	• Suppress erosion • Install sewers • Maintenance of watercourses 3rd category (e.g. mowing, sludge removal)
VVPW	River management of non-navigable watercourses 2nd and 3rd category located within the polder areas	• Protection of polder areas against water inconvenience • Achieve a favourable water economy for agriculture, nature conservation, fishing, … in the polder areas
VLM	• Co-ordination of project regarding land consolidation, land use and nature planning • Manure bank • The Support Centre GIS Flanders	• Collection and making available of card material about Flanders (VHA atlas, ecological typology of watercourses, DTM, …) • Design of small-scale water treatment plants • Ecological design of watercourses (re-meandering, bank and buffer strips, ecological river restoration, …)

Although the organization of river management in Flanders is relatively complex, the 'Decree Integrated Water Management', which mainly implements the principles of the EU Water Framework Directive, gives a basis for an integrated assessment and management of river basins in Flanders (Van Bockstal, 2002). The Decree entered into force the 24th of November 2003 and assigned the organization of the integrated water management in Flanders to the Coordination Committee for Integrated Water Management (CIW) in which delegated managers of VMM, AMINAL, AWZ, AROHM, VVP, VVSG and VVPW interact to obtain the best integrated management solution. In this way, the different stakeholders take part in these debates (water quantity managers, land-use planners, wastewater collection and treatment managers, drinking water production companies, ecologists, …).

Integrated water management assumes that the water system itself should determine water policy rather than the administrative borders. Therefore, the 'Decree Integrated Water Management' starts from a territorial approach. In this way, water systems are divided into river basins and river basin districts, according to the WFD. Within the Flemish Region, four major river basins can be distinguished: the Scheldt river basin, the Meuse river basin, the Yzer river basin and the Bruges Polders (Fig. 1.1). They are grouped into two river basin districts: the Scheldt (including the Scheldt river basin, the Yzer river basin and the Bruges Polders) and the Meuse river basin district. Both river basin districts cross the Flemish borders. The Flemish part is subdivided into eleven river (drainage) basins (Fig. 1.1), each containing several sub-basins.

At each of these levels river basin management plans have to be created. These management plans should be geared to one another. At the level of the river basin districts, the International Scheldt Commission (ISC) and the International Meuse Commission (IMC) are responsible for the development of a single river basin management plan for both the entire international Scheldt and Meuse river basin district. According to the WFD, these should be achieved by the end of 2009. The Flemish contribution to both management plans are coordinated by the CIW. In addition, the CIW explains in a policy document the vision of the Flemish Government on integrated water management including water quality, durable water consumption, flood control, … For each of the eleven river (drainage) basins in Flanders, a river basin committee has been appointed. Their river basin management plans should be achieved by the end of 2006. Finally, sub-basin management plans (also called DuLo (sustainable and local) water plans) should be created at the sub-basin level.

To reach the objectives of the WFD and the 'Decree Integrated Water Management', numerous investment programmes will have to be carried out by river managers during the coming years. A well-considered decision support of river management will be currently needed in Flanders during the coming years in order to reach an improved (= good) quality and a high biodiversity value in the running waters. This will include: (1) setting priorities in river management; (2) reaching a cost-efficient river management and (3) assessing mitigation and restoration options/actions in rivers. This decision support system needs to offer a broad view of the impact of river management on all ecosystem components and at different scales of importance for the biological communities (Adriaenssens, 2004).

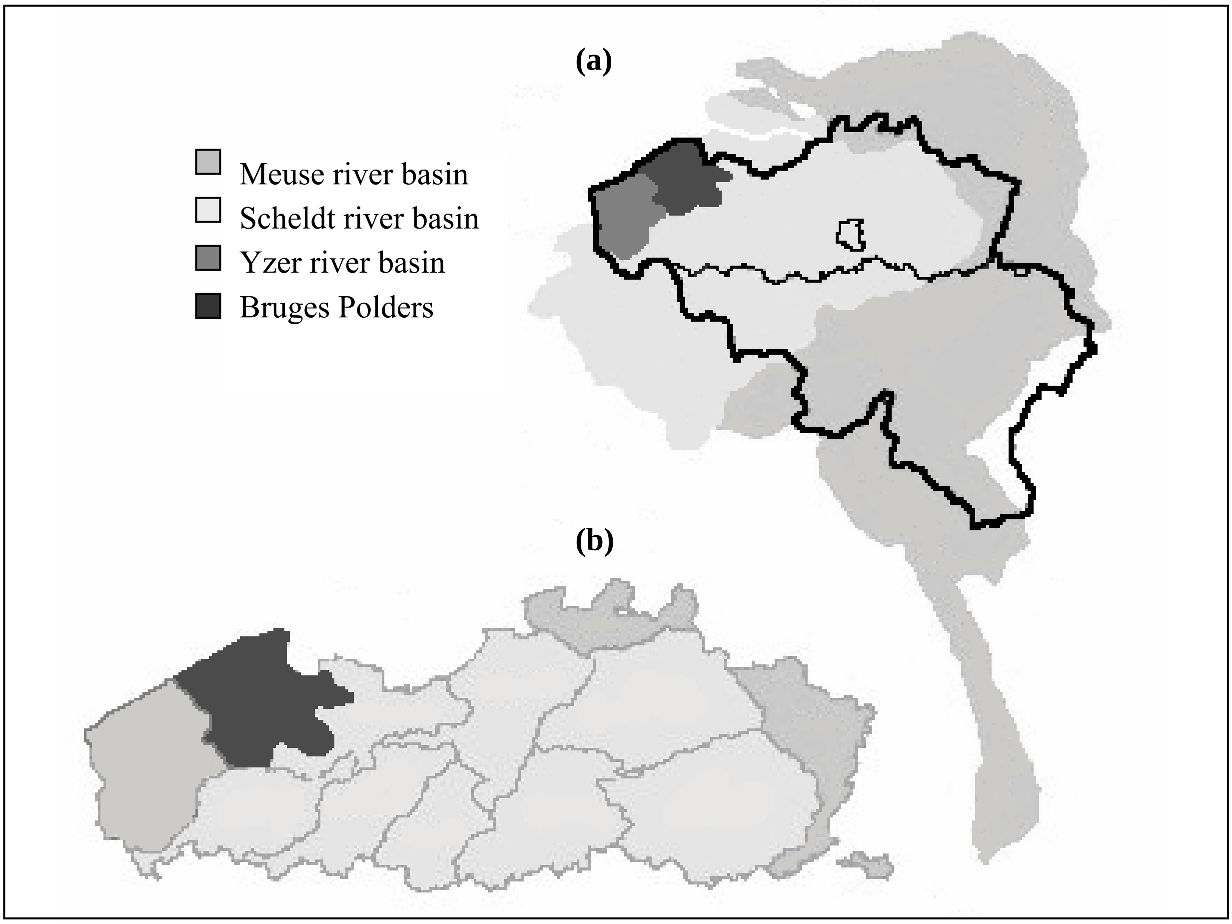


Fig. 1.1. Location of the four river basin districts across Flanders (a) and the division into eleven river (drainage) basins (b).

1.2. Decision support in river management

At present, ecological decision support of river management in Flanders is mainly based on biological assessment methods: the Belgian Biotic Index (BBI) (De Pauw and Vanhooren, 1983; De Pauw and Vannevel, 1991) and the Belgian Sediment Index (BSI) (De Pauw and Heylen, 2001). Both are based on diversity and pollution tolerance of macroinvertebrate taxa. Besides, the Index of Biotic Integrity (IBI) is used, based on a multimetric index for fish communities (Belpaire *et al.,* 2000; Breine *et al.*, 2004). Although such methods are useful for summarizing and presenting data, they only make use of a small proportion of the information contained in the data set. The BBI for example is not differentiated per river type but is applied in a uniform way for all watercourses in Flanders (Fig. 1.2). In addition to these regional constraints, the existing indices do not include sufficient diverse response variables, which can distinguish among different types of impacts (Goethals and De Pauw, 2001).

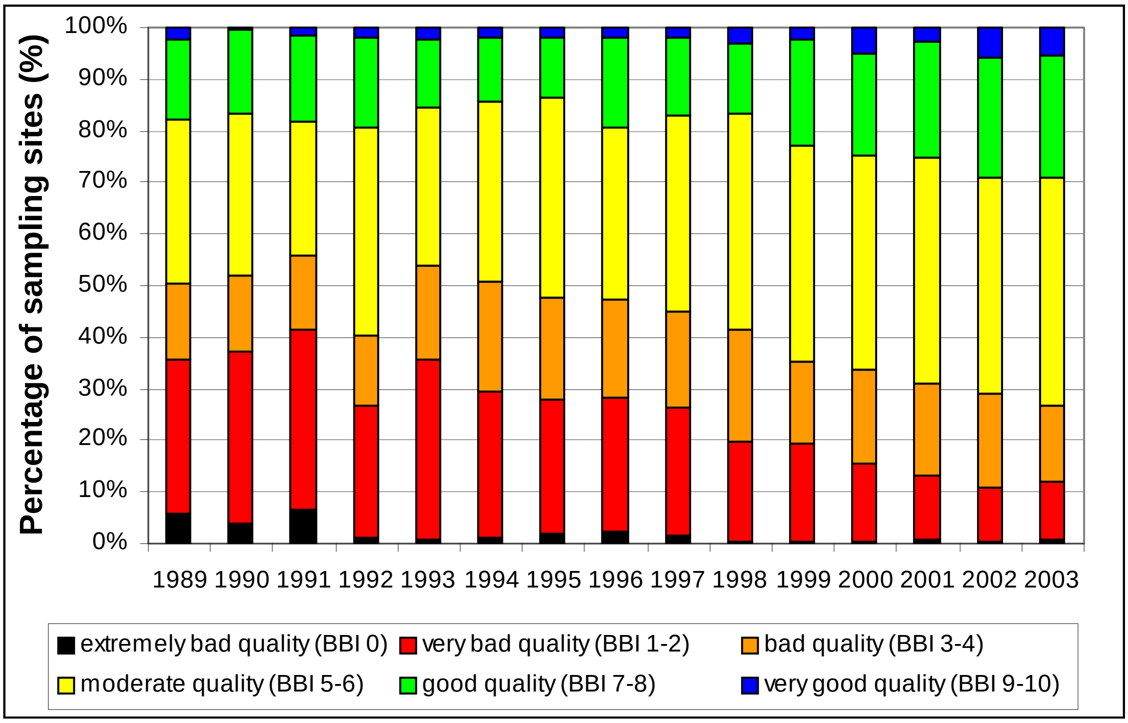


Fig. 1.2. Evolution of the biological quality of running waters based on the Belgian Biotic Index (BBI) (Flanders, period 1989 – 2003) (MIRA-T, 2004).

Integrated water management however supposes a thorough foundation when the juridical armamentarium is applied, water management plans are prepared or restoration measures have to be realized in the field. After all, restoration of deteriorated river systems for example can entail drastic social and economical consequences. In order to take valuable measures to restore disturbed water systems, it is not enough to perform ad hoc monitoring, launch individual project-based studies and apply the current assessment methods to evaluate them. Models on the contrary can act as interesting tools to support river quality monitoring (e.g. the development of effective and efficient monitoring networks) and assessment (e.g. to reduce the subjectivity of the assessment measures, to allow a better interpretation of the results, to increase the insight necessary to improve assessment systems), and decision-making in river management (Goethals and De Pauw, 2001).

In Flanders, several governmental administrations and institutions involved in water management develop and apply modelling techniques to support their decisions. In addition, a number of universities, consultancies and institutes develop their own models, mostly acting under government order. An overview of the currently used river models applied by the different governmental administrations and institutions is given in Table 1.2 (Cauwenberghs, 2003).

Table 1.2. Overview of the currently used river models applied by the different governmental administrations and institutions in Flanders (based on Cauwenberghs, 2003)

Model users	**Model**	**MODEL APPLICATION**
AMINAL Water	PDM	Hydrological modelling
	ISIS, InfoWorks RS	Hydraulic modelling
	FloodWorks	Operational river basin modelling
AQUAFIN	Hydroworks, InfoWorks CS	Sewer system modelling
	WEST, CHAT	Modelling of Waste Water Treatment Plants
AWZ	NAM	Hydrological modelling
	Mike 11, Delft 3D, ISIS	Hydraulic modelling
	Floodwatch	Operational modelling
	Delft 3D	Morphological and sediment transport modelling
IN	ITORS-Vl, NICHE-Vl	Eco-hydrological modelling
	Multsed	Sediment transport modelling
VLM	ISIS	Hydraulic modelling
VMM	SIMCAT, Pegase	Water quality modelling
	SENTWA	Nutrient transport modelling
	SEPTWA	Pesticide transport modelling
Province of Antwerp	PDM, ISIS, InfoWorks-RS	Hydraulic modelling
Province of Flemish Brabant	PDM, ISIS, InfoWorks-RS	Hydraulic modelling
Province of East Flanders	ISIS	Hydraulic modelling

In addition to the water quantity (hydraulic, hydrological) and quality models applied so far, also predictive models will be necessary to relate the different components within a system and get insight into the effect of the variables on each other. In particular models that can provide a predictive link between management actions and ecosystem response need to be developed. One way to assist predicting the ecological effect of specific management options is to develop habitat models, based on various hypotheses as to how environmental factors control the distribution of species.

Predictive modelling by means of habitat suitability models recently gaining importance as a tool in environmental management has been reviewed by Guisan and Zimmerman (2000). Only two ecological models however are actually used by the Flemish government: ITORS-Vl and NICHE-Vl (Cauwenberghs, 2003). ITORS-Vl predicts the habitat suitability for several macrophytes and vegetation types dependent on groundwater in the river valleys of the Kempen, Hageland and Loam region based on dynamical and chemical characteristics linked with shallow groundwater. This model is based on a multiple logistic regression method (IN, 2002). On the other hand, NICHE-Vl predicts the development and composition of groundwater dependent vegetation based on the habitat characteristics soil type, hydrology, nutritional richness and acidity.

A decision support system for river management based on predictive habitat suitability models can be defined as a system that is able to evaluate the suitability of different habitats for different species of wildlife and simulate management activities that modify features and obtain results from simulated activities in terms of changes in habitat suitability and the associated management costs (Adriaenssens, 2004). It should be an interactive tool, it should enable resource managers the ability to develop ‘what if’ scenarios, and should allow for graphically displaying changes in habitat suitability resulting from simulated alterations within a GIS (Garcia and Armbruster, 1997).

1.3. Habitat suitability modelling based on macroinvertebrates

1.3.1. Introduction

Predictive ecological models for environmental management are part of a much broader environmental modelling approach. The final objective of environmental modelling is to link hydrological, geomorphologic and ecological models to each other. The model can then be used to predict the effects of different land uses and water management scenarios on the selected ecosystems (Nilsson *et al.,* 2003). To describe these environmental processes, including both natural and human caused influences, a wide variety of modelling techniques or models exist. Ecological models, however, can be described as taking the ecology, more specifically the biota (and their interactions) as the main component of the models. These models try to reveal the influence of environmental variables on living organisms by linking abiotic with biotic variables.

Models developed with the aim of describing the habitat of a certain species in a predictive way are referred to as habitat suitability models. In the habitat suitability modelling approach, one wants to either express the suitability of a habitat for a specific species or use the information concerning the physical habitat of a certain species in order to predict the species' absence/presence or abundance. Often these models are also referred to as 'habitat distribution models' (e.g. Guisan and Zimmerman, 2000) or simply 'habitat models' (e.g. Yamada *et al.*, 2003).

The analysis of species-environment relationships, in particular the relationships between macroinvertebrates and their environment, has always been a main issue in ecology. Since the focus of this study is on macroinvertebrates, in the next paragraph an overview of the advantages and disadvantages of using macroinvertebrates in river monitoring, assessment and modelling is given. Afterwards, habitat suitability models based on macroinvertebrates will be discussed.

1.3.2. Advantages and disadvantages of macroinvertebrates

Predictive ecological models for use in river management can differ in biological endpoint. This can both depend on the conservation value of a specific group of organisms or either on their functionality as biological indicators of river conditions. For example, The EU WFD formulates its biological endpoints, referred to as biological quality elements for rivers, into five elements: phytoplankton, phytobenthos, macrophytes, macroinvertebrates and fish. Among these biological communities living in running waters, macroinvertebrates are good candidates to monitor ecological changes caused by human impacts (De Pauw and Hawkes, 1993; Karr and Chu, 1999). They 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 without magnification. In fact most of them are larger than 1 mm (e.g. Sladecek, 1973; Cummins, 1975; Wiederholm, 1980; De Pauw and Vanhooren, 1983; Rosenberg and Resh, 1993; Tachet et al., 2002).

The majority of aquatic macroinvertebrates has a benthic life and inhabits the bottom substrates (sediments, debris, logs, macrophytes, filamentous algae, …). For this reason, literature about biological water quality assessment methods 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. Sladecek, 1973; Hawkes, 1979; Helawell, 1986; Metcalfe, 1989; Rosenberg and Resh, 1993; Hering *et al.*, 2004). Macroinvertebrates are ubiquitous and abundant throughout the entire 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. canalization, impoundment, river regulation). 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. In this way, macroinvertebrates are ideally suited to predict and evaluate the effect of river restoration scenarios. In addition, there has been a tradition of macroinvertebrate monitoring in Flanders in relation to river quality assessment. More than twenty years of monitoring history has led to the development of river and river sediment quality assessment tools based on macroinvertebrates, namely the Belgian Biotic Index (BBI) (De Pauw and Vanhooren, 1983) and the Belgian Sediment Index (BSI) (De Pauw and Heylen, 2001). Water quality assessment based on macroinvertebrates forms as such the core of tools that is used by the government to assess and manage rivers nowadays, aside the Index of Biotic Integrity (IBI) that relies on fish communities (Belpaire *et al.,* 2000).

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 (in contrast to fishes) 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 also has its limitations (De Pauw and Hawkes, 1993). 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. Factors other than water quality are also important determinants of benthic communities. Of these the related factors of flow velocity and nature of the substrate 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. 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).

1.3.3. Habitat suitability models based on macroinvertebrates

Models can either be descriptive or predictive, i.e. they describe the functioning of a natural system or they make predictions of future functioning. To describe the natural functioning of rivers and streams, a wide variety of modelling techniques or models exist. In this study, the focus is on modelling techniques and models that link abiotic data with habitats for macroinvertebrates. Habitat models can be subdivided into two main categories (Harby *et al.*, 2004):

- process-based population or bio-energetic models;
- empirical-based habitat suitability models.

Process-based population or bio-energetic models incorporate knowledge on population dynamics of species and/or energy budgets for feeding, growth or other functions to describe biological processes. These models can either be linked upon the results of a physical habitat model or be directly linked with certain data describing the physical and physiographic environment. Bio-energetic models are a special type of biological process models where optimal species location is based on energy budgets (Jorgensen, 2002b). These models compute how much energy an organism uses as a function of flow velocity or turbulence and of food intake. The budget of energy intake and energy loss due to the current denotes the optimal location for the organism.

However, the stress of this work is on the development of habitat suitability models. As stated by Guisan and Zimmerman (2000), habitat suitability models are in general empirical. These empirical habitat suitability models are based on a description of the abiotic environment of species that is subsequently linked with the biotic system of flora and fauna described with the habitat concept. Univariate or multivariate functions link abiotic characteristics with habitat suitability. Univariate functions consider individual characteristics, while multivariate analysis takes into account the interaction of physical variables and determines the species response to the cumulative effect of a number of environmental characteristics. Several habitat suitability models based on the multivariate approach have been applied in water quality assessment using macroinvertebrates (Norris and Georges, 1993). Examples are (Harby *et al.*, 2004):

- regression (Montgomery and Peck, 1982) (e.g. multiple and logistic regression);
- ordination techniques using indirect and direct gradient analyses (Jongman *et al.*, 1995) (e.g. Principle Component Analysis, Canonical Correspondence Analysis);
- Artificial Neural Networks (Lek and Guégan, 1999);
- fuzzy rule-based functions (Zadeh, 1965);
- classification and regression trees (Breiman *et al.*, 1984), often referred to as decision trees when discussing both methods (Quinlan, 1986).

Until the last decade of the 20th century, limnology was mainly descriptive and diagnostic and less oriented towards prediction. However, bio-manipulation experiments not only highlighted the functioning (bottom up versus top down) of freshwater ecosystems, but also demonstrated that the achieved predictive capacity was sufficient for successful management of freshwater ecosystems (Zalewski and Roberts, 2003). At the end of the last century, RIVPACS (River Invertebrate Prediction And Classification System) (Armitage *et al.*, 1983; Wright *et al.,* 1984; Wright *et al.*, 1993; Wright *et al.*, 2000) has led to an increasing interest of the use of predictive models for river management. RIVPACS was the first model with a predictive capacity for use in river management. It was developed for assessing the ecological quality of rivers in Great Britain and based on the presence of macroinvertebrate communities

(Wright *et al.,* 1984). The RIVPACS philosophy is to develop statistical relationships, using the multivariate statistical technique of multiple discriminant analysis, between the fauna and the environmental characteristics of a large set of high quality reference sites that can be used to predict the macroinvertebrate fauna which is expected to occur at any site in the absence of pollution or other environmental stresses. The observed fauna at new test sites can then be compared with their site-specific expected fauna to derive indices of ecological quality (Clarke *et al.,* 2003). Based on the RIVPACS approach, other similar models have been developed: AUSRIVAS (Australian River Assessment System) in Australia (Norris and Norris, 1995; Smith *et al.*, 1999; Davies, 2000) and BEAST (Benthic Assessment of Sediment) in North America (Reynoldson *et al.,* 1995). However, the RIVPACS type model is empirical and descriptive and its primary use is as a tool for estimating and monitoring the ecological quality of river sites. RIVPACS cannot be used with any confidence as a dynamic model to predict the impact of environmental changes because data from impaired sites are not included (De Pauw, 2000; Clarke *et al.,* 2003).

Since the new millennium, a shift in use from multivariate statistical (based on data distribution functions, e.g. regression and ordination) to soft computing (based on heuristic search methods, e.g. artificial neural networks, decision trees) techniques has started. Reviews concerning the recently used techniques for the purpose of ecological modelling in general are given by Jorgensen (2002a), Recknagel (2002), Verdonschot and Nijboer (2002a), Adriaenssens (2004) and Goethals (2005). A more general review on the theoretical basis of most of the techniques can be found in Berthold and Hand (1999). The review of Verdonschot and Nijboer (2002a) can be situated within the European PAEQANN-project. The main applied objective of this project was to propose a set of predictive tools, mainly based on Artificial Neural Networks, for water management and water policies in order to allow assessment of the ecological quality and perturbations of river ecosystems. Since Artificial Neural Network models are also applied in this PhD study, this method will be described and reviewed in more detail in Chapter 2. Emphasis will be given on their use for the prediction of macroinvertebrates in rivers.

Chapter 2
State of the art of Artificial Neural Networks (ANNs) to predict macroinvertebrates in rivers

---

Parts of this chapter are based on:

Dedecker, A.P., Gabriels, W., Lek, S., Goethals, P.L.M. and De Pauw, N. (submitted). Applications of artificial neural networks predicting macroinvertebrates in freshwaters. *Journal of Applied Ecology*.

2.1. Introduction

Artificial Neural Networks (ANNs) are non-linear mapping structures that can be applied for predictive modelling and classification. The choice of the type of network depends on the nature of the problem to be solved. The most popular ANNs are multi-layer feed-forward neural networks with the backpropagation algorithm, i.e. backpropagation networks (Rumelhart *et al.*, 1986; Hagan *et al.*, 1996) and Kohonen Self-Organizing Maps (SOMs) (Kohonen, 1982). As the latter are mainly interesting for clustering data, they will not be further discussed in this review on predictive models for macroinvertebrates. The backpropagation 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 presence/absence or abundances of a number of macroinvertebrate taxa (such as Gammaridae (Crustacea, Amphipoda), Baetidae (Insecta, Ephemeroptera), Chironomidae (Insecta, Diptera)) which are being predicted based on a number of environmental variables such as dissolved oxygen, flow velocity, percentages of clay, silt and sand in the sediment, river depth, pH, …

2.2. General description of ANNs

The backpropagation network typically comprises three types of neuron layers: an input layer, one or more hidden layers and an output layer each including one or more neurons (Fig. 2.1). In a backpropagation network neurons from one layer are connected to all neurons in the following layer, but no lateral connections within any layer, nor feed-back connections are possible. With the exception of the input neurons, which only connect one input value with its associated weight values, the net input for each neuron is the sum of all input values xn, each multiplied by its weight wji, and a bias term bj which may be considered as the weight from a supplementary input equalling one (Fig. 2.2):

$$a_j = \sum w_{ji} x_i + b_j \qquad (3.1)$$

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

$$y_j = f(a_j) \tag{3.2}$$

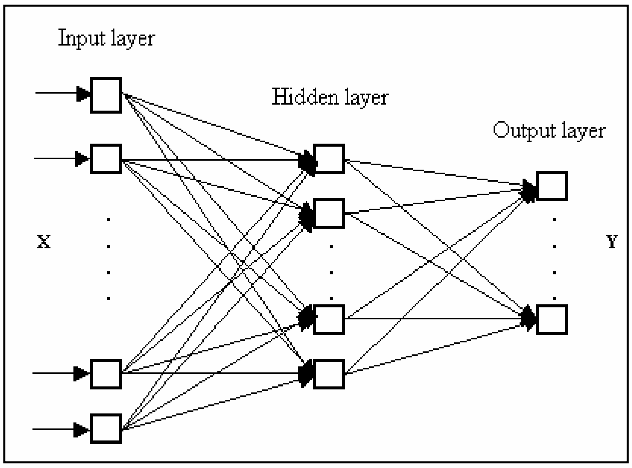


Fig. 2.1. Schematic illustration of a three-layered feed forward neural network including one input layer, one hidden layer and one output layer.

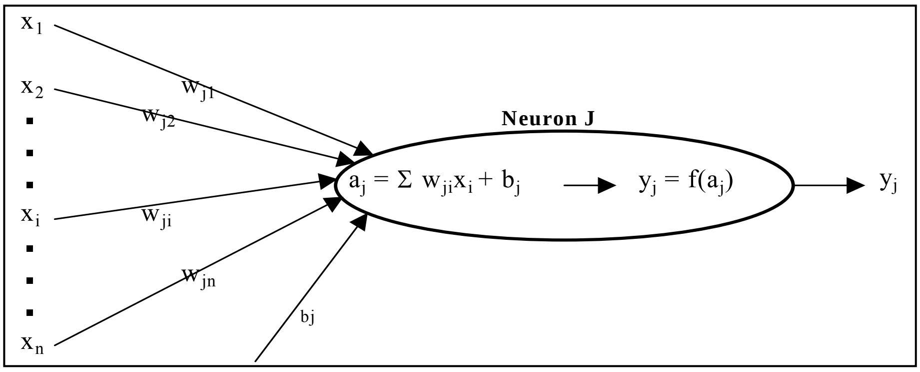


Fig. 2.2. Scheme of a neuron in a backpropagation network receiving input values from n neurons, each associated with a weight, as well as a bias b_j. The resulting output value y_j is computed according to the presented equations.

Before training, the values of the weights and biases are initially set to small random numbers. Subsequently, a set of input/output vector pairs is presented to the network. For each input vector, the output vector is calculated by the neural network model, and an error term is calculated for the outputs of all hidden and output neurons, by comparing the calculated output vector and the actual output vector. 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. A more detailed description can be found in Lek and Guégan (1999).

This chapter analyses ANN development and application studies to predict macroinvertebrates communities in aquatic ecosystems, as these communities have been proven to be good indicators for the assessment of surface waters as illustrated in the previous chapter. Based on this overview, suggestions to improve model development, assessment and application in ecological river management are presented.

2.3. Development of predictive ANNs

2.3.1. Data analysis and processing

2.3.1.1. 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 standardized. 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 (Chon *et al.*, 2001, 2002; Gabriels *et al.*, 2002; Gabriels *et al.*, 2005; Obach *et al.*, 2001; Park *et al.*, 2003a, 2003b; Schleiter *et al.*, 1999; Schleiter *et al.*, 2001; Wagner *et al.*, 2000) proportionally normalized the data between zero and one [0 1] in the range of the maximum and minimum values. Dedecker *et al.* (2004), Dedecker *et al.* (2005d), Gabriels *et al.* (2002) and Goethals (2005) on the other hand, rescaled the variables to be included within the interval [−1 1].

2.3.1.2. Band width

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.* (2005b) tested the sensitivity and robustness of the ANN models when data, containing variables with values beyond the range of the data for initial training, was added. Therefore, they created a virtual dataset based on ecological expert knowledge to introduce 'extreme' values to the model. The obtained results indicated that 'extreme' outputs in the test set were 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.

2.3.2. 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).

According to Walczak and Cerpa (1999), three steps 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 generalization performance. ANN input variables should not be highly correlated. Correlated variables degrade ANN performance by interacting with each other as well as 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). For macroinvertebrates, this can be done on the basis of spiking tests in situ or with artificial river systems for instance. A first 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). When a significant correlation ($P<0.01$) was found between two variables, Brosse *et al.* (2003) removed the one accounting for less variation in the single-scale models.

In addition, there are distinct advantages in using analytical techniques to help determine the inputs for ANN models (Maier and Dandy, 2000). However, these methods can merely be applied when large datasets are available. Beauchard *et al.* (2003), Obach *et al.* (2001), Schleiter *et al.* (1999) and Schleiter *et al.* (2001) 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, ... 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. Obach *et al.* (2001), Schleiter *et al.* (1999, 2001) and Wagner *et al.* (2000) 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. D'heygere *et al.*, 2005b; Obach *et al.*, 2001; Schleiter *et al.*, 2001). This technique automatically selects the relevant input variables (Goldberg, 1989).

2.3.3. Model architecture

According to Haykin (1999), generalization 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. Input variable selection was already discussed in the previous section.

2.3.3.1. Learning method

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-based predictions. 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 generalize 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 function 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 optimize 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 sigmoid 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 minimized 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) and takes into account the previous weight update. 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.

2.3.3.2. 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.

2.3.3.3. 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 aspect 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 generalize 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 feed-forward 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 2.1 shows the rules that suggest the number of hidden neurons based on the number of input (N_i) and/or output (N_o) nodes.

Table 2.1. 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

Some authors suggest rules to determine the necessary number of training samples (S) based on the number of connection weights. Given the number of training samples is fixed, inverting these rules gives an indication of the maximum number of connection weights to avoid overfitting (Table 2.2).

Table 2.2. 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

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).

2.3.4. Model validation

2.3.4.1. Validation

To test the model performance, a set with data independent from the training set is required (Lek and Guégan, 1999; Maier and Dandy, 2000). In the testing phase, the input patterns are fed into the network and the desired output patterns are compared with those given by the ANN model. The agreement or disagreement of these two sets gives an indication of the performance. As mentioned before, the data used for testing has to be within the range of the data used for training. It is also imperative that the training and test sets are representative of the same population. The optimal solution is to have two independent databases (Lek and Guégan, 1999). In this way, the first can be used for training and the second for testing of the model (e.g. Mastrorillo *et al.*, 1998; Obach *et al.*, 2001). However, when limited data are available, it might be necessary to split the available data into a training and a test set. A frequently used procedure, is the *k*-fold cross-validation method (e.g. Dedecker *et al.*, 2002; Dedecker *et al.*, 2004, 2005b, c, d; D'heygere *et al.*, 2005b). In this case, the data set is equally divided into *k* parts. The ANN model may be trained with *(k-1)* parts, and tested with the remaining part. Beauchard *et al.* (2003), Brosse *et al.* (2001), Brosse *et al.* (2003) and Guégan *et al.* (1998) for example used the 'leave-one-out' cross-validation method (Efron, 1983). This procedure is a generalization of the *k*-fold cross-validation, where *k* equals the sample size. This procedure is also appropriate when the amount of data is limited and/or when each sample is likely to have 'unique information' (Efron, 1983). Moreover, it has been found to be efficient for ANN modelling.

2.3.4.2. Performance measures

Based on the output, different performance measures can be distinguished. When presence/absence of the macroinvertebrates is predicted, the percentage of Correctly Classified Instances (CCI) is frequently used 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 2.3), proposed to assess the performance of presence/absence models (Table 2.4), Fielding and Bell (1997) and Manel *et al.* (1999) recommended Cohen's kappa as a reliable performance measure, since the effect of prevalence on Cohen's kappa appeared to be negligible (e.g. Dedecker *et al.*, 2004, 2005d; D'heygere *et al.*, 2005b).

Table 2.3. 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	+	TP	FP
	-	FN	TN

Table 2.4. Measures based on the confusion matrix to assess the performance of presence/absence models (after Fielding and Bell, 1997). CCI is the percentage Correctly Classified Instances, NMI is the normalized mutual information statistic and N is the total number of instances

Performance measure	**CALCULATION**
CCI	$\frac{(TP+TN)}{N}$
Misclassification rate	$\frac{(FP+FN)}{N}$
Sensitivity	$\frac{TP}{(TP+FN)}$
Specificity	$\frac{TN}{(FP+TN)}$
Positive predictive power	$\frac{TP}{(TP+FP)}$
Negative predictive power	$\frac{TN}{(FN+TN)}$
Odds-ratio	$\frac{(TP \times FP)}{(FN \times TN)}$
Cohen's kappa	$\frac{[(TP+TN)-(((TP+FN)(TP+FP)+(FP+TN)(FN+TN))/N)]}{[N-(((TP+FN)(TP+FP)+(FP+TN)(FN+TN))/N)]}$
NMI	$\frac{[-TP.\ln(TP)-FP.\ln(FP)-FN.\ln(FN)-TN.\ln(TN)+(TP+FP).\ln(TP+FP)+(FN+TN).\ln(FN+TN)]}{[N.\ln(N)-((TP+FN).\ln(TP+FN)+(FP+TN).\ln(FP+TN))]}$

When the output of the ANN 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 2.5).

Table 2.5. 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)	$\sqrt{\dfrac{1}{N}\sum(P-O)^2}$
Mean Squared Error (MSE)	$\dfrac{1}{N}\sum(P-O)^2$

2.3.5. 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. 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 (Brosse *et al.*, 2003; Chon *et al.*, 2001; Dedecker *et al.*, 2002, 2005d; Guégan *et al.*, 1998; Hoang *et al.*, 2001, 2002; Laë *et al.*, 1999; Marshall *et al.*, 2002; Mastrorillo *et al.*, 1997a; Olden and Jackson, 2002) 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). 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.*, 2005b, c). This method can thus be seen as a local sensitivity analysis. Gevrey *et al.* (2003), Dedecker *et al.* (2005b, c) and Beauchard *et al.* (2003) used the 'PaD' method (Dimopoulos *et al.*, 1995; Dimopoulos *et al.*, 1999) which consists in a calculation of the partial derivatives of the output according to the input variables. Several authors (Brosse *et al.*, 1999, 2001, 2003; Dedecker *et al.*, 2005b, c; Gevrey *et al.*, 2003; Mastrorillo *et al.*, 1997b; Olden and Jackson, 2002) applied Garson's algorithm (Garson, 1991; Goh, 1995). This algorithm is based on a computation using the connection weights. Gevrey *et al.* (2003) and Dedecker *et al.* (2005b, c) 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 randomization 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.

2.3.6. Model optimization

Traditionally, optimal network geometries have been found by trial and error (Maier and Dandy, 2000). However, a number of systematic approaches for determining optimal network geometry have been proposed, including pruning and constructive algorithms. The basic thought of pruning algorithms is to start with a network that is large enough to capture the desired input-output relationship and to subsequently remove or disable unnecessary weights and/or neurons. A review of pruning algorithms is given by Reed (1993). Constructive algorithms approach the problem of optimising the number of hidden layer neurons from the opposite direction to pruning algorithms. The smallest possible network is used at the start.

Hidden layer neurons and connections are then added one at a time in an attempt to improve model performance. A review of constructive algorithms is given by Kwok and Yeung (1997a). Several disadvantages of these approaches are mentioned in the literature (Maier and Dandy, 2000). For example, the networks generally have to be trained several times, i.e. each time a hidden neuron is added or deleted (Kwok and Yeung, 1997b). It has also been suggested that the pruning and constructive algorithms are susceptible to becoming trapped in structural local optima (Angeline *et al.*, 1994). In addition, they 'only investigate restricted topological subsets rather than the complete class of network architectures' (Angeline *et al.*, 1994). Algorithms based on evolutionary programming and genetic algorithms have been proposed to overcome these problems and have been used successfully to determine optimal network architecture (e.g. Fang and Xi, 1997; Kim and Han, 2000; Zhao *et al.*, 2000; Wicker *et al.*, 2002). Evolutionary approaches are significantly different from the previous techniques described. They produce more robust solutions because they use a population of networks in the search process. A complete review of the use of evolutionary algorithms in neural networks is given by Yao (1993). Beside the optimization of the network geometry, input variable selection can also be seen as model optimization. This has already been discussed in Section 2.3.2.

2.4. Applications of macroinvertebrate predictions using ANNs in water management

Table 2.6 gives an overview of articles discussing case studies on the prediction of macroinvertebrates by means of ANNs. A total of 27 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 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. Several of the modellers not mentioning the software package use their own code, implemented in an existing modelling environment such as MATLAB. Evidently, the software package used should not influence the modelling results although neither the use of own programming nor existing software is an absolute guarantee that small errors will not occur, which means that any system should be used with care.

The number of input variables ranged from 3 to 39, usually between 5 and 15. These variables included geographical, seasonal and habitat quality parameters (sinuosity, vegetation, …) as well as physical-chemical properties (dissolved oxygen, water temperature, pH, nutrient concentrations, COD, …) and characteristics of toxicity. The performance of neural networks with more input variables was not necessarily higher, as shown in some studies (e.g. Walley and Fontama, 1998). The target variables were usually presence/absence (10 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 feed-forward connection type, in some cases combined with a Self-Organising Map. 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 feed-forward 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.

The 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’ (7 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 optimization were hardly used to optimize network geometry.

The transfer functions, where specified, were of the sigmoid transfer function type.

The data were generally rescaled to the interval [-1 1] or [0 1]. Maier and Dandy (2000) 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 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 or the proportion (PI) 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. 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.

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.

Table 2.6. Overview of publications discussing case studies on the prediction of macroinvertebrates by means of artificial neural networks

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	?	A, P, Lo, R, DISTs, SDA	richn	Morocco, Algeria, Tunisia	FF	BP	7-4-1	210-1 (leave-one-out)	Empirically	STF	?	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)	?	STF	?	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)	?	STF	?	r, MSE
Céréghino *et al.*, 2003	?	A, SO, DISTs, T	richn	Adour-Garonne river basin (France)	FF	BP	4-5-1	130-25	Trial and error	?	?	r
Chon *et al.*, 2001	?	MI, FV, D, OM, S	comm	Yangjae stream (Korea)	RTRC	RL	(7+4)-13-7	?	Trial and error	?	[0 1]	r
Chon *et al.*, 2002	?	MI	dens	Yangjae stream (Korea)	FF	BP	(5-25)-(8-30)-5	?	Empirically	STF	[0 1]	r
					ERC	BP	5-30-5					
		MI, FV, D, OM, S	comm		RTRC	RL	(7+4)-13-7					
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	?	CCI
Dedecker *et al.*, 2004	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, CK
Dedecker *et al.*, 2005b	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)	?	STF	IN? OUT[log (abu+1)]	r
Dedecker *et al.*, 2005c	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)	?	STF	IN? OUT[log (abu+1)]	r

Dedecker *et al.*, 2005d	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-?-1	108-12 (10-fold)	Trial and error	STF	[-1 1]	CCI, CK
D'heygere *et al.*, 2005b	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)	?	?	?	CCI, CK, 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	?	IN[-1 1], OUT[0, 1]	r, CCI
Gabriels *et al.*, 2005	?	Day, W, D, FV, S, pH, DO, Cond, Ni, Ph	pr/ab	Flemish river sediment (Belgium)	FF	BP	12-?-(1, 92)	294-49 (7-fold)	Trial and error	?	[-1 1]	CCI, CK
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	?	CCI
Goethals, 2005	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	pr/ab and abu	Zwalm river basin (Belgium)	FF	BP	24-10-1	119-60	?	STF	?	CCI, CK, r
		Day, W, D, FV, S, T, pH, DO, Cond, TOX, TOC, OM, Ni, Ph, Met		Flemish river sediment (Belgium)				228-114				
Hoang *et al.*, 2001	?	A, SO, R, SoilT, VEG, S, T, …	pr/ab	Queensland streams (Australia)	FF	BP	39-15-37	650-167	?	STF	?	CCI
Hoang *et al.*, 2002	?	SO, Lo, Ni, …	pr/ab	Queensland streams (Australia)	FF	BP	?	?	?	STF	?	CCI
Marshall *et al.*, 2002	?	A, SO, R, SoilT, VEG, S, T, …	pr/ab	Queensland streams (Australia)	FF	BP	39-15-37	650-167	?	STF	?	CCI
Obach *et al.*, 2001	?	T, DI, P	abu	Hesse (Germany)	FF	Mod BP	?	?	?	?	[0 1]	R², RMSE
					FF	GRNN						
					FF	LNN						
					SOM	RBF	?-120			N/A		
Park *et al.*, 2001	?	Ex		Suyong river (Korea)	SOM	KLR	?	?	?	N/A	[0.01 0.99]	r
		Comm	Ex		FF	BP	?-5-?			STF		

Park *et al.*, 2003a	?	EPTC A, SO, DISTs, T	EPTC	Adour-Garonne river basin (France)	SOM FF	KLR BP	?-140 4-?-1	130-25	?	N/A ?	[0 1]	r
Park *et al.*, 2003b	?	S, VEG, DO, W, Cond, T, FV, D, NO_3^-, PO_4^{3-}, NH_4^+, …	richn, SH	The Netherlands	Forward only	CPN	?	500-164	?	?	[0 1]	r
Schleiter *et al.*, 1999	?	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	?	150-150 200-100 225-75	?	?	[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	?	45-6	?	?	[0 1]	R^2, RMSE, CVE
Wagner *et al.*, 2000	?	T, P, DI	abu	Breitenbach (Germany)	FF	BP	?	216-54	?	?	[0 1]	R^2
Walley and Fontama, 1998	?	Coord, DISTs, SL, Alk, DI, A, S, W, D	ASPT, NFAM	UK	FF	BP	13-6-6-1	307-307 (2-fold)	?	?	No, log (DISTs), log(SL)	r

?=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; Met=metals; 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=feed-forward; 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; CK=Cohen's kappa; CVE=cross-validation error; 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

Actually, there is almost no insight into the practical usefulness of ANN models in decision support systems. Most articles only discuss the development of the models and evaluate these 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 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 dynamics in the river systems (e.g. flow velocities, water temperatures). 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 often are related to the environment in which they are released (bio-availability, accumulation, …). 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 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 optimize the parameters that control neural network architecture.

Presently, there is a lack of comparative papers (e.g. Skidmore *et al.*, 1996; Manel *et al.*, 1999) in which more than two statistical methods have been applied to the same data set. Most published ecological modelling studies use only one of the many techniques that may be properly 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 that show the advantages and disadvantages of different methods in a particular context (Guisan and Zimmermann, 2000).

When looking at the different soft computing techniques, they all seem to have particular strengths and weaknesses. ANNs can give for instance well performing models, but the integration of expert knowledge is difficult. Fuzzy logic can be used to develop models merely on expert knowledge, but the number of input variables is very limited, because the rule sets become very complex when more than five input variables are used. Bayesian Belief Networks have the interesting characteristic to be able to reveal how different variables interact, on the other hand, the information necessary to build these networks and to set up the variable distributions is also huge.

Therefore, based on the rather limited set of case studies in which several methods were compared, it is up to now nearly impossible to have clear insight into when to use what kind of method. On top of this, also several methods such as Bayesian Belief Networks and fuzzy logic were rarely applied yet in ecology, so the methods themselves need further exploration. Important to note is that many different measures are used to evaluate model performance, which makes it difficult to compare different studies Additionally, it must be mentioned that the practical applicability of the models for decision support (ecological knowledge extraction, predictions) is also a crucial quality aspect.

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.*, 2004).

Ceccaroni *et al.* (2004) stress the importance of ontology 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 about those processes which 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 as well be a feasible option. A major issue in this context is the selection of the most convenient inference techniques, hereby considering a realistic process description, relevant outputs for the users and a low simulation time. Also here an interesting role can be played by ANN models to link several variables of the different system components. In particular in the field of hydrological modelling, several studies proved that also ANN models can provide good results combined with short calculation times.

In particular for river and lake restoration management, there is a need for tools to guide the investments necessary to meet a good ecological status as set by the European Water Framework Directive. Although many ecological modelling methods exist for several decennia, their practical application to support river and lake management is rather limited. Many river managers are still unaccustomed using these ecological models. One of the reasons is that the required knowledge to use these methods is not straightforward and readily available. In addition these managers are not always involved in the model development process resulting in a reserved attitude towards these models and their results. Although there is quite some experience gained with data driven models to predict macroinvertebrates, several key-questions remain unanswered with regard to the optimal architecture, the input variables and the ecological relevance of the models.

Therefore, in this thesis (Chapter 6) the optimal ANN model architecture for the prediction of five macroinvertebrates, having a different sensitivity to pollution and prevalence in the study area (Chapter 4), based on environmental river characteristics is investigated. The use of sensitivity analyses is probably a major need to increase the credibility of these often called 'black box' methods to ecologists and river managers. Additionally, the contribution of ecologically relevant input variables, is another crucial aspect that needs more research. The latter two bottlenecks are tackled in this thesis (Chapter 7) applying three input variable contribution methods to the ANN models. In this way, the 'black box' models could be clarified and the ecological relevance of the input variables could be assessed.

Chapter 3
Modelling migration behaviour of macroinvertebrates

3.1. Migration behaviour of macroinvertebrates in running waters

3.1.1. Introduction

In a stream, many activities take place over a broad range of spatial and temporal scales. The benthic macroinvertebrates have actually been described as being in a state of 'continuous redistribution' (Townsend and Hildrew, 1976). One obvious reason for movement relates to the flowing nature of the water in rivers and streams. The drag force of flow can cause dislodgement and subsequent downstream transport of individuals (Williams and Williams, 1993). However, also other physical, chemical and biological factors are involved (Williams and Feltmate, 1992), hence movements are not purely passive in nature. For macroinvertebrates, movements include downstream drift in the water (= passive), swimming or crawling from upstream or downstream (= active), and for insects aerial movements by adult stages along or between river channels. Macroinvertebrates are generally considered to have high dispersal capabilities, but these migration and mobility abilities do vary considerable among invertebrate taxa (Giller and Malmqvist, 1998).

A difference should however be made between small-scale and medium- to large-scale movements (Giller and Malmqvist, 1998). Many macroinvertebrates appear to show small-scale daily changes in their distribution on individual stones, tending to occupy lower surfaces during the day, but upper surfaces at night (Elliott, 1971a; Glozier and Culp, 1989). This activity is usually seen for grazer/scraper species (such as many mayflies) and is associated with daily patterns in feeding. Short-term small-scale movements also include the response to the presence of predators or competitors. Movements into refugia during periods of peak flow are of a similar scale. Studies on individual fish have also shown daily movements and swimming between pools and riffles (Greenberg and Giller, 2000). Movements of individuals over relatively longer time periods and larger spatial scales, give some idea of the potential size of the home range of the individual (the area over which the macroinvertebrate normally migrates) and perhaps the net movement over the lifetime of an individual (e.g. Erman, 1986). Large-scale dispersal between streams and rivers normally requires some terrestrial or aerial migration and is thus largely restricted to insects. The extent of such movements for insects has been relatively poorly studied, but is clearly the most important route of colonization of severely disturbed river basins or new channels (Giller and Malmqvist, 1998).

As discussed before, migration behaviour and in particular the upstream and downstream movement of animals within and between habitats is of specific ecological significance and is essential for the maintenance of gene flow and genetic diversity, and also for colonization and re-colonization of potential habitats (Clobert *et al.*, 2001; Bullock *et al.*, 2002).

In the following parts, the major aspects of migration like drift and active upstream and downstream movements are discussed in more detail.

3.1.2. Downstream migration behaviour of macroinvertebrates: drift and active movements

In running waters, the drift of macroinvertebrates is one of the most ubiquitous and widely studied phenomenon (Waters, 1965; Brittain and Eikeland, 1988). The term 'invertebrate drift' describes the downstream dispersal in the water of benthic invertebrates that usually live on or amongst the substratum of stream and rivers (Elliott, 2002a, b, c). Drift enables organisms to escape unfavourable conditions and gives them the potential to colonize new habitats (Brittain and Eikeland, 1988).

Drift can be classified into four categories based on Waters (1972) and Brittain and Eikeland (1988):

- <u>constant drift</u>: continuous, accidental displacement;
- <u>distributional drift</u>: a method of dispersal, especially in the very young stages soon after egg hatching;
- <u>catastrophic drift</u>: movements resulting from major physical and chemical disturbances (e.g. high discharge);
- <u>behavioural drift</u>: periodical, resulting from daily patterns of activity or avoidance of predators, competitors or other stressors. The latter form of behavioural drift is often termed 'active drift', to distinguish it from drift due to general activity.

Many biotic and abiotic variables have been cited as factors influencing drift: current/discharge, photoperiod (daily or seasonal patterns), water chemistry, benthic densities, predators and life-cycle stage (Brittain and Eikeland, 1988). Passive drift can occur through simple accidental erosion by the force of moving water during the course of normal activity and increasing shear stress. Other physical stresses such as rapid flow reduction and pollution also increase drift. One problem in determining what causes drift is that it can be difficult to prove whether entry to the drift is accidental (= passive) or not (= active), which in turn relates to the various causes of drift (Giller and Malmqvist, 1998). In some of these responses, active entry is probable, as it provides an immediate escape mechanism (e.g. escape from predators and competitors). The use of drift may also provide a useful mechanism of actively searching for new food patches or preferred habitat (Kohler, 1985). Any behaviour that increases the probability of exposure to the eroding current can increase the chance of passive drift entry, however. This may occur if macroinvertebrates increase their activity on the top surfaces of stones. Drift entry is almost certainly determined to an extent by accidental dislodgement resulting from such daily changes in foraging movements and activity (Giller and Malmqvist, 1998).

3.1.3. Active upstream migration behaviour of macroinvertebrates

Upstream migration of many different taxa has been reviewed by Söderström (1987). Unlike drift, upstream movements are against the current direction and are thus always non-accidental but active. The major reasons for upstream migration seem to be:

- search for new habitat and food (Bishop and Hynes, 1969; Hildrew, 1996);
- avoidance of unfavourable abiotic conditions (Hayden and Clifford, 1974; Olsson and Söderström, 1978);
- search for suitable emergence, pupation or mating sites (Hultin *et al.*, 1969; Goedmaker and Pinkster, 1981);
- a compensation for drift in order to maintain a basal population within a certain habitat (Hayden and Clifford, 1974; Goedmaker and Pinkster, 1981).

3.1.4. Aerial migration behaviour of macroinvertebrates

This type of migration is largely restricted to the adult stages of insects. Aerial migration is for insects relatively poorly studied, but is clearly the most important route of colonization of severely disturbed river basins or new channels. The rates of recolonization depend on the size of the disturbed area and the proximity of a source population and hence distance that these organisms must travel (Giller and Malmqvist, 1998). These rates of colonization also depend on the taxa, largely because the taxa themselves vary in their propensity to move and modes of movement. Newly opened stream channels connected to other river stretches for example can be rapidly colonized by small-scale movements (drift or active upstream migration). Isolated streams or restored river sections far from potential source populations however can only be recolonized by means of aerial migration.

3.1.5. The 'drift paradox'

Discussing the downstream drift and upstream movement of macroinvertebrates without mentioning the 'drift paradox' is however inappropriate. The 'drift paradox' is indeed a well-known and frequently discussed concept in the ecological literature. The 'drift paradox' arises because the upper reaches of the 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; Giller and Malmqvist, 1998). The compromise has generally been that upstream compensation 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 does one explain the persistence of the many species that are commonly found to drift but do not have an aerial adult stage (Humphries and Ruxton, 2002)? Speirs and Gurney (2001) and Humphries and Ruxton (2002) demonstrated that long-distance flight by aerial adults is not required to prevent extinction of upstream reaches. It can be achieved by very small upstream movements of individuals along the substrate.

3.2. Models to simulate migration behaviour

Models are seen as useful tools to simulate migration behaviour of all kinds of fauna and flora. During the preceding years, several models are developed for the dispersal of mammals (e.g. Andrén, 1994; Ferreras, 2001; Verbeylen et al., 2003), birds (e.g. Graham, 2001; Green and Alerstam, 2002; Erni et al., 2003; Tankersley and Orvis, 2003), insects (e.g. Kareiva, 1983; Pitcher and Taylor, 1998; Baguette et al., 2003; Chardon et al., 2003), plants (e.g. Jung et al., 2004), … However, these studies were mainly describing movements of terrestrial organisms. The objectives of these migration studies were very diverse: biological conservation, ecological research, examining the effect of human impacts, optimizing the models themselves, … Migration studies concerning aquatic organisms have mainly been focussing on fish. Several examples can be found in the literature (e.g. Schonfisch and Kinder, 2002; Hubbard et al., 2004; Humston et al., 2004; Magnusson et al., 2004; Mommsen, 2004; Nykanen et al., 2004; Steel et al., 2004). These models have generally been designed from a commercial point of view.

Although there is a lot of knowledge available describing migration behaviour of macroinvertebrates, studies modelling their migration possibilities are rare. Literature, covered by the ISI Web of Knowledge[SM] from 1972 till 2005, was reviewed. This search resulted in only six papers describing the application of models to simulate migration behaviour of macroinvertebrates (Hayes *et al.*, 2003; Largier, 2003; Ellien *et al.*, 2004; Englund and Hambäck, 2004a, b; Weber *et al.*, 2004), which are reviewed in Table 3.1.

Table 3.1. Overview of the studies discussing the migration of macroinvertebrates based on literature, covered by the ISI Web of Knowledge[SM] from 1972 till 2005

Reference	**Model type**	**Migration component modelled (modelled species)**	**Objective of the model**	**Model scale**
Ellien *et al.*, 2004	2D hydrodynamic model	Dispersion (*Pectinaria koreni*)	Analyse the processes involved in the transport and dynamics of larval populations	Bay of Seine
Englund and Hambäck, 2004a, b	1D and 2D analytical and random walk model	Upstream and downstream migration (*Baetis* sp. and *Asellus aquaticus*)	Analyse the patch-scale-dependence of migration rates	Micro-meso habitat scale
Hayes *et al.*, 2003	2D hydraulic model	Drift	Predict the food available to drift feeding brown trout	Reach scale
Largier, 2003	Advection-diffusion model	Dispersion (coastal populations)	Determine the larval dispersal distances and larval origins in oceanography	Coastal zones
Weber *et al.*, 2004	3D fluid dynamic model	Drift (*Gammarus pulex*, Simuliidae, Baetidae)	Predict the food available to drift feeding fishes	Reach scale

The studies describing the migration of river macroinvertebrates examined only aquatic dispersion (drift, up and downstream migration). In general, the models were very detailed applying for example a particle tracking technique in which the macroinvertebrates were treated as particles that enter the stream (e.g. Weber et al., 2004). Additionally, the developed models could only be applied over a small scale (micro-meso habitat scale to reach scale). However, if models have to be used in river restoration or conservation management (i.e. to investigate the connectivity between population patches or the possibility to migrate from a source population to recolonize a restored river section), also migration over land / through the air covering a broader scale should be included. To this end, these models were less appropriate and more robust ones are required.

Recently however, some studies used 'least-cost' modelling as an approach to incorporate detailed geographical information as well as behavioural aspects in a measure of connectivity and migration (e.g. Michels *et al.*, 2001; Schadt *et al.*, 2002; Adriaensen *et al.*, 2000, 2003; Chardon *et al.*, 2003). Based on this type of models, upstream and downstream migration through the river, as well as migration through the air (based on the species' movement behaviour through the water and the air, and the river and environmental characteristics influencing the migration) could be integrated. In this way, it was opted for to apply this type of models to reach the objectives of this thesis. A detailed description of the specific characteristics and requirements of these models is further discussed in Chapter 8.

Chapter 4
Study area and data collection

<u>Parts of this chapter are based on</u>:

Dedecker, A.P., Janssen, K., Wouters, H., Thas, O., Goethals, P.L.M. and De Pauw, N. (in preparation). Explaining the variability in data collection for river assessment purposes. *Freshwater Biology*.

4.1. Introduction

In this chapter, the applied data collection techniques are discussed, both for biotic and abiotic monitoring of running waters in the study area. Two datasets were gathered in this study area: (1) the Zwalm river basin (Flanders, Belgium), referred to as the Zwalm river basin dataset, and (2) the brooks Verrebeek and Dorenbosbeek and the upstream part of the Zwalm river basin, referred to as the 'short distance' monitoring network dataset. These monitoring data were used to fulfil the following aims:

- **Development, optimization and validation of habitat suitability models based on ANNs**.
  For the development of the ANN models, the measured data were used to define the optimal size of training and test set and to determine the optimal model architecture for both datasets. In addition the monitoring data were used to validate the habitat suitability performance (Chapter 6).

- **Ecological interpretation of ANNs applying input variable contribution methods**.
  When using the data for information extraction, the main objective in this study was to determine the most important environmental variables that influence the macroinvertebrates and which are of interest for river management. To find out the importance, three input variable contribution methods were applied (Chapter 7).

- **Development of migration models for macroinvertebrates as an extension of the habitat suitability models**.
  Because in general, habitat suitability models do not include spatial and temporal relationships, migration models were developed and validated based on sensitivity analysis to include migration dynamics of the predicted organism and migration barriers along the river. To this end, the 'short distance' monitoring network dataset was used (Chapter 8, Chapter 9).

- **Prediction and evaluation of river restoration scenarios**.
  Finally, the collected data were used to predict and evaluate river restoration scenarios in the Zwalm river basin (Chapter 10).

4.2. General description of the Zwalm river basin

The Zwalm river basin is part of the hydrographical basin of the Upper-Scheldt in Flanders, Belgium (Fig. 4.1). The basin drains an area of about 11,650 ha, the Zwalm River itself has a length of 22 km. The Zwalm river basin is divided into two administrative zones, referred to as VHA zones, and runs through different municipalities. Concerning river management in Flanders, responsibilities are divided over different administrations based on the given category of the watercourse. Responsible administrations include the Flemish government (represented by AMINAL), the provinces and the municipalities (Fig. 4.1). This means that river management of the river basin is partitioned at different levels, although consultation between the responsible administrations and parties concerned is done by the river basin (management) committees.

According to the river typology of the Flemish watercourses defined by Jochems *et al.* (2003), streams within the Zwalm river basin are of the river type small and large streams (Fig. 4.2).

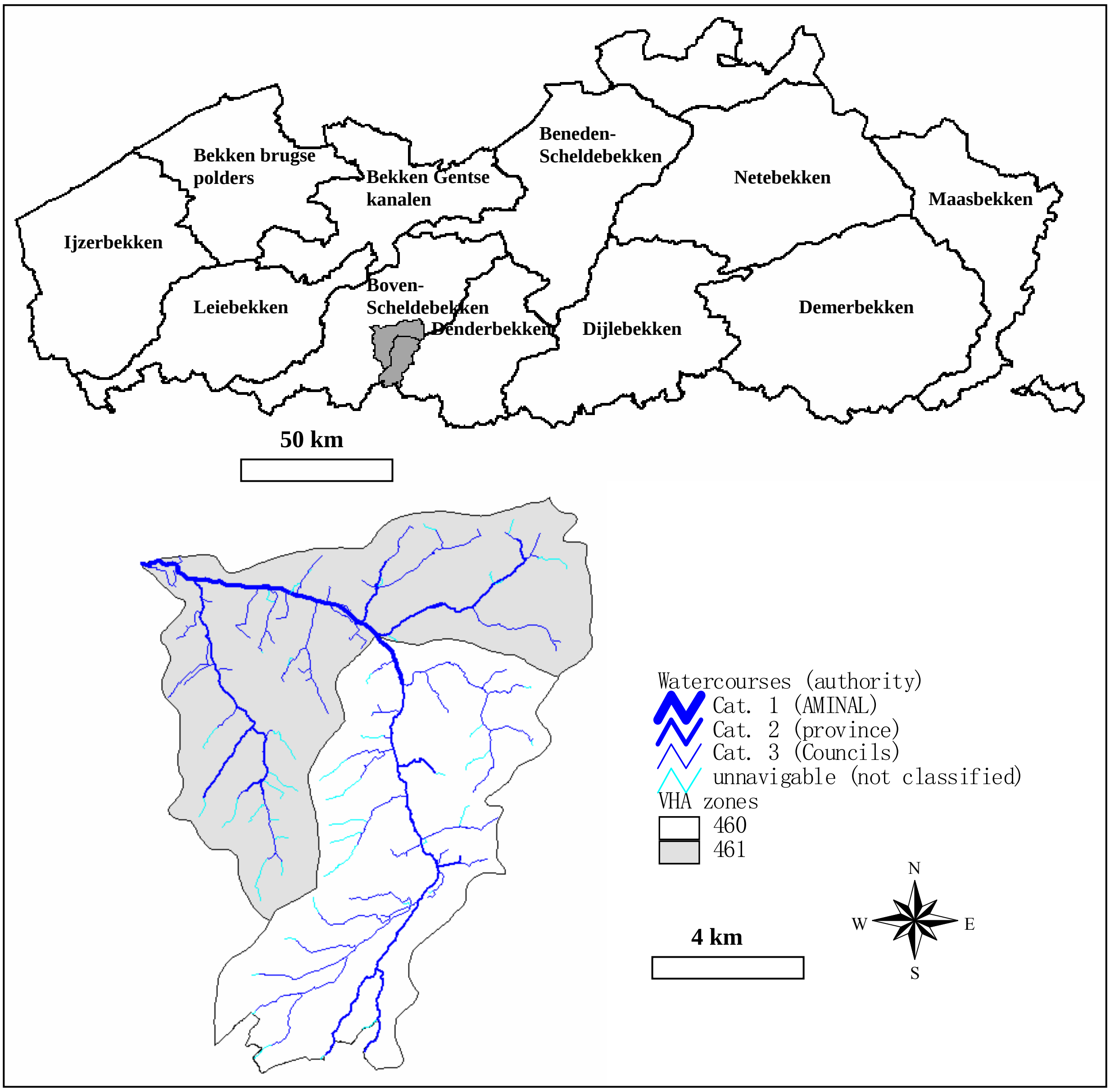


Fig. 4.1. The Zwalm river basin, located in the Upper-Scheldt (cf. Boven-Schelde) basin in Flanders (Belgium).

The definition of a river type is dependent on the eco-region, the river basin surface area and the stream order as given in Fig. 4.3. These small and large streams, often referred to as brooks can be found within the eco-region ‘sand and sand loam area’ and are of a high stream order in the river basin, in contrast to the rivers.

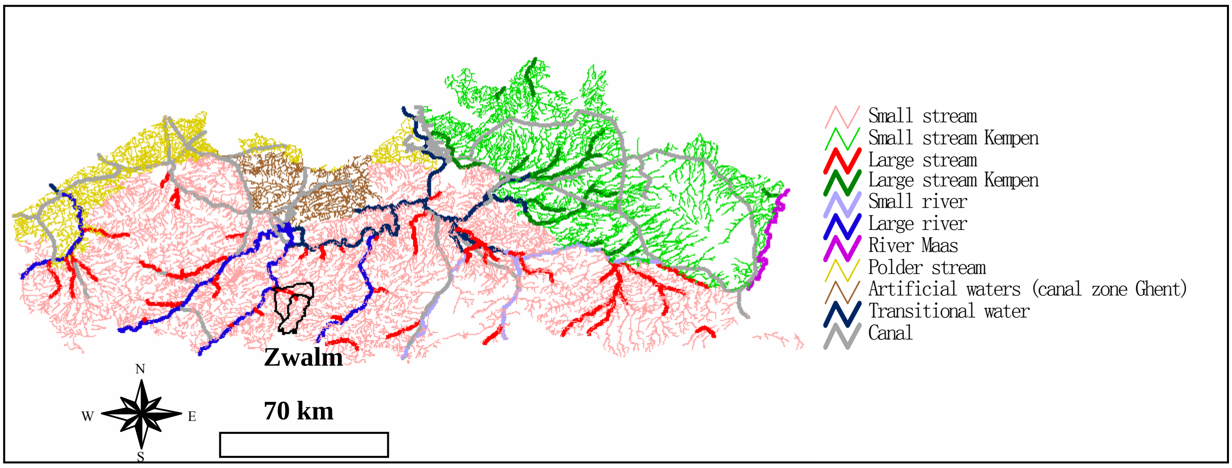


Fig. 4.2. River typology of watercourses in Flanders (Jochems *et al.*, 2003) with an indication of the Zwalm river basin.

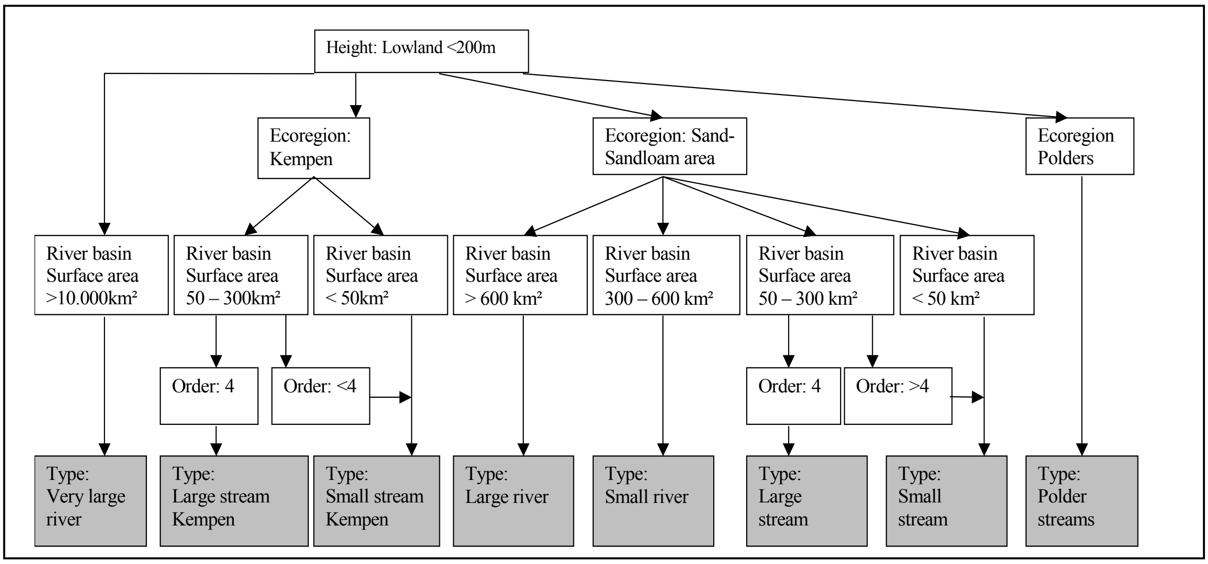


Fig. 4.3. Hierarchical construction of the river typology according to system B as defined by the WFD (Jochems *et al.*, 2003).

Since 1999, the water quality in the Zwalm river basin has considerably improved due to investments in sewerage and wastewater treatment plants during the preceding years (VMM, 2003). Nevertheless, several parts of the river are still polluted by untreated urban wastewater and by diffuse pollution originating from agricultural activities (Goethals and De Pauw, 2001). Although Flanders is in general a rather flat region, the Zwalm river basin is characterized by a number of differences in altitude, making it a quite unique ecosystem. 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 still exist (e.g. weirs for water quantity control, artificial embankments, …). A description of the major sources of stress due to human activities in the Zwalm river basin is given in Table 4.1. In the context of drawing up an ecological inventory and the development of a vision, Konings and Meire (2003) monitored the current status of the structural habitat quality of the watercourses in the Zwalm river basin. To this end, river stretches of about 100 m in length were monitored by means of the structural index developed by Bervoets and Schneiders (1998). This index is based on a scoring system that assesses the presence of hollow banks, pools and riffles and the degree of meandering in a river reach. The results for the Zwalm river are shown in Fig. 4.4 (MIRA-T, 2004).

Table 4.1. Description of the main disturbances caused by human activities in the Zwalm river basin (Goethals and De Pauw, 2001)

Human impacts
Physical-chemical disturbances
Point sources:
• Effluents WWTPs (urban and industrial)
• Combined sewer overflows
• Sewer systems
• Accidents (mainly leaking fuel storage tanks)
• Feeding of animals, fishing
Diffuse sources:
• Agriculture
• Traffic
• Scattered housings
Structural and morphological disturbances
• Water quantity management (weirs, artificial embankments)
• Transport infrastructure
• Physical pollution (wood debris, large wastes)
Biological disturbances
• Fishing
• Rat traps
• Sampling related to monitoring
• Fish stocking (angling management, pond overflows)
• Game hunting

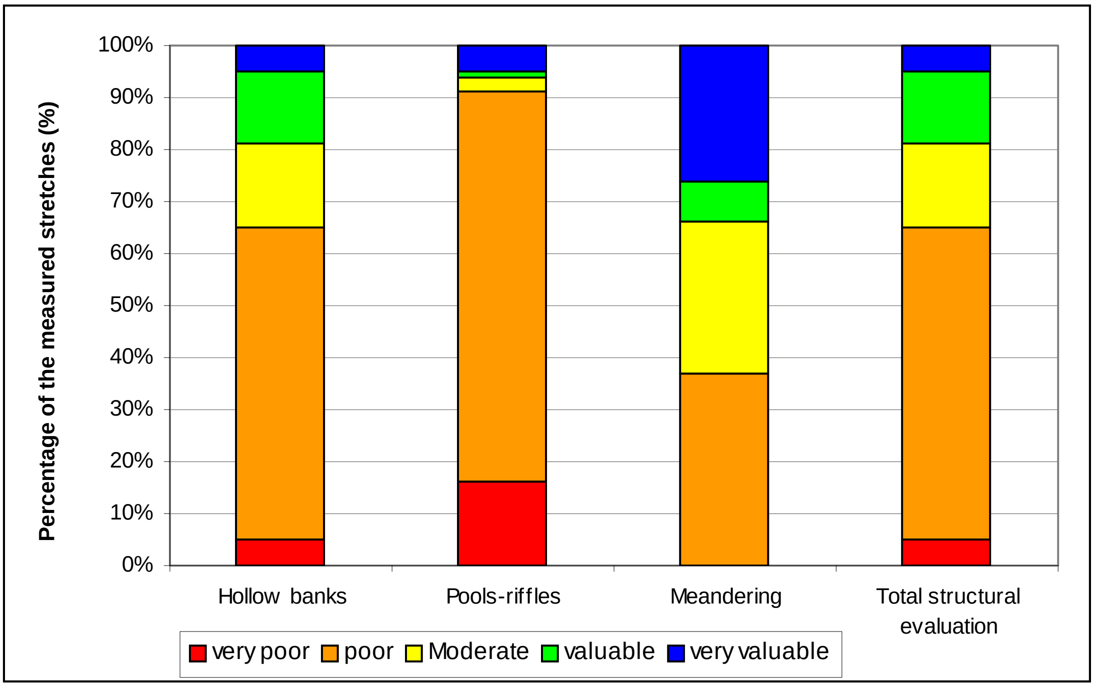


Fig. 4.4. Structural evaluation of the Zwalm river based on Konings and Meire (2003) in MIRA-T (2004).

The significant human impact on the Zwalm river basin can be demonstrated by the Belgian Biotic Index (BBI) (Fig. 4.5). The BBI method uses macro-invertebrates as indicators for the level of pollution (De Pauw and Vannevel, 1991). 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 index values are interpreted as follows and can be mapped by means of a colour code:

- 0 = extremely bad quality (black);
- 1 - 2 = very bad quality (red);
- 3 – 4 = bad quality (orange);
- 5 – 6 = moderate quality (yellow);
- 7 – 8 = good quality (green);
- 9 – 10 = very good quality (blue).

From the BBI maps, one can read off a good to moderate quality in the upper stream parts, and a bad quality in the more downstream parts of the Zwalm river basin. Only a limited number of the measured sites reaches the quality standard set by VLAREM II (VLAREM, 1999), which means at least a 'good quality' corresponding with a BBI value of 7 and more.

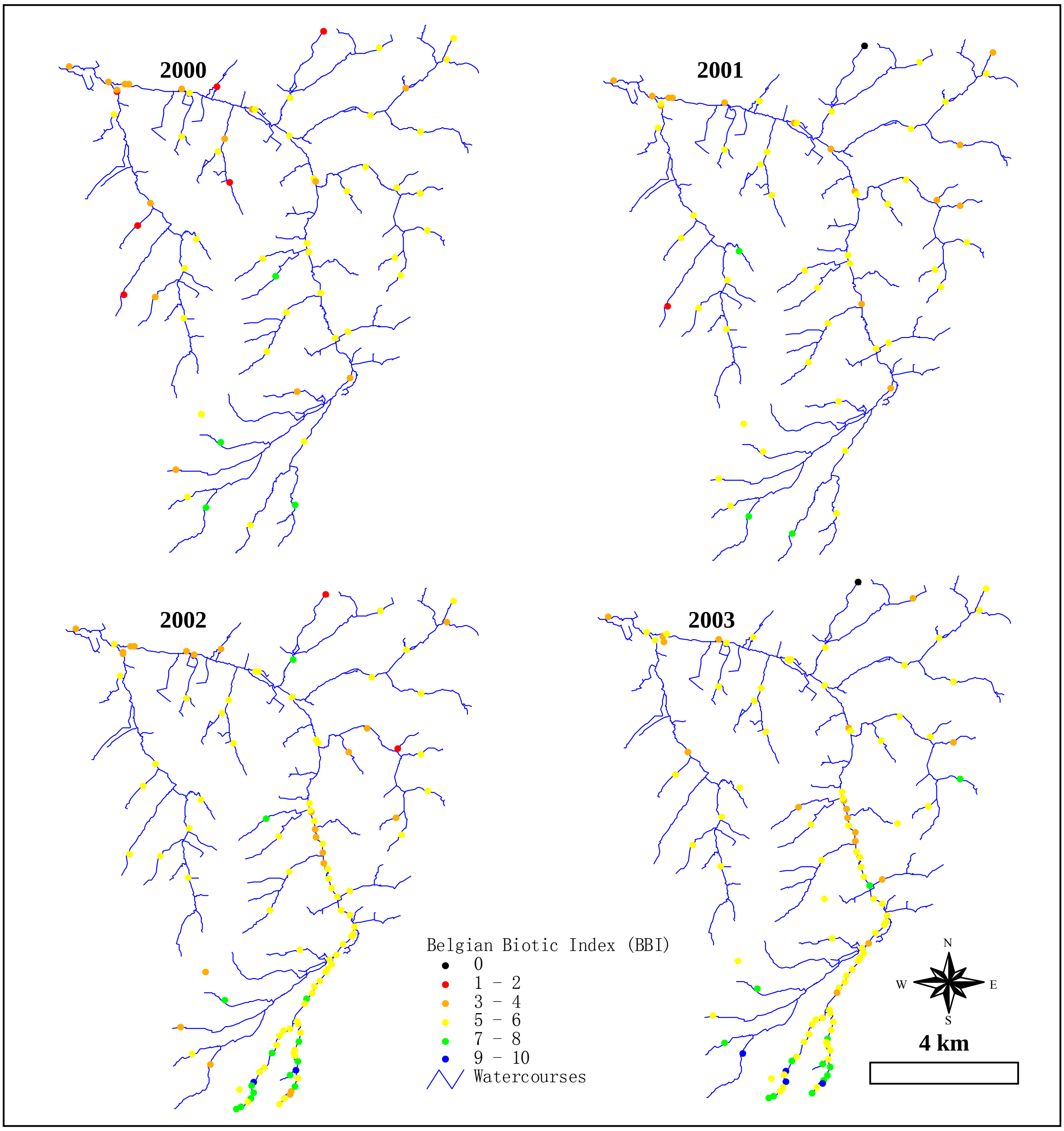


Fig. 4.5. Illustration of the water quality in the Zwalm river basin based on the Belgian Biotic Index (BBI). 60 sites were sampled on a yearly basis (2000-2003). Additionally, 52 sites were sampled in 2002 and 2003 within the 'short distance' monitoring network.

4.3. Data collection

4.3.1. Monitoring strategy in rivers for habitat suitability modelling

Although a lot of data have been gathered in Flanders on river systems, there are still some gaps to fill before these data meet the requirements of our modelling objectives. First of all, the data are distributed over different institutes in Flanders using various format types, other co-ordinate systems, ... The existing monitoring network of the Flemish Environment Agency (VMM) for example could not be used because the monitoring approach currently adopted in Flanders is not adjusted to the requirements of the data needed within the aim of this thesis focussing on habitat suitability modelling. The VMM has selected specific sampling sites through Flanders to reveal effects of pollution and monitor long-term trends in water quality. In this way, sites are selected that are expected to be influenced by domestic, agricultural or industrial activities. Also, sites near wastewater treatment plants are included. The general trend of VMM is to monitor sites with an expected degradation in water quality (VMM, 2003). This biased monitoring network is as such not stratified over a certain environmental gradient, and sites are somehow clustered.

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

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

Furthermore, to be efficient, a sampling strategy needs to be based on those gradients that are believed to exert major control over the distribution of species. These gradients should be considered primarily for the sampling, because otherwise vital information will limit model accuracy, in particular when data driven models, such as artificial neural networks, have to be developed. For habitat suitability modelling, systematic sampling along the most important environmental gradients is therefore preferred in contrast to random sampling. 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 that are degraded. Also the more pristine sites in the upper reaches should be sampled. These sites will reveal what is feasible from an ecological point of view.

An additional reason for setting up the Zwalm river basin monitoring network consisted of a need to give more attention to the physical environment of the macroinvertebrate taxa under study. This comprises the measurements and observations of physical characteristics and habitat characteristics, not taken into consideration by the water quality monitoring network of the VMM. Moreover, within the VMM network, a significant part of the physical-chemical sampling is obtained in a different period than the biotic sampling, causing a temporal distortion within the dataset. Variability in data, natural as well as human-induced, in particular with regard to physical-chemical sampling, reveals the need for continuous measurements to identify the outer limits. Although the importance of on-line water quality measurements has been stressed during an FWO project on the river Dender in Flanders (Vandenberghe *et al.,* 2000), still it is a very labour-intensive and as such costly method. Automated measurement stations for the on-line quality monitoring have as such not been considered in the context of this thesis. In addition, no heavy metals and organic micro-pollutants were part of the analysis. Because of the limited industrial activities in the Zwalm river basin, these compounds probably played a minor role for the river ecology and could thus be excluded from the data collection.

4.3.2. Sampling sites

Within the Zwalm river basin, 60 sampling sites were selected (Fig. 4.6, left) based on the following criteria:

- sites should include different levels of water and habitat quality;
- sites should include different types of land uses;
- sites should cover different river types;
- sites should be distributed over the entire river basin;
- sites should be easy to reach.

This stratified monitoring network was sampled during August and September 2000, 2001, 2002 and 2003. In this way, 240 sets of observations were available. In 2001 however, the artificial substrates got lost in one site and in 2003, two sites were too dry to take macroinvertebrate samples which means that no biological data were available for these three sites. The final database consisted thus only of 237 instances.

In addition, a 'short distance' monitoring network has been set up (Dedecker *et al.*, 2005a) including 52 additional sites which were sampled during August and September 2002 and 2003 (Fig. 4.6, right). For this monitoring network, a part of the Zwalm river basin of about 12 km was selected, which consisted of the brooks Verrebeek, Dorenbosbeek and the upstream part of the Zwalm river. This part contained river sites characterized by structural and morphological disturbances (weirs for water quantity control, artificial embankments, watering places for the cattle, culverted river sections, ...), while others nearly met reference conditions (forests with good meandering, hollow river banks, pool/riffle variation, …). The selected part of the river basin was located in a region with different types of land use (urban, agricultural and industrial regions).

4.3.3. Monitoring of environmental variables

The measurements of the structural characteristics were partially based on Schneiders *et al.* (1999) and comprised the visual observation of the bank structure, the development of hollow banks, the presence of pools and riffles, and the meanders in the watercourse. These variables are respectively divided into four and six categories to take the different possibilities of structural river habitat into account as optimal as possible. To illustrate the meaning of the different categories of these variables used in this thesis, a description in combination with some pictures clarifying their meaning is presented in Tables 4.3, 4.4, 4.5 and 4.6. Flow velocity, distance to source, stream order (Strahler, 1957), width and depth were measured as well as the granulometric characteristics of the sediment. Physical-chemical data that were measured included pH, conductivity, water temperature, suspended solids, dissolved oxygen, nitrate, ammonium, total nitrogen, phosphate, total phosphorous and COD (Table 4.2).

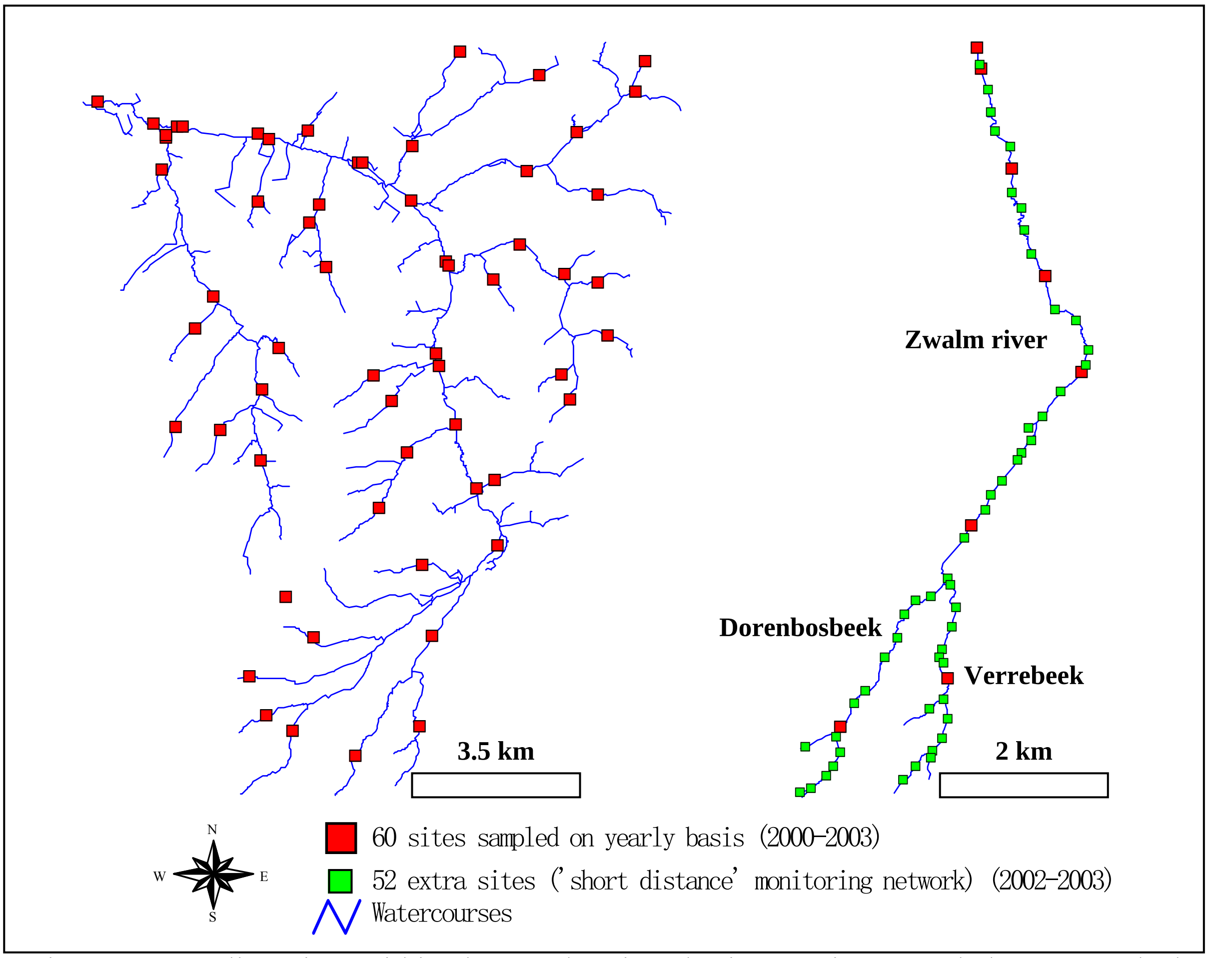


Fig. 4.6. Sampling sites within the Zwalm river basin, 60 sites sampled on a yearly basis (2000-2003) during August and September (left) and 52 additional sites sampled in August and September 2002 and 2003 within the brooks Verrebeek, Dorenbosbeek and the upstream part of the Zwalm river (right).

Table 4.2. Environmental variables measured in the Zwalm river basin

Variables	Measuring units	Measuring method and equipment
Temperature	°C	Oximeter WTW Cellox 330
pH	- log [H+]	pH meter Consort P114
Conductivity	µS/cm	WTW TetraCon® 325
Ammonium	mg NH_4^+-N/l	Spectrophotometrical (Spectroquant®, Merck)
Nitrate	mg NO_3^--N/l	Spectrophotometrical (Spectroquant®, Merck)
Total nitrogen	mg N/l	Spectrophotometrical (Dr Lange® LCK 238)
Ortho phosphate	mg PO_4^{3-}-P/l	Spectrophotometrical (Spectroquant® phosphorus (PMB), Merck)
Total phosphorus	mg P/l	Spectrophotometrical (Spectroquant® total phosphate 1453, Merck)
Chemical Oxygen Demand	mg COD/l	Spectrophotometrical (Spectroquant® COD, Merck)
Suspended solids	mg/l	Membrane filtration (PALL Supor® 800)
Dissolved oxygen	mg/l	Oximeter WTW Cellox 330
Depth	cm	Measured at highest depth
Width	cm	Measured at highest width
Flow velocity	m/s	Hydrometric propeller (Laboratory of Hydraulics, Ghent University)
Fraction pebbles	% surface bottom (> 2 cm)	Visual observation
Fraction gravel	% surface bottom (2 mm – 2 cm)	Sampled by means of a Van Veen grab, dried and measured in the lab
Fraction sand	% surface bottom (50 µm – 2 mm)	Sampled by means of a Van Veen grab, dried and measured in the lab
Fraction loam/clay	% surface bottom (< 50 µm)	Sampled by means of a Van Veen grab, dried and measured in the lab
Embankment	3 categories (0 (absent), 1 (partial), 2 (total))	Visual observation according to Schneiders *et al. (*1999)
Meandering	6 categories (1 (well developed) to 6 (absent))	Visual observation according to Schneiders *et al. (*1999)
Hollow banks	6 categories (1 (well developed) to 6 (absent))	Visual observation according to Schneiders *et al. (*1999)
Pools-riffles	6 categories (1 (well developed) to 6 (absent))	Visual observation according to Schneiders *et al. (*1999)
Distance to mouth	m	ArcView GIS 3.2a
Stream order (= Strahler order)	4 categories (1 to 4)	Topographic map (scale 1/25000)

Table 4.3. Development of meandering pattern

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

Table 4.4. Development of pool-riffle pattern

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

Table 4.5. Development of hollow river banks

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

Table 4.6. Bank structure

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

To allow for the development of the migration models (Chapter 8), a second monitoring campaign ('short distance' monitoring network) was set up in August and September 2002 and continued in August and September 2003. This type of model development required a different and more intensive monitoring approach. Therefore, the selected river parts (the brooks Verrebeek and Dorenbosbeek and the upstream part of the Zwalm river) were split up in stretches of 50 m, each marked with an upstream and downstream (X,Y) co-ordinate. In addition to the environmental variables in Table 4.2, an inventory of the structural and morphological characteristics along the selected part of the Zwalm river basin was made. In each river stretch of 50 m, the dominant type of land use (wooded area, housings, industrial sites, arable or grazing land, …) was monitored as well as the occurrence of domestic, industrial or agricultural discharges, and the presence of buffer strips along the river (type and distance to the river), natural or artificial river banks and meanders, hollow river banks and pool-riffle patterns.

4.3.4. Monitoring of macroinvertebrates

In shallow river stretches, the handnet method was used. The macroinvertebrates were collected by means of kick-sampling (Fig. 4.7) with a standard handnet consisting of a metal frame holding a conical net (mesh-size 350 µm) (IBN, 1984). The handnet is held in a vertical position on the river bottom. The bottom material immediately upstream of the net is turned over by foot. In this way, the dislodged animals are carried into the net by the current. Additional hand sampling of the vegetation, stones and other substrates is required to collect the attached species. The objective of the sampling consists in collecting the most representative diversity of macroinvertebrates at the site examined (De Pauw and Vanhooren, 1983). The sampling method is based on a multi-habitat design, where major habitats are sampled according to their proportional distribution within a sampling reach and consisted of five minutes active sampling in a 10 m reach of the watercourse.

Before the weirs and at the mouth of the Zwalm river, artificial substrates were used (De Pauw and Vanhooren, 1983; De Pauw *et al.,* 1994) because the river stretches were deeper and not wadable (seven sites of the yearly measured monitoring network, one site of the 'short distance' monitoring network). Artificial substrates were made of a polyethylene netting (potato bag), filled with pieces of brick (Fig. 4.7) and exposed in the field for about four weeks. Retrieval of the colonized substrates was done by carefully lifting bags from the bottom, and placing them in a bucket (Fig. 4.7).

In the laboratory, the coarse debris was removed and the organisms retained on different sieves transferred to white plastic sorting trays (30 x 50 cm). The organisms were sorted out and transferred to small flasks and grouped roughly according to major taxonomic groups. The organisms were preserved in 70% denatured alcohol (De Pauw and Vanhooren, 1983). After separation, the marcroinvertebrates were identified under a stereoscopic dissection microscope (magnification 10 to 50 times). The identification for the systematic units was performed to the levels as determined by De Pauw and Vanhooren (1983) (Table 4.7).

Fig. 4.7. Left and middle: handnet and kick-sampling with handnet, right: artificial susbstrates and retrieval of artificial substrates.

Table 4.7. Taxonomic groups and the identification levels as defined by De Pauw and Vanhooren (1983)

Taxonomic group	Identification level
Plathelminthes	genus
Oligochaeta	family
Hirudinea	genus
Mollusca	genus
Crustacea	family
Plecoptera	genus
Ephemeroptera	genus
Trichoptera	family
Odonata	genus
Megaloptera	genus
Hemiptera	genus
Coleoptera	family
Diptera	family
	Chironomidae group *thummi-plumosus*
	Chironomidae group *non-thummi-plumosus*
Hydracarina	presence

4.3.5. Scaling aspects of the study

The scale at which macroinvertebrates in watercourses exhibit the greatest variation is at one scale at which important physical-chemical gradients or biotic interactions control the assemblage composition (Li *et al.,* 2001). By recording and analyzing macroinvertebrate responses at different scales, it can be determined explicitly whether the scale of sampling influences interpretation of community structures (Townsend *et al.,* 1997a). Within this study, scaling aspects were considered as described by Adriaenssens *et al.* (2003). These scaling considerations mainly have an implication for the monitoring of physical variables and habitat characteristics. The considered scales in this study are given in Table 4.8.

Table 4.8. A scale-based classification for river monitoring (based on Maddock, 1999)

Scale	**Characteristics**	**Measured variables**
Drainage basin	On the basis of hydrological features, a drainage basin can be assigned and characterized.	• distance to source
Stream type	Stream typologies based on near-natural reference conditions are an essential basis for developing assessment systems.	• river width • stream order
Reach or segment scale	At the reach or segment scale, the environment is relatively stable and biota are determined by the overall features of the region, its topography and altitude and its geomorphic or land use pattern.	• physical-chemical measurements in general (pH, nutrients, dissolved oxygen, conductivity, organic matter, …)
Macro- or mesohabitat	At this scale of approach, habitat features of different river sections are deemed influential.	• meandering • pool-riffle pattern • hollow banks • flow velocity
Microhabitat	Habitat that includes the distribution of hydraulic and structural features comprising the actual living space of the organism.	• sediment characteristics

Temporal variability and seasonal fluctuations in macroinvertebrate composition caused by diurnal and life-cycle changes in organism behaviour or development, and seasonal or annual changes in the environment, are not considered in this thesis. This was avoided taking all samples during the same period of the year (August-September). In this way, this type of variation in the dataset could be eliminated.

4.4. Variability in the collection of macroinvertebrate data

There is a growing requirement, in Europe and worldwide, to assess the quality of our rivers, which are under conflicting pressures due to human demands for water and the requirements of the freshwater biota. In this way, there is a need to monitor river water quality through a comprehensive programme, not only through chemical but also through biological monitoring (Clarke *et al.*, 2003). The European Water Framework Directive (WFD) (EU, 2000) (2000/60/EC) recognizes this need for biological monitoring. Implementation of this WFD requires that the river monitoring programmes in each member state have to be operational before the end of 2006.

The WFD requires that a variety of taxonomic groups, including macroinvertebrates, be studied in assessing the ecological status of rivers. There are several good reasons why macroinvertebrates are useful as indicators of the ecological quality of rivers as discussed in Section 1.3.2. Therefore, macroinvertebrates are the most widely used group of organisms for determining the biological water quality (Wils *et al.*, 1994).

Because single samples may be sufficient to indicate environmental problems over the previous few months, macroinvertebrate sampling is an ideal procedure for monitoring large numbers of sites. In particular, the community composition and the taxonomic richness observed in macroinvertebrate samples collected using standard protocols are considered to be sensitive indicators of alterations in aquatic ecosystems (Cairns and Pratt, 1993). The fact that the various invertebrate species comprising a sample have differing abilities to cope with environmental stresses (e.g. organic pollution, heavy metal pollution, modified flow regime, loss of habitat richness) provides a method for identifying the type of stress operating at a given site (Walley and Fontama, 2000; Brabec *et al.*, 2004; Buffagni *et al.*, 2004). Although these surveys are often very extensive in their geographical coverage, most methods for the assessment of the ecological quality of rivers are based on single standard samples of macroinvertebrates. It is therefore important to take into account the expected levels of sampling variation for different types and qualities of sites and derive a variability measure that can be used in procedures to provide confidence limits for estimates of ecological quality and for assessments of temporal or spatial change in quality (Clarke *et al.*, 2002). Also Annex V, Section 1.3 of the WFD (EU, 2000) requires that 'Estimates of the level of confidence and precision of the results provided by the monitoring programmes shall be given in the plan'. Uncertainty is definitely one of the key topics in environmental monitoring, assessment and management. This is a considerable challenge but also an opportunity for scientists involved in freshwater biomonitoring to increase the practical application of their research.
The aim of this study was to describe the sampling variation obtained during the monitoring of macroinvertebrates in the Zwalm river basin. Therefore, five sampling sites were selected

covering different water quality classes and having a similar habitat over a distance of about hundred meter. In this way, one could assume that the macroinvertebrate communities within these hundred meter were very similar. These sampling sites were located in the river Zwalm downstream of the weir ‘Ter Biestmolen’ (X = 102416, Y = 175185), the river Zwalm in Michelbeke (X = 107734, Y = 169215), in the river Boekelbeek (X = 103105, Y = 171951) and in the brooks Verrebeek (X = 107013, Y = 163287) and Dorenbosbeek (X = 106640, Y = 164612) (Fig. 4.8).

During spring 2004, within each of the five river sections, six standard macroinvertebrate samples were taken as described in Section 4.3.4. The six hand net samples were taken by two different operators (A and B). In this way, the within and between-operator variation could be examined. The six replicas on one site were taken at the same day, eliminating seasonal variation between the samples. Furthermore, attention was paid to the weather conditions. To this end, it was avoided to monitor after periods of heavy rainfall. After sample processing, all organisms were identified to the taxonomic level (genus or family level) as specified in De Pauw and Vannevel (1991). The sample processing and identification was carried out by one person to minimize this type of error. At the end, the number of taxa, the number of taxa found more than once and the BBI (De Pauw and Vanhooren, 1983) were calculated for each sample.

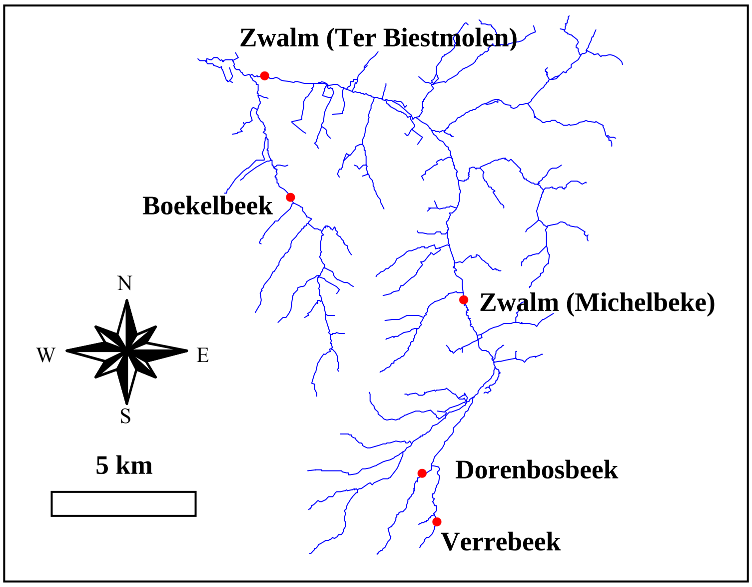


Fig. 4.8. Location of the selected sampling sites for studying the variability analysis of macroinvertebrates.

Fig. 4.9, 4.10 and 4.11 show respectively the average number of taxa found, the average number of taxa found more than once and the BBI obtained by operator A, operator B and the six samples together for the five sampling sites. Also the 95% confidence interval of the average is indicated. In general, the number of taxa found and the number of taxa found more than once in each sample varied. Generally, there was also a variation for these measures between operator A and B. In addition, not all the macroinvertebrates found in one sites were observed in each of the six samples. For the river Boekelbeek for example, Lumbricidae (Annelida, Oligochaeta), *Glossiphonia* (Annelida, Hirudinea), *Sialis* (Insecta, Megaloptera) and Limoniidae (Insecta, Diptera) were not found in each sample. The sampling results revealed that if water quality improved (e.g. the brooks Verrebeek and Dorenbosbeek), the number of taxa which were not found in each of the six samples also increased. In addition, the most sensitive taxa seemed most difficult to catch since these are not that abundant (e.g. *Cordulegaster* (Insecta, Odonata), *Nemoura* (Insecta, Plecoptera), Limnephilidae and Sericostomatidae (Insecta, Trichoptera)). The variation in sampling results has obviously an impact on the BBI. For one site, different scores were obtained, except for the river Boekelbeek for which they were the same.

Although variations in the number of taxa, the number of taxa found more than once and the BBI between operator A and B and between the six replicas were found, these differences were not significant ($p \geq 0.05$) (Dedecker *et al.*, 2005e).

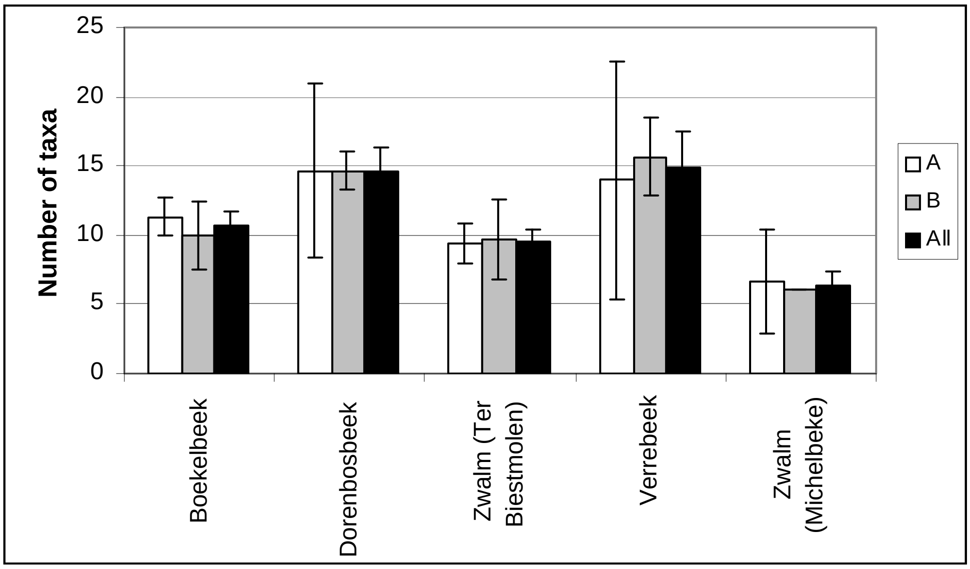


Fig. 4.9. Comparison of the average 'number of taxa found' obtained by operator A, operator B and the six samples together for the sites located in the river Zwalm downstream of the weir 'Ter Biestmolen', the river Zwalm in Michelbeke, the river Boekelbeek and the brooks Verrebeek and Dorenbosbeek. The 95% confidence interval of the average is shown.

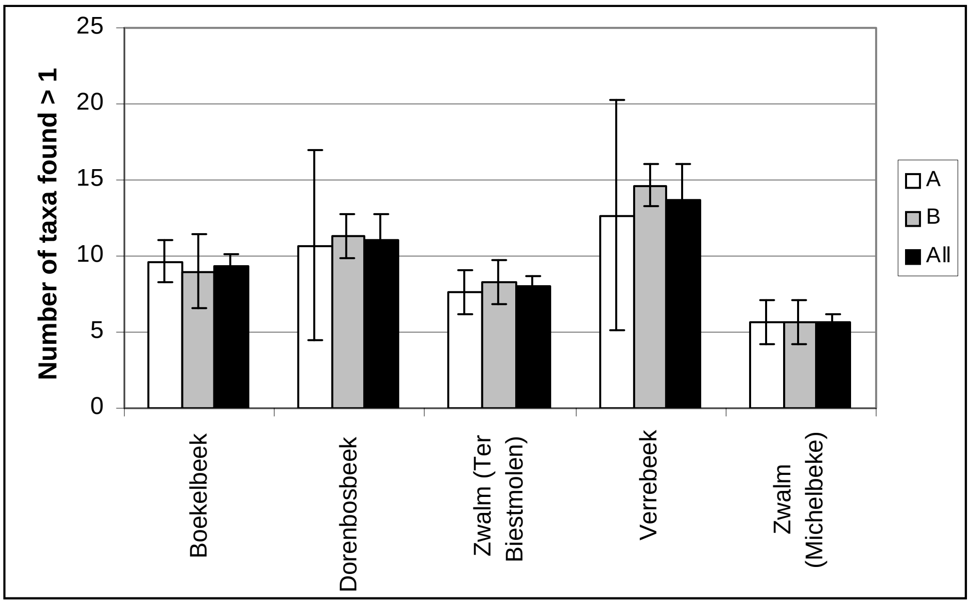


Fig. 4.10. Comparison of the average 'number of taxa found more than once' obtained by operator A, operator B and the six samples together for the sampling sites located in the river Zwalm downstream of the weir 'Ter Biestmolen', the river Zwalm in Michelbeke, the river Boekelbeek and the brooks Verrebeek and Dorenbosbeek. The 95% confidence interval of the average is shown.

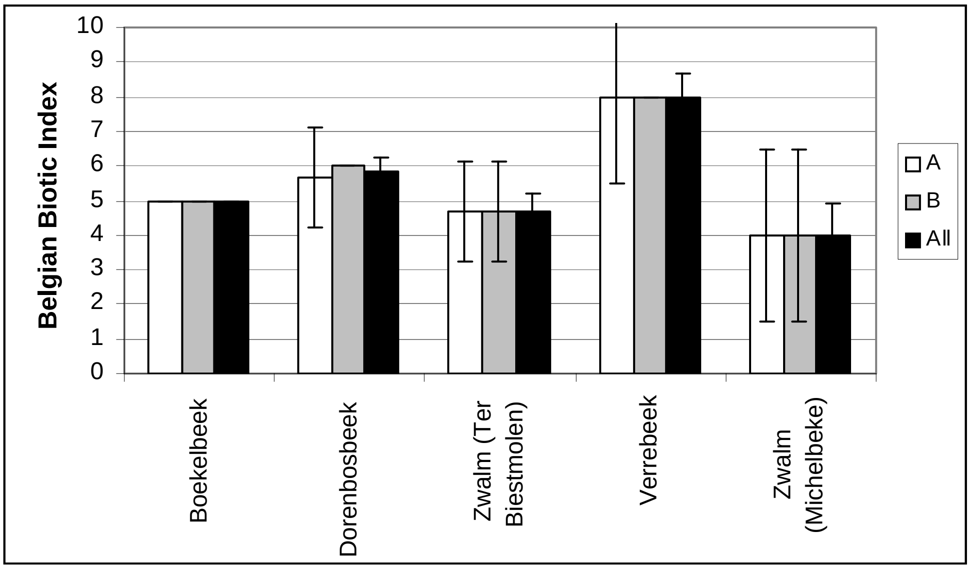


Fig. 4.11. Comparison of the average 'Belgian Biotic Index' (BBI) obtained by operator A, operator B and the six samples together for the sampling sites located in the river Zwalm downstream of the weir 'Ter Biestmolen', the river Zwalm in Michelbeke, the river Boekelbeek and the brooks Verrebeek and Dorenbosbeek. The 95% confidence interval of the average is shown.

Based on these results, one may conclude that it is essential to take into account estimates of the uncertainty and errors when river systems are monitored and assessed or when monitoring data is used to develop habitat suitability models based on data driven techniques as proposed in this thesis. One major source of error in the observed macroinvertebrate fauna is because of sampling variation. Usually however, it is too costly to take and identify replicate samples at each site. In the end, this study can nevertheless deliver interesting insights for further academic research (e.g. reliability of the monitoring results used for the development and validation of ecological models as mentioned before) and for the environmental administrations such as the Flemish Environment Agency, reporting on water quality issues (e.g. annual environment reports, the Environment and Nature Report (MIRA), …).

4.5. Information collection on the habitat preferences of Tubificidae, Gammaridae, Asellidae, *Baetis* and Limnephilidae

4.5.1. Introduction

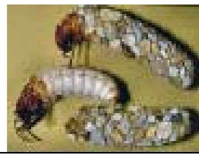
Based on ordination analysis (Hill, 1979; ter Braak and Smilauer, 1998), Sørensen Similarity Ratio Clustering (Sørensen, 1948; van Tongeren, 1986) and Self-Organizing Maps (Kohonen, 1982), it was concluded by Adriaenssens (2004) that no clear community structure could be detected for the Zwalm dataset. Since the analysis of the Zwalm dataset did not result in the description of a clear community structure, macroinvertebrate communities were concluded not to be suitable as output variables in habitat suitability models. For this reason, an approach based on taxon level prediction was required. The possibility of using specific macroinvertebrate taxa as indicators has been reviewed, based on their optional bio-indicator value as well as their presence in the Zwalm river basin.

Because of their highly variable presence in both ecological databases of the Zwalm and their use as bio-indicator in river quality assessment (e.g. De Pauw and Vannevel, 1991; MacNeil *et al.*, 2002), Tubificidae (Annelida, Oligochaeta), Asellidae (Crustacea, Isopoda), Gammaridae (Crustacea, Amphipoda), *Baetis* (Insecta, Ephemeroptera) and Limnephilidae (Insecta, Trichoptera) were selected as taxa (Table 4.9). Table 4.10 gives the taxonomic classification of the five taxa.

Table 4.9. Overview of the tolerance to pollution based on De Pauw and Vannevel (1991) and the percentage of occurrence of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae in the Zwalm river basin dataset (237 sites) and the 'short distance' monitoring network dataset (120 sites)

	Tolerance to pollution	**Zwalm river basin**	**'Short distance' monitoring network**
Limnephilidae	very low	12.2 %	17.5 %
Baetis	low	24.5 %	43.3 %
Asellidae	intermediate	50.6 %	54.2 %
Gammaridae	low - intermediate	73.0 %	85.8 %
Tubificidae	high	89.5 %	92.5 %

Table 4.10. Taxonomic classifications (source: http://www.itis.usda.gov) and pictures of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae

	Tubificidae	**Asellidae**	**Gammaridae**	***Baetis***	**Limnephilidae**
Kingdom	Animalia	Animalia	Animalia	Animalia	Animalia
Phylum	Annelida	Arthropoda	Arthropoda	Arthropoda	Arthropoda
Subphylum		Crustacea	Crustacea	Hexapoda	Hexapoda
Class	Clitellata	Malacostraca	Malacostraca	Insecta	Insecta
Subclass	Oligochaeta	Eumalacostraca	Eumalacostraca	Pterygota	Pterygota
Superorder		Peracarida	Peracarida		
Order	Haplotaxida	Isopoda	Amphipoda	Ephemeroptera	Trichoptera
Suborder	Tubificina	Asellota	Gammaridea	Pisciforma	
Superfamily		Aselloidea			Limnephiloidea
Family	Tubificidae	Asellidae	Gammaridae	Baetidae	Limnephilidae
Genus	e.g. *Tubifex*	e.g. *Asellus*	e.g. *Gammarus*	*Baetis*	e.g. *Limnephilus*
Species		e.g. *Asellus aquaticus*	e.g. *Gammarus pulex*		
Picture					

In the next paragraphs, a general description of the main environmental factors determining the distribution of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae is given. This information will be used for the practical ecological validation of the data driven ANN 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 datasets of all kinds, ... Nevertheless, some concordant characteristics were found and are mentioned below.

4.5.2. Tubificidae

Tubificidae are most commonly found in soft sediments rich in organic matter, and several species characteristically live in sites that receive organic pollution (Peckarsky *et al.*, 1989). Like all aquatic oligochaetes, Tubificidae respire cutaneously, but a unique feature of this family is that some species (e.g. *Tubifex tubifex*) can tolerate anoxic conditions (Peckarsky *et al.*, 1989; De Pauw and Vannevel, 1991).

4.5.3. Asellidae

Two *Asellus* species (*Asellus aquaticus* and *Asellus meridianus*) were present in the samples of the Zwalm river basin. These species have almost no apparent differences in ecological preferences, although *Asellus aquaticus* is thought to be more resistant to pollution than *Asellus meridianus* (Gledhill *et al.,* 1976; Chambers, 1977; Cuppen, 1980; Gongrijp, 1981; Verdonschot, 1990). *Asellus aquaticus* is very resistant to low oxygen conditions (Hawkes, 1979; Verdonschot, 1990) and can even stand percentages lower than 40 % (De Pauw and Vannevel, 1991). According to Holland (1976), *Asellus aquaticus* tolerates dissolved oxygen levels as low as 1.5 mg/l and is highly abundant at 5.8 mg/l. *Asellus aquaticus* is tolerant against organic conditions and often replaces *Gammarus* species in highly enriched organic conditions (Hawkes, 1979; Verdonschot, 1990). *Asellus aquaticus* prefers waters with a varied detritus layer. Asellidae are mentioned to behave as indifferent to water velocity (Bayerisches Landesamt für Wasserwirtschaft, 1996; Verdonschot, 1990), while others mention the preference for downstream sections characterized by low flow velocities (Steenbergen, 1993; Tachet *et al.*, 2002). 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 watercourses with higher width (Steenbergen, 1993). Peeters (2001) mentions a moderate sensitivity towards metal contamination, in comparison to other macroinvertebrate taxa.

4.5.4. Gammaridae

Of the family of the Gammaridae, only one representative was present in the watercourses of the Zwalm river basin, as is *Gammarus pulex*. *Gammarus pulex* appears in all kinds of water: lakes, headwaters, river tributaries, canals, … (Holthuis, 1956; Karaman and Pinkster, 1977; Hawkes, 1979; Verdonschot, 1990; Peeters 2001). Because of their very good swimming abilities (Brehm and Meijering, 1990) however, this species prefers small running waters with rather high flow velocity (Verdonschot, 1990; Bayerisches Landesamt für Wasserwirtschaft, 1996). *Gammarus pulex* is almost non-tolerant for low oxygen conditions (Wesenberg-Lund, 1982) and can only tolerate low oxygen concentrations when water temperatures are low. It generally prefers well-oxygenated localities and temperatures well below 20°C (Gledhill *et al.,* 1993). According to Macan (1961), *Gammarus pulex* tolerates dissolved oxygen down to

2.7 mg/l and is highly abundant at 7.4 mg/l or above. *Gammarus pulex* is suppressed by high organic conditions (Hawkes, 1979) and is in-tolerant at organic sewage (Whitehurst and Lindsey, 1990; Whitehurst 1991a, Whitehurst, 1991b), though it can stand moderate organic pollution (Gledhill *et al.,* 1976; Gledhill *et al.,* 1993). Organic pollution is normally a result of discharges from sewers, industrial effluents and agricultural run-off. Especially agricultural activities are responsible for organic pollution in the Zwalm river basin. In that context, the *Gammarus*/*Asellus* 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 (Holland, 1976; MacNeil *et al.*, 2002). *Gammarus pulex* prefers substrate-heterogeneity (Tolkamp, 1980), especially detritus substrates or detritus mixed with sand or gravel or leaf material (Tolkamp, 1982). Gammaridae are more sensitive to high conductivity values, caused by agricultural activities and treated or untreated wastewater effluents, than Asellidae, but at conductivity values above 1000 µS/cm, both macroinvertebrate taxa experience adverse influences (Steenbergen, 1993). *Gammarus pulex* is less tolerant than *Asellus aquaticus* to inorganic pollutants (Martin and Holdich, 1986). *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 4.11). The latter author found out that *Gammarus pulex* occurs in ranges between pH 4.7 to 11.6.

Table 4.11. 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)	0.10	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 - >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

4.5.5. *Baetis*

Baetis are generally restricted to running waters (Elliott *et al.*, 1988). Based on Verdonschot (2000a, 2000b) and De Loose *et al.* (1995), *Baetis* can be found in both small and large brooks. They prefer moderate to fast running waters (Bayerisches Landesamt für Wasserwirtschaft, 1996; Verdonschot, 2000a, 2000b; RIZA, 2000; Tachet *et al.*, 2002). *Baetis* are also sensitive to various forms of pollution (Elliott *et al.*, 1988). Ephemeroptera larvae are among the most sensitive of aquatic insects for acidification (Elliott *et al.*, 1988). As pH decreases, they progressively disappear (Tachet *et al.*, 2002). In a survey of 600 streams in Sweden, the optimal pH range was 6.0 – 7.5 (Elliott *et al.*, 1988). However *Baetis rhodani*, which was also found in the Zwalm river basin, was recorded in Swedish streams at pH 4.50 while *Baetis rhodani* and other *Baetis* spp. occurred at mean pH 4.70 and 4.45 in two streams in the English Peak District (Elliott *et al.*, 1988). It is possible that the adverse effects of high acidity were reduced by relatively higher concentrations of calcium, alkalinity, sodium, potassium and chloride.

4.5.6. Limnephilidae

In general, Limnephilidae prefers low contents of nutrients (ortho-phosphate, total phosphorus, nitrate and ammonium) and minerals (potassium, calcium, magnesium and chloride) and high oxygen concentrations (De Pauw and Vannevel, 1991; Steenbergen, 1993). Based on Steenbergen (1993), Limnephilidae has a preference for wide and deeper streams. However, some Limnephilidae species prefer small brooks while others have a preference for large brooks (De Loose *et al.*, 1995). Some Limnephilidae species are mentioned to behave as indifferent to water velocity, while others prefer slow to moderate running waters (Bayerisches Landesamt für Wasserwirtschaft, 1996; Verdonschot, 2000a, 2000b; Tachet *et al.*, 2002).

4.6. Factors affecting the migration behaviour of Gammaridae, *Baetis* and Limnephilidae

In the next paragraphs, an overview of the factors affecting the migration movements of Gammaridae, *Baetis* and Limnephilidae is given. For Tubificidae, no migration model was constructed, because they were already observed in ± 90 % of the sites. Also for Asellidae, no migration model was constructed since no information was found about active or passive (= drift) migration behaviour. It can be assumed that the movements of Asellidae occur at a very local scale (microhabitat scale). In this way, the development of a migration model for Asellidae was thus also irrelevant.

4.6.1. Gammaridae

Gammaridae show a very active migration pattern (Rawer-Jost *et al.*, 1999) and especially *Gammarus pulex*, the only Gammaridae species found in the Zwalm river basin, is a frequent component of the invertebrate drift (Elliott, 2002a). Based on this author, flow velocity could be considered as the major environmental factor affecting the quantity of drift of *Gammarus pulex*. Although, literature quantifying the factors affecting the upstream/downstream movement of *Gammarus pulex*, was hard to find, Elliott (2002c) was able to describe the significant relationship between drift and flow velocity on the basis of field experiments with 10 different water velocities: the mean distance ($\bar{x}$, m) travelled by *Gammarus pulex* in the drift was significantly related to the mean flow velocity (V, m/s) by the power function

$$\bar{x} = 7.82 \times V^{0.963} \tag{5.1}$$

Also several other authors described the importance of flow velocity on the upstream and downstream migration of *Gammarus pulex*. According to Hultin (1971), a moderate increase in flow velocity increased the activity of upstream movement of *Gammarus pulex*, while a strong increase suppressed the activity of upstream movement (Goedmaker and Pinkster, 1981). Experimentally, Hughes (1970) found that by changing the water velocity most of the upstream movements occurred at the lower velocities. Estimates for distances of upstream movements in *Gammarus pulex* (which has no aerial dispersal stage) can be up to 14 m per day (Elliott, 1971a). Temperature, oxygen concentration, season and length of the day or night seemed to be of additional importance. Seasonal fluctuations of *Gammarus pulex* in downstream drift have been recorded in several studies (Waters, 1962; Hughes, 1970; Goedmaker and Pinkster, 1981; Allan and Malmquist, 1989; Williams and Williams, 1993). However, Elliott (2002a) found no consistent seasonal pattern in drift rates, implying that temperature had no significant effect. In general, there was a nocturnal daily pattern of downstream drift with peaks soon after dusk and just before dawn (Waters, 1962; Elliott, 2002a). Because Gammaridae are generally more active at night, also the upstream movement against the current frequently occurred at night (Hughes, 1970; Goedmaker and Pinkster, 1981; Söderström,1987; Williams and Williams, 1993; Rawer-Jost *et al.*, 1999). Elliott (2002a) found a slight positive relationship with mean temperature and a slight negative relationship with the length of the night. However, neither relationship was significant. Meijering (1972) on the other hand found that *Gammarus pulex* had a stronger tendency to drift at low temperatures close to 0°C. According to Müller (1966), a rise in temperature caused an increase in the drifting of *Gammarus pulex*, while Meijering (1972) found that decreasing temperatures stimulated upstream movement.

Also the presence of predatory fish or invertebrates is a factor that may affect drift. In their review of 22 studies, Wooster and Sih (1995) found that predatory fish had variable effects on drift rates that sometimes increased, sometimes decreased, or were unaffected, whereas predatory invertebrates often increased drift rates more frequently than expected by chance. In the study of Elliott (2002a), the nocturnal increase in drift could not be interpreted as an antipredator behaviour. Although predatory fishes and invertebrates are probably catching *Gammarus pulex* in the Zwalm river basin, it is difficult to see how this could affect drift. Therefore, one cannot conclude that the drift of *Gammarus pulex* is the consequence of an antipredator behaviour in this stream.

Because *Gammarus pulex* has no aerial life stage, it is more likely that the migration behaviour is significantly more important than for flying insects and impacted by the design of weirs or culverted river sections. During upstream movement, these barriers could limit available habitat (Rawer-Jost *et al.*, 1999; Vaughan, 2002). Minckley (1964) reported that small dams prevented the upstream migration of *Gammarus pulex* in England. Also the Zwalm river basin is impacted by six weirs for water quantity control and several culverted river sections (Goethals and De Pauw, 2001). To improve the upstream migration in these situations, fish by-passes could be installed. Their effects on benthic invertebrates have been studied by Rawer-Jost *et al.* (1999), who investigated whether benthic invertebrates successfully use two types of fish by-passes, a boulder ramp and a concrete bypass, for upstream movements. Their results indicated that the boulder ramp allowed for the upstream migrations, whereas the concrete bypass was more difficult to ascend. Currently, only one weir in the Zwalm river basin is provided with a concrete fish by-pass, but the stream velocity and height of the steps are probably too high to allow the counter current migration of gammarids. In this way, weirs could be considered as a serious migration barrier in the Zwalm river. Also culverted river sections can be a serious obstacle for migrating *Gammarus pulex*. Especially culverts having an outflow above the downstream water level, limit upstream passage.

4.6.2. *Baetis*

Literature quantifying the factors affecting the upstream/downstream movement and migration through the air of *Baetis* was hard to find. Elliott (2002b) however re-analyzed data from an earlier study (Elliott, 1971b) to determine the time and distance spent in the drift by 17 different taxa, including *Baetis*. A significant relationship between drift and flow velocity was observed by Elliott (2002b). The mean distance ($\bar{x}$, m) travelled by *Baetis* in the drift was significantly related to the mean flow velocity (V, m/s) by the function:

$$\bar{x} = 8.97 \times V + 0.11 \tag{5.2}$$

Besides flow velocity, also water depth and substrate type appeared to be paramount factors determining the time spent in the drift. If macrophytes were present, a reduction of 50 % of the time spent in the drift was observed for *Baetis*. The average time spent in the drift was 9.4 (± 0.3) s in a site with a mainly stony substrate, while an average time of 4.9 (± 0.02) s was observed if macrophytes were present (Elliott, 2002b). An earlier study of Corkum and Clifford (1979) on larvae of *Leptophlebia cupida* (Insecta, Ephemeroptera) in an artificial stream, already revealed that macrophytes reduce the time spent in the drift.

Information about the active migration of *Baetis* is provided by Elliott (2003). During a 24 h period, 33.1 (± 1.2) % of the initial *Baetis* population moved actively, either in the upstream or the downstream direction. Elliott (2003) concluded that the maximum distances *Baetis* was actively moving upstream and downstream were respectively 5.5 and 1.5 m. Contrary to the distance covered by *Baetis* in the drift, flow velocity has no effect on the distances during active movements (Elliott, 2002b). Also temperature does not seem to have a significant effect on the active dispersal (Elliott, 2003). On the other hand, the presence of pebbles does have a positive effect.

Several studies discussed the mechanisms of *Baetis* to avoid fish and invertebrate (e.g. by Plecoptera larvae (Peckarsky *et al.*, 1994)) predation. Nocturnal drift is one of the methods often referred to (e.g. Kohler and McPeek 1989; Tikkanen *et al.*, 1994). Although predatory fishes and invertebrates are probably catching *Baetis* in the Zwalm river basin, it is difficult to see how this could affect drift. Therefore, one cannot conclude that the drift of *Baetis* is the consequence of an antipredator behaviour in this stream.

Gyselman (1979) stressed that the drift of *Baetis* is density-dependent. However, most of the studies reveal that dispersal of *Baetis* is density-independent (e.g. Reisen and Prins, 1972; Corkum *et al.*, 1977; Bohle, 1978; Ciborowski, 1983; Statzner and Mogel, 1985; Humphries, 2002; Elliott, 2003).

4.6.3. Limnephilidae

One of the few studies providing information on the migration behaviour of Limnephilidae is Elliott (2003). During a 24 h period, 20.1 (± 0.9) % of the initial Limnephilidae (in that study: *Potamophylax cingulatus*) population moved actively, either in the upstream or the downstream direction. Elliott (2003) observed that Limnephilidae moved maximum 3.5 m in both the upstream and downstream direction. These results were confirmed for another Limnephilidae larvae (*Chyrandra centralis*) in a study by Erman (1986). In addition, the active migration is affected by the presence of macrophytes and the type of the river banks since Limnephilidae are able to crawl along the banks for small distances (Wissinger *et al.*, 2003). Elliott (2002b) showed on the other hand that Limnephilidae were seldomly found in the drift. In this way, it can be assumed that the passive downstream migration is of less importance than the active migration.

Several studies were performed on the aerial dispersion possibilities of Limnephilidae. Kovats *et al.* (1996), Petersen *et al.* (1999) and Winchester *et al.* (2002) showed that 90 % of the adult organisms were caught within a distance of 20 m from the river. They observed that less than 10 % of the organisms flew further inland.

Predation by salamanders seems to be important in some regions (Wissinger *et al.*, 1999). However, this seems of no importance in the Zwalm river basin. In addition, aggressive behaviour towards other Limnephilidae and cannibalism have been observed by Wissinger *et al.* (1996, 1999). Although this behaviour could be important in the Zwalm river basin, it is difficult to see how this could affect the migration. Therefore, it was not taken into account in the present study.

Chapter 5
Data analysis and data preparation

5.1. Introduction

This chapter is dedicated to the exploratory analysis and preparation of the environmental input and biological output data. Since Artificial Neural Network (ANN) models are merely built on the data presented during the training, these models are unable to make reliable predictions beyond this data range. Therefore, a first and essential step before model development and application, is getting insight into the range of inputs and outputs, what determines also the maximum and practical application range of the models. This can be of major importance in relation to river restoration management. In addition, the mutual correlation between input variables and between input and output variables is calculated to help identify 'noise' variables. A visual relation analysis between input and output variables is conducted to get insight into outliers, the data clusters, missing or scarce variable combinations in certain ranges, … Also the geographical distribution of the variables over the study area is presented on the basis of a Geographical Information System to identify the hot spots for river restoration. These data analyses give advice regarding new or additional data that need to be collected and how this should be done to be able to develop more appropriate models and to gather meaningful information for river management afterwards. In addition, errors related to database constructions (e.g. typing errors) could be detected and corrected in advance.

5.2. Material and methods

5.2.1. Bandwidth and distribution of input and output variables

Lek and Guégan (1999) stated that data driven models, including ANN models, are built solely on 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 testing data contain values outside the range of those used for training (Maier and Dandy, 2000). Insight into 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 applications can start.

5.2.2. Correlation between input and output variables

A first 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 absolute correlation value of 0.20 or higher indicates a probable noise source to the ANN (Walczak and Cerpa, 1999). From an ecological point of view however, relationships based on correlations between environmental variables should be considered with caution, because these correlations do not necessarily involve relevant causal ecological processes. The removal of input variables can also be overruled for practical reasons (river managers could be interested in particular simulations for which certain variables are essential) or by use of ecological expert knowledge.

5.2.3. Visual relation analysis between input and output variables

A visual relation analysis between the input and output variables can be beneficial to get insight into outliers, the data clusters, missing or scarce variable combinations in certain ranges, … (Goethals, 2005). As such, these methods can be very interesting in delivering insight into 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 visualization methods can be very useful 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 last years and became standard tools in most data mining and analysis software packages. Part of these analyses were based on the WEKA software (Waikata Environment for Knowledge Analysis) (Witten and Frank, 2000), which is a collection of machine learning algorithms for data mining tasks. This type of plots was based on the use of classes to look at the distributions. However, also the distribution of the output classes (presence/absence) was directly plotted. Based on these 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 chapters when the results are evaluated. On the other hand, input data of both datasets (237 and 120 sites) were graphically visualized to get better insight into the distribution and the outliers of the environmental variables and to see the distribution over the different years. Also the quality standards for the appropriate variables (VLAREM, 1999) were plotted to verify which part of the data complied with these quality standards. To get a better overview of the spatial variation of the data over both study areas and to identify the hot spots for river restoration, ArcView GIS 3.2a, a product of the Environmental Systems Research Institute (ESRI), was applied for visualization.

5.3. 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, medians and standard deviations of the environmental input variables (Table 5.1 and 5.2). Analysis of these parameters for the output variables was rather irrelevant because the outputs are only expressed by 0 and 1 representing respectively the absence and presence of the five macroinvertebrate taxa Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae. An overview of the prevalence of these taxa is given in Table 4.9. Preferably, these analyses can be combined with visualization graphs as presented in Fig. 5.1 to 5.4. 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 compared with the average, minimum and maximum) can as well give a good indication in this context.

Based on these summary statistics, several outliers (very high values and standard deviation and a big difference between average and median) could already be detected in both databases. In the Zwalm river basin dataset (Table 5.1), outliers could be observed for the variables suspended solids, chemical oxygen demand, depth, width and most of the nutrients. In the dataset of the 'short distance' monitoring network (Table 5.2), a few outliers could be distinguished for the variables suspended solids, chemical oxygen demand, depth, width and to a smaller extent for some nutrients. The high standard deviations and maxima in comparison with the median values for the variables depth and width are strongly related to the presence of the weirs along the Zwalm river. A major consequence is that these sites are very difficult to predict on the basis of data driven models like ANNs although these weirs are essential elements to consider concerning the Water Framework Directives.

Table 5.1. Minima, maxima, averages, medians and standard deviations of the input variables that were used in the Zwalm river basin database containing 237 measuring sites (2000-2003)

Variable (Measuring unit)	**Minimum**	**Maximum**	**Average**	**Median**	**Standard deviation**
Temperature (°C)	6.4	20.9	14.7	14.8	2.2
pH	6.72	9.06	7.67	7.69	0.31
Conductivity (µS/cm)	10	2190	778	781	231
Ammonium (mg NH_4^+-N/l)	0.0	6.0	0.9	0.5	1.1
Nitrate (mg NO_3^--N/l)	0.2	15.8	5.4	4.8	3.4
Total nitrogen (mg N/l)	2.5	77	12.0	9.7	9.7
Ortho phosphate (mg PO_4^{3-}-P/l)	0.0	5.0	0.5	0.3	0.7
Total phosphorus (mg P/l)	0.1	12.4	0.7	0.4	1.5
COD (mg COD/l)	7.0	918.0	25.8	20.0	59.3
Suspended solids (mg/l)	0.0	949.0	44.2	20.0	82.5
Dissolved oxygen (mg/l)	0.06	11.60	6.83	7.10	2.20
Depth (cm)	3	190	32	18	37
Width (cm)	22	1100	227	127	235
Flow velocity (m/s)	0.00	1.92	0.31	0.22	0.31
Fraction pebbles (%)	0.0	100.0	37.5	25.0	38.9
Fraction gravel (%)	0.0	67.7	12.4	6.0	17.5
Fraction sand (%)	0.0	87.8	20.2	14.9	21.3
Fraction loam/clay (%)	0.0	82.2	24.7	12.4	24.9
Embankment (class variable)	0	2	-	-	-
Meandering (class variable)	1	6	-	-	-
Hollow banks (class variable)	1	6	-	-	-
Pools-riffles (class variable)	1	6	-	-	-
Distance to mouth (m)	1541	19865	9986	10063	5078
Stream order	1	4	2	2	1

Table 5.2. Minima, maxima, averages, medians and standard deviations of the input variables that were used in the 'short distance' monitoring network database containing 120 measuring sites (2002-2003)

Variable (Measuring unit)	Minimum	Maximum	Average	Median	Standard deviation
Temperature (°C)	8.0	22.3	16.9	17.2	2.0
pH	6.78	8.27	7.78	7.79	0.28
Conductivity (µS/cm)	207	1079	588	557	141
Ammonium (mg NH_4^+-N/l)	0.0	2.7	0.5	0.3	0.6
Nitrate (mg NO_3^--N/l)	0.2	12.2	3.3	2.9	2.1
Total nitrogen (mg N/l)	2.6	20.5	7.2	6.3	3.1
Ortho phosphate (mg PO_4^{3-}-P/l)	0.0	1.7	0.3	0.2	0.2
Total phosphorus (mg P/l)	0.0	1.7	0.3	0.2	0.3
COD (mg COD/l)	3.0	918.0	26.8	18.0	83.5
Suspended solids (mg/l)	0.0	945.0	40.8	15.0	122.4
Dissolved oxygen (mg/l)	2.50	12.80	6.12	5.71	1.86
Depth (cm)	2	190	24	16	25
Width (cm)	28	750	187	140	131
Flow velocity (m/s)	0.01	0.87	0.20	0.16	0.14
Fraction pebbles (%)	0.0	100.0	27.3	10.0	33.5
Fraction gravel (%)	0.0	63.4	14.1	4.4	18.3
Fraction sand (%)	0.0	87.8	31.3	24.2	26.7
Fraction loam/clay (%)	0.0	85.4	25.8	12.5	26.4
Embankment (class variable)	0	2	-	-	-
Meandering (class variable)	1	6	-	-	-
Hollow banks (class variable)	1	6	-	-	-
Pools-riffles (class variable)	1	6	-	-	-
Distance to mouth (m)	10063	20549	16229	17242	3176
Stream order	1	4	3	3	1

5.3.2. Correlation between input and output variables

A second step in data analysis is to verify how related some variables might be on the basis of their correlation coefficient (r) (Table 5.3 and 5.4). Correlation coefficients with an absolute value of at least 0.20 and lower than 0.50 are marked in yellow, higher values are marked in orange.

In the Zwalm river basin dataset (Table 5.3), quite a high r could be observed between the structural habitat characteristics (embankment, meandering, pools-riffles, hollow banks). Meandering is however clearly related to hollow river banks (r = 64). These correlations are rather logic, because many artificial structures are combined (e.g. channel straightening with bank reinforcement). However, this is not always the case, and many exceptions exist. Models without this set of variables might lead to practical limitations of these. Also an expected good relation existed between width, depth and stream order. Here a variable reduction might be considered. However, also wider shallow streams exist in the Zwalm, especially downstream of the weirs. For the same reason, flow velocity was less correlated with the variable width for example as expected. Increased correlation values are identified for several nutrients (ammonium, nitrate, total nitrogen, phosphate and total phosphorus). Because these nutrients can originate from different pollution sources, it may be interested to keep them in.

In the ‘short distance’ dataset (Table 5.4), similar variables were correlated. However, the correlation values were generally higher because a lot of ‘exceptions’ were not included in this dataset (e.g. sites upstream and downstream of the weirs, heavily polluted sites with a rather good habitat quality, …). Also the correlation between distance to mouth and stream order was more expressed in this dataset since streams having a low stream order near the mouth are not included anymore.

Table 5.3. Correlation matrix of the 24 environmental variables and presence/absence of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae based on the dataset of Zwalm river basin (2000-2003) ($0.20 \leq r < 0.50$ = yellow, $0.50 \leq r$ = orange)

	Width	**Embankment**	**Meandering**	**Pools-riffles**	**Hollow banks**	**Depth**	**Flow velocity**	**pH**	**Temperature**	**Dissolved oxygen**	**Conductivity**	**Suspended solids**	**Ammonium**	**Nitrate**	**Total nitrogen**	**Phosphate**	**Total phosphorus**	**COD**	**Pebbles**	**Gravel**	**Sand**	**Loam/clay**	**Distance to mouth**	**Stream order**	**Tubificidae**	**Asellidae**	**Gammaridae**	***Baetis***	**Limnephilidae**
Width	1.00	-0.14	0.21	0.30	0.22	0.78	0.05	0.02	0.05	0.03	-0.12	-0.04	-0.13	-0.14	-0.22	-0.11	-0.10	-0.02	-0.06	-0.09	-0.03	0.27	-0.47	0.78	0.17	0.54	-0.30	-0.11	-0.22
Embankment		1.00	0.41	0.29	0.30	-0.12	-0.02	0.02	-0.07	0.00	0.07	-0.09	0.12	-0.10	0.07	0.19	0.16	0.09	-0.07	-0.04	-0.23	-0.16	-0.12	-0.12	-0.05	0.04	-0.06	0.04	-0.10
Meandering			1.00	0.40	0.64	0.10	0.06	-0.03	0.00	-0.10	0.12	-0.09	0.15	-0.08	0.05	0.10	0.12	0.02	0.07	-0.19	-0.28	-0.06	-0.32	0.19	-0.09	0.25	-0.18	-0.06	-0.21
Pools-riffles				1.00	0.49	0.32	0.00	-0.05	-0.01	-0.03	-0.20	-0.09	-0.11	-0.24	-0.15	0.00	0.05	0.06	-0.17	-0.02	0.03	-0.11	-0.06	0.32	0.00	0.22	-0.18	-0.01	-0.11
Hollow banks					1.00	0.10	0.16	0.09	0.04	-0.09	0.14	-0.07	0.08	0.00	0.01	0.09	0.09	0.02	0.06	-0.09	-0.16	-0.14	-0.46	0.34	-0.05	0.36	-0.18	-0.10	-0.32
Depth						1.00	-0.06	-0.14	-0.01	-0.06	-0.15	-0.03	-0.05	-0.16	-0.18	-0.11	-0.08	0.02	-0.22	-0.09	0.17	0.35	-0.35	0.60	0.15	0.39	-0.32	-0.16	-0.18
Flow velocity							1.00	0.14	-0.36	0.40	-0.18	0.18	-0.14	-0.03	-0.20	-0.17	-0.19	-0.06	0.39	-0.12	-0.27	-0.27	-0.10	0.24	0.12	0.04	0.11	0.03	-0.09
pH								1.00	-0.08	0.39	0.17	0.07	-0.16	0.26	0.04	-0.05	-0.03	0.10	0.11	-0.03	-0.08	-0.12	-0.20	0.07	-0.06	0.05	-0.04	-0.02	0.02
Temperature									1.00	-0.39	0.09	-0.31	0.04	0.10	0.10	0.01	0.01	0.09	0.09	-0.02	-0.06	-0.02	-0.03	0.09	-0.05	0.14	0.09	0.20	0.07
Dissolved oxygen										1.00	-0.34	0.10	-0.40	0.23	-0.35	-0.47	-0.44	0.03	0.10	0.10	-0.10	-0.18	0.09	0.06	-0.03	-0.07	0.17	0.07	0.17
Conductivity											1.00	-0.05	0.29	0.17	0.52	0.47	0.40	0.01	0.02	-0.18	-0.10	0.16	-0.34	-0.19	-0.07	-0.03	-0.23	-0.05	-0.17
Suspended solids												1.00	0.01	-0.09	0.05	0.13	0.16	0.01	-0.05	-0.04	0.02	0.16	0.01	-0.11	-0.04	-0.11	-0.11	-0.18	0.02
Ammonium													1.00	-0.09	0.35	0.41	0.38	0.03	-0.11	-0.21	-0.02	0.22	-0.05	-0.22	0.00	-0.14	-0.24	-0.19	-0.20
Nitrate														1.00	0.23	-0.23	-0.24	-0.12	-0.01	-0.03	-0.06	0.02	-0.15	-0.23	-0.15	-0.23	0.16	0.17	0.10
Total nitrogen															1.00	0.68	0.54	0.05	0.00	-0.12	-0.08	0.12	-0.07	-0.31	-0.14	-0.23	-0.09	-0.02	-0.08
Phosphate																1.00	0.81	0.09	-0.04	-0.11	-0.05	0.10	-0.07	-0.16	-0.07	-0.07	-0.19	-0.14	-0.13
Total phosphorus																	1.00	0.07	-0.10	-0.09	-0.02	0.11	-0.07	-0.15	-0.16	-0.03	-0.21	-0.12	-0.12
COD																		1.00	0.01	-0.05	0.07	-0.03	0.09	0.03	0.04	0.06	-0.16	0.07	-0.04
Pebbles																			1.00	-0.31	-0.59	-0.64	0.00	0.22	0.00	0.12	0.17	0.15	0.05
Gravel																				1.00	-0.03	-0.05	0.26	-0.10	0.08	0.00	0.14	0.04	0.18
Sand																					1.00	0.28	0.13	-0.07	0.11	-0.12	-0.08	-0.08	0.04
Loam/clay																						1.00	-0.19	-0.10	0.00	-0.06	-0.27	-0.20	-0.15
Distance to mouth																							1.00	-0.39	-0.07	-0.37	0.32	0.09	0.36
Stream order																								1.00	0.20	0.60	-0.22	-0.04	-0.26

Tubificidae	1.00	0.18	-0.02	0.10	-0.08
Asellidae		1.00	-0.16	-0.12	-0.20
Gammaridae			1.00	0.19	0.20
Baetis				1.00	0.06
Limnephilidae					1.00

Table 5.4. Correlation matrix of the 24 environmental variables and presence/absence of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae based on the 'short distance' monitoring network (2002-2003) (0.20 ≤ *r* < 0.50 = yellow, 0.50 ≤ *r* = orange)

	Width	**Embankment**	**Meandering**	**Pools-riffles**	**Hollow banks**	**Depth**	**Flow velocity**	**pH**	**Temperature**	**Dissolved oxygen**	**Conductivity**	**Suspended solids**	**Ammonium**	**Nitrate**	**Total nitrogen**	**Phosphate**	**Total phosphorus**	**COD**	**Pebbles**	**Gravel**	**Sand**	**Loam/clay**	**Distance to mouth**	**Stream order**	**Tubificidae**	**Asellidae**	**Gammaridae**	***Baetis***	**Limnephilidae**
Width	1.00	0.02	0.27	0.25	0.21	0.80	-0.08	-0.20	0.19	-0.54	0.60	-0.14	0.61	0.41	0.36	0.39	0.38	0.02	0.11	-0.20	-0.19	0.25	-0.85	0.78	0.23	0.59	-0.51	-0.27	-0.35
Embankment		1.00	0.59	0.61	0.61	0.10	0.16	0.08	0.09	0.08	0.02	-0.10	-0.03	-0.09	-0.12	-0.03	-0.02	0.08	0.33	0.14	-0.27	-0.38	-0.10	0.11	0.07	0.22	0.01	0.22	-0.16
Meandering			1.00	0.68	0.75	0.20	0.11	-0.05	0.10	-0.10	0.21	-0.07	0.12	0.12	-0.01	0.06	0.10	0.05	0.35	-0.16	-0.24	-0.19	-0.33	0.36	0.04	0.39	-0.21	0.06	-0.19
Pools-riffles				1.00	0.61	0.24	0.08	-0.11	0.12	-0.11	0.28	-0.17	0.11	-0.16	-0.08	0.01	0.03	0.06	0.30	-0.40	-0.13	-0.07	-0.36	0.39	0.25	0.38	-0.16	0.01	-0.32
Hollow banks					1.00	0.16	0.25	-0.08	0.09	-0.15	0.24	0.07	0.19	0.25	0.11	0.19	0.22	0.06	0.35	-0.13	-0.26	-0.18	-0.34	0.39	-0.03	0.45	-0.10	0.00	-0.22
Depth						1.00	-0.23	-0.32	0.09	-0.51	0.35	-0.13	0.47	0.10	0.16	0.20	0.34	0.08	-0.09	-0.13	-0.12	0.35	-0.62	0.53	0.19	0.42	-0.57	-0.25	-0.25
Flow velocity							1.00	-0.22	-0.03	-0.19	-0.06	0.17	0.06	0.08	-0.03	0.22	0.15	-0.07	0.34	0.04	-0.25	-0.33	-0.10	0.19	0.06	0.15	0.26	0.21	-0.10
pH								1.00	0.15	0.66	-0.10	-0.22	-0.19	-0.02	-0.02	-0.20	-0.28	0.07	0.05	0.20	0.04	-0.25	0.18	-0.17	-0.04	-0.13	0.12	0.09	0.09
Temperature									1.00	-0.25	0.13	-0.23	0.12	-0.03	0.09	0.14	0.18	0.07	0.13	-0.08	-0.09	-0.01	-0.23	0.16	0.06	0.17	-0.09	-0.15	-0.21
Dissolved oxygen										1.00	-0.26	0.09	-0.40	-0.24	-0.20	-0.30	-0.31	0.11	-0.05	0.25	0.09	-0.24	0.55	-0.53	-0.13	-0.40	0.16	0.28	0.36
Conductivity											1.00	-0.18	0.56	0.41	0.46	0.33	0.28	0.03	0.20	-0.40	-0.22	0.31	-0.76	0.63	0.22	0.54	-0.46	-0.34	-0.24
Suspended solids												1.00	0.14	-0.05	0.13	0.57	0.54	0.10	-0.07	0.17	-0.01	-0.02	0.27	-0.27	-0.01	-0.21	0.02	-0.13	0.00
Ammonium													1.00	0.35	0.50	0.71	0.71	0.03	0.22	-0.15	-0.25	0.12	-0.64	0.53	0.09	0.47	-0.52	-0.29	-0.29
Nitrate														1.00	0.53	0.25	0.28	-0.04	0.15	-0.06	-0.11	0.00	-0.41	0.41	-0.19	0.37	-0.11	-0.21	-0.11
Total nitrogen															1.00	0.43	0.43	0.08	0.16	-0.16	-0.14	0.11	-0.40	0.30	0.02	0.28	-0.16	-0.28	-0.28
Phosphate																1.00	0.82	0.13	0.18	-0.04	-0.21	0.05	-0.43	0.39	0.18	0.31	-0.30	-0.26	-0.33
Total phosphorus																	1.00	0.12	0.12	-0.01	-0.17	0.06	-0.32	0.23	0.01	0.24	-0.36	-0.35	-0.25
COD																		1.00	0.09	-0.05	0.01	-0.07	-0.02	0.04	0.04	0.10	-0.24	0.07	-0.07
Pebbles																			1.00	0.04	-0.64	-0.59	-0.26	0.29	0.04	0.39	0.00	0.18	-0.10
Gravel																				1.00	-0.30	-0.39	0.36	-0.34	-0.18	-0.20	0.09	0.16	0.17
Sand																					1.00	0.08	0.30	-0.25	0.12	-0.40	0.12	-0.09	0.08
Loam/clay																						1.00	-0.27	0.17	0.06	0.12	-0.21	-0.33	-0.12
Distance to mouth																							1.00	-0.92	-0.29	-0.78	0.45	0.22	0.42

Stream order	1.00	0.34	0.80	-0.31	-0.08	-0.45
Tubificidae		1.00	0.18	-0.12	-0.13	-0.45
Asellidae			1.00	-0.33	-0.01	-0.41
Gammaridae				1.00	0.26	0.06
Baetis					1.00	0.22
Limnephilidae						1.00

5.3.3. Visual relation analysis between input and output variables

In combination with the summary statistics in Section 5.3.1., a third type of data analysis based on visualization graphs was applied. Based on these graphs, an additional insight is gained into the distribution, the outliers and the spatial and temporal variation of the data.

The first part of these analyses was performed with the WEKA software. The outcomes are presented in Appendix 1. The example for Limnephilidae in the Zwalm river basin is given in Fig. 5.1. For most variables a logical relation could be observed, e.g. Limnephilidae was mainly found when the values of width, stream order, distance to mouth and nutrient concentrations were low, while Asellidae mainly occurred when the values of width, stream order and distance to mouth were high. In addition, one could easily derive the distributions of the different variables. For some variables, small differences within these distributions were detected between both datasets, e.g. for the variable meandering. One could also observe that several combinations of input variable ranges were less represented. Therefore in addition to removing variables due to correlations, and instances because of outliers, one can consider to remove instances to make the distributions over all classes and values of the input variables more even. A transformation of the input variables can be considered as well. However, the dataset was kept as natural as possible, mainly to see whether the models can also cope with this bottleneck or not.

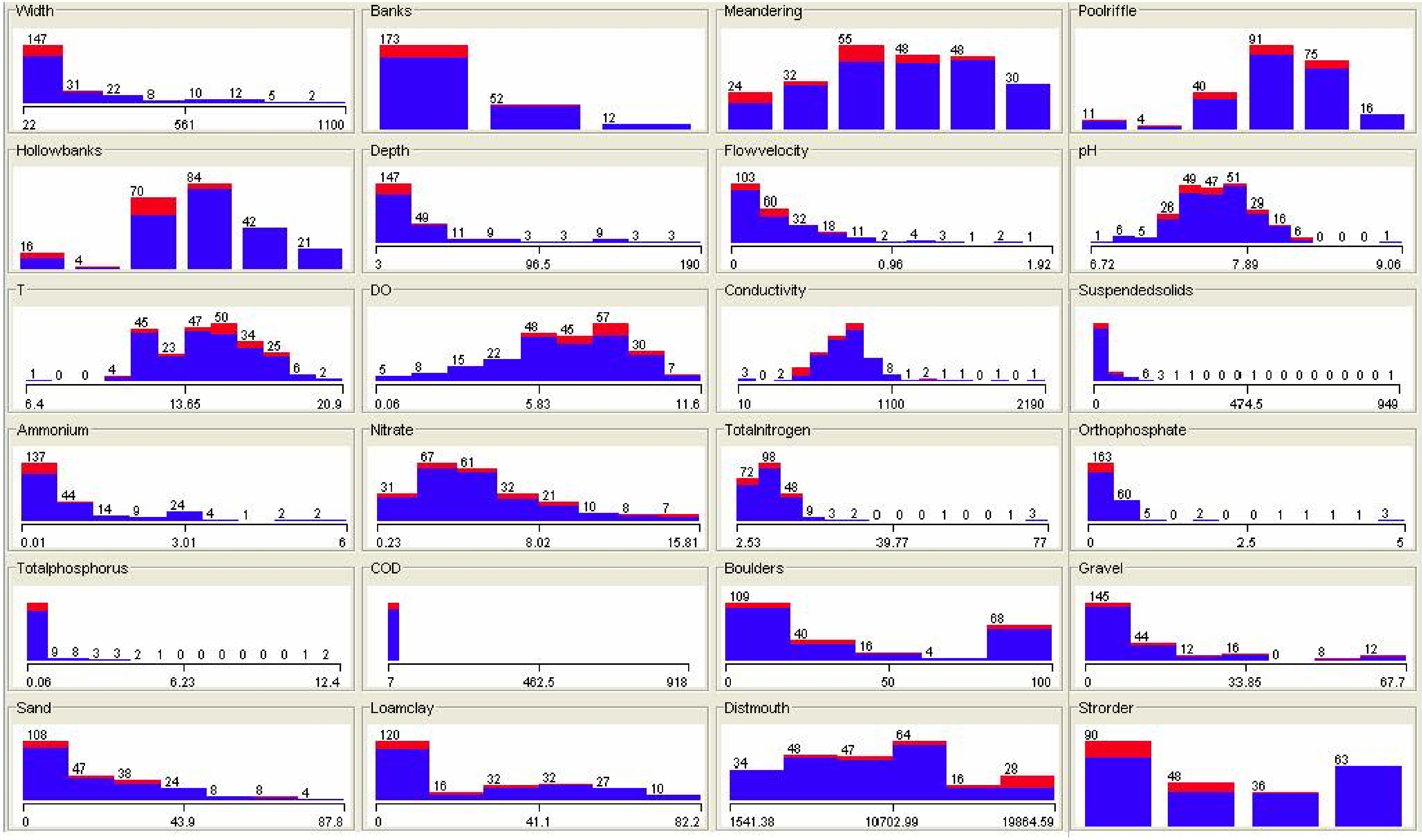
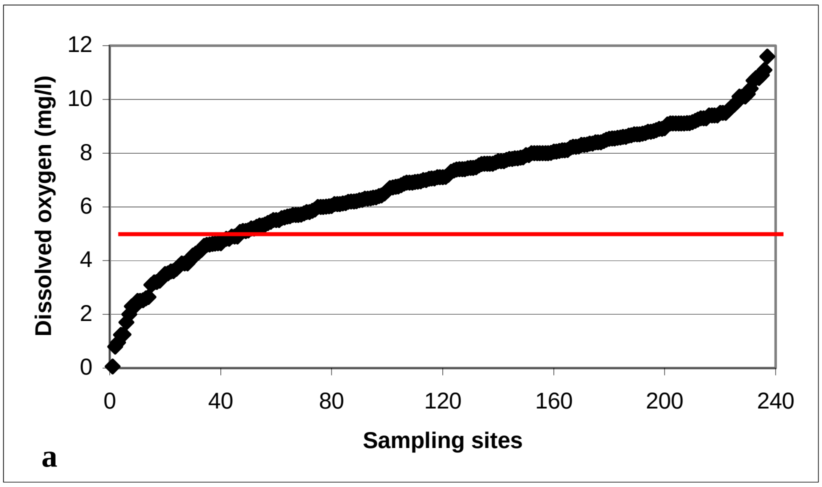


Fig. 5.1. Data relation visualization graphs for Limnephilidae presence/absence in the Zwalm river basin (in total 237 instances) in relation to the 24 environmental variables (Limnephilidae absent in 208 instances (blue), Limnephilidae present in 29 instances (red)).

Based on the second type of plots (the example for the variable dissolved oxygen is given in Fig. 5.2. All the plots for the dataset of the entire Zwalm river basin and the ‘short distance’ monitoring network are presented in Appendix 2). Similar conclusions could be drawn as in Section 5.3.1. In the Zwalm river basin dataset, outliers could be detected for the variables suspended solids, chemical oxygen demand, depth, width and part of the nutrients. In the dataset of the ‘short distance’ monitoring network, a few outliers could be observed for the variables suspended solids, chemical oxygen demand, depth, width and to a smaller extent for some nutrients.

To get better insight into the temporal variation over the different years, variables are plotted as illustrated in Fig. 5.3 for the variable dissolved oxygen (plots of the appropriate variables are presented in Appendix 3 for the dataset of the entire Zwalm river basin and the ‘short distance’ monitoring network). To observe spatial variation in both datasets and to detect hot spots for river restoration, values of the input and output variables were plotted on the maps of the study area. In Fig. 5.4, the maps for the input variable dissolved oxygen and the macroinvertebrate Asellidae (year 2002) are given for both study area (all plots are presented in Appendix 4).

Based on Appendix 3, major temporal trends were found for the variables temperature (generally: 2000 < 2001 < 2002, 2003), pH (generally: 2002 < 2003), dissolved oxygen (generally: 2002 < 2003), depth (generally: 2003 < 2002) and flow velocity (generally: 2003 < 2002) whereas for the other environmental variables no unambiguous tendencies could be observed. Paramount spots of pollution in the Zwalm river basin were detected in the northern part, near the city of Zottegem, and along the Zwalm river itself (Appendix 4). However, most of the upper reaches are still unpolluted. This is reflected in the habitat suitability of the Limnephilidae, an indicator of good water quality, which was exclusively found in these headwaters. Asellidae was mainly observed in the wider, deeper watercourses of the Zwalm whereas Gammaridae preferred apparently the smaller, fast flowing streams. For *Baetis*, no major trend could be detected while the tolerant taxon Tubificidae was found all over the river basin.



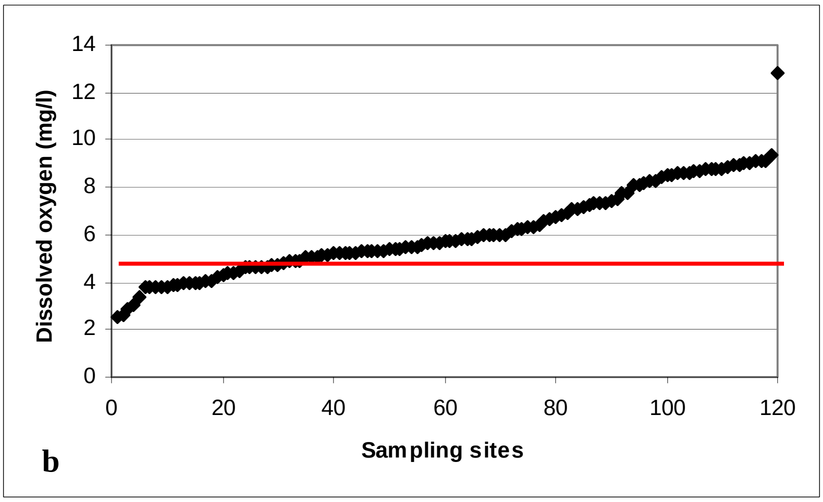


Fig. 5.2. Distribution of the variable dissolved oxygen (mg/l), with indication of the quality standard in red, over the 237 (a, Zwalm river basin) and 120 (b, 'short distance' monitoring network) sampling sites.

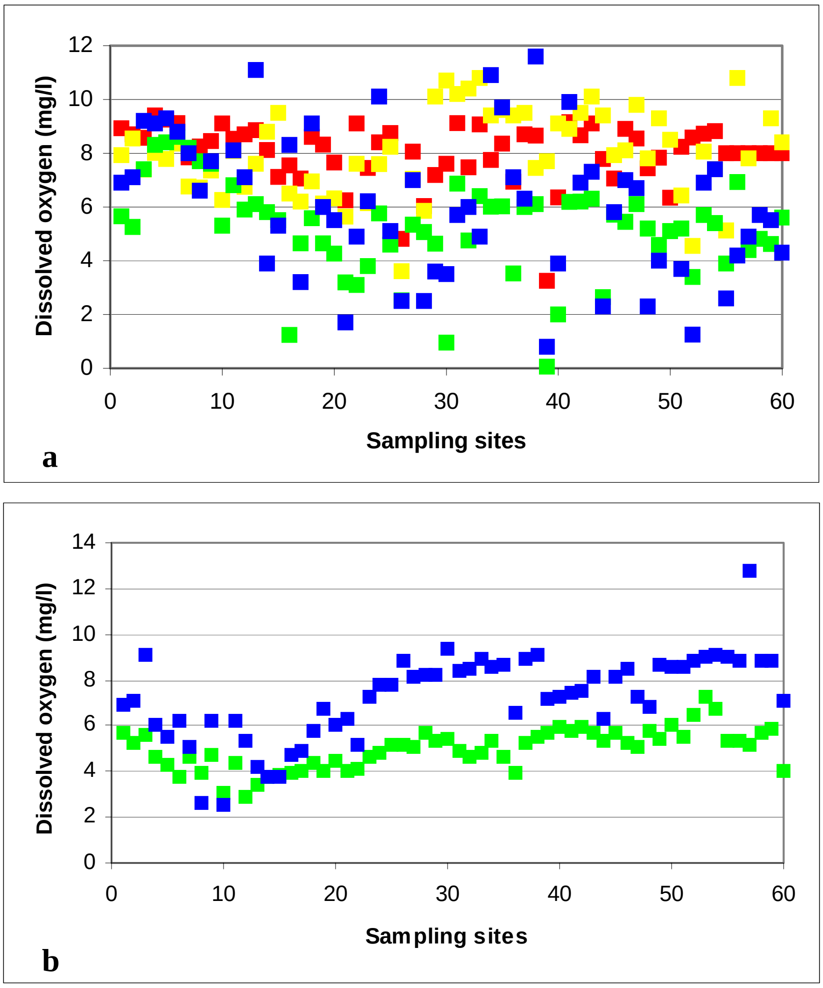


Fig. 5.3. Spatial variation of the variable dissolved oxygen (mg/l) over the respectively 4 years (a, Zwalm river basin, 237 sampling sites) and 2 years (b, 'short distance' monitoring network, 120 sampling sites) (red = 2000; yellow = 2001; green = 2002; blue = 2003).

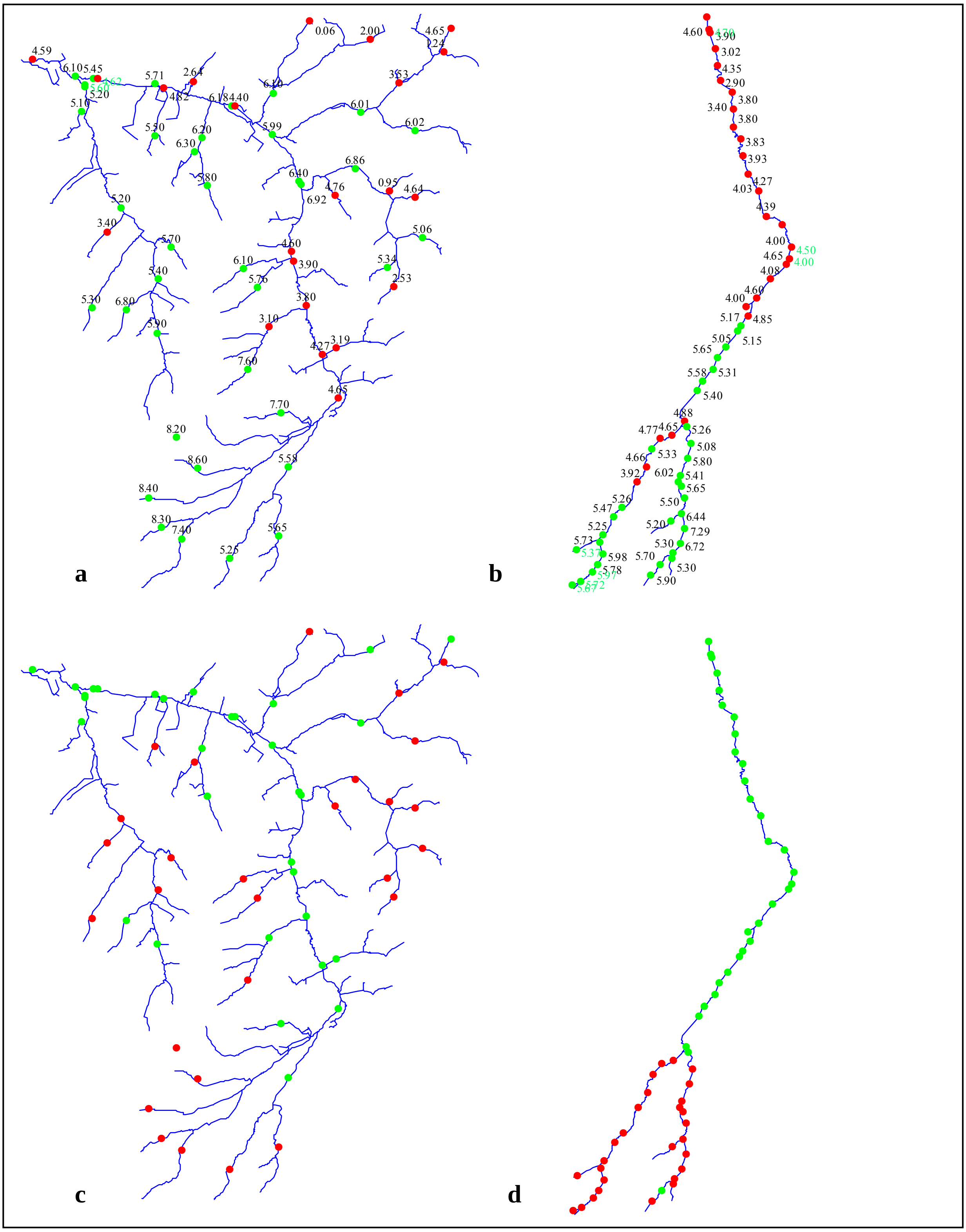


Fig. 5.4. Spatial distribution of the input variable dissolved oxygen (mg/l) (a = the Zwalm river basin, b = the ‘short distance’ monitoring network; green dot complies with the quality standard, red dot does not comply with the quality standard) and the macroinvertebrate Asellidae (c = the Zwalm river basin, d = the ‘short distance’ monitoring network; green dot = present, red dot = absent) for the year 2002.

5.4. Discussion

No standard procedures for preliminary data analyses were described in most articles reviewed in Chapter 2. 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).

An *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. In the Zwalm river basin dataset, at one site the artificial substrates disappeared in 2001 and two sites were totally overgrown in 2003. As a consequence, no biological measurement were available for these sites. Therefore, 237 instead of 240 instances could only be used. But also the reason behind the data collection played a key role. The Zwalm river basin 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 constraints played a major role why certain variables (metals, organic micropollutants, variables important for bio-availability calculations, certain hydraulic measurements) were not included in the database (Goethals, 2005). In addition, this also influenced the amount of instances (60 field observations per year was the absolute maximum with this set of variables). On the other hand, the 'short distance' monitoring network was developed with the aim of constructing migration models. For this purpose, a more intensive and dense sampling campaign was set up. 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 kind of variables play a major role on the ecosystems and can be important for river managers.

The presence of outliers led directly to a tough decision, whether to leave these measurements (instances) in the database or not, because they can surely 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.* (2005d) tested the sensitivity and robustness of the ANN models on Asellidae when data, containing variables beyond the range of the data for training, were added. To this end, the authors created a virtual dataset based on ecological expert knowledge to introduce these 'extreme' values in the model. According to this study, the presence/absence of Asellidae in the 'extreme' test set was predicted significantly better when the number of 'extreme' examples in the training set increased, while 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 summary statistics and the data visualization plots in Section 5.3.1. and 5.3.3., this study by Dedecker *et al.* (2005d) could be an argument for keeping the outliers in. But also to make the models applicable in the widest extent of theoretical (e.g. model optimization) and practical (e.g. effect prediction of river restoration scenarios) 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 or not 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 the dataset need too much preparation, the methods will probably become less attractive.

Both datasets of the Zwalm river basin consisted of variables with a high standard deviation and high maxima in comparison with the median values. However, by rechecking in the field one became 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.

According to Walczak and Cerpa (1999), any r with an absolute value of 0.20 or higher indicates a probable noise source in particular to ANN models. They advice to consider the removal of one of these variables without adversely affecting the ANN performance. However, there might be practical reasons to leave these correlated variables in, such as ecologically not relevant correlations (not causal, 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. specific channel modifications). This exercise is in this respect very rewarding, 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 (Goethals, 2005).

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 worth studying: 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).

The number of variables used in both databases (24) is relatively high compared to most articles referred to in the review presented in Chapter 2, 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 thesis, they were all retained, mainly for reasons of 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.

Because the environmental input variables span different ranges as illustrated in this chapter, they have to be standardized before the modelling can be started. In this way, all variables receive equal attention during the training process. The methods used for standardization will be discussed in Section 6.2.1.

5.5. Conclusions

Several factors played a key role in the final set of variables (and the amount of instances) that are presented to the ANN models. 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 of the Zwalm river basin was built with the aim of model development whereas the 'short distance' monitoring network was developed with the purpose of constructing migration models. However, 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) increased over 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 thesis, they were finally all kept in.

The results obtained by these data analyses can be used to interpret the ANN modelling results in the next chapters. In addition, they could be very useful to detect hot spots for river restoration and as a result to collect new or additional data and how this should be done to develop more appropriate models and to gather meaningful information for river management afterwards. To ensure all variables received equal attention during the ANN training process, the variables were standardized. The latter will be discussed in the next chapter.

Chapter 6
Development of Artificial Neural Network models for the prediction of macroinvertebrates

<u>Parts of this chapter are based on</u>:

Dedecker, A.P., Goethals, P.L.M., Gabriels, W. and De Pauw, N. (2004). Optimization 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.

6.1. Introduction

The general objective of this study was the development, analysis and optimization of habitat suitability models based on Artificial Neural Networks (ANNs) for the prediction of macroinvertebrates. It was shown that machine learning techniques such as Artificial Neural Networks (ANNs) basically mimic aspects of biological information processing for data modelling and could be useful in ecology (Recknagel, 2001). The prediction of aquatic communities, such as macroinvertebrates (as discussed in Chapter 2), fishes, macrophytes, algae, … by means of ANN models has recently been discussed by several authors (e.g. Wagner et al., 2000; Hoang et al., 2001; Maier and Dandy, 2001; Park et al., 2001; Reyjol et al., 2001; Scardi, 2001; Wei et al., 2001; Wilson and Recknagel, 2001; Jorgensen et al., 2002; Olden and Jackson, 2002; Ibarra et al., 2003; Lee et al., 2003; Park et al., 2003a). The traditional view of an ANN is of a programme that emulates biological neural networks and 'learns' to recognize patterns by being trained on a set of sample data from the domain. Learning through training and subsequently the ability to generalize is the unique perceived source of intelligence in an ANN. According to Haykin (1999), generalization 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 size of the training set and the architecture are the only two factors that can be influenced in the modelling process, making it crucial steps, which should be considered carefully. However, it is stressed that the ANN architecture is generally highly problem dependent (Maier and Dandy, 2000). For this reason, it is necessary to develop and optimize the ANNs to obtain the best model configuration that gives lower error with minimal computing time. Design of optimal neural networks is challenging in that there exists a large number of alternative ANN physical architectures and learning methods. 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 variables, selection of a learning method, design of the number of hidden layers and selection of the number of hidden neurons for each layer.

The specific aim of this study was to discuss the development and optimization of different neural network models to obtain the best model configuration for the prediction of the macroinvertebrate taxa specified in Section 4.5: Tubificidae, Asellidae, Gammaridae, *Baetis*, Limnephilidae. Based on the criteria mentioned by Haykin (1999) and Walczak and Cerpa (1999), the following actions were considered:

1) selection of relevant environmental variables as input for macroinvertebrate prediction (see Chapter 5);
2) choosing the backpropagation algorithm as appropriate learning method. Because of its generality (robustness) and ease of implementation, this algorithm is the best choice for the majority of ANN systems. In addition, the ANNs used for prediction in ecology are generally trained with the backpropagation method (Cherkassky and Lari-Najafi 1992; Maier and Dandy 2000) (Table 2.6);
3) searching for the optimal size of training and test set since the size of the training and test set influences respectively the generalization capability of the model and accuracy of the model performance. Therefore, nine cross-validation methods (2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation) were tested (Section 6.3.1);
4) searching for the optimal ANN architecture for the five selected macroinvertebrates in both databases to use the optimal training and test size concerning two questions (Section 6.3.2):

    - How many hidden layers should exist in the ANN architecture?
    - How many neurons should be present in the hidden layer(s)?

    Additionally, the optimal ANN architecture was analysed, examining the effect of annual testing. ANN models for the entire Zwalm river basin (237 sampling sites) were trained with measured input and output data from 3 years, while data of the remaining independent year was used for testing. Similarly, the ANN models for the 'short distance' monitoring network (120 sampling sites) were trained with data from one year and tested with data of the other year.

6.2. Material and methods

The neural network models were implemented in the software package MATLAB 6.1. for MS Windows^{TM} (Matworks$^{®}$) (according to Gevrey *et al.*, 2003). MATLAB is an interactive computer programme that serves as a convenient 'laboratory' for computations involving 'matrices'. A general description of ANN model development is given in Chapter 2. In the present chapter, attention is paid to the specific model settings used in this study.

6.2.1. Data processing

In both datasets, the different environmental variables span different ranges (see Chapter 5). In order to ensure that all variables receive equal attention during the training process, the input variables were standardized based on the following equation:

$$V_n = \frac{V_O - \overline{V_O}}{\sigma_{V_O}} \tag{7.1}$$

in which V_0 and V_n are respectively the old and new value of the variable for a sampling point, $\overline{V_O}$ and σ_{V_O} are respectively the average and the standard deviation of that variable in the original dataset. The output values (0 and 1, respectively the absence and presence of the macroinvertebrates) were not rescaled since they are already adapted to the logarithmic sigmoid transfer function ($f(x) = \frac{1}{1+\exp^{(-x)}}$) used in the output layer. The continuous network output is mapped to 0 and 1 using a threshold of 0.5.

6.2.2. Model architecture

The modelling method used was a multi-layer feed-forward neural network based on the principles of the backpropagation algorithm (Rumelhart *et al.*, 1986) which applies examples of data with known outputs. As mentioned in Chapter 2, a backpropagation neural network typically comprises three types of neuron layers: an input layer, one or more hidden layers and an output layer each including one or more neurons. As shown in Fig. 2.1, neurons from one layer are connected to all neurons in the following layer, but no lateral connections within any layer, nor feed-back connections are possible. The network consisted of 24 input neurons, each representing an environmental variable. The output layer comprises one neuron, indicating the presence or absence of the macroinvertebrate taxon. The number of hidden neurons in the hidden layer(s) depended on the problem to be solved:

- to define the optimal size of training and test set, one hidden layer comprising 10 neurons was used extracted from the rules of thumb summarized in Chapter 2;
- to define the optimal model architecture for the five selected macroinvertebrates in both databases, several number of layers and neurons were analysed: six three-layered and four four-layered networks with respectively [2], [5], [10], [15], [20], [25] and [5 5], [5 10], [10 10], [10 20] neurons in the hidden layer(s) were tested. The same number of hidden neurons was analysed to investigate the effect of annual testing.

6.2.3. Backpropagation algorithm

The main principle of the backpropagation algorithm is that the (connection) weights and biases in the network are updated in order to decrease the error term, 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. The backpropagation algorithm can be divided in three major steps which are discussed below in detail (Fig. 6.1):

- Step 1: the input vectors are presented to the network and are transferred through the network from the input to the output layer to obtain the predicted output;
- Step 2: the predicted and the observed output are compared and the error term is calculated;
- Step 3: correction of the weights and biases by modification of the connection values backwards through the network, this is the ‘backpropagation of the error’.

The algorithm discussed in detail here is the algorithm in case of a network with one hidden layer. The indices e, i and s are used to indicate respectively the input layer, the hidden layer and the output layer. w is used for the connection weights. Before training, the values of the weights and biases are initially set to small random numbers ([-0.3 0.3]).

Step 1:

The information x_e is transferred from the input to the hidden layer across the connection weights w_{ei} included between both layers. b_e is the bias which is added to the neurons from the input layer. The net input for the neurons of the hidden layer is:

$$net_{ei} = \sum_i w_{ei} x_e + b_e \tag{7.2}$$

The net input of the hidden neurons is fed into a logarithmic transfer function:

$$x_i = f\left(net_{ei}\right) \tag{7.3}$$

The information x_i, as output of the hidden neurons, is transferred from the hidden to the output layer across the connection weights w_{is} included between both layers. b_i is the bias which is added to the neurons from the hidden layer. The net input for the neuron of the output layer is:

$$net_{is} = \sum_s w_{is} x_i + b_i \tag{7.4}$$

The net input of the output neuron is fed into a logarithmic transfer function which results in the predicted output $\hat{x}_s$:

$$\hat{x}_s = f\left(net_{is}\right) \tag{7.5}$$

B_e and B_i have an input arbitrarily fixed at 1 and the weights associated with these biases have to be trained like normal weights. Biases can be considered as a constant value added to the network and permit to increase the quality of the obtained predictions.

Step 2:

The predicted ($\hat{x}_s$) and the observed (x_s) output are compared and the error term is calculated:

$$err = x_s - \hat{x}_s \tag{7.6}$$

Several error measures can be applied. In this study however, the 'Sum Squared Error' (SSE) is used:

$$SSE = \sum (x_s - \hat{x}_s)^2 \tag{7.7}$$

<u>Step 3</u>:

The connection weighs w_{is} between the hidden and the output layer are updated:

$$w_{is}(t+1) = \Delta w_{is} + \alpha w_{is}(t) \tag{7.8}$$

in which α is the 'momentum' (the momentum term takes into account the previous weight update), t is the value of the iteration and

$$\Delta w_{is} = \eta \delta_{is}(t+1) x_i (t+1) \tag{7.9}$$

in which η is the 'learning rate' (the learning rate is directly proportional to the size of the steps taken in weight update) and

$$\delta_{is} = f'(net_{is}) err \tag{7.10}$$

The connection weights w_{ei} between the input and the hidden layer are updated:

$$w_{ei}(t+1) = \Delta w_{ei} + \alpha w_{ei}(t) \tag{7.11}$$

in which $\Delta w_{ei} = \eta \delta_{ei}(t+1) x_e (t+1)$ (7.12)

in which $\delta_{ei} = f'(net_{ei}) \left(\sum \delta_{is} w_{is}\right)$ (7.13)

The settings of the ANN parameters were all default and determined on the basis of experience: learning rate = 0.001, incremental learning rate = 1.05, decreasing learning rate = 0.75, momentum = 0.95 and the transfer function in the hidden and output layer = logarithmic sigmoid transfer function.

A complete calculation cycle of the network is called 'iteration (t)' or 'epoch'. These three steps, which form one iteration, are repeated several times. A certain number of iterations is needed to obtain an acceptable error level.

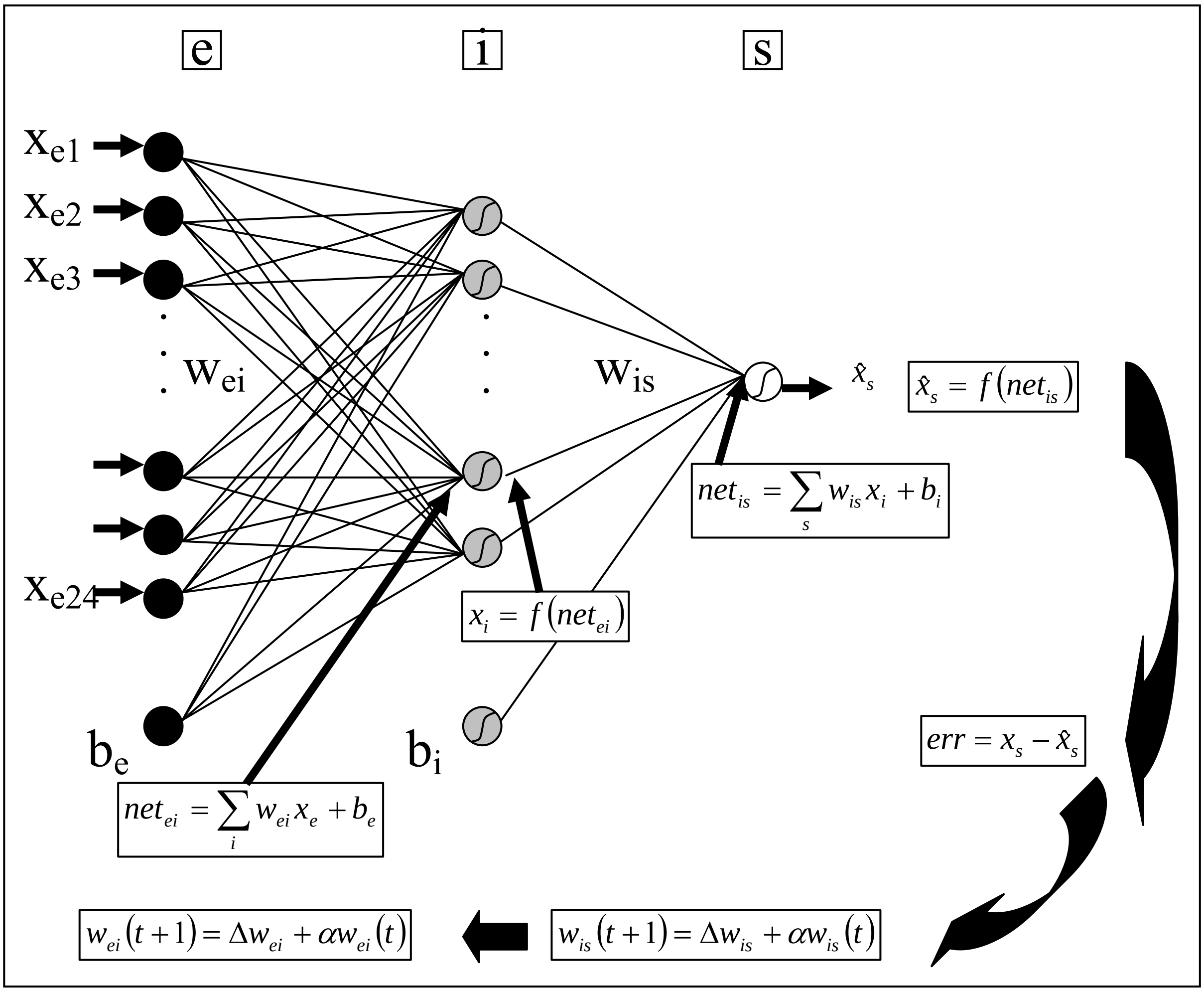


Fig. 6.1. A calculation cycle of the backpropagation algorithm. The information x_e is transferred from the input to the output layer across the connection weights. The obtained error between the predicted ($\hat{x}_s$) and the observed (x_s) output is used to update the values of the connection weights backwards through the network.

6.2.4. Model validation

The models were evaluated on the basis of two performance measures: the percentage of Correctly Classified Instances (CCI) and Cohen's kappa (CK). For this one requires the derivation of matrices of confusion that identified true positive (TP) (= known distributional habitats correctly predicted as present), true negative (TN) (= sites where the species has not been found and that are classified by the model as absent), false positive (FP) (= a measure of sites of absence (or 'pseudo-absence) incorrectly predicted present, commission error) and false negative (FN) (= sites of known distributions predicted absent by the model, omission error) cases predicted by each model (Fielding and Bell, 1997; Manel et al., 2001). In that way, observed presence/absence patterns were tabulated against those predicted (Table 6.1).

Table 6.1. 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	+	TP	FP
	-	FN	TN

The first performance measure that was calculated was the percentage of Correctly Classified Instances (CCI):

$$CCI = \frac{(TP + TN)}{(TP + FP + FN + TN)} \times 100 \qquad (7.14)$$

There is however clear evidence that this CCI is affected by the frequency of occurrence of the test organism being modelled (Fielding and Bell, 1997; Manel *et al.*, 2001). Among the different measures, which are based on the confusion matrix, Fielding and Bell (1997) and Manel *et al.* (2001) recommended Cohen's kappa (Cohen, 1960) as a reliable performance measure, since the effect of prevalence on Cohen's kappa appeared to be negligible (e.g. Dedecker *et al.*, 2004). It is a simply derived statistic that measures the proportion of all possible cases of presence or absence that are predicted correctly by a model after accounting for chance predictions:

$$CK = \frac{\left[(TP+TN)-(((TP+FN)(TP+FP)+(FP+TN)(FN+TN))/N)\right]}{\left[N-(((TP+FN)(TP+FP)+(FP+TN)(FN+TN))/N)\right]} \tag{7.15}$$

where N is the total number of Instances. A CCI of at least 70.0 % and CK higher than 0.40 (Landis and Koch, 1977; Fielding and Bell, 1999) were considered as good classifications.

The model validation was based on stratified *k*-fold cross-validation. For *k*-fold cross-validation the data are split into *k* folds or partitions. Each fold in turn is used for testing while the rest is used for training. That is, use *k-1* folds for training and 1 folds for testing, and repeat the procedure *k* times so that in the end, every instance has been used exactly once for testing. Since the size of the training and test set influences respectively the generalization capability of the model and accuracy of the model performance, the optimal size of training and test set was searched for. Therefore, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation was tested. To avoid biased training, the amount of instances in which Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae were present was similar in all training and test sets. Additionally, before training of the neural network, the data were randomly shuffled in the training datasets.

In order to check the stability of each model, the training of the network was repeated *k* times, according to the *k*-fold cross-validation. Based on this repetition, the average performances (CCI and CK) were calculated. The *k*-fold cross-validation allowed also for calculating the standard deviation. This standard deviation gave an indication of the stability of each model.

The objective of the training part is to find a balance between memorization and generalization. Memorization means that the network is able to produce a correct output based on the input data used for training. On the other hand, generalization means that the network is able to produce an acceptable output based on similar but not identical input data used for training. The more the network is trained, the better the training data is memorized, i.e. a smaller SSE based on the training set. Therefore, an independent test dataset is used to stop training. During the training procedure, the error shows at first a decrease on the test set as well as on the training set, as the network generalizes from the training data to the underlying input/output function. After some training steps, the network starts overfitting. Consequently the error on the test set increases, while the error on the training set keeps decreasing (Fig. 6.2). The training procedure can therefore be stopped at the iteration ‘u’ with the smallest error on the validation set, because at this point the network is expected to yield the best generalization performance (known also as 'early stopping') (Berthold and Hand, 1999).

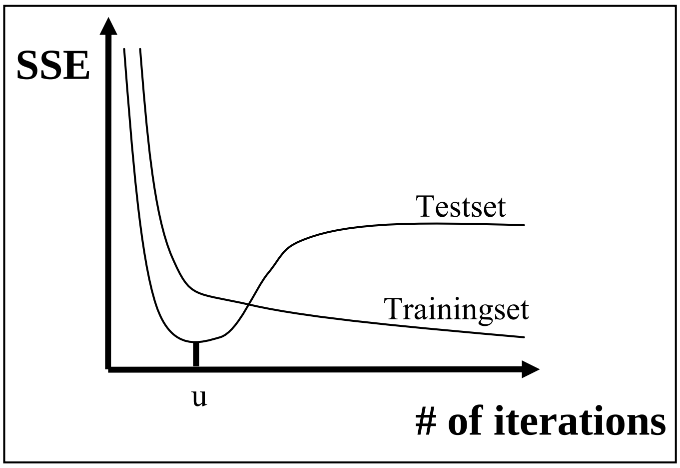


Fig. 6.2. Error plot on the training set and on the test set as a function of the number of iterations. The training procedure can be stopped at the iteration 'u'.

6.3. Results

6.3.1. Determination of the optimal training and test size

6.3.1.1. Determination of the optimal training and test size based on the dataset of the Zwalm river basin

The predictive performances based on the CCI and the CK of the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae are shown in Table 6.2. For Tubificidae, the performances based on the CCI were very similar for all cross-validation methods and gave an indication of very good model prediction capacity (CCI between 88.2 ± 3.3 % and 90.3 ± 2.8 %). Based on the CK however, weak models were generated (CK below 0.20). The highest performance was obtained based on the 8-fold cross-validation. However, the model stability based on the standard deviation of the CK was very low. In this way, it could be decided that the 4-fold cross-validation was more appropriate. For Asellidae, reliable models were obtained for all the cross-validation methods based on both performance measures. The best CCI (79.8 ± 6.3 %) and CK (0.60 ±

0.13) were however achieved using 4-fold cross-validation. For Gammaridae, the best performances were generated using 10-fold cross-validation (CCI = 80.6 ± 6.9 %; CK = 0.44 ± 0.25). However, the standard deviation for this cross-validation method was much higher than for the 4-fold cross-validation while the performances were only slightly lower (CCI = 80.0 ± 1.5 %; CK = 0.43 ± 0.07). Although for *Baetis* rather good model predictions were achieved based on the CCI (higher than 70.0 %), the CK indicated that the major part of the good predictions were based on chance. All CK values were lower than 0.20. While the performances of the different cross-validation methods were very similar based on the CCI, the optimal training and test size for *Baetis* based on the CK was obtained by the 4-fold cross-validation which generated a CK of 0.18 ± 0.14. Similar results were achieved for the rare taxon Limnephilidae. Very good performances were obtained based on the CCI (CCI between 85.6 ± 2.5 % and 89.0 ± 2.1 %) while the CK (CK between 0.09 ± 0.16 and 0.28 ± 0.05) gave the indication of a rather poor to moderate model prediction capacity and stability. Moderate model performance and an acceptable model stability however was obtained by the 4-fold cross-validation (CK = 0.28 ± 0.05).

Table 6.2. Analysis of the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae for the entire Zwalm river basin (237 sites). Model evaluation is based on the percentage of Correctly Classified Instances (CCI) and Cohen's kappa (CK), model stability on the standard deviation (Stdev)

	k	**Average CCI**	**Stdev CCI**	**Average CK**	**Stdev CK**
Tubificidae	2	89.5	0.5	0.00	0.00
Tubificidae	3	89.9	0.0	0.10	0.17
Tubificidae	4	88.2	3.3	0.10	0.12
Tubificidae	5	89.5	0.1	0.00	0.00
Tubificidae	6	89.0	1.2	0.07	0.12
Tubificidae	7	89.5	1.5	0.04	0.10
Tubificidae	8	90.3	1.1	0.12	0.24
Tubificidae	9	89.9	1.8	0.05	0.16
Tubificidae	10	90.3	2.8	0.08	0.25
Asellidae	2	76.0	4.0	0.52	0.08
Asellidae	3	75.1	6.0	0.50	0.12
Asellidae	4	79.8	6.3	0.60	0.13
Asellidae	5	77.7	7.4	0.55	0.15
Asellidae	6	77.7	7.7	0.55	0.15
Asellidae	7	76.8	7.3	0.54	0.14
Asellidae	8	78.2	9.3	0.56	0.19
Asellidae	9	78.2	9.2	0.56	0.18
Asellidae	10	78.2	11.3	0.56	0.23
Gammaridae	2	77.2	3.7	0.39	0.14
Gammaridae	3	76.4	0.7	0.37	0.07
Gammaridae	4	80.0	1.5	0.43	0.07
Gammaridae	5	80.2	4.6	0.42	0.18
Gammaridae	6	78.9	4.4	0.39	0.15
Gammaridae	7	80.6	6.3	0.42	0.24
Gammaridae	8	78.5	4.7	0.37	0.15
Gammaridae	9	78.9	4.2	0.38	0.17
Gammaridae	10	80.6	6.9	0.44	0.25
Baetis	2	75.5	0.1	0.00	0.00
Baetis	3	75.5	1.5	0.06	0.09
Baetis	4	75.6	6.2	0.18	0.14
Baetis	5	76.4	1.7	0.10	0.14
Baetis	6	75.6	5.7	0.17	0.18
Baetis	7	75.1	4.6	0.17	0.17
Baetis	8	76.4	3.3	0.15	0.20
Baetis	9	76.8	2.9	0.16	0.20
Baetis	10	74.7	2.9	0.13	0.18
Limnephilidae	2	87.3	1.3	0.16	0.07
Limnephilidae	3	87.3	1.3	0.09	0.16
Limnephilidae	4	87.3	1.7	0.28	0.05
Limnephilidae	5	87.3	3.3	0.24	0.17
Limnephilidae	6	86.9	1.9	0.11	0.13
Limnephilidae	7	87.3	2.8	0.13	0.21
Limnephilidae	8	85.6	2.5	0.11	0.22
Limnephilidae	9	85.7	6.5	0.21	0.32
Limnephilidae	10	89.0	2.1	0.19	0.25

6.3.1.2. Determination of the optimal training and test size based on the dataset of the 'short distance' monitoring network

The predictive performances based on the CCI and the CK of the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae for the 'short distance' monitoring network are presented in Table 6.3. Based on the CCI, very good predictions were obtained for Tubificidae. A CCI between 92.5 ± 1.2 % and 95.1 ± 3.9 % for respectively 2- and 7-fold cross-validation was reached. The CK however ranged between 0.00 ± 0.00 and 0.29 ± 0.49 for the same cross-validation method. Contrary to the CCI, the low CK indicated a bad prediction capacity of the ANN models. Although the 7-fold cross-validation had a slightly higher CCI and CK and the 9-fold cross-validation a slightly higher CK (0.28 ± 0.46), the optimal *k*-fold cross-validation could be decided at 4 (CCI = 93.3 ± 0.0 %; CK = 0.27 ± 0.32) since a much lower standard deviation was obtained. The best performance for Asellidae was obtained using the 4-fold cross-validation. The average CCI (94.2 ± 1.7 %) as well as the average CK (0.88 ± 0.03) were the highest while the standard deviation of both performance measures was lower than for the other cross-validation methods. Excellent average performances and low standard deviations pointed out that very good models were obtained. For Gammaridae, the performance based on the CCI was high for all the cross-validation methods (between 90.0 ± 2.4 % and 93.3 ± 7.1 %). Also based on the CK, the predictions could be considered as good (higher than 0.40). Although the CK of the 8-fold cross-validation was slightly better (0.63 ± 0.42) than for the 4-fold cross-validation (0.62 ± 0.13), it could be decided that the optimal training and test size for Gammaridae was obtained applying the 4-fold cross-validation based on the standard deviation which was much lower. For *Baetis*, the best performances were obtained using the 10-fold cross-validation (CCI = 78.3 ± 10.5 %; CK = 0.55 ± 0.21). Although the performance was slightly better for the 10-fold cross-validation, the optimal cross-validation method was the 4-fold cross-validation since the model was much more stable based on the lower standard deviation. For Limnephilidae, similar results were obtained. The 10-fold cross-validation method had a better CCI (88.3 ± 8.1 %) in comparison with the 4-fold cross-validation (86.7 ± 5.4 %), but the obtained CK for the 4-fold cross-validation was higher (0.49 ± 0.20 in comparison with 0.43 ± 0.43 for the 10-fold cross-validation). In addition, more stable models were generated applying the 4-fold cross-validation.

Table 6.3. Analysis of the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae for the 'short distance' monitoring network (120 sites). Model evaluation is based on the percentage of Correctly Classified Instances (CCI) and Cohen's kappa (CK), model stability on the standard deviation (Stdev)

	k	**Average CCI**	**Stdev CCI**	**Average CK**	**Stdev CK**
Tubificidae	2	92.5	1.2	0.00	0.00
Tubificidae	3	92.5	0.0	0.12	0.21
Tubificidae	4	93.3	0.0	0.27	0.32
Tubificidae	5	92.5	1.9	0.22	0.30
Tubificidae	6	94.2	2.0	0.24	0.37
Tubificidae	7	95.1	3.9	0.29	0.49
Tubificidae	8	94.2	4.3	0.24	0.47
Tubificidae	9	93.3	4.4	0.28	0.46
Tubificidae	10	93.3	5.3	0.27	0.44
Asellidae	2	90.8	5.9	0.82	0.12
Asellidae	3	90.0	5.0	0.80	0.10
Asellidae	4	94.2	1.7	0.88	0.03
Asellidae	5	92.5	1.9	0.85	0.04
Asellidae	6	92.5	2.7	0.85	0.05
Asellidae	7	93.4	2.0	0.87	0.04
Asellidae	8	93.3	5.0	0.87	0.10
Asellidae	9	93.3	2.5	0.87	0.05
Asellidae	10	93.3	7.7	0.86	0.16
Gammaridae	2	90.0	2.4	0.57	0.07
Gammaridae	3	90.0	2.5	0.49	0.19
Gammaridae	4	91.7	3.3	0.62	0.13
Gammaridae	5	91.7	4.2	0.57	0.33
Gammaridae	6	91.7	4.1	0.54	0.30
Gammaridae	7	92.5	6.6	0.58	0.42
Gammaridae	8	93.3	7.1	0.63	0.42
Gammaridae	9	91.6	5.9	0.51	0.41
Gammaridae	10	92.5	6.1	0.59	0.37
Baetis	2	67.5	3.5	0.34	0.09
Baetis	3	68.3	13.8	0.35	0.30
Baetis	4	77.5	1.7	0.54	0.04
Baetis	5	72.5	14.3	0.44	0.29
Baetis	6	70.0	7.7	0.39	0.17
Baetis	7	77.5	8.6	0.54	0.18
Baetis	8	67.5	11.5	0.33	0.24
Baetis	9	75.8	12.3	0.52	0.23
Baetis	10	78.3	10.5	0.55	0.21
Limnephilidae	2	82.5	1.2	0.00	0.00
Limnephilidae	3	84.2	8.0	0.37	0.31
Limnephilidae	4	86.7	5.4	0.49	0.20
Limnephilidae	5	84.2	3.5	0.29	0.20
Limnephilidae	6	85.8	6.6	0.41	0.27
Limnephilidae	7	82.5	4.6	0.18	0.20
Limnephilidae	8	85.8	5.6	0.36	0.32
Limnephilidae	9	84.3	6.4	0.30	0.26
Limnephilidae	10	88.3	8.1	0.43	0.43

6.3.2. Determination of the optimal model architecture

This study aimed at determining the effect of ANN model architecture when analysing the relationship between river characteristics and the presence/absence of the five macroinvertebrate taxa Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae. Derived from the results of Section 6.3.1 ‘Determination of the optimal training and test size’ (= optimal number of folds), model validation was based on 4-fold cross-validation for both datasets (the entire Zwalm river basin and the ‘short distance’ monitoring network). The model architectures tested for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae were six three-layered and four four-layered networks with respectively [2], [5], [10], [15], [20], [25] and [5 5], [5 10], [10 10], [10 20] neurons in the hidden layer(s).

6.3.2.1. Determination of the optimal model architecture based on the dataset of the Zwalm river basin

The percentage of Correctly Classified Instances and the Cohen’s kappa are shown in Table 6.4. For all macroinvertebrates, model performance based on the CCI was very similar and independent of the model architecture tested. The percentage of CCI ranged respectively between 88.2 % and 90.3 %, 76.4 % and 79.8 %, 77.2 % and 80.2 %, 71.3 % and 77.2 % and 84.4 % and 89.0 % for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae. Based on the CCI, the predictive results were good for Asellidae, Gammaridae and *Baetis* and even excellent for Tubificidae and Limnephilidae. However the CKs for the latter two organisms (respectively between 0.00 and 0.24 and between 0.10 and 0.34) indicated that these high CCI values were for a major part related to their relatively high and low prevalence (Tubificidae was present in 89.5 % and Limnephilidae in 12.2 % of the instances) in the Zwalm river basin 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 measures. Although satisfying results were obtained for *Baetis* based on the CCI, similar conclusions as for Tubificidae and Limnephilidae could be drawn based on the CK (between 0.01 and 0.18). This means again that a good CCI can be achieved without using any information from the environmental variables. The CK value for Asellidae (CK between 0.53 and 0.60) and Gammaridae (CK between 0.33 and 0.47) on the other hand, revealed that reliable models were obtained for both organisms (CCI > 70 % and CK > 0.40) for the major part of the model architectures.

Comparing the different model architectures for each organism, performance measures were very similar using the dataset of the entire Zwalm river basin. In this way, no clear conclusions could be drawn according to the optimal number of hidden neurons. Although the predictive performance for Tubificidae was in general very low based on CK, the optimal number of hidden neurons was 5. However, the large standard deviation of the CK over the four folds pointed out that this model was very unstable. Similar results were obtained for *Baetis* and Limnephilidae. The best, however very unstable models were generated using respectively 10 and 5 hidden neurons. For Asellidae and Gammaridae, optimal model architectures were respectively those containing 10 and 15 hidden neurons.

Table 6.4. Determination of the optimal model architecture for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae based on the dataset of the Zwalm river basin (237 sites). Model validation is based on 4-fold cross-validation. Model evaluation is based on the percentage of Correctly Classified Instances (CCI) and Cohen's kappa (CK), model stability on the standard deviation (Stdev)

	# of neurons	**Average CCI**	**Stdev CCI**	**Average CK**	**Stdev CK**
Tubificidae	2	89.5	0.7	0.00	0.00
Tubificidae	5	90.3	2.1	0.12	0.24
Tubificidae	10	88.2	3.3	0.10	0.12
Tubificidae	15	89.5	0.7	0.00	0.00
Tubificidae	20	89.5	0.7	0.05	0.10
Tubificidae	25	88.6	1.6	0.03	0.06
Tubificidae	5 - 5	89.5	0.7	0.00	0.00
Tubificidae	5 - 10	89.9	1.3	0.07	0.13
Tubificidae	10 - 10	89.5	0.7	0.00	0.00
Tubificidae	10 - 20	89.5	0.7	0.00	0.00
Asellidae	2	78.5	7.6	0.57	0.24
Asellidae	5	78.1	8.2	0.56	0.26
Asellidae	10	79.8	6.3	0.60	0.20
Asellidae	15	77.7	9.0	0.55	0.29
Asellidae	20	78.9	8.2	0.58	0.26
Asellidae	25	77.2	6.9	0.55	0.22
Asellidae	5 - 5	76.4	4.5	0.53	0.14
Asellidae	5 - 10	76.8	5.6	0.54	0.18
Asellidae	10 - 10	78.1	9.1	0.56	0.29
Asellidae	10 - 20	76.8	7.0	0.54	0.22
Gammaridae	2	77.7	5.1	0.43	0.14
Gammaridae	5	78.9	2.3	0.41	0.09
Gammaridae	10	80.0	1.5	0.43	0.07
Gammaridae	15	80.2	5.6	0.47	0.15
Gammaridae	20	78.0	2.5	0.40	0.08
Gammaridae	25	79.3	2.2	0.39	0.08
Gammaridae	5 - 5	78.9	3.3	0.42	0.12
Gammaridae	5 - 10	77.2	1.1	0.41	0.02
Gammaridae	10 - 10	77.7	5.1	0.33	0.26
Gammaridae	10 - 20	78.1	4.9	0.34	0.25
Baetis	2	74.7	2.4	0.01	0.03
Baetis	5	75.5	4.9	0.12	0.18
Baetis	10	75.5	6.2	0.18	0.14
Baetis	15	74.7	2.9	0.10	0.12
Baetis	20	73.8	5.2	0.15	0.12
Baetis	25	77.2	2.9	0.16	0.19
Baetis	5 - 5	72.1	4.1	0.05	0.08
Baetis	5 - 10	74.3	3.0	0.03	0.06
Baetis	10 - 10	73.8	2.1	0.08	0.10
Baetis	10 - 20	71.3	7.8	0.16	0.14
Limnephilidae	2	87.3	0.9	0.21	0.25
Limnephilidae	5	89.0	1.0	0.34	0.37
Limnephilidae	10	87.3	1.7	0.28	0.09
Limnephilidae	15	86.1	2.9	0.28	0.25
Limnephilidae	20	87.3	0.9	0.11	0.23
Limnephilidae	25	84.4	5.1	0.27	0.31
Limnephilidae	5 - 5	88.2	0.1	0.10	0.32
Limnephilidae	5 - 10	86.5	4.6	0.17	0.34
Limnephilidae	10 - 10	85.6	1.7	0.25	0.27
Limnephilidae	10 - 20	86.5	3.7	0.29	0.35

6.3.2.2. Determination of the optimal model architecture based on the dataset of the 'short distance' monitoring network

In Table 6.5 the CCI and CK are presented. The predictive performance based on the CCI was acceptable for *Baetis* (in the neighbourhood of about 70.0 %) and even very good for the remaining taxa (CCI value between 92.5 % and 94.2 %, 90.8 % and 94.2 %, 88.3 % and 92.5 %, 77.5 % and 86.7 % for respectively Tubificidae, Asellidae, Gammaridae and Limnephilidae). The CCI was strikingly constant over the four folds for all taxa, with a few exceptions for *Baetis*. Although the predictive success was slightly better based on the dataset of the 'short distance' monitoring network in comparison with the dataset of the entire Zwalm river basin, similar conclusions could be drawn here based on the CK. Poor to moderate results were obtained for Tubificidae, Limnephilidae and *Baetis* while for Asellidae, the results based on the CK were excellent (higher than 0.80 for all model architectures). Also for Gammaridae the predictions were good based on the CK, with one exception for the four-layered network with 5 neurons in both hidden layers where the CK drops to 0.29.

The optimal number of hidden neurons could be decided at 15 or 20 for Tubificidae, 10 or 10 – 10 for Asellidae, 10 – 10 for Gammaridae, 10 for *Baetis* and 10 for Limnephilidae.

Table 6.5. Determination of the optimal model architecture for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae based on the dataset of the 'short distance' monitoring network (120 sites). Model validation is based on 4-fold cross-validation. Model evaluation is based on the percentage of Correctly Classified Instances (CCI) and Cohen's kappa (CK), model stability on the standard deviation (Stdev)

	# of neurons	**Average CCI**	**Stdev CCI**	**Average CK**	**Stdev CK**
Tubificidae	2	93.3	0.0	0.16	0.31
Tubificidae	5	93.3	0.0	0.27	0.32
Tubificidae	10	93.3	0.0	0.27	0.32
Tubificidae	15	94.2	1.7	0.32	0.37
Tubificidae	20	94.2	1.7	0.32	0.37
Tubificidae	25	93.3	0.0	0.16	0.31
Tubificidae	5 - 5	93.3	0.0	0.27	0.32
Tubificidae	5 - 10	92.5	1.7	0.00	0.00
Tubificidae	10 - 10	93.3	0.0	0.16	0.31
Tubificidae	10 - 20	92.5	1.7	0.25	0.28
Asellidae	2	92.5	1.7	0.85	0.03
Asellidae	5	93.3	2.7	0.87	0.06
Asellidae	10	94.2	1.7	0.88	0.03
Asellidae	15	93.3	2.7	0.87	0.06
Asellidae	20	91.7	4.3	0.83	0.09
Asellidae	25	92.5	3.2	0.85	0.06
Asellidae	5 - 5	92.5	1.7	0.85	0.03
Asellidae	5 - 10	92.5	3.2	0.85	0.06
Asellidae	10 - 10	94.2	1.7	0.88	0.03
Asellidae	10 - 20	90.8	3.2	0.82	0.06
Gammaridae	2	91.7	6.4	0.55	0.40
Gammaridae	5	92.5	4.2	0.62	0.23
Gammaridae	10	91.7	3.3	0.62	0.13
Gammaridae	15	92.5	4.2	0.65	0.17
Gammaridae	20	91.7	3.3	0.60	0.12
Gammaridae	25	91.7	3.3	0.62	0.13
Gammaridae	5 - 5	88.3	3.3	0.29	0.35
Gammaridae	5 - 10	92.5	3.2	0.65	0.16
Gammaridae	10 - 10	92.5	6.3	0.66	0.23
Gammaridae	10 - 20	90.8	4.2	0.54	0.21
Baetis	2	66.7	9.8	0.33	0.18
Baetis	5	70.0	11.2	0.39	0.21
Baetis	10	77.5	1.7	0.54	0.04
Baetis	15	74.2	5.7	0.47	0.12
Baetis	20	70.8	7.4	0.41	0.14
Baetis	25	70.8	7.9	0.41	0.15
Baetis	5 - 5	62.5	6.9	0.20	0.17
Baetis	5 - 10	70.8	12.0	0.42	0.22
Baetis	10 - 10	69.2	12.9	0.39	0.23
Baetis	10 - 20	69.2	8.8	0.39	0.16
Limnephilidae	2	80.0	2.7	0.17	0.22
Limnephilidae	5	84.2	4.2	0.13	0.26
Limnephilidae	10	86.7	5.4	0.49	0.20
Limnephilidae	15	84.2	3.2	0.28	0.32
Limnephilidae	20	84.2	5.7	0.35	0.30
Limnephilidae	25	81.7	7.9	0.26	0.22
Limnephilidae	5 - 5	82.5	1.7	0.00	0.00
Limnephilidae	5 - 10	80.8	5.0	0.01	0.02
Limnephilidae	10 - 10	81.7	3.3	0.06	0.11
Limnephilidae	10 - 20	77.5	4.2	0.14	0.09

6.3.2.3. Determination of the optimal model architecture based on annual testing

Several studies (e.g. Recknagel *et al.*, 1997; Maier *et al.*, 1998; Bowden *et al.*, 2002) have revealed that good predictions and generalization ability of ANN models are obtained if the available dataset is divided into a training and test set based on annual data, i.e. using data from one year for testing and the remaining data for training. Based on this, the ANN models for the entire Zwalm river basin were trained with measured input and output data from 3 years, while data of the remaining independent year was used for testing. Similarly, the ANN models for the 'short distance' monitoring network were trained with data from one year and tested with data of the other year.

In Table 6.6 and 6.7, the CCI and CK are presented respectively for the dataset of the entire Zwalm river basin (237 sites, 4 years) and the dataset of the 'short distance' monitoring network (120 sites, 2 years). For the first dataset, similar results based on the CCI were obtained as for the 4-fold cross-validation. In general however, the models were less stable derived from the standard deviation of the CCI. Based on the CK, slightly better predictions were achieved applying the 4-fold cross-validation while the standard deviation for both model types was rather similar. Comparing the predictive results of the 2-fold cross-validation including 10 neurons in the hidden layer with those of the annual testing based on the dataset of the 'short distance' monitoring network, CCI and CK were very similar, with one major exception for Tubificidae, where no reliable model could be found based on annual testing (CK = 0.00) while the CK for the 2-fold cross-validation was 0.28. However, the latter model was very unstable based on the standard deviation (0.39). The major part of the remaining models were more stable when validated on the basis of the 2-fold cross-validation.

Table 6.6. Determination of the optimal model architecture based on annual testing for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae based on the dataset of the Zwalm river basin (237 sites, 4 years). Model evaluation is based on the percentage of Correctly Classified Instances (CCI) and Cohen's kappa (CK), model stability on the standard deviation (Stdev)

	# of neurons	Average CCI	Stdev CCI	Average CK	Stdev CK
Tubificidae	2	89.4	3.9	0.00	0.00
Tubificidae	5	89.4	3.9	0.00	0.00
Tubificidae	10	89.4	3.9	0.00	0.00
Tubificidae	15	89.4	3.9	0.04	0.09
Tubificidae	20	89.4	3.9	0.06	0.12
Tubificidae	25	89.8	3.8	0.06	0.11
Tubificidae	5 - 5	88.2	5.0	0.12	0.19
Tubificidae	5 - 10	89.4	3.9	0.04	0.09
Tubificidae	10 - 10	88.2	2.6	0.14	0.17
Tubificidae	10 - 20	89.4	3.9	0.00	0.00
Asellidae	2	77.6	3.7	0.56	0.07
Asellidae	5	75.9	3.7	0.52	0.08
Asellidae	10	79.7	5.3	0.59	0.10
Asellidae	15	78.5	4.7	0.57	0.10
Asellidae	20	80.2	3.9	0.61	0.08
Asellidae	25	78.9	4.1	0.58	0.08
Asellidae	5 - 5	78.5	2.0	0.57	0.04
Asellidae	5 - 10	78.5	3.7	0.57	0.08
Asellidae	10 - 10	70.0	13.3	0.40	0.29
Asellidae	10 - 20	80.2	6.8	0.61	0.13
Gammaridae	2	76,8	2.6	0.39	0.05
Gammaridae	5	76.4	6.3	0.34	0.14
Gammaridae	10	77.6	6.1	0.37	0.14
Gammaridae	15	76.3	7.3	0.30	0.15
Gammaridae	20	78.4	7.8	0.34	0.23
Gammaridae	25	78.0	9.2	0.38	0.19
Gammaridae	5 - 5	79.3	6.1	0.39	0.09
Gammaridae	5 - 10	77.2	5.7	0.36	0.14
Gammaridae	10 - 10	75.9	2.1	0.29	0.20
Gammaridae	10 - 20	76.8	8.4	0.32	0.25
Baetis	2	75.1	16.4	0.10	0.19
Baetis	5	65.0	5.1	0.06	0.02
Baetis	10	70.1	5.8	0.03	0.07
Baetis	15	75.1	16.4	0.09	0.10
Baetis	20	75.9	15.2	0.05	0.07
Baetis	25	74.6	12.7	0.02	0.04
Baetis	5 - 5	73.8	16.9	0.00	0.01
Baetis	5 - 10	75.5	15.4	0.00	0.00
Baetis	10 - 10	73.8	17.6	0.05	0.06
Baetis	10 - 20	76.3	14.9	0.04	0.05
Limnephilidae	2	87.7	2.7	0.00	0.00
Limnephilidae	5	88.6	1.6	0.12	0.23
Limnephilidae	10	87.8	3.1	0.20	0.17
Limnephilidae	15	88.6	2.6	0.20	0.25
Limnephilidae	20	87.8	2.5	0.19	0.25
Limnephilidae	25	88.2	3.5	0.29	0.25
Limnephilidae	5 - 5	87.7	2.7	0.00	0.00
Limnephilidae	5 - 10	87.7	2.7	0.00	0.00
Limnephilidae	10 - 10	87.7	2.7	0.00	0.00
Limnephilidae	10 - 20	87.7	2.7	0.00	0.00

Table 6.7. Determination of the optimal model architecture based on annual testing for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae based on the dataset of the ‘short distance’ monitoring network (120 sites, 2 years). Model evaluation is based on the percentage of Correctly Classified Instances (CCI) and Cohen’s kappa (CK), model stability on the standard deviation (Stdev)

	# of neurons	**Average CCI**	**Stdev CCI**	**Average CK**	**Stdev CK**
Tubificidae	2	90.8	3.5	0.15	0.21
Tubificidae	5	92.5	1.2	0.00	0.00
Tubificidae	10	94.2	1.2	0.28	0.39
Tubificidae	15	92.5	1.2	0.00	0.00
Tubificidae	20	92.5	1.2	0.00	0.00
Tubificidae	25	95.0	0.0	0.47	0.12
Tubificidae	5 - 5	92.5	1.2	0.00	0.00
Tubificidae	5 - 10	86.7	7.1	0.16	0.13
Tubificidae	10 - 10	90.8	3.5	0.15	0.21
Tubificidae	10 - 20	92.5	1.2	0.00	0.00
Asellidae	2	90.0	0.0	0.80	0.00
Asellidae	5	91.7	0.0	0.83	0.00
Asellidae	10	90.8	1.2	0.82	0.03
Asellidae	15	89.2	1.2	0.78	0.03
Asellidae	20	90.8	1.2	0.82	0.02
Asellidae	25	89.2	1.2	0.78	0.03
Asellidae	5 - 5	90.0	0.0	0.80	0.00
Asellidae	5 - 10	90.8	1.2	0.82	0.03
Asellidae	10 - 10	90.0	0.0	0.80	0.00
Asellidae	10 - 20	92.5	1.2	0.85	0.02
Gammaridae	2	89.2	1.2	0.32	0.45
Gammaridae	5	93.3	2.4	0.68	0.03
Gammaridae	10	91.7	2.4	0.62	0.02
Gammaridae	15	91.7	0.0	0.64	0.03
Gammaridae	20	92.5	3.5	0.66	0.12
Gammaridae	25	92.5	1.2	0.64	0.10
Gammaridae	5 - 5	90.0	0.0	0.43	0.29
Gammaridae	5 - 10	85.0	4.7	0.44	0.04
Gammaridae	10 - 10	85.0	4.7	0.46	0.19
Gammaridae	10 - 20	88.3	0.0	0.28	0.40
Baetis	2	66.7	2.4	0.36	0.00
Baetis	5	65.0	4.7	0.31	0.11
Baetis	10	68.3	0.0	0.37	0.01
Baetis	15	66.7	4.7	0.37	0.04
Baetis	20	59.2	3.5	0.19	0.08
Baetis	25	57.5	5.9	0.18	0.16
Baetis	5 - 5	62.5	5.9	0.26	0.13
Baetis	5 - 10	62.5	3.5	0.26	0.09
Baetis	10 - 10	57.5	5.9	0.19	0.17
Baetis	10 - 20	73.3	2.4	0.47	0.05
Limnephilidae	2	82.5	5.9	0.00	0.00
Limnephilidae	5	82.5	5.9	0.00	0.00
Limnephilidae	10	80.8	3.5	0.04	0.06
Limnephilidae	15	82.5	5.9	0.00	0.00
Limnephilidae	20	80.0	2.4	0.03	0.04
Limnephilidae	25	80.8	3.5	0.10	0.14
Limnephilidae	5 - 5	82.5	5.9	0.00	0.00
Limnephilidae	5 - 10	73.3	9.4	0.17	0.07
Limnephilidae	10 - 10	79.2	1.2	0.17	0.24
Limnephilidae	10 - 20	79.2	1.2	0.14	0.20

6.4. Discussion

Nowadays, models are an important tool in modern water management. However, when models are used in a more flexible way, the risk of injudicious use may increase as a result of errors in the programme, incomplete manuals or mistakes made by the model designer himself. Careless use of input data, insufficient calibration and validation, departing from wrong model assumptions, … are mistakes which can lead to unreliable models (STOWA/RIZA, 1999). Therefore, enough time has to be spend on model design, before models can be used for practical applications in order to avoid these pitfalls.

To compare different model set-ups, models have to be evaluated based on reliable performance measures. Based on international literature however, a serious gap concerning model evaluation has been observed. The performance measures used to evaluate ANN models was mainly based on the overall prediction success (= percentage of Correctly Classified Instances, CCI). Only a few researchers used alternative evaluation methods (Fielding and Bell, 1997; Manel *et al.*, 2001; Anderson *et al.*, 2003). Based on 87 international ecological publications evaluating the performance of presence/absence models on birds, mammals, macroinvertebrates, fish, … during the period 1989-1999, Manel *et al.* (2001) indicated that in 52 % of the cases no evaluation method was used, in 43 % model evaluation was merely based on prediction success, while only in 3 % of the publications Cohen's kappa was used. Based on the review on Artificial Neural Networks predicting macroinvertebrates in rivers and lakes (Chapter 2), only five presence/absence models used the CCI as performance measure whereas three models applied a combination of CCI and CK.

From an ecological point of view, the CCI is the most logical performance measure because of its easy interpretation towards water managers and policy makers. This evaluation method indicates the number of sites in the test set that is predicted correctly. In this way, it is a simple measure for the accuracy of the prediction. There is clear evidence though, that the CCI is influenced by the frequency of occurrence of the organism being modelled (e.g. Fielding and Bell, 1997; Manel *et al.*, 2001; Dedecker *et al.*, 2004; D'heygere *et al.*, 2005b). The problem with rare taxa is that there is little information to allow the neural network model to learn when these taxa are present. In this way the models tend to 'learn' that very rare taxa are always absent. The same difficulty occurs with very common taxa. Here the models 'learn' that very common taxa are always present. This means that a good CCI can be achieved without using any information from the environmental variables. Among the different measures, which are based on the confusion matrix, Fielding and Bell (1997) and Manel *et al.* (2001) recommended the CK (Cohen, 1960) as a reliable performance measure, since the effect of prevalence on the CK appeared to be negligible. It is a simply derived statistic that measures the proportion of all possible cases of presence or absence that are predicted correctly by a model after accounting for chance predictions

This bottleneck in presence/absence predictions is illustrated in the present study predicting the habitat suitability of Tubificidae and Limnephilidae. These taxa were respectively found in 89.5 and 92.5 % and 12.2 and 17.5 %, based on respectively the dataset of the entire Zwalm river basin and the 'short distance' monitoring network. For all models tested in this chapter (determination of the optimal training and test size (optimal number of folds) and determination of the optimal model architecture based on regular folds and annual testing), the predictive performance was very good based on the CCI. Derived from the CK however, poor to moderate models were generated. For both taxa, high CCI percentages can be associated for a major part to their relatively low and high prevalence, and the related ease to make good qualifications, even without the extraction of information from the environmental variables presented as inputs to the networks.

In this way, predictions can be misleading if only the overall prediction success is used as performance measure. Therefore, a combination of both the CCI (easy to interpret from an ecological point of view) and the CK (from an mathematical point of view) can be advised to evaluate model performance.

To evaluate the developed models, part of the dataset has to be used for model testing which has not been used to train the model. In this way, the training dataset is reduced which can influence the generalization capacity of the model (Haykin, 1999). Conversely, a smaller test dataset reduces the accuracy of the model performance. In this context, a consideration has to be made between generalization capacity on the one hand and model accuracy on the other hand. Witten and Frank (2000) describe 10-fold cross-validation as standard method. This means that 10 % of the data is used for testing and 90 % for training. This is repeated ten times. In this way, each part of the dataset is used once for testing. Seen the relatively small datasets, and the related unreliably small test sets it involves, this method seems less suitable.

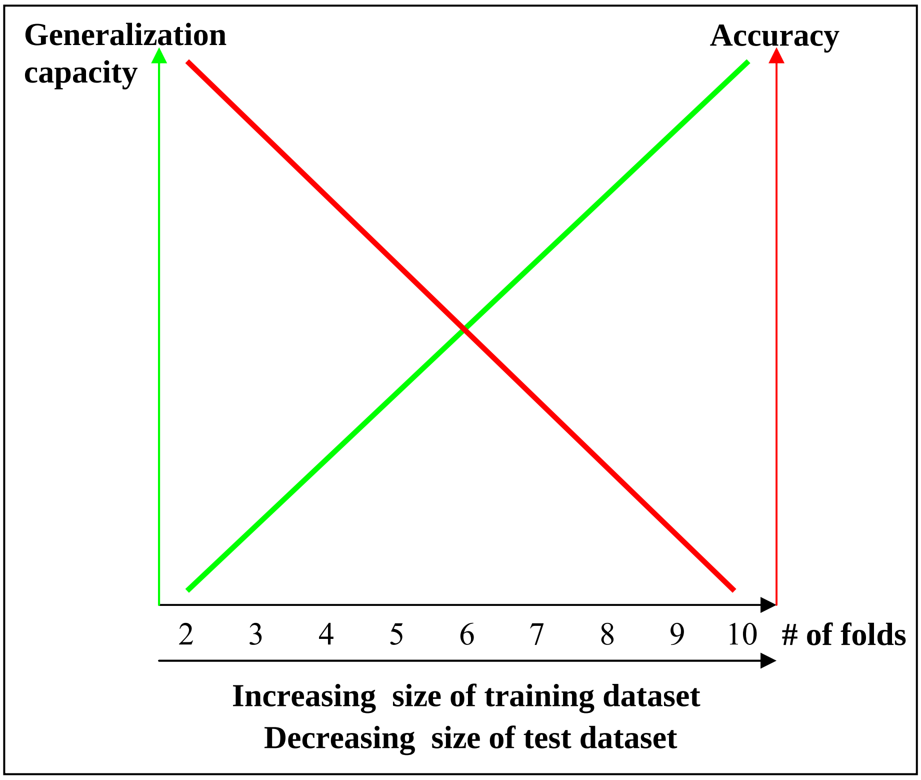


Fig. 6.3. Impact of the size of the training and test dataset (= number of folds) on the generalization capacity and accuracy of the model predictions.

In addition, if for example a very common taxon has to be modelled and if at random a fold is split off containing very few present instances, this has important consequences on model training and testing. To avoid this problem in the present thesis, the relative prevalence of the taxon was similar in all training and test sets. Taking these thoughts into consideration, the optimal size of training and test set (= the optimal number of folds) was searched for. Therefore, nine cross-validation methods (2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation) were tested. Based on the average and standard deviation of the CCI and CK, the best possible number of folds was four for all five taxa for the entire Zwalm river basin as well as for the ‘short distance’ monitoring network. To this end, one may conclude that 4-fold cross-validation is a good consideration between model generalization capacity and accuracy. Therefore, model validation during further research was based on 4-fold cross-validation. In this manner, also the work load (data preparation, model training and testing) could be limited in comparison with 10-fold cross-validation for instance.

Traditionally, optimal network architectures (or geometries) have been determined by trial and error (Maier and Dandy, 2000). However, a number of systematic approaches for determining optimal network geometry have been proposed, including pruning and constructive algorithms. The basic thought of pruning algorithms is to start with a network that is large enough to capture the desired input-output relationship and to subsequently remove or disable unnecessary weights and/or neurons. A review of pruning algorithms is given by Reed (1993). Constructive algorithms approach the problem of optimizing the number of hidden layer neurons from the opposite direction to pruning algorithms. The smallest possible network is used at the start. Hidden layer neurons and connections are then added one at a time in an attempt to improve model performance. A review of constructive algorithms is given by Kwok and Yeung (1997a). However, several disadvantages of these approaches are mentioned in literature (Maier and Dandy, 2000). For example, the networks generally have to be trained several times, i.e. each time a hidden neuron is added or deleted (Kwok and Yeung, 1997b). It has also been suggested that the pruning and constructive algorithms are susceptible to becoming trapped in structural local optima (Angeline *et al.*, 1994). Algorithms based on evolutionary programming and genetic algorithms have been proposed to overcome these problems and have been used successfully to determine optimal network architecture (e.g. Fang and Xi, 1997; Kim and Han, 2000; Zhao *et al.*, 2000; Wicker *et al.*, 2002). Evolutionary approaches are significantly different from the previous techniques described. They produce more robust solutions because they use a population of networks in the search process. A complete review of the use of evolutionary algorithms in neural networks is given by Yao (1993). However, this is beyond the scope of this thesis and is no further discussed. In this way, it was decided to use trial and error to optimize the neural network architecture for the five selected macroinvertebrates in both databases. In total, ten network architectures were tested to obtain a reliable estimation of the best architecture: six three-layered and four four-layered networks with respectively [2], [5], [10], [15], [20], [25] and [5 5], [5 10], [10 10], [10 20] neurons in the hidden layer(s).

Network architecture is generally known to be highly problem dependent (Maier and Dandy, 2000). However, comparing the different model architectures for each organism, predictive performances were very similar. Also the guidelines, mentioned in Chapter 2, do not ensure optimal network geometry. In addition, there is quite a high variability in the number of neurons suggested by the various rules, making them less reliable. In this way, no straightforward conclusions towards the five taxa could be drawn according to the optimal number of hidden layers and neurons. On the other hand, comparing the optimal number of hidden neurons for both datasets, a small difference could be detected. More neurons (roughly twice as much for all taxa) were needed based on the 'short distance' monitoring network. The reason could be that the number of training samples was respectively fixed at 178 and 90, for the datasets of the entire Zwalm river basin and the 'short distance' monitoring network, using 4-fold cross-validation. The importance of striking a balance between having sufficient free parameters (weights) to enable representation of the function to be approximated and having too many free parameters, which can result in overfitting, is well known and has been discussed widely in the literature (e.g. Maren *et al.*, 1990; Rojas, 1996). Since overfitting was always avoided using the principle of 'early stopping' and the number of training samples was twice as less when using the dataset of the 'short distance' monitoring network, more free parameters were possibly needed to learn the relations between input and output adequately.

Although less data were available, better prediction results were generally obtained for the dataset of the 'short distance' monitoring network. An important reason can be that five weirs (ten sampling sites a year) were not included in the latter dataset. Upstream and downstream of the weirs, the Zwalm river is characterized by a modification of the flow channel. Upstream the weir, the river is drastically deepened, creating exceptional conditions, unnatural for this type of water bodies. Just in front, the depth can be nearly around two meters depending on the control level of the weir and the amount of sediments accumulated at the site. Also the flow velocity is reduced drastically, creating an almost stagnant water body immediately upstream of the weir (Belconsulting, 2003). On the other hand, the depth is very low and the flow velocity is very high downstream of the weir. This situation is also unusual for these wider rivers in the Zwalm river basin. In this way, the situation upstream and downstream of these weirs is very difficult to predict.

Several studies (e.g. Recknagel *et al.*, 1997; Maier *et al.*, 1998; Bowden *et al.*, 2002) have revealed that good predictions and generalization ability of ANN models are obtained if the available dataset is divided into a training and test set based on annual data, i.e. using data from one year for testing and the remaining data for training. Based on this, the ANN models for the entire Zwalm river basin were trained with measured input and output data from 3 years, while the data of the remaining independent year were used for testing. Similarly, the ANN models for the 'short distance' monitoring network were trained with data from one year and tested with data of the other year. When annual testing was compared with the usual cross-validation, the average CCIs and CKs were in general very similar for all taxa. In contrast, the standard deviation was mostly higher when annual testing was applied. The higher instability can be explained by the unequal prevalence of the taxa each year. Since this is one of the major parameters affecting the model's performance, the model stability can be highly influenced.

6.5. 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 the habitat suitability of a species based on the 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 to managers and engineers dealing with rivers and channels (Lek *et al.*, 1996a; Mastrorillo *et al.*, 1997a, b; Guégan *et al.*, 1998).

ANN models can in this context play an interesting role to find general trends on habitat suitability of macroinvertebrate taxa. This study aimed at developing, analyzing and optimizing the ANN models for the prediction of the habitat suitability (presence/absence) of five macro-invertebrate taxa Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae in the Zwalm river basin.

The results illustrated the convenience of using two performance measures. The overall prediction success (= percentage of Correctly Classified Instances, CCI) is from an ecological point of view the most logical performance measure because of its easy interpretation towards water managers and policy makers. There is clear evidence however, that the CCI is influenced by the frequency of occurrence of the organism being modelled. Therefore, a second performance measure (Cohen's kappa, CK) was used, since the effect of prevalence on the CK appeared to be negligible.

Because the size of the training and test set influences respectively the generalization capability of the model and accuracy of the model performance, the optimal size of training and test set was searched for. Based on the average and standard deviation of the CCI and CK, the best possible number of folds was four (= 4-fold cross-validation) for both datasets used.

Network architecture is generally known to be highly problem dependent. However, comparing the different model architectures for each taxon, predictive performances were very similar. In this way, no straightforward conclusions towards the five taxa could be drawn according to the optimal number of hidden layers and neurons.

Chapter 7
Application of input variable contribution methods to Artificial Neural Network models for selection of habitat suitability variables

---

<u>Parts of this chapter are based on</u>:

Dedecker, A.P., Goethals, P.L.M., D'heygere, T., Gevrey, M., Lek, S. and 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*.

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

7.1. Introduction

It has been widely demonstrated that environmental conditions can strongly influence the occurrence of macroinvertebrates (e.g. Marshall et al., 2002). Habitat selection of river invertebrates depends upon the interaction of numerous physical, chemical, structural as well as biological factors. Abiotic factors, in particular those related to disturbance and pollution, clearly determine the composition of invertebrate communities (De Pauw and Hawkes, 1993; D'heygere et al., 2003; Reice et al., 1990; Resh et al., 1988; Townsend et al., 1997). Biota are directly and indirectly affected by these abiotic factors and in most cases optimum conditions are different for each species. 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 Cogernino, 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). As a result, the actual knowledge is still insufficient to completely understand the habitat preferences of the river macroinvertebrates. This leads to problems when using macroinvertebrate communities for surveillance purposes (Fontoura and De Pauw, 1994) and river management (Goethals and De Pauw, 2001).

To gain more insight into the habitat preferences of the different taxa and the relationships between these local environmental conditions and the occurrence of macroinvertebrates (presence/absence or abundance), empirical models could be developed. These models would allow for extracting the major relations between the environment and these aquatic invertebrates and could be used to predict their occurrence under altered conditions (Reynoldson et al., 1997; Wright, 1995). Artificial Neural Network (ANN) models have recently been applied in the context of this problem. These ANN models however do not allow for gaining direct insight into the habitat preferences of the species or taxa. For this reason, ANN models have been labelled as being a 'black box' (Lek and Guégan, 1999; Olden and Jackson, 2002). Nonetheless, several authors have explored and proposed different techniques to illustrate the role of the environmental variables in ANN models as discussed in Chapter 2 (e.g. Beauchard et al., 2003; Dedecker et al., 2005b, c; Gevrey et al., 2003; Marshall et al., 2002; Olden and Jackson, 2002). In most papers, these techniques have been used to select the relevant input variables to predict the aquatic communities. Traditionally, optimal combinations between environmental input and output variables have been determined on the basis of expert knowledge (Dedecker et al., 2004), 'trial and error' (Gabriels et al., 2002), sensitivity analyses (Lek et al., 1996b) or optimization algorithms, such as genetic algorithms (D'heygere et al., 2005a). The purpose of these techniques is not only to improve the network performance by selecting the relevant input variables, but also to know the contribution of each input variable to the output. These techniques allow for specifying the major river characteristics that describe the preferred habitat of a particular taxon. This can deliver important additional information to assess and interpret ecological phenomena and to enhance indirectly the understanding of relationships between human impacts and river fauna. The ANN models can also indicate which variables are of major importance to monitor and assess. The quality and appropriate choice of monitoring variables is indeed one of the most crucial factors in the assessment process of river systems (Goethals and De Pauw, 2001).

In previous studies (Dedecker, 2005b, c; Goethals, 2005), the following six contribution methods as described by Gevrey *et al.* (2003) have been applied to determine the influence of each input variable and its contribution to the occurrence of Asellidae and Gammaridae:

- the 'Partial Derivatives' ('PaD') method consists of a calculation of the partial derivatives of the output in relation to the input variables (Dimopoulos *et al.*, 1995, 1999);
- the 'Weights' method is a computation using the connection weights of the backpropagation Artificial Neural Networks (Garson, 1991; Goh, 1995; Olden and Jackson, 2002);
- the Perturb method analyses the effect of a perturbation of the input variables on the output variable (Yao *et al.*, 1998; Scardi and Harding, 1999);
- the 'Profile' method is a successive variation of one input variable while the others are kept constant at a fixed set of values (Lek *et al.*, 1995, 1996a, b; Marshall *et al.*, 2002);
- the Classical Stepwise method is an observation of the change in the error value when an elimination (backward) step of the input variables is applied;
- the Improved Stepwise method involves the network being trained and fixed step by step, with one input variable on its mean value to note the consequences on the error.

These six contribution methods, which were applied on a database of the Zwalm river basin containing 179 sampling sites taken between 2000 and 2002, seemed to give similar results.

Nevertheless, their diverse computation led to differences in sensitivity and stability of the methods. The 'PaD', 'Weights' and 'Profile' method were the most sensitive techniques. Especially the 'PaD' and the 'Profile' method were able to distinguish minor and major contributing environmental variables for both species, this in contrast with the Perturb, Stepwise and the Improved Stepwise method. Similar results were found by (Gevrey *et al.*, 2003), who tried to predict the relation between ten environmental variables and trout density.

Based on these preliminary results, it was decided to apply in the present study only the 'PaD', 'Weights' and 'Profile' method to support the selection of the major river characteristics in order to describe the preferred habitats of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae and explain their distribution in both datasets (Zwalm river basin and the 'short distance' monitoring network). The three input variable contribution methods are described in more detail in Section 7.2.

7.2. Material and methods

Although many methods (and terms) exist for variable selection and sensitivity analysis (e.g. Witten and Frank, 2000), only a limited set, consisting of three methods that had already proven to be convenient in ecological modelling studies, was applied on the dataset of the Zwalm river basin and the 'short distance' monitoring network. The three 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 Characteristics on the Structure of Aquatic Communities (Algae, Benthic and Fish Fauna)'.

A detailed overview of the specific ANN model settings used to obtain the contribution curves is given in Section 6.2. The variable contribution methods applied in this study are described below. These methods were performed in addition to each constructed ANN model in Chapter 6 in order to examine the effects of the size of training and test set, model architecture, … on the relative importance of the environmental input variables on the habitat suitability of the macroinvertebrates. The contribution methods discussed here are those applied in cases of networks with one hidden layer.

7.2.1. The 'PaD' method

This method makes a classification of the relative contributions of each variable to the network output, as a result of small changes of these input variables. For this classification, the partial derivatives of the ANN output with respect to the input are calculated (Dimopoulos *et al.*, 1995, 1999). For a network with n_e inputs, one hidden layer with n_i neurons and one output (i.e. $n_s = 1$), the partial derivatives of the output function $\hat{x}_{sj}$ with respect to input x_{ej}, (with j = 1, …, N and N the total number of observations) are (parameter terms are based on Fig. 6.1):

$$d_{je} = S_j \sum_{i=1}^{n_i} w_{is} I_{ij} (1 - I_{ij}) w_{ei} \tag{8.1}$$

This is valid on the assumption that a logarithmic sigmoid function is used for the activation. When S_j is the derivative of the output neuron with respect to its input, I_{ij} is the response of the *i*th hidden neuron, w_{is} and w_{ei} are the weights between the output neuron and *i*th hidden neuron, and between the *e*th input neuron and the *i*th hidden neuron.

As such, the relative contribution of the ANN output to the data set with respect to an input is obtained. It is calculated by a sum of the square derivatives for each input variable:

$$SSD_e = \sum_{j=1}^{N} (d_{je})^2 \tag{8.2}$$

One SSD (Sum of Square Derivatives) value is obtained per input variable, what allows the 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.

7.2.2. The '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 *i*, the absolute values of the connection weights between the hidden and the output layer (w_{is}) are multiplied by the absolute values of the connection weights between the hidden and the input layer (w_{ei}). This is repeated for each input neuron. In this way, the product P_{ei} was obtained:

$$P_{ei} = |w_{is}| \times |w_{ei}| \tag{8.3}$$

2. P_{ei} is divided by the sum of all the input variables, i.e.

For i = 1 to n_i,

For e = 1 to n_e,

$$Q_{ei} = \frac{|P_{ei}|}{\sum_{e=1}^{n_e} |P_{ei}|} \tag{8.4}$$

end,

end.

3. For each input neuron, divide the sum of the Q_{ei} for each hidden neuron by the sum for each hidden neuron of the sum for each input neuron of Q_{ih}, multiply by 100. The relative importance of all output weights attributable to the given input variable is then obtained.

For e = 1 to n_e

$$RI(\%)_e = \frac{\sum_{i=1}^{n_i} Q_{ei}}{\sum_{i=1}^{n_i}\sum_{e=1}^{n_e} Q_{ei}} \times 100 \tag{8.5}$$

end.

7.2.3. The 'Profile' method

The general idea of this method is to study each input variable successively when the others are blocked at fixed values (Lek *et al.*, 1995, 1996a, b). The principle of this algorithm is to construct a fictitious matrix pertaining to the range of all input variables. For this, 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. In this work, a scale of 12 was used between the minimum and maximum of the input variables.

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 way, for each input variable it can be derived if an increase in its value has a positive (curve with an increasing slope), a negative (curve with a decreasing slope) or not any (curve with a horizontal slope) effect on the probability of presence of the macroinvertebrate taxon.

7.2.4. Evaluation of the contribution method stability

In order to check the stability of each model, the training of the network was repeated *k* times, according to the *k*-fold cross-validation, and the relative contributions of the input variables on the output obtained for each method and each trained network were noted. Based on this repetition, the average contribution of every variable for each method was calculated. The *k*-fold cross-validation allowed also for calculating the standard deviation. This standard deviation gave an indication of the stability of each contribution method.

7.3. Results

7.3.1. Overall ranking of the environmental variables

In Table 7.1 to 7.5, the overall ranking of the environmental input variables for the Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae presence/absence ANN models (24-10-1) based on the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation methods is presented for both the Zwalm river basin and the 'short distance' monitoring network. To make an overall ranking of the variables, the sum was made of the rankings per contribution method ('PaD', Weighs and 'Profile'), and based on this sum, the overall ranking was determined. Based on Table 7.1 to 7.5, one can generally deduce that it is difficult to find major trends over the five taxa, both datasets and the nine cross-validation methods. The first can be explained by different ecological preferences of the taxa (cf. ecological expert knowledge in Chapter 4 and the data analysis in Chapter 5). The dissimilarity between both databases can be related to the different scales of measurements. In the Zwalm river basin, samples were spread over the entire river basin (large scale), while in the 'short distance' monitoring network samples were taken closer together (small scale) and the gradient from pure headwaters to more polluted main streams was much more obvious. Differences between the nine cross-validation methods can be associated with the different predictive performances obtained as discussed in Chapter 6, because different methods can probably use slightly other variables to predict the output. However, for most of the taxa the key environmental variables for the presence/absence could be detected.

In the Section 7.3.2, the effect of the three contribution methods is examined in more detail. In addition, key variables for the presence/absence of the five taxa are analyzed and discussed from an ecological point of view. The graphs obtained by the best performing ANN models, as derived from Chapter 6 (4-fold cross-validation for all taxa combined with the optimal network architecture), are presented.

Table 7.1. Comparison of the overall rankings of the input variables derived from the three contribution methods ‘Weights’, ‘PaD’ and ‘Profile’ for the Tubificidae presence/absence ANN models (24-10-1) on the basis of the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation methods

	Zwalm river basin									**‘Short distance’ monitoring network**								
k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Width	17	15	8	8	5	6	14	2	4	2	18	22	15	8	13	12	11	13
Embankment	20	13	20	19	23	18	22	22	24	14	15	17	14	24	18	16	22	24
Meandering	15	7	9	9	8	11	9	7	9	6	6	13	16	23	17	11	20	20
Pools-riffles	19	9	19	18	24	23	24	18	23	13	12	6	8	5	7	9	9	8
Hollow banks	16	11	4	24	19	15	18	15	14	7	4	8	9	3	5	8	2	7
Depth	3	23	7	15	22	2	13	10	18	19	8	12	4	9	6	5	7	11
Flow velocity	2	5	2	3	11	3	3	3	1	24	14	7	17	12	15	13	12	10
pH	22	20	6	12	16	4	16	16	12	8	17	24	21	18	21	21	14	21
Temperature	8	16	18	10	13	19	20	20	16	11	2	14	12	10	19	18	17	22
Dissolved oxygen	14	22	17	20	12	22	23	19	10	5	9	16	19	20	22	23	15	16
Conductivity	5	4	14	14	15	12	17	17	21	3	13	10	5	13	8	7	8	5
Suspended solids	7	19	22	11	3	10	12	8	15	12	19	15	13	22	16	14	23	14
Ammonium	9	14	3	1	14	13	1	12	7	16	10	9	10	7	9	17	6	9
Nitrate	1	18	16	2	1	1	5	1	3	10	11	5	2	2	3	3	3	1
Total nitrogen	4	8	13	16	9	7	10	14	11	21	16	18	18	14	14	10	16	18
Phosphate	12	24	24	22	17	21	6	23	19	18	7	4	6	15	10	6	13	12
Total phosphorus	10	2	10	5	7	5	11	4	6	22	24	11	11	11	11	15	10	6
COD	11	12	21	17	18	20	15	13	17	17	20	23	24	21	20	19	24	17
Pebbles	21	21	23	23	20	24	19	21	22	23	21	19	23	17	24	24	18	19
Gravel	6	10	5	6	6	14	4	6	2	20	23	21	22	16	12	22	21	15
Sand	18	6	12	13	10	16	7	11	13	9	3	3	3	4	1	2	4	4
Loam/clay	13	17	15	21	21	17	21	24	20	15	22	20	20	19	23	20	19	23
Distance to mouth	24	3	1	7	2	8	2	9	8	1	5	2	7	6	4	4	1	3
Stream order	23	1	11	4	4	9	8	5	5	4	1	1	1	1	2	1	5	2

Table 7.2. Comparison of the overall rankings of the input variables derived from the three contribution methods ‘Weights’, ‘PaD’ and ‘Profile’ for the Asellidae presence/absence ANN models (24-10-1) on the basis of the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation methods

	Zwalm river basin									**‘Short distance’ monitoring network**								
k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Width	1	2	1	1	1	1	1	1	1	2	8	8	6	2	2	11	3	4
Embankment	9	14	18	17	17	22	22	18	18	21	14	23	17	23	21	14	20	15
Meandering	14	8	8	9	9	9	8	10	13	15	19	14	12	17	18	18	15	16
Pools-riffles	12	12	9	15	16	20	24	14	23	16	11	12	7	10	12	7	12	10
Hollow banks	3	4	3	3	3	3	5	3	3	11	5	17	18	14	16	16	14	19
Depth	7	5	11	4	5	4	3	4	5	7	20	7	9	11	6	5	4	7
Flow velocity	17	16	22	19	22	23	18	21	20	17	22	22	22	21	24	24	21	22
pH	13	21	16	18	24	17	20	16	11	22	15	24	20	24	23	19	23	21
Temperature	24	6	6	6	13	7	11	8	8	23	21	19	21	22	20	20	24	20
Dissolved oxygen	8	13	15	11	11	11	14	13	12	20	16	6	13	15	19	17	16	17
Conductivity	23	20	20	24	21	12	17	22	19	6	3	11	14	4	10	12	8	14
Suspended solids	15	11	21	23	18	14	23	20	24	5	6	9	3	8	8	6	7	5
Ammonium	10	15	12	10	8	10	10	12	9	19	12	4	10	9	9	8	10	8
Nitrate	5	7	7	8	6	6	6	7	7	10	17	15	15	12	4	13	9	11
Total nitrogen	6	9	4	7	7	8	7	6	6	9	9	10	8	6	11	15	11	12
Phosphate	19	18	19	20	20	24	19	23	22	12	24	21	23	20	17	21	19	23
Total phosphorus	16	19	13	21	12	21	15	17	15	18	10	13	11	13	13	9	13	13
COD	20	10	10	16	15	18	13	15	16	13	18	16	19	18	14	23	17	18
Pebbles	18	22	23	14	19	15	12	11	14	4	4	5	4	7	5	3	5	6
Gravel	21	24	14	12	14	13	9	9	10	24	23	18	24	19	22	22	22	24
Sand	11	17	17	13	10	19	21	19	17	8	7	3	5	5	7	4	6	3
Loam/clay	22	23	24	22	23	16	16	24	21	14	13	20	16	16	15	10	18	9
Distance to mouth	4	3	5	5	4	5	2	5	4	3	2	1	2	3	3	2	2	2
Stream order	2	1	2	2	2	2	4	2	2	1	1	2	1	1	1	1	1	1

Table 7.3. Comparison of the overall rankings of the input variables derived from the three contribution methods ‘Weights’, ‘PaD’ and ‘Profile’ for the Gammaridae presence/absence ANN models (24-10-1) on the basis of the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation methods

	Zwalm river basin									**‘Short distance’ monitoring network**								
k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Width	1	8	9	7	7	16	11	9	12	7	8	10	8	15	10	10	15	12
Embankment	24	23	23	22	24	24	24	21	21	19	15	15	18	20	18	15	10	14
Meandering	22	20	18	17	20	22	19	18	18	9	7	7	7	7	8	6	6	4
Pools-riffles	21	16	19	19	16	17	18	15	20	15	14	11	9	12	6	7	7	9
Hollow banks	13	15	17	18	12	14	16	11	19	21	20	24	24	24	21	24	23	23
Depth	8	4	7	5	3	3	3	2	6	1	1	1	1	1	1	1	1	1
Flow velocity	12	12	13	12	14	7	13	10	10	3	4	3	2	3	5	5	9	5
pH	17	13	12	15	18	18	15	13	14	18	18	23	15	18	19	20	20	20
Temperature	5	5	6	6	4	6	4	4	5	16	16	14	16	16	14	17	13	13
Dissolved oxygen	14	22	14	20	15	11	17	14	15	11	5	8	12	9	9	12	4	8
Conductivity	2	1	4	1	1	4	1	7	4	4	3	4	5	8	7	8	11	6
Suspended solids	6	9	8	13	9	13	8	23	11	23	22	13	17	17	17	14	16	17
Ammonium	4	2	2	3	6	5	6	5	3	2	2	5	3	2	2	2	3	2
Nitrate	9	11	15	10	11	8	12	12	8	10	12	9	11	13	11	13	12	10
Total nitrogen	20	19	22	24	23	23	23	24	22	13	21	16	23	14	15	18	18	15
Phosphate	23	17	24	16	17	19	20	16	17	8	9	12	13	6	12	9	8	11
Total phosphorus	11	6	11	11	19	15	10	17	13	5	10	6	6	5	3	4	2	3
COD	15	10	1	2	5	2	5	3	2	6	6	2	4	4	4	3	5	7
Pebbles	19	24	20	23	22	20	21	22	23	17	19	19	19	21	24	21	22	24
Gravel	16	21	16	14	13	10	14	19	16	24	24	20	21	23	23	23	24	22
Sand	18	18	21	21	21	21	22	20	24	14	13	21	20	19	20	22	19	19
Loam/clay	7	7	5	9	8	9	7	6	7	12	11	22	10	10	16	16	14	16
Distance to mouth	3	3	3	4	2	1	2	1	1	22	23	18	14	11	13	11	17	21
Stream order	10	14	10	8	10	12	9	8	9	20	17	17	22	22	22	19	21	18

Table 7.4. Comparison of the overall rankings of the input variables derived from the three contribution methods 'Weights', 'PaD' and 'Profile' for the *Baetis* presence/absence ANN models (24-10-1) on the basis of the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation methods

	Zwalm river basin									**'Short distance' monitoring network**								
k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Width	11	6	17	8	21	19	20	12	19	15	23	24	9	21	16	18	16	18
Embankment	18	16	14	17	13	7	6	10	13	6	15	13	21	14	14	16	14	10
Meandering	16	18	18	19	18	17	14	20	20	18	24	22	24	24	24	24	24	23
Pools-riffles	23	14	12	24	19	13	10	15	14	16	8	3	8	7	9	8	5	7
Hollow banks	10	10	8	6	12	4	4	8	7	22	22	17	23	23	23	22	22	20
Depth	3	7	7	13	4	6	5	5	6	7	11	5	3	5	7	6	1	2
Flow velocity	17	19	15	16	10	5	12	7	8	1	2	8	7	8	6	7	10	8
pH	15	15	6	14	8	10	21	17	9	10	21	10	14	13	10	12	8	9
Temperature	1	1	2	1	2	2	2	2	2	17	12	9	5	9	8	9	12	11
Dissolved oxygen	24	11	20	10	17	23	19	19	23	2	3	6	4	6	3	4	4	3
Conductivity	2	20	19	11	16	21	22	21	18	5	4	1	2	1	1	3	2	1
Suspended solids	13	4	1	2	1	1	1	1	1	11	6	11	15	16	11	10	13	12
Ammonium	5	3	3	3	5	8	7	3	3	20	9	19	22	17	15	17	18	21
Nitrate	12	2	4	4	3	3	3	4	4	8	1	4	10	3	5	2	6	4
Total nitrogen	19	17	16	15	22	15	17	16	21	14	10	21	12	15	21	14	19	19
Phosphate	4	8	9	12	7	11	8	11	10	13	16	15	19	22	19	13	23	15
Total phosphorus	21	13	11	20	14	18	16	14	15	4	7	7	1	2	2	1	7	6
COD	7	12	10	9	9	12	9	9	11	3	19	12	17	11	12	20	15	16
Pebbles	8	9	13	7	11	14	13	13	12	19	13	20	20	12	20	15	20	22
Gravel	22	22	23	23	23	20	18	23	16	21	20	14	18	20	18	23	21	24
Sand	9	24	22	21	15	22	23	18	22	24	17	16	13	18	17	21	17	17
Loam/clay	6	5	5	5	6	6	11	6	5	9	14	23	11	10	22	11	11	13
Distance to mouth	14	21	21	18	20	16	15	24	17	12	18	18	16	19	13	19	9	14
Stream order	20	23	24	22	24	24	24	22	24	23	5	2	6	4	4	5	3	5

Table 7.5. Comparison of the overall rankings of the input variables derived from the three contribution methods 'Weights', 'PaD' and 'Profile' for the Limnephilidae presence/absence ANN models (24-10-1) on the basis of the 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- and 10-fold cross-validation methods

	Zwalm river basin									**'Short distance' monitoring network**								
k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Width	16	11	11	12	10	8	10	6	11	6	7	4	7	6	5	5	6	5
Embankment	23	19	15	24	23	17	21	23	21	12	16	18	12	17	21	19	14	14
Meandering	9	8	20	17	14	14	15	22	16	20	22	21	22	20	19	17	16	16
Pools-riffles	17	23	24	23	17	24	24	24	22	8	5	9	5	5	7	6	7	6
Hollow banks	2	2	1	2	2	2	2	2	2	21	23	14	16	22	15	15	17	19
Depth	10	5	3	10	6	5	6	9	14	18	6	7	8	12	9	10	15	8
Flow velocity	20	9	16	5	8	12	11	10	17	7	12	11	19	11	16	13	18	15
pH	19	20	13	21	24	18	13	14	15	14	19	8	11	10	17	11	12	9
Temperature	7	12	5	6	4	3	5	3	6	2	9	15	9	19	10	14	9	20
Dissolved oxygen	11	6	19	7	12	6	9	11	12	11	4	1	1	2	3	2	4	1
Conductivity	13	3	12	11	15	10	18	17	10	24	15	23	18	21	18	23	13	23
Suspended solids	12	16	7	14	13	16	12	15	18	19	17	19	23	23	22	16	19	22
Ammonium	5	4	4	3	3	4	4	1	4	22	11	16	13	15	12	18	11	16
Nitrate	15	22	17	13	11	19	19	13	9	17	21	22	20	14	23	22	24	21
Total nitrogen	14	14	21	19	16	23	17	16	19	9	1	2	2	3	1	1	3	2
Phosphate	6	15	10	20	18	11	14	12	8	4	2	3	3	1	2	4	1	3
Total phosphorus	24	13	18	15	22	15	16	20	13	10	10	12	14	9	6	12	8	11
COD	18	18	23	18	21	22	23	19	24	16	20	20	17	24	14	24	22	17
Pebbles	21	21	22	16	20	20	20	18	20	23	13	17	21	16	24	20	20	13
Gravel	8	17	8	8	7	13	7	8	5	5	18	13	15	13	13	9	21	12
Sand	22	24	14	22	19	21	22	21	23	13	24	24	24	18	20	21	23	24
Loam/clay	4	10	9	9	9	9	8	7	7	15	14	10	10	8	11	7	10	10
Distance to mouth	1	1	2	1	1	1	1	4	1	3	8	6	4	7	8	8	5	7
Stream order	3	7	6	4	5	7	3	5	3	1	3	5	6	4	4	3	2	4

7.3.2. Effect of the environmental characteristics on the habitat suitability based on the three selected contribution methods

7.3.2.1. The Zwalm river basin

In Fig. 7.1 to 7.10, the outcomes of three input variable contribution methods (‘PaD’, ‘Weights’ and ‘Profile’) for the presence/absence models of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae are presented.

Based on the Zwalm river basin dataset, the variables ammonium (followed by stream order and distance to mouth) and nitrate (followed by distance to mouth and total phosphorus) had the highest contribution for Tubificidae based on respectively the ‘Weights’ and the ‘PaD’ method. Their contributions were respectively 5.32 % (ammonium), 5.21 % (stream order) and 5.04 % (distance to mouth) and 12.09 % (nitrate), 8.86 % (distance to mouth) and 6.89 % (total phosphorus). The standard deviation represented by the error bars, was relatively high for most of the variables based on the ‘PaD’ method. This indicated that the ‘PaD’ method was not very stable over the four folds, when the contributions were expressed as values. However, the ranking of the relative contribution of the variables was stable over the different folds. The standard deviation obtained with the ‘Weights’ method was in general rather small. Based on the ‘Profile’ method, no clear distinction between the different variables could be observed for Tubificidae. However, the curves of total phosphorus and nitrate could be distinguished if zoomed in. An increase of the total phosphorus and nitrate concentration induced a decrease of the probability of presence of Tubificidae. The range between the minimum and maximum of the other input variables was very small, i.e. they were of minor importance.

For Asellidae, stream order, width and distance to mouth were the major input variables based on both the ‘PaD’ and the ‘Weights’ method. Their contributions were 7.77 % and 22.23 %, 6.20 % and 16.78 %, and 5.78 % and 12.42 % for respectively the ‘Weighs’ and ‘PaD’ method. Although the standard deviations were smaller in comparison with Tubificidae, the same conclusions could be drawn for the stability of both methods. Based on the ‘Profile’ method, the variables width, hollow banks and stream order were best expressed. Because it covered the largest range, width can be seen as the environmental variable with the greatest effect on the presence/absence of Asellidae based on the ‘Profile’ method. An increase of the river width meant an increase of the probability of the presence of Asellidae. Similarly, an increase of the value of hollow banks and stream order led to an increase of the probability of presence of Asellidae.

Based on the ‘Weights’ and ‘PaD’ method, distance to mouth (respectively 5.30 % and 10.82 %) and temperature (respectively 5.27 % and 10.09 %) were the major contributing variables for Gammaridae. The standard deviation of the ‘PaD’ method was high for the most important variables, while rather small on the basis of the ‘Weights’ method. Based on the ‘Profile’ method, water quality related variables were of major importance. The variables conductivity, COD and ammonium were best expressed. The probability of presence of Gammaridae decreased if the concentration of these variables increased.

For *Baetis*, suspended solids, ammonium, temperature and nitrate had the highest influence on the prediction of the output. Their contributions were 6.75 % and 20.56 %, 5.68 % and 10.78 %, 5.85 % and 10.34 %, and 5.26 % and 8.97 % based on respectively the ‘Weights’ and the ‘PaD’ method. Both methods were rather stable as derived from the error bars. Based on the ‘Profile’ method, minor distinction between the different variables could be observed for *Baetis*. However, an increase of suspended solids induced a decrease in the probability of presence.

For both the ‘Weights’ and the ‘PaD’ method, stream order (respectively 7.74 % and 15.73 %) and hollow banks (respectively 6.68 % and 11.11 %) were the key variables predicting Limnephilidae. Based on the ‘Profile’ method, suspended solids was the only variable which could be distinguished. An increase in concentration led to a decrease of the probability of presence of Limnephilidae.

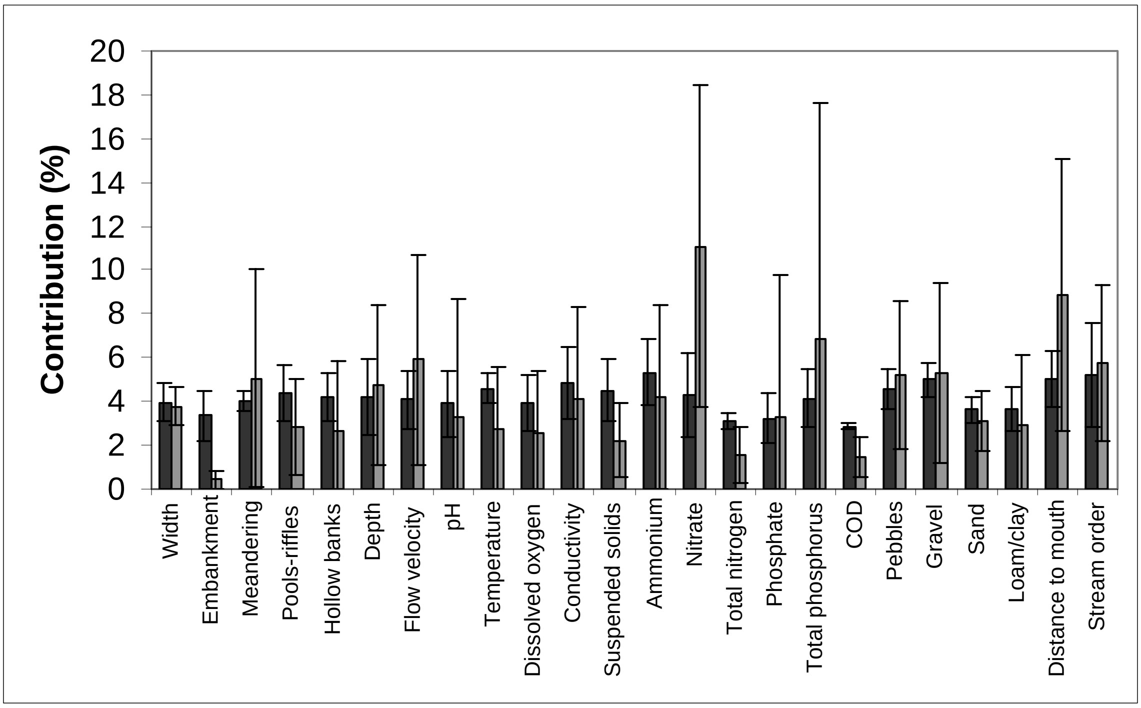


Fig. 7.1. Contribution of the environmental variables used in the 24-5-1 ANN model for Tubificidae based on the 'PaD' (= grey bar) and the 'Weights' (= black bar) method (Zwalm river basin dataset).

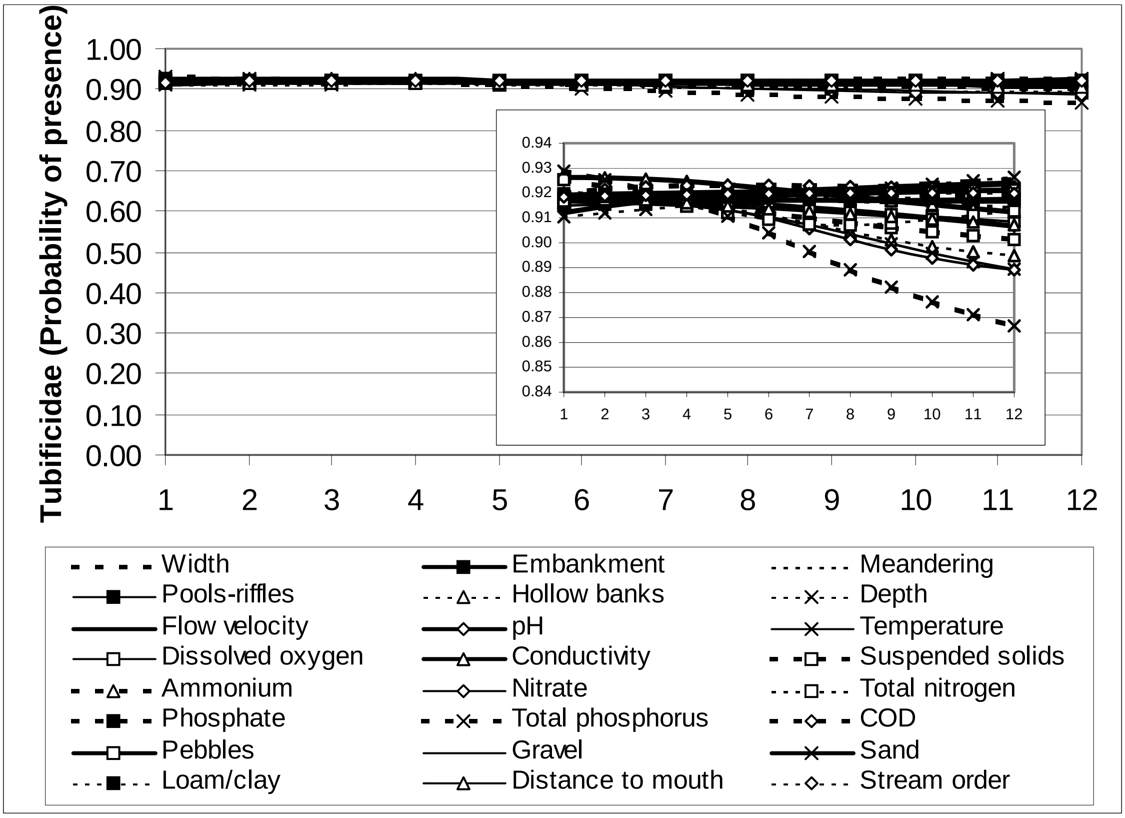


Fig. 7.2. Contribution of the environmental variables used in the 24-5-1 ANN model for Tubificidae based on the 'Profile' method (Zwalm river basin dataset).

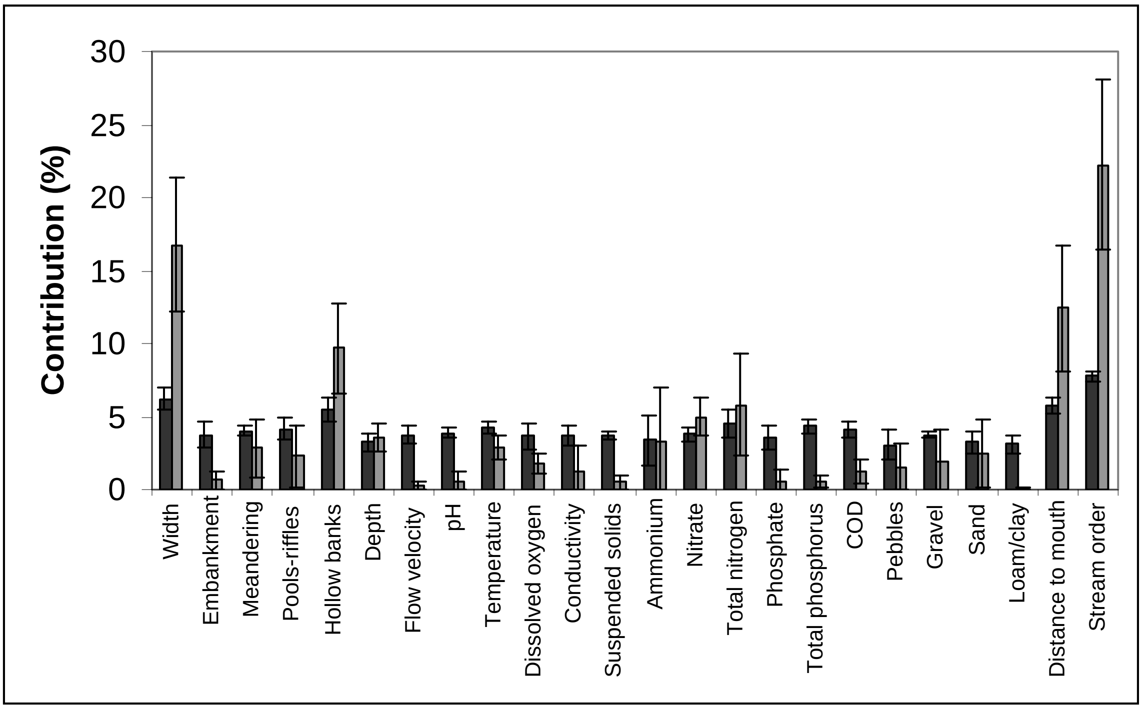


Fig. 7.3. Contribution of the environmental variables used in the 24-10-1 ANN model for Asellidae based on the 'PaD' (= grey bar) and the 'Weights' (= black bar) method (Zwalm river basin dataset).

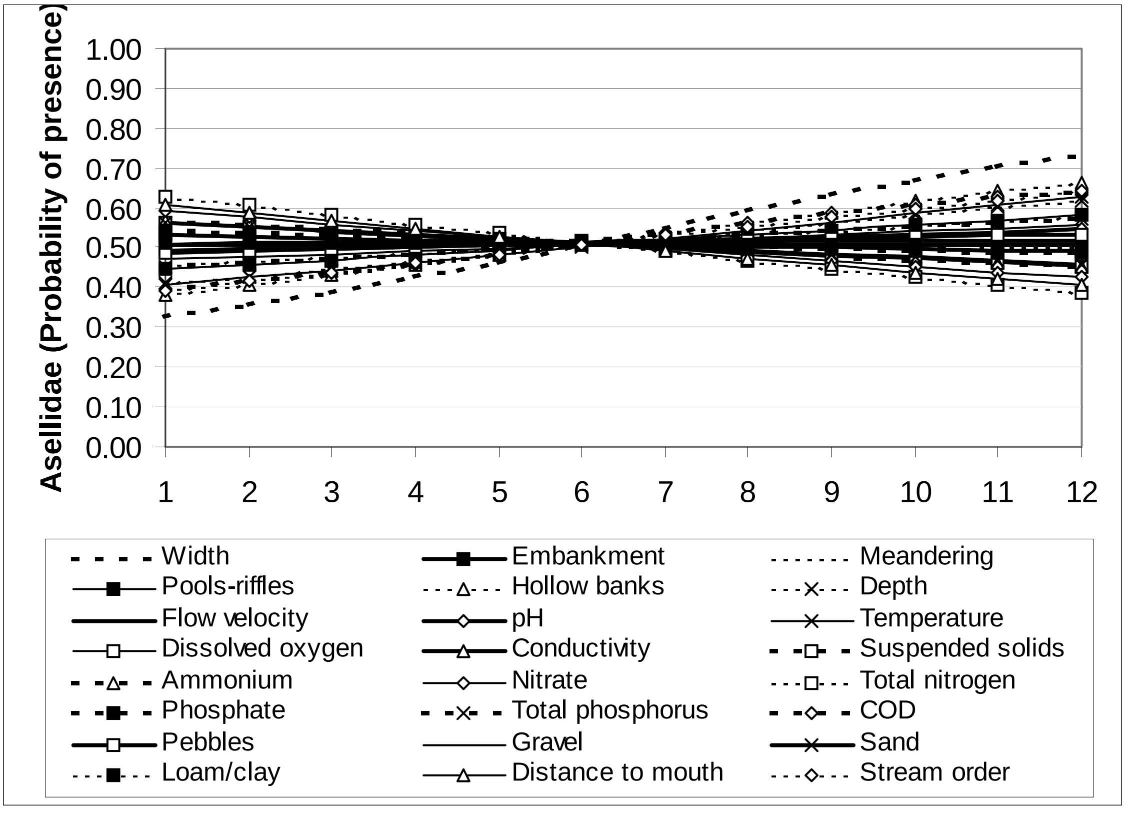


Fig. 7.4. Contribution of the environmental variables used in the 24-10-1 ANN model for Asellidae based on the 'Profile' method (Zwalm river basin dataset).

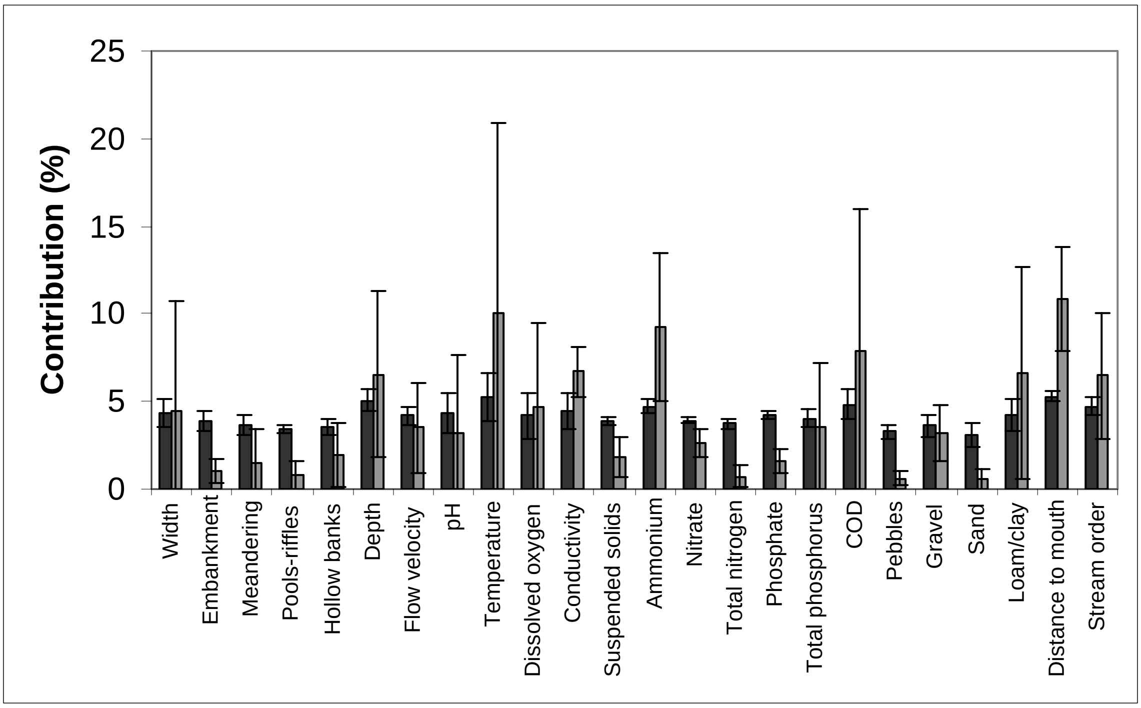


Fig. 7.5. Contribution of the environmental variables used in the 24-15-1 ANN model for Gammaridae based on the 'PaD' (= grey bar) and the 'Weights' (= black bar) method (Zwalm river basin dataset).

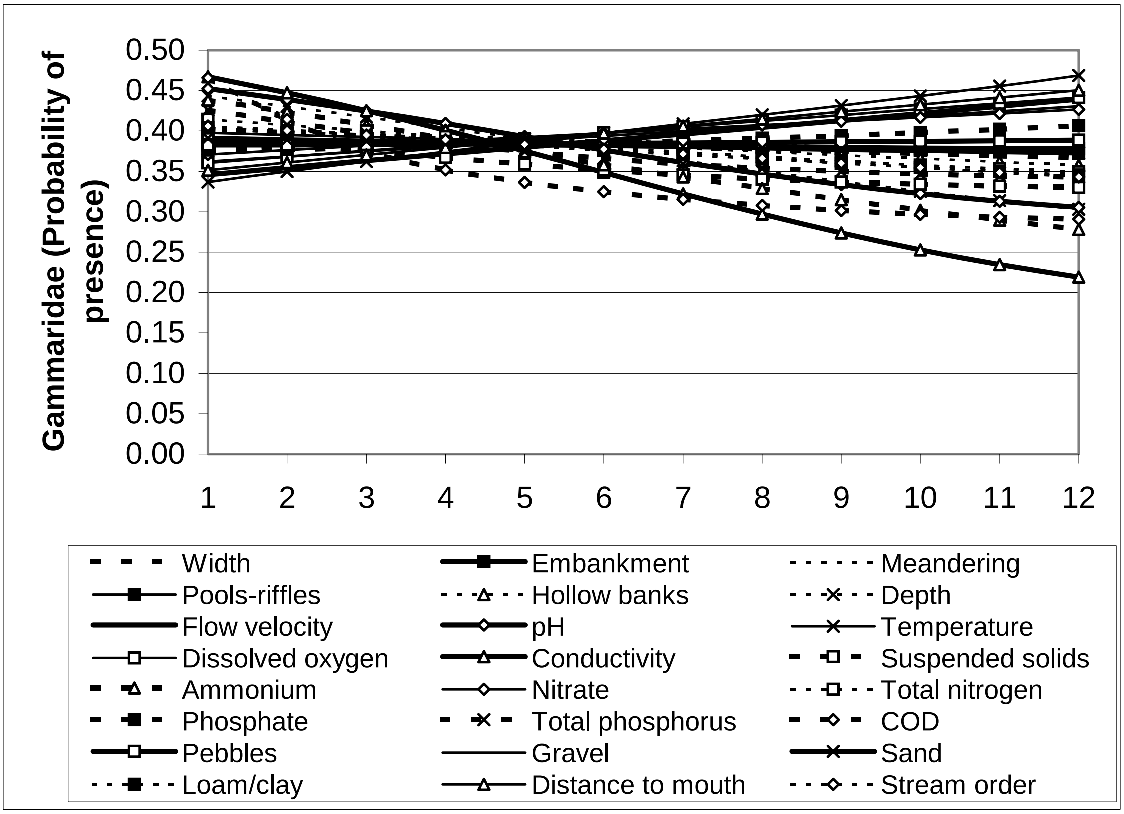


Fig. 7.6. Contribution of the environmental variables used in the 24-15-1 ANN model for Gammaridae based on the 'Profile' method (Zwalm river basin dataset).

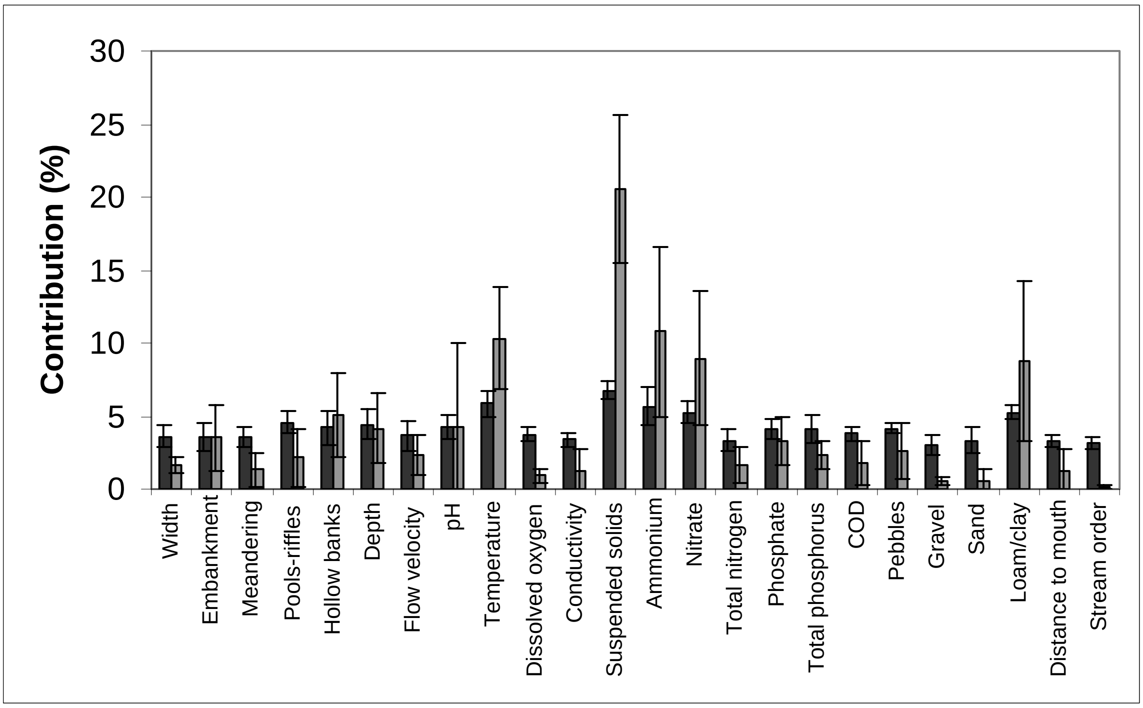


Fig. 7.7. Contribution of the environmental variables used in the 24-10-1 ANN model for *Baetis* based on the 'PaD' (= grey bar) and the 'Weights' (= black bar) method (Zwalm river basin dataset).

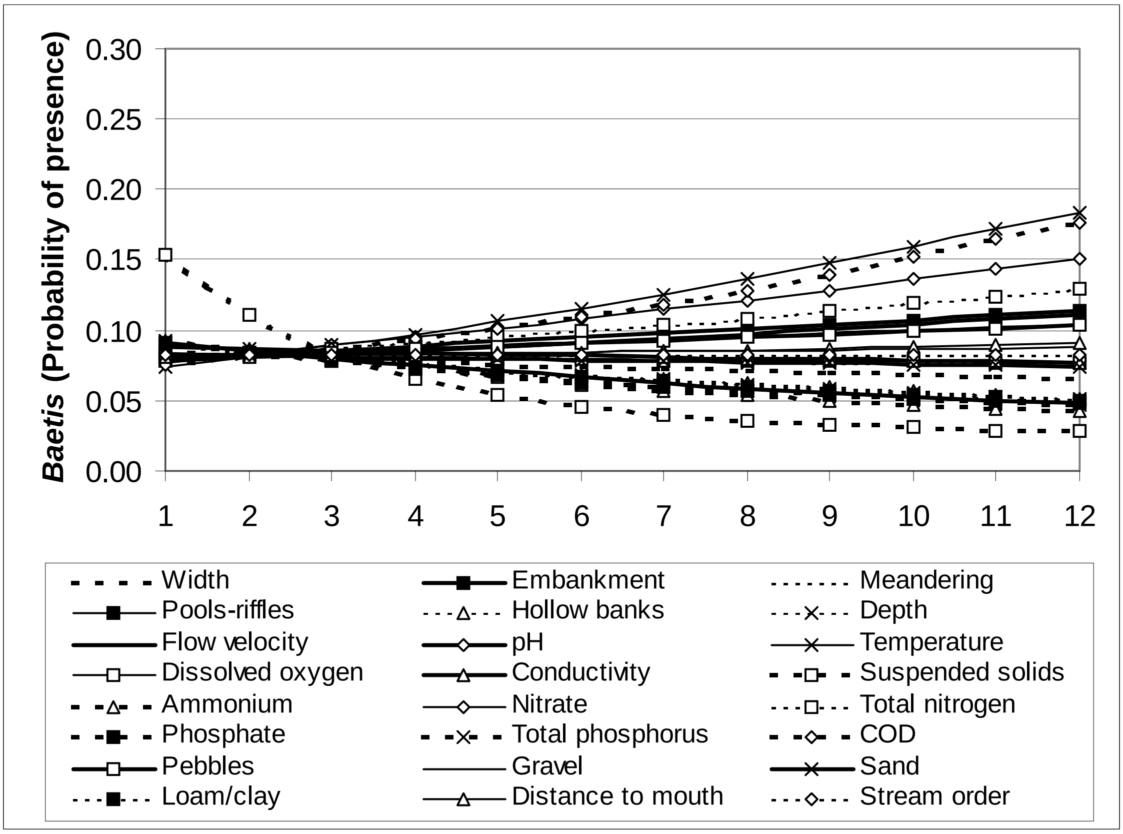


Fig. 7.8. Contribution of the environmental variables used in the 24-10-1 ANN model for *Baetis* based on the 'Profile' method (Zwalm river basin dataset).

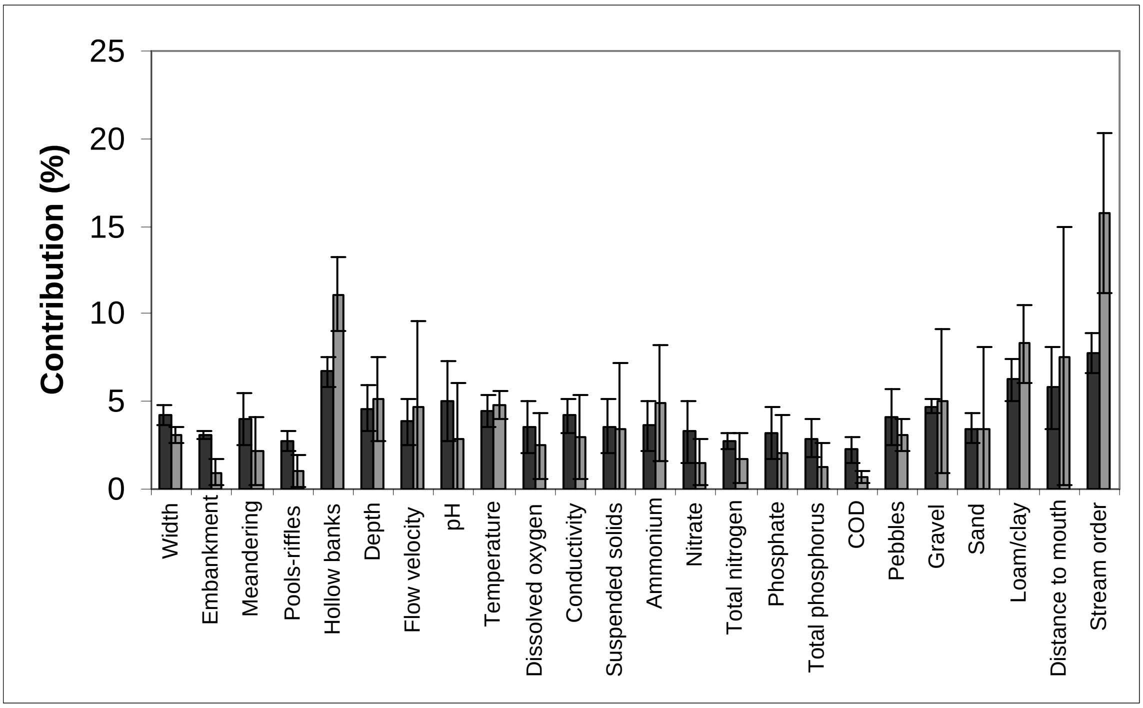


Fig. 7.9. Contribution of the environmental variables used in the 24-5-1 ANN model for Limnephilidae based on the 'PaD' (= grey bar) and the 'Weights' (= black bar) method (Zwalm river basin dataset).

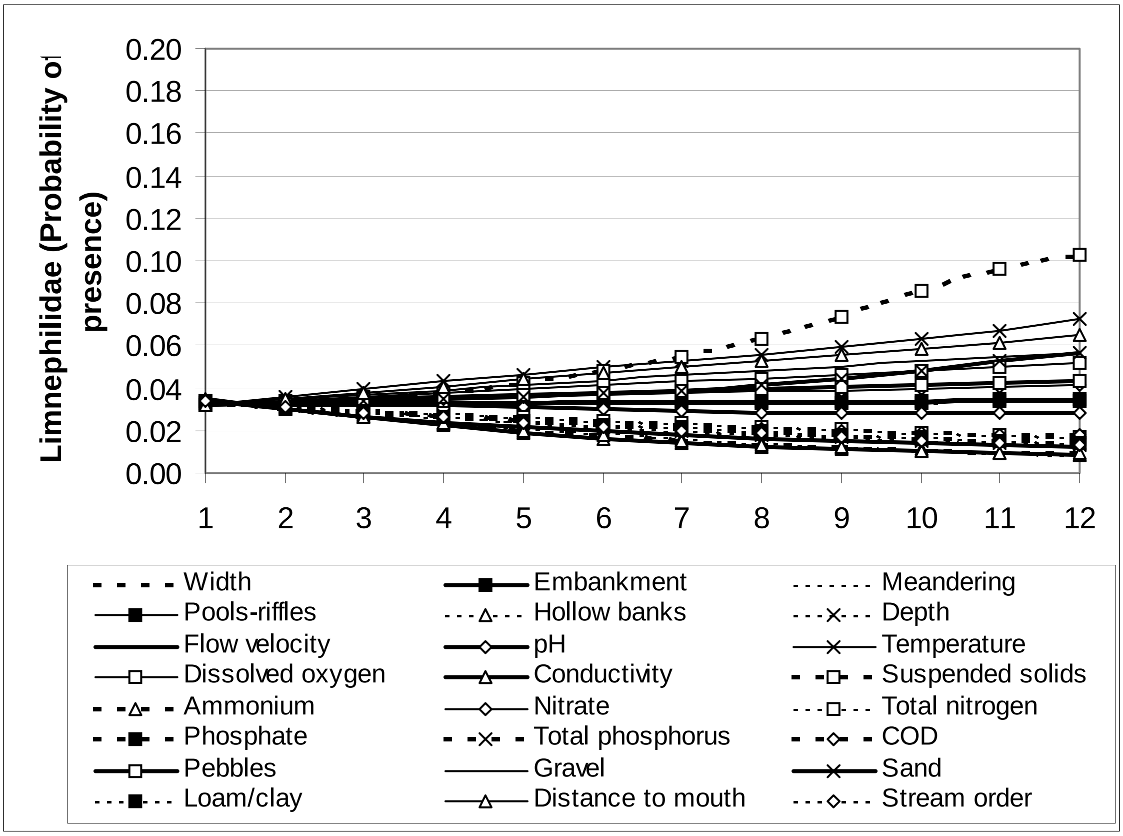


Fig. 7.10. Contribution of the environmental variables used in the 24-5-1 ANN model for Limnephilidae based on the 'Profile' method (Zwalm river basin dataset).

7.3.2.2. 'short distance' monitoring network

In Fig. 7.11 to 7.20, the results of the 'PaD', 'Weights' and 'Profile' methods for the presence/absence models of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae are presented.

Based on the 'short distance' monitoring network database, stream order and sand were key variables to predict the presence/absence of Tubificidae on the basis of both the 'Weights' and the 'PaD' method. Stream order and sand have contributions of respectively 5.96 % and 13.80 %, and 5.82 % and 10.87 % for the 'Weights' and the 'PaD' method. The standard deviation represented by the error bars, was relatively high for most of the variables based on the 'PaD' method. This indicated that the 'PaD' method was not very stable over the four folds, when the contributions were expressed as values. However, the ranking of the relative contribution of the variables was stable over the different folds. The standard deviation obtained with the 'Weights' method was in general rather small. Based on the 'Profile' method, the only variable which was expressed well was nitrate. An increase in nitrate concentration led to a decrease of the probability of presence.

For Asellidae, the most important variables based on the three methods were distance to mouth and stream order. They had a contribution of respectively 8.64 % and 22.21 %, and 7.62 % and 18.57 % based on the 'Weights' and 'PaD' method. The other variables were of minor importance for the prediction of the presence/absence. Based on the 'Profile' method, an increase of distance to mouth and stream order induced respectively decrease and an increase in the probability of presence of Asellidae.

For Gammaridae, the major variables were depth (respectively 7.16 % and 18.44 %) and ammonium (respectively 6.12 % and 12.86 %) for both the 'Weights' and 'PaD' method. In addition, COD, flow velocity and total phosphorus had a paramount effect on the prediction of the output. Based on the 'Profile' method, depth, COD and flow velocity were best expressed. Gammaridae preferred more shallow streams with a relatively high flow velocity and low COD concentrations.

For *Baetis*, key variables were stream order (6.59 %) and nitrate (6.40 %) based on the 'Weights' method while the variables stream order (15.90 %), pools-riffles (10.79 %), conductivity (10.50 %) and dissolved oxygen (9.83 %) had the highest contribution based on the 'PaD' method. On the basis of the 'Profile' method, conductivity, depth, nitrate and total phosphorus had the highest effect on the probability of presence of *Baetis*.

Based on the 'Weights' method, dissolved oxygen (6.53 %), distance to mouth (5.53 %) and total nitrogen (5.47 %) were the most important variables for Limnephilidae. On the other hand, dissolved oxygen (16.10 %), stream order (10.29 %) and total nitrogen (9.92 %) were of major importance based on the 'PaD' method. Similar to the previous taxa, high standard deviations for the most important variables were observed based on the 'PaD' method whereas rather small standard deviations were detected for the 'Weights' method. Derived from the 'Profile' method, phosphate, total nitrogen and dissolved oxygen had the highest impact on the probability of presence. An increase in dissolved oxygen concentration led to an increase of the probability of presence. On the contrary, an increase in phosphate and total nitrogen concentration induced a decrease in Limnephilidae.

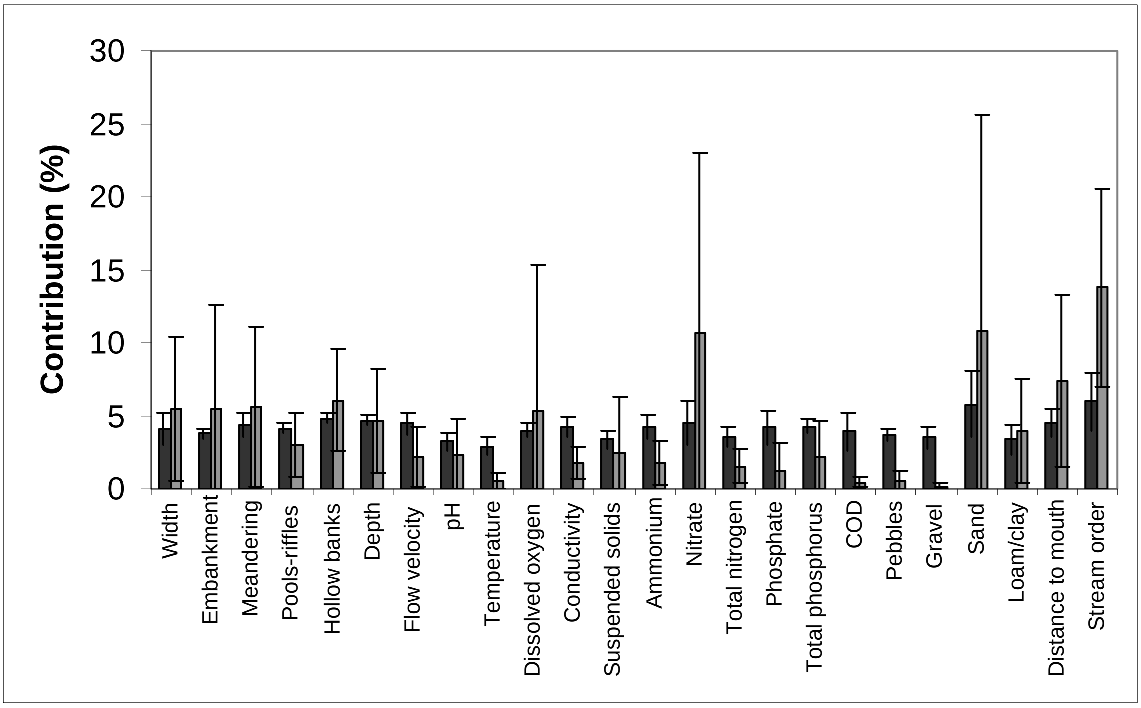


Fig. 7.11. Contribution of the environmental variables used in the 24-15-1 ANN model for Tubificidae based on the 'PaD' (= grey bar) and the 'Weights' (= black bar) method ('short distance' monitoring network dataset).

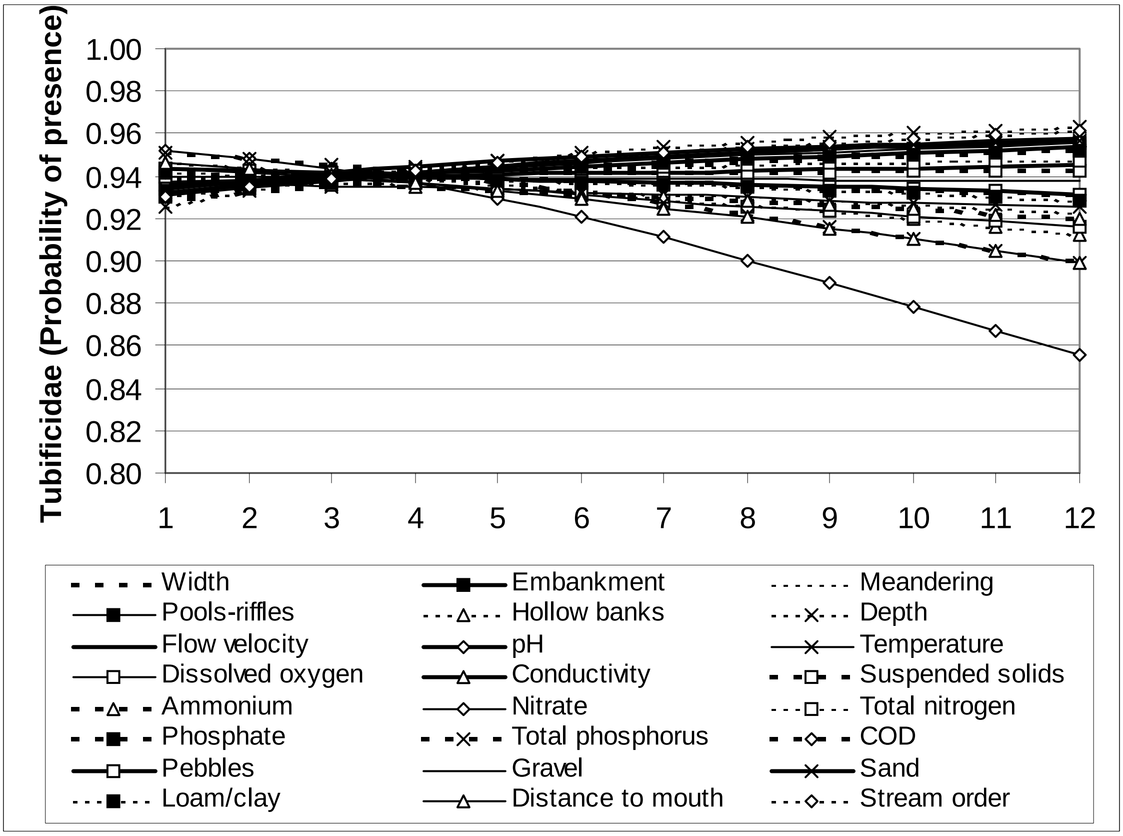


Fig. 7.12. Contribution of the environmental variables used in the 24-15-1 ANN model for Tubificidae based on the 'Profile' method ('short distance' monitoring network dataset).

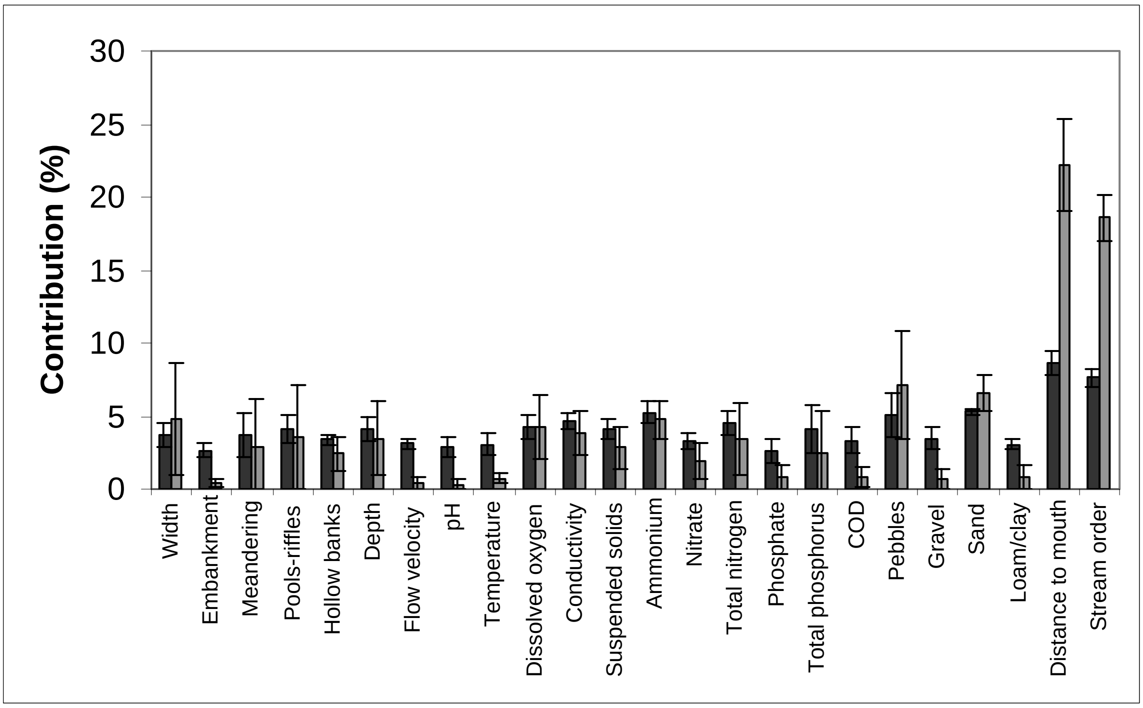


Fig. 7.13. Contribution of the environmental variables used in the 24-10-1 ANN model for Asellidae based on the ‘PaD’ (= grey bar) and the ‘Weights’ (= black bar) method (‘short distance’ monitoring network dataset).

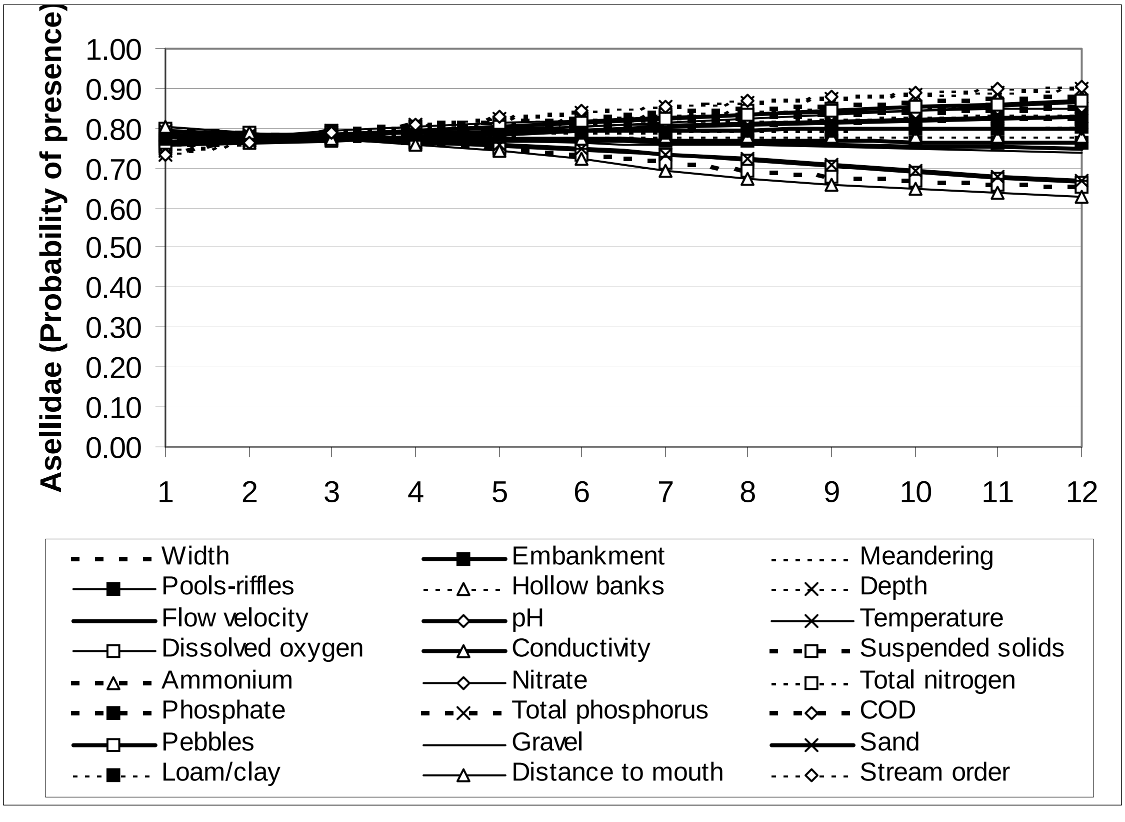


Fig. 7.14. Contribution of the environmental variables used in the 24-10-1 ANN model for Asellidae based on the ‘Profile’ method (‘short distance’ monitoring network dataset).

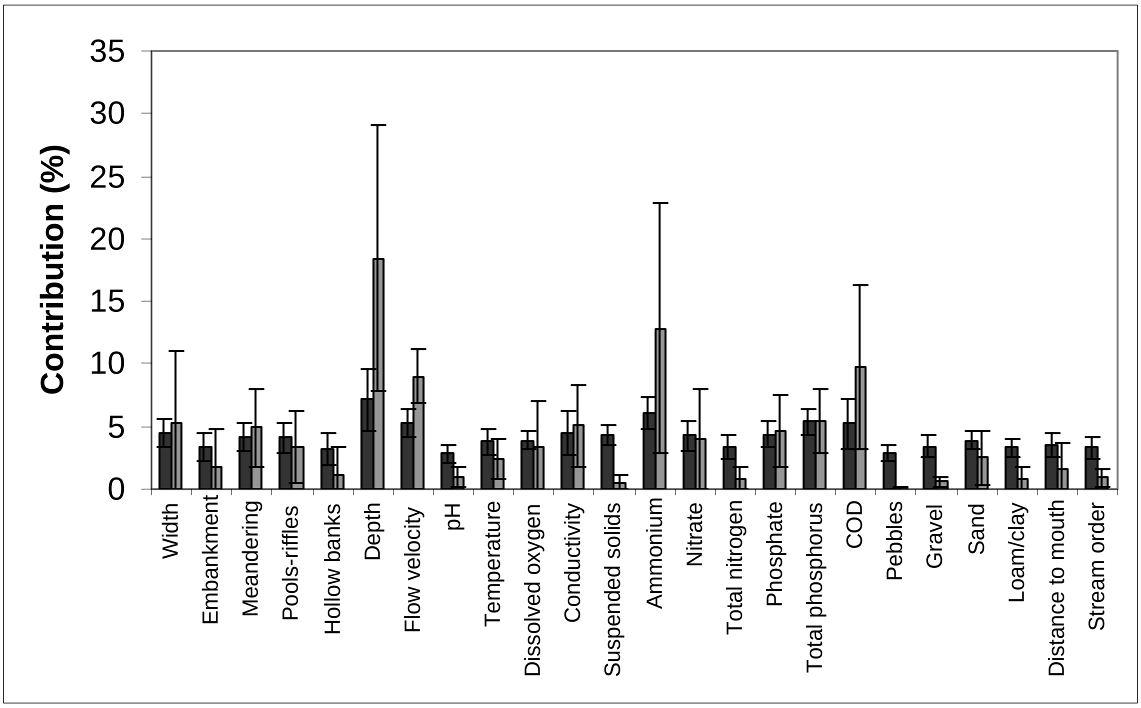


Fig. 7.15. Contribution of the environmental variables used in the 24-10-10-1 ANN model for Gammaridae based on the 'PaD' (= grey bar) and the 'Weights' (= black bar) method ('short distance' monitoring network dataset).

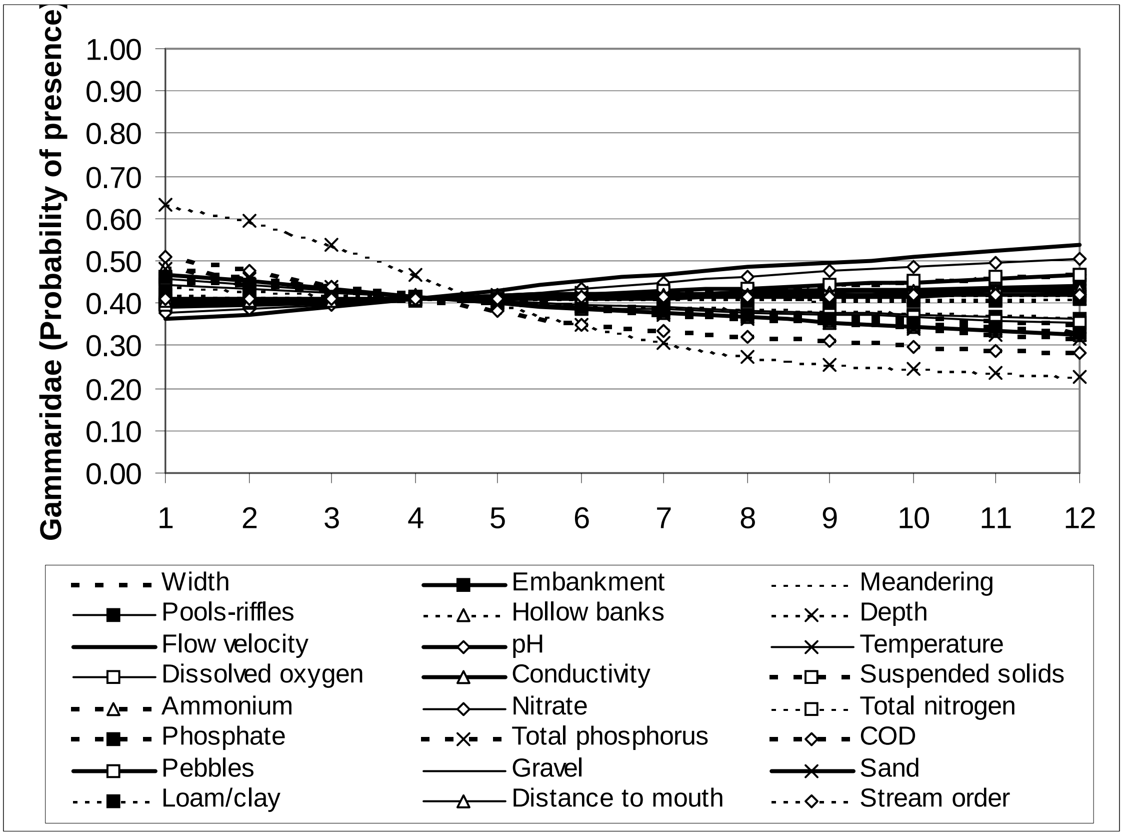


Fig. 7.16. Contribution of the environmental variables used in the 24-10-10-1 ANN model for Gammaridae based on the 'Profile' method ('short distance' monitoring network dataset).

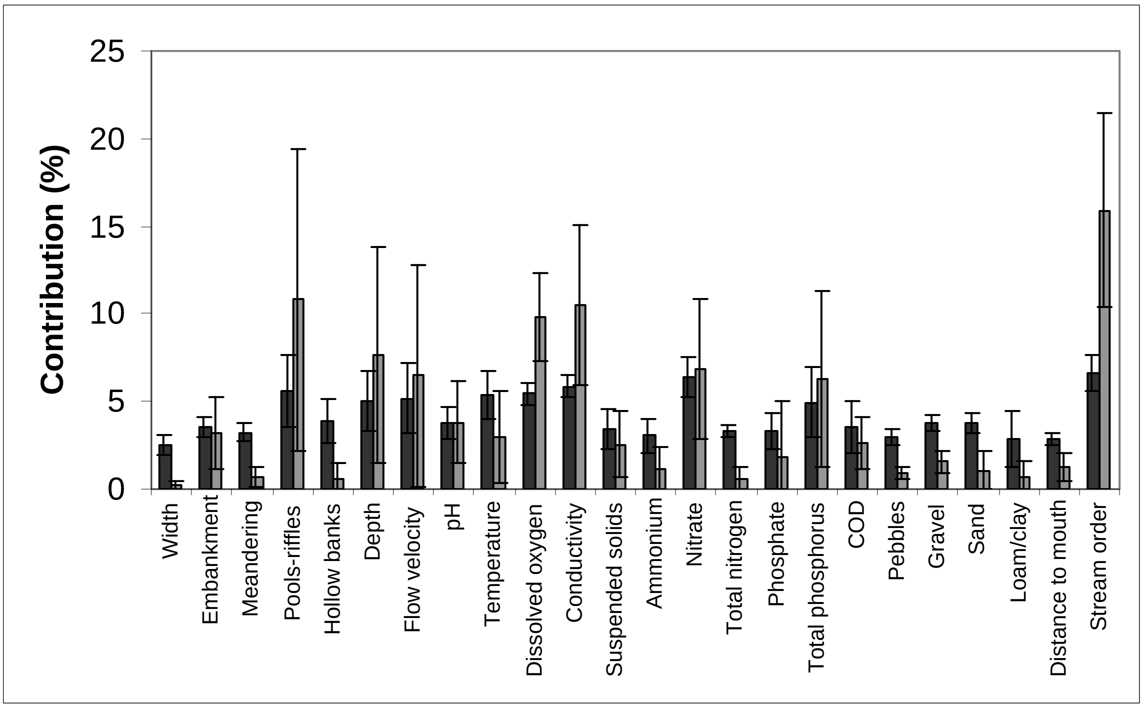


Fig. 7.17. Contribution of the environmental variables used in the 24-10-1 ANN model for *Baetis* based on the 'PaD' (= grey bar) and the 'Weights' (= black bar) method ('short distance' monitoring network dataset).

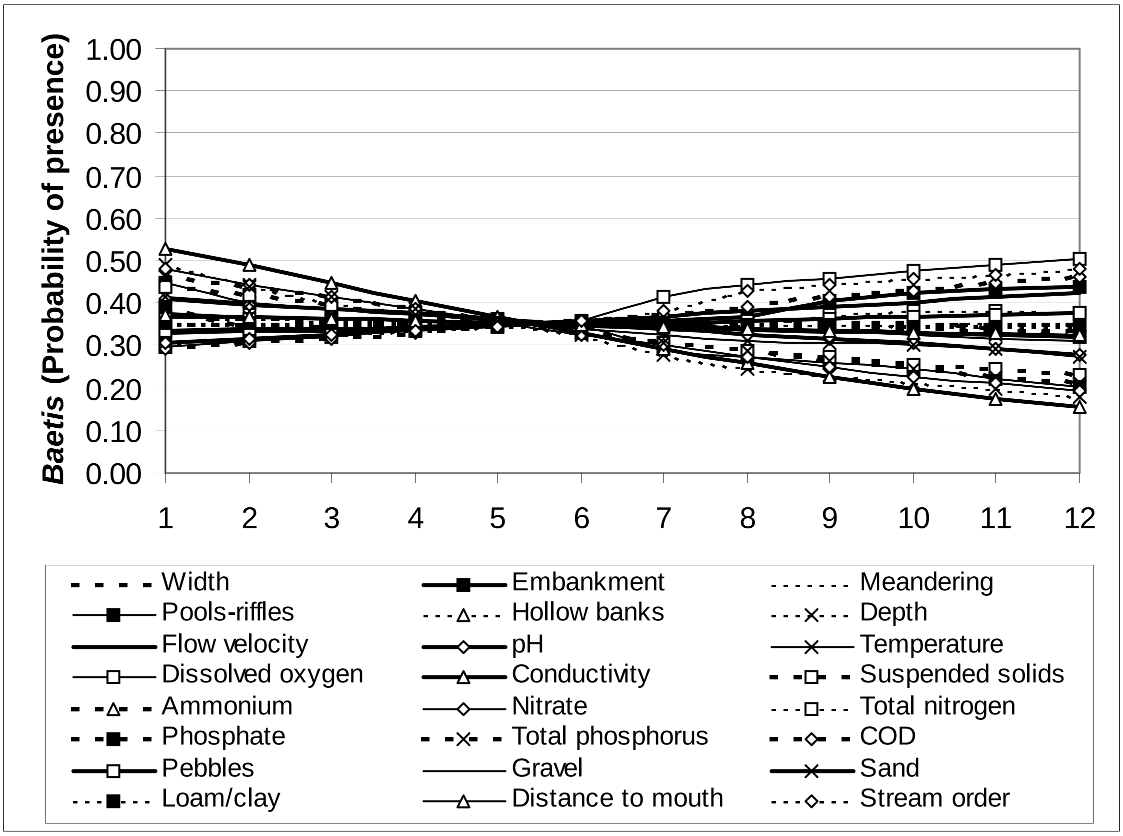


Fig. 7.18. Contribution of the environmental variables used in the 24-10-1 ANN model for *Baetis* based on the 'Profile' method ('short distance' monitoring network dataset).

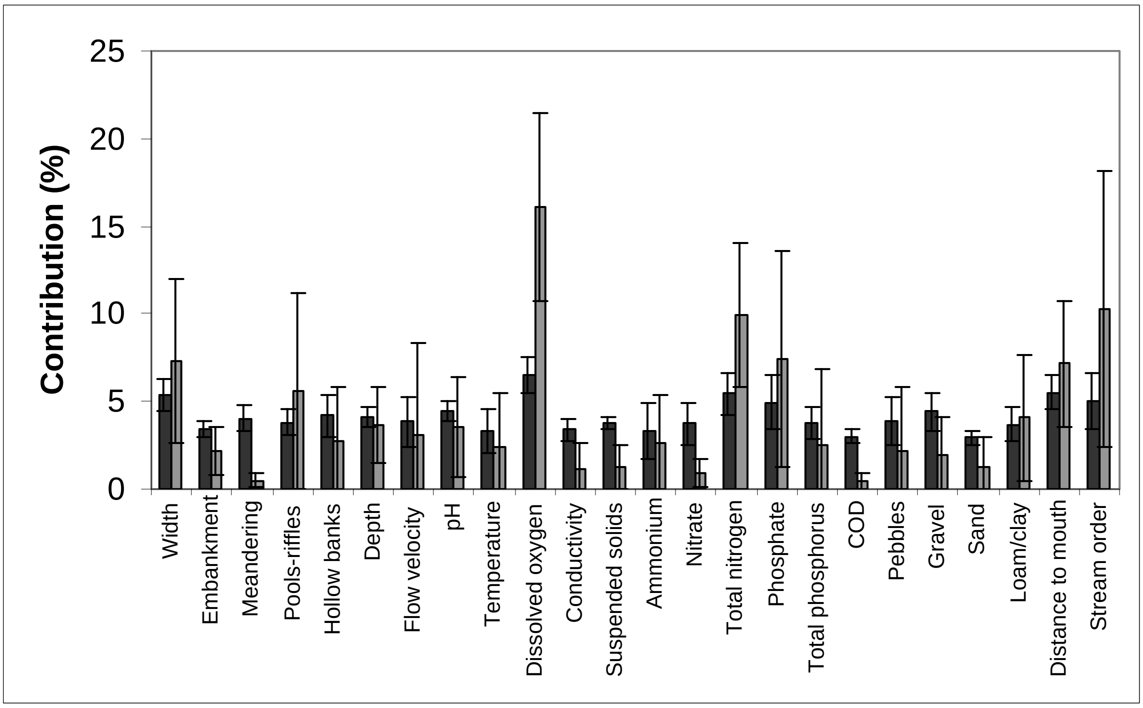


Fig. 7.19. Contribution of the environmental variables used in the 24-10-1 ANN model for Limnephilidae based on the 'PaD' (= grey bar) and the 'Weights' (= black bar) method ('short distance' monitoring network dataset).

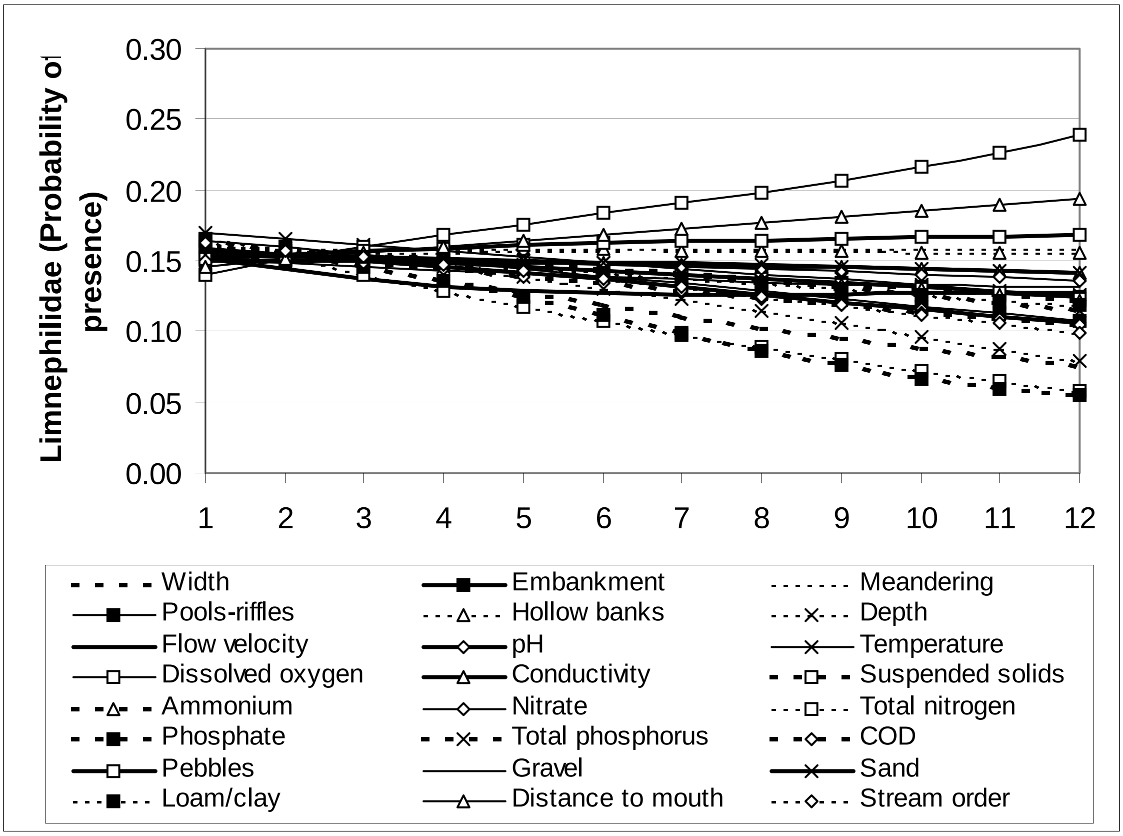


Fig. 7.20. Contribution of the environmental variables used in the 24-10-1 ANN model for Limnephilidae based on the 'Profile' method ('short distance' monitoring network dataset).

7.3.2.3. Stability of the contribution methods

As mentioned in the material and methods' section, the stability of the contribution methods was checked based on the *k* repetitions, according to the *k*-fold cross-validation. Based on this repetition, the average contribution of every variable for each method was calculated. In addition, the standard deviation was determined. This standard deviation gave an indication of the stability of each contribution method. In the previous paragraphs, the standard deviation is represented by the error bars for both the 'Weights' and 'PaD' method.

In general, the standard deviation was relatively high for most of the variables based on the 'PaD' method, especially for the most important variables. This indicated that the 'PaD' method was not very stable over the four folds, when the contributions were expressed as values. However, the ranking of the relative contribution of the variables was stable over the different folds. The standard deviation obtained with the 'Weights' method was in general rather small, demonstrating this method was more stable.

The stability of the 'Profile' method was determined in the same way as for the 'Weights' and 'PaD' method: the average of the four curves was taken and the standard deviation was derived. The standard deviation of the 'Profile' curves however was difficult to visualize in the figures above since all variables were presented in one chart for practical reasons. In Fig. 7.21, an example of the stability of the 'Profile' method is illustrated for Asellidae for the four major variables (width, hollow banks, stream order and total nitrogen) of the Zwalm river basin dataset. This is repeated for the other taxa, variables and dataset.

Although, the four profiles (based on the four cross-validation steps) presented a similar trend, they varied slightly in shape, slope and covered range, resulting in small standard deviation flags. However, this variation was rather small as could be deduced from the examples in Fig. 7.21. Similar results were obtained for the other 'Profile' methods performed on the other taxa, variables and dataset.

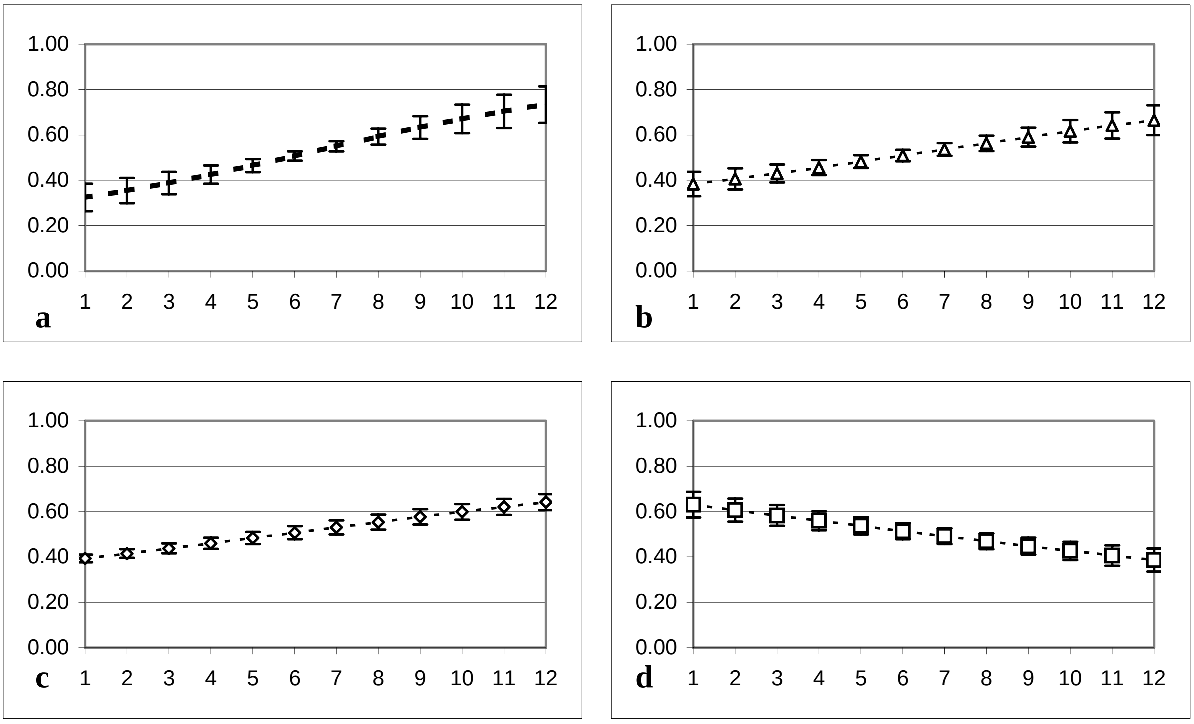


Fig. 7.21. Contribution and stability of the four major environmental variables (a = width, b = hollow banks, c = stream order, d = total nitrogen) used in the 24-10-1 ANN model for Asellidae based on the ‘Profile’ method (Zwalm river basin dataset).

7.4. Discussion

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 species or populations based on habitat characteristics is an challenging 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.*, 1997b; Guégan *et al.*, 1998). To this end, ANN models have been shown to reveal a high predictive power (e.g. Lek and Guégan, 1999). These models are in particular very powerful in dealing with non-linear relationships between for example environmental variables. ANN models are also recognized to make reliable predictions of river invertebrates as discussed in Chapter 2.

The use of macroinvertebrates can also be justified since many studies have revealed that these organisms are important biological indicators of water quality (Klemm *et al.*, 2002). Moreover, the European Union Water Framework Directive states that invertebrate fauna should be part of the ecological assessment system for rivers (EU, 2000). Macroinvertebrates reflect the biological integrity of the aquatic ecosystem (Davis and Simon, 1995; Klemm *et al.*, 1990; Rosenberg and Resh, 1993). Numerous macroinvertebrate species having relatively long life cycles of one year or more, are especially important biological indicators of site conditions over time, and respond rapidly and predictably to changes in water quality and habitat changes. Additionally, some groups are tolerant and are found in polluted environments, while other groups are intolerant and respond to either specific stressors or a wide array of stressor and anthropogenic disturbances (Davis and Simon, 1995; Karr and Chu, 1999; Klemm *et al.*, 1990; Meyer, 1997; Rosenberg and Resh, 1993). In this way, Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae were selected as representative taxa because of their variable presence in the Zwalm river basin, and their use as bio-indicators in river quality assessment (e.g. De Pauw and Vannevel, 1991; MacNeil *et al.*, 2002; Tachet *et al.*, 2002).

The lack of illustrative power of these ANN models however is a major concern to ecologists since the interpretation of statistical models is desirable for gaining knowledge of the relationships driving ecological phenomena (Olden and Jackson, 2002). In this way, ANN models have been labelled as 'black box'. To make them more transparent and enlarge their explanatory capacity, different methods like contribution or sensitivity analysis are being developed (e.g. Dimopoulos *et al.*, 1995, 1999; Garson, 1991; Goh, 1995; Lek *et al.*, 1996a, b; Olden and Jackson, 2002; Scardi and Harding, 1999). In this study, three contribution methods were selected and applied to the ANN models: the 'Weights', the 'PaD' and the 'Profile' method. These methods seemed to outperform the other methods according to former studies based on macroinvertebrates in the same study area (Dedecker *et al.*, 2005b, c; Goethals, 2005). The selected contribution methods helped to identify environmental factors influencing the habitat suitability of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae.

The 'Profile' method 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 probability of presence. 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 can help to identify impact-specific indicator taxa and enhance the capacity to monitor, assess and mitigate the effects of human activities on river ecosystems. The other methods ('PaD' and 'Weights') were merely able to classify the variables by order of their importance, in other words, to reveal their contribution to the output.

The 'Weights' and 'PaD' method seemed to give similar results concerning the order of importance of the paramount input variables. The 'PaD' method however seemed to distinguish more clearly minor and major contributing environmental variables in comparison to the 'Weights' methods. Similar results were found by Gevrey *et al.* (2003), Dedecker *et al.* (2005b, c) and Goethals (2005). In those studies, the relation between environmental variables and respectively trout spawning site density, *Gammarus* and *Asellus* was predicted. These results are 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 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.

The results of the 'Profile' method were in general quite different in comparison to the 'Weights' and 'PaD' method. The impact of minimum and maximum values of the input variables is very important for this curves and was used to determine the rank of the variables. However, differences between minimum and maximum value were often very small, providing sometimes a biased rank based on this technique. More important is the trend of the different variables itself. A major outcome of these techniques is that only a few variables (about four) really seem to play a role in the models, while the effect of the other variables seems almost nil for the predictions (very horizontal curves). The charts obtained for Tubificidae were an exception since no clear trend could be observed for any variable. The performance (especially Cohen's kappa) of these models was also very poor. In general, they were not able to find major trends in the data and predicted Tubificidae in most cases as present what could be explained by the obtained 'Profile' curves. This type of graphs is therefore crucial to find out how ecologically sound the models are and what their meaning can be for practical simulations. Only when the variables of interest to managers take crucial part in the predictions, the models will be useful for decision support in river management.

Thus, the combination of the 'PaD' method and the 'Profile' method gives a clear idea of the ecological meaning of the models and their practical relevance for decision support of river management.

The outcomes of the three contribution methods are not always very constant over the different subsets, as can be deduced from the standard deviation flags in the graphs. The highest standard deviation was found for the 'PaD' method, especially for the key input variables. The standard deviations obtained with the 'Weights' and the 'Profile' method however were in general rather small, demonstrating that these methods were more stable. These often large standard deviation can be a results of outliers. Perhaps these could also be reduced by making stratified subsets based on these major variables (in addition to the output variables).

From an ecological point of view, pollution related variables, such as ammonium, conductivity, COD and total phosphorus, as well as variables defining the river type (e.g. distance to mouth) and the physical habitat (e.g. temperature, depth and flow velocity) were essential to describe the habitat of Gammaridae in the Zwalm river basin according to the different contribution methods. In particular, the input variables ammonium and COD and to a smaller extent conductivity and total phosphorus, were of major importance in predicting the habitat suitability of Gammaridae in the Zwalm river basin and the 'short distance' monitoring network. Based on the 'Profile' method, an increase in the concentration of ammonium, COD and conductivity were correlated with a decrease in the probability of presence of Gammaridae. This is confirmed by Hawkes (1979) who stated that Gammaridae is suppressed by high organic loads and by Whitehurst and Lindsey (1990) who stated that Gammaridae is intolerant to organic sewage. Organic pollution is normally a result of discharges from sewers, industrial effluents and agricultural run-off. Especially agricultural activities are responsible for organic pollution in the Zwalm river basin. Also dissolved oxygen expresses an important notion of organic pollution present in the rivers. Although dissolved oxygen had less influence on the prediction of the output, the 'Profile' method indicated that the habitat suitability of Gammaridae improved when dissolved oxygen concentration increased. That is affirmed by Wesenberg-Lund (1982) who stated that Gammaridae are almost non-tolerant to low oxygen conditions. However, these conditions were hardly present anymore in the Zwalm and pollution problems are more related to increased nutrient levels. Also a minor relationship between conductivity and the probability of presence of Gammaridae was found in the Zwalm river basin during the period 2000-2003. Steenbergen (1993) mentioned that Gammaridae are sensitive to high conductivity values. In relation to freshwaters, this variable may describe pollution caused by agricultural activities and treated or untreated wastewater effluents. At conductivity values above 1000 µS/cm, Gammaridae should experience negative influences. However, conductivity values higher than 1000 µS/cm were very rare in this study area (Chapter 5). This could explain why conductivity was not identified as an important variable in the present study. Modelling studies were made in Flanders, Belgium, on the basis of datasets from several catchments during other periods. For example Adriaenssens *et al.* (2005), who used international expert knowledge from literature, concluded that the environmental variable conductivity explained a major part of the abundance based on fuzzy knowledge-based models. 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. Although pollution related variables seemed to be of high importance to predict the presence/absence of Gammaridae, habitat characteristics such as distance to mouth and depth also had a high influence on the output. Contrary to what was expected, flow velocity was only considered a relatively relevant variable in the 'short distance' monitoring network. Based on the Moog (1995) and Bayerisches Landesamt für Wasserwirtschaft (1996) however, Gammaridae prefer rather fast running streams, due to its very good swimming abilities (Brehm and Meijering, 1990). Also in the headwaters of the Zwalm river basin Gammaridae were more abundant then in the lower parts. As a consequence, Gammaridae have potential as a good indicator organism in those headwaters, since this habitat is more suited by nature for this taxon. In this way, the methods for testing the contributions of the different input variables facilitate the selection of the suitable habitats in which certain species can or cannot act as an indicator organism for the assessment and management of rivers. This study demonstrated that Gammaridae 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.

When analysing the ecological relevance of the models for Asellidae, it seems that the river continuum concept (Vannote *et al.*, 1980) is very well confirmed by the data (distance to mouth and stream order variables). For Asellidae, the environmental variables describing the stream type (width, stream order and distance to mouth) were, beside hollow banks, the most significant variables in the Zwalm river basin. In the 'short distance' monitoring network, only distance to mouth and stream order were detected as significant variables. Based on the 'Profile' method, an increase of the river width resulted in an increase of the probability of presence of Asellidae for both datasets. This is confirmed by field experience (see Chapter 5) and by Steenbergen (1993). In a study by Goethals *et al.* (2001), classification trees were used to predict the presence/absence of Asellidae. The only rule generated by the classification trees was: 'if width is more than 3.5 meters, then Asellidae are present, while absent in the more narrow streams'. Logically, the highly correlated variables stream order and distance to mouth (see Chapter 5) were also very important. Similar ecological interpretations could be drawn from these variables: an increase of depth and stream order led to an improvement of the habitat suitability of Asellidae. On the other hand, an increase of distance to mouth resulted in a decrease of the probability of presence. Contrary to what was expected, the flow velocity was not strongly correlated with the variables width, depth, stream order and distance to mouth as could be derived from Table 5.3 and 5.4. Low flow velocities could be expected if width, depth and stream order increased and distance to mouth decreased. Due to several weirs along the Zwalm river however, this relationship was not that obvious because fast flowing parts downstream of each weir occurred which could contain Asellidae accidentally drifting from the weir. In this way, flow velocity was not considered an important variable to predict the presence/absence of Asellidae in the Zwalm catchment. This finding is confirmed by the Bayerisches Landesamt für Wasserwirtschaft (1996), which states that Asellidae behave indifferently along a gradient of flow velocity. Based on the different contribution methods applied on both datasets, habitat characteristics seemed to be more important than the impact of physical and chemical variables for Asellidae. In this way, Asellidae only have potential as a good indicator organism in broader streams, because the habitat of the headwaters is less suited by nature for this taxon. In the end, these methods for testing the contributions of the different input variables facilitate the selection of the suitable habitats in which certain taxon can or cannot act as an indicator organism for the assessment and management of rivers.

Concerning their value as indicator taxa, 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 beside the habitat characteristics indicated by the different contribution methods, also the pollution related physical-chemical variables can be important to explain the habitat suitability of Asellidae. This underlines the need of relevant datasets for habitat preference studies. It also underlines that predictive ecological models developed with data driven techniques should be used with enough care for practical decision support in river restoration management as illustrated in Goethals *et al.* (2002).

For Tubificidae, ANN models were less reliable based on both datasets as could be derived from the predictive results in Chapter 6. Biased predictions were mainly caused by the high prevalence of Tubificidae in both datasets as was earlier mentioned. In this way, it was also very difficult to extract the major environmental variables based on the contribution methods and to make conclusions about environmental preferences of Tubificidae. Based on the 'Weights' method, contributions were equally distributed over the different input variables. On the basis of the 'PaD' method, some key variables could be recognized for the Zwalm river basin (nitrate, distance to mouth and total phosphorus) and the 'short distance' monitoring network (nitrate, sand and stream order). However, the standard deviation represented by the error bars, was relatively high for most of these variables. This indicated that the 'PaD' method was not very stable over the four folds, when the contributions were expressed as values. The ranking of the relative contribution of the variables on the other hand was stable over the different folds. Based on the 'Profile' method, no clear distinction between the different variables could be observed for Tubificidae. The effect of the major part of the environmental variables seemed almost nihil for the predictions (cf. the nearly horizontal curves).

Baetis is sensitive to various forms of pollution (Elliott *et al.*, 1988) and is in this way an indicator for moderate to good water quality (De Pauw and Vannevel, 1991). This was partly confirmed by field experience in the Zwalm river basin where *Baetis* was mainly found in the upstream parts of the streams. Therefore, it was little surprising that pollution related variables (e.g. ammonium and nitrate for the Zwalm river basin and conductivity, nitrate and dissolved oxygen for the 'short distance' monitoring network) were of major importance based on the contribution methods. However, also variables describing the physical habitat were important: for example suspended solids and temperature had a relatively high contribution based on the Zwalm river basin dataset while pools-riffles and depth, beside stream order, gave relatively high contributions to the prediction of the presence/absence of *Baetis* for the 'short distance' monitoring network. In this way, one might conclude that beside the pollution related variables, also the habitat characteristics can be important to explain the presence/absence of *Baetis*. However, the results based on the Zwalm river basin dataset has to be qualified since rather poor predictive results were obtained (Chapter 6).

Similar to Tubificidae, rather poor predictive performances were obtained for Limnephilidae, especially for the Zwalm river basin dataset (Chapter 6). Biased predictions for Limnephilidae were mainly caused by their low prevalence in both datasets. As could be derived from the data analyses in Chapter 5, Limnephilidae were only found in the upstream brooks with a good habitat and water quality. This was also reflected by the contribution methods. Based on the 'Profile' method, the probability of presence increased for both datasets if stream order decreased, indicating that the chance to predict Limnephilidae as present in streams with stream order 1 is higher. Based on Steenbergen (1993) however, Limnephilidae have a preference for wider and deeper streams. This could be explained by the fact that these streams do not have good water and habitat quality in the Zwalm. Based on the dataset of the 'short distance' monitoring network, pollution related input variables such as dissolved oxygen, total nitrogen and phosphate were distinguished by their higher contribution. Based on the 'Profile' method, an increase of dissolved oxygen induced an increase of the probability of presence while an increase of phosphate and total nitrogen led to a decrease. This is confirmed De Pauw and Vannevel (1991) and by Steenbergen (1993) who stated that Limnephilidae prefer low contents of nutrients and high oxygen concentrations.

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

7.5. Conclusions

The three input variable contribution methods (the 'Weights', 'PaD' and 'Profile' method) applied to Artificial Neural Network models can be useful to select essential variables for macroinvertebrate taxa. In this way, the choice of ecologically significant variables to describe the species' habitat(s) and to include these 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 into 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.

However, it was often difficult to find major trends for the five taxa, the three contribution methods and the different folds. The first two can be explained by different ecological preferences of the taxa and by the different aspects the three contribution methods deal with. The instability over the different folds is perhaps related to the relative small size of the datasets (respectively 237 and 120 instances) in combination with a high variability of the sites, the high number of input variables or outliers in the measurements. This will therefore need further research based on larger datasets and use of sub-samplings.

In spite of the above remarks, it could be concluded that 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 of river management.

Chapter 8
Development of migration models for macroinvertebrates as an extension of the habitat suitability models

---

<u>Parts of this chapter are based on</u>:

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

Dedecker, A.P., Van Melckebeke, K., Goethals, P.L.M. and De Pauw, N. (submitted). Development of migration models for macroinvertebrates in the Zwalm River basin as a tool in river assessment and restoration management. *Ecological Modelling*.

8.1. Introduction

As discussed in the previous chapters, Artificial Neural Networks (ANNs) have increasingly been used in aquatic sciences to analyze ecological patterns (e.g. Lek and Guégan, 1999; Maier and Dandy, 2000; Gevrey *et al.*, 2003; Dedecker *et al.*, 2004). These models are suited for dealing with ecological data that are known to be very complex and may vary and co-vary in non-linear fashions (Lek *et al.*, 1996b). On the other hand, the use of ANN models to predict habitat suitability is still considered as empirical (Guisan and Zimmermann, 2000). In general, empirical models make the assumption being 'perfectly mixed', where all habitats are freely and equally accessible. However, when accessibility is restricted, it is necessary to control for the effects of it before conclusions about preference and habitat suitability can be drawn because variations in spatial distribution of a species can no longer be attributed entirely to its preference (Matthiopoulos, 2003). For restoration actions for instance, ANN models can predict a positive effect on habitat suitability for certain species. However, it is possible that the restored river section cannot be recolonized through the fact that the distance from the nearest source population is unbridgeable or that migration barriers limit accessibility. These species can thus not reach the restored river although its habitat is improved and even suitable. Dedecker *et al.* (2005d) illustrated this problem. After water quality improvement, the caddisfly Limnephilidae was predicted present at the restored river sections based on the habitat suitability model. However, existing source populations were located more than five kilometres away from these potential new habitats. In this way, recolonization of the restored sites was very unlikely. Nevertheless, this could not be retrieved from the ANN model. For this reason, ANN models need to be optimized to result in more reliable simulations of river restoration scenarios. Specifically, spatial and temporal expert-rules could be included. Therefore, knowing the migration kinetics of downstream drift and upstream migration of a species and the presence of migration barriers along the river (weirs, culverted river sections, …) might deliver important additional information on the effectiveness of the restoration plans, and also on the timing of the expected effects. The development of migration models would enable to investigate the connectivity between population patches or the possibility to migrate from a source population to recolonize a restored river section.

Recently, in some studies 'least-cost' modelling was used as an approach to incorporate detailed geographical information as well as behavioural aspects to measure connectivity and migration (e.g. Michels et al., 2001; Schadt et al., 2002; Adriaensen et al., 2000, 2003; Chardon et al., 2003). This modelling tool is receiving growing attention in applied species-management projects as well as in research. Tool boxes based on the 'least-cost' algorithm are available in the most current GIS packages (e.g. ArcGIS, this study; Idrisi (Michels et al., 2001)) and in some specialized programmes (e.g. CONNEC (Gulinck et al., 1993)). The algorithm underlying this approach is similar to the method proposed by Knaapen et al. (1992). In their model, every unit (grid or raster cell) on a map was assigned a resistance value according to its facilitating/hindering effects on the considered migration process. This value was used to calculate the connectivity between a source cell and a target cell, by adding the values of all cells crossed.

The aim of this study was to develop migration models for the crustacean Gammaridae, the mayfly *Baetis*, and the caddisfly Limnephilidae in the Zwalm river basin. In Section 4.6, an overview of the factors affecting the migration behaviour of these taxa is given. For Tubificidae, no migration model was constructed. This was less appropriate since Tubificidae were already observed in 90 % of the sites. Also for Asellidae, no migration model was constructed since no information about active or passive (= drift) migration behaviour could be found. It can be assumed that the movements of Asellidae take place on a very local scale (microhabitat scale). The migration models for Gammaridae, *Baetis* and Limnephilidae were based on the Cost Weighted Distance function. In the next paragraphs, data processing and migration model development (e.g. functions and algorithms used) are discussed. In addition, the determination of the resistance values used to quantify the upstream and downstream migration possibilities through the river and the air (if an aerial live stage is available) is described for the three taxa. In the result and discussion section, the obtained migration models are presented and discussed.

8.2. Material and methods

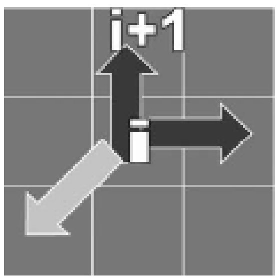
8.2.1. Data processing and migration model development

To allow for the development of the migration models a 'short distance' monitoring network was set up as introduced in Chapter 4. This type of model development required a more intensive monitoring approach. Therefore, the selected river parts (the brooks Verrebeek and Dorenbosbeek and the upstream part of the Zwalm river) were split up in stretches of 50 m, each marked with a code and an up and downstream (X,Y) co-ordinate. An inventory of the structural and morphological characteristics along the selected part of the Zwalm river basin was made. In each river stretch of 50 m, the dominant type of land use (wooded area, housings, industrial sites, arable or grazing land, …) was monitored as well as the occurrence of domestic, industrial or agricultural discharges, and the presence of buffer strips along the river (type and distance to the river), natural or artificial river banks and meanders, hollow river banks and pool-riffle patterns. Based on these characteristics, 60 sites were selected in the brooks Verrebeek, Dorenbosbeek and the upstream part of the Zwalm river. Each site represented a river section of about 250 m. At each site, macroinvertebrates and 24 environmental characteristics were collected as described in Chapter 4.

Geographically referred data for implementation of the watercourses in a GIS were obtained from the Flemish Land Agency (VLM) who digitalized the watercourses as linear elements in the Flemish Hydrological Atlas (MVG *et al.*, 2000). Landscape information on the Zwalm river basin was also available in GIS vector format (polygons for the land use) from the VLM (Fig. 8.2). However, two problems occurred. Firstly, the brooks Verrebeek and Dorenbosbeek have not been digitalized near the sources. Also a number of tributaries of these watercourses were not digitally available in the Flemish Hydrological Atlas. Based on the (X,Y) co-ordinates of the 50 m river stretches measured on-site, these files were completed (Fig. 8.2). Secondly, the digitalization of the watercourses in the Flemish Hydrological Atlas has been based on 'segments'. The starting and end points of these segments have been defined as the source or the mouth of a watercourse, the confluence of watercourses or where they split off, places where the quality objectives of the surface water change, … Because these segments were not small enough, it was impossible to link the data obtained from the inventory of the 50 m stretches, to the digital files. To solve this problem, the segments in the Flemish

Hydrological Atlas were divided in stretches of 50 m (Fig. 8.2). Based on the inventory of the visual characteristics during the first phase of the monitoring, the biological communities and the physical-chemical variables could be extrapolated between the sampling sites, because these were supposed to be representative of the stretches up and downstream. In this way, the values could be set at the same level.

The migration models were developed applying ArcGIS 8.3, a product of the Environmental Systems Research Institute (ESRI) (ESRI, 2001). One of the available extensions is ArGIS Spatial Analyst. The Cost Weighted Distance tool of this extension was used to develop the migration models. This function finds the least accumulative cost for migrating from each cell of the resistance layer to the nearest, cheapest source. The Cost Weighted Distance tool is based on the least cost algorithm. To move from cell N_i to cell N_{i+1}, the cumulative resistance is calculated as the resistance to reach cell N_i plus the average resistance to move through cell N_i and N_{i+1}. The function is based on an eight-neighbour-cell algorithm. As a result, also diagonal movements are allowed for. In case of diagonal directions, the cost is multiplied by $\sqrt{2}$ to compensate for the longer distance (Fig. 8.1).



$$N_{i+1} = N_i + \left(R_i + R_{i+1}\right)/2$$

$$N_{i+1} = N_i + \sqrt{2}\left(R_i + R_{i+1}\right)/2$$

Fig. 8.1. The algorithm underlying the Cost Weighted Distance function. i = source cell; i+1 = target cell; N_i = accumulated resistance to reach cell i; R_i = the resistance to migrate through cell i.

Before the Cost Weighted Distance function could be used, the vector files had to be converted into raster files (Fig. 8.2). The resolution of one raster or grid cell was set al 2 x 2 m. To use the Cost Weighted Distance function, two GIS layers were needed: a source and a resistance layer. The source layer indicates the source populations of the modelled organisms. The resistance layer indicates both the resistance value and the geographical position and orientation of all relevant landscape components. The resistance value of each cell was based on the determining variables affecting the migration of the organism (e.g. land use of the surrounding environment, flow velocity, migration barriers such as weirs, culverted river sections, …). If several criteria in different measurement systems were concerned, they had to be reclassified to a common scale since they could not be compared relatively to one another. In addition, the resistances were divided in a number of discrete classes. In this way, each cell of the resistance layer got a value representing the resistance for the organism to migrate through that cell.

The produced Cost Weighted Distance raster presents the least accumulated cost of getting from each cell to the nearest source population, but it does not provide the way how to get there. The direction raster however gives a map, identifying the route to take from any cell, along the least-cost path, back to the nearest source population. The algorithm for computing the direction raster assigns a code to each cell that identifies which one of its neighbouring cells is on the least-cost path back to the nearest source. Based on the information provided by the direction raster, a second tool, the shortest path, can be applied to compute and visualize the least-cost route from a chosen destination to the source population. In this study, both tools were combined to visualize the possibilities of the organisms to disperse from the source populations.

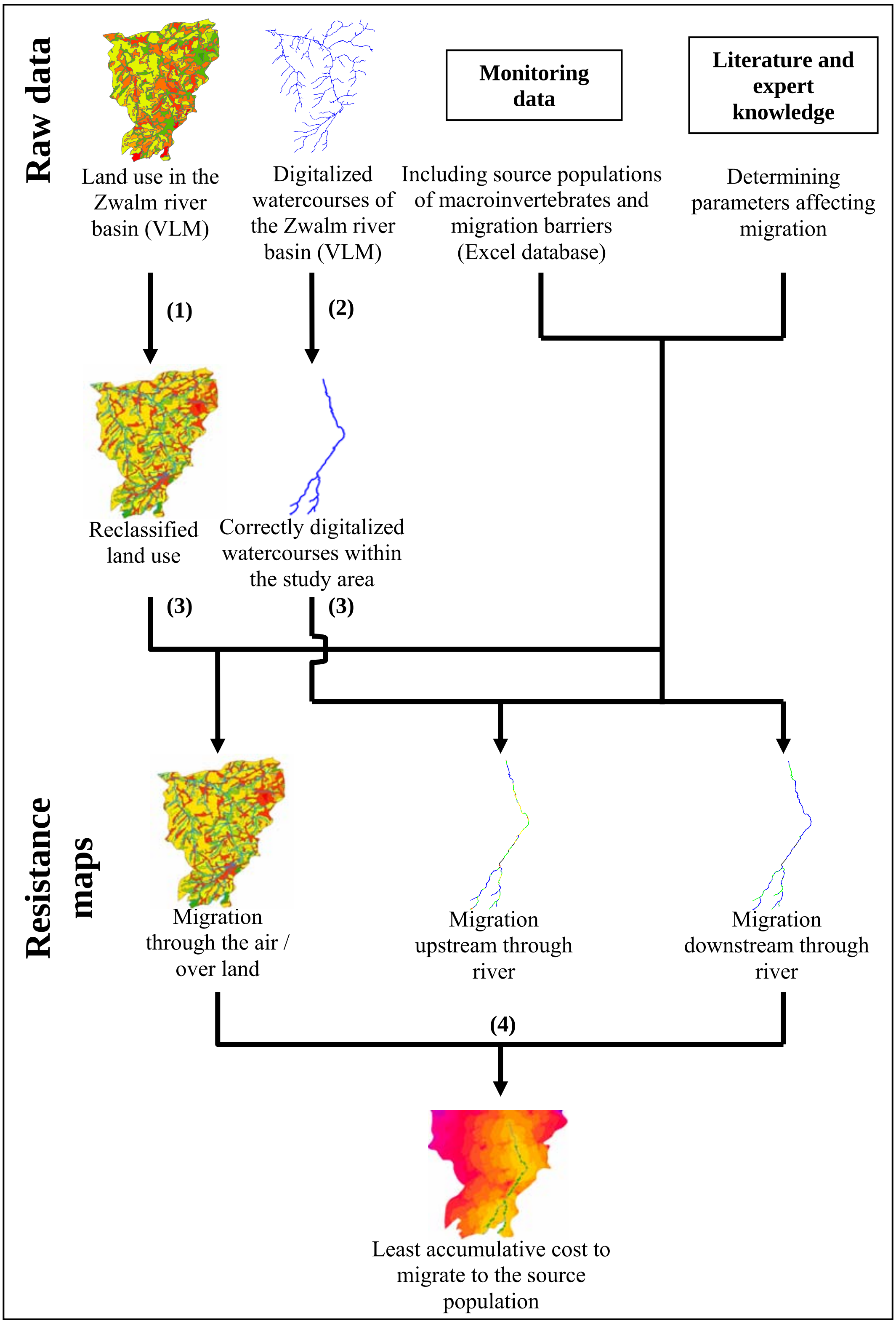


Fig. 8.2. Steps in the development of a migration model for macroinvertebrates in the Zwalm river basin. (1) = reclassify the land use; (2) = divide 'segments' in 50 m stretches and complete digitalized watercourse file; (3) = convert vector to raster file; (4) = perform the Cost Weighted Distance function applying the extension ArGIS Spatial Analyst of ArcGIS 8.3.

8.2.2. Determination of resistance values

Three resistance layers were used: one for migration through the air / over land and two for migration through the river (up and downstream). Since no published data were available on the resistance values of the different determining environmental features for macroinvertebrates, resistance values (= resistance/friction that a species experiences to travel through, along a certain landscape (element), river section, …, the resistance value is an integer) for migration through the air / over land ($R_{air/land}$) were assigned based on expert knowledge of the species. For *Baetis*, resistance values for different land use classes were chosen in such a way that resistance for movement through wooded area, meadow area, arable land or nature reserve ($R_{air/land}$ = 2) was higher than for watercourses and buffer strips (if present) ($R_{air/land}$ = 1), but clearly lower than for hindering landscape covers such as urban or industrial areas ($R_{air/land}$ = 20). $R_{air/land}$ for culverted river sections and weirs was set respectively at 10-20 and 2. The resistance values for Limnephilidae were based on the same values. However, studies have shown that 90 % of the adult organisms of Limnephilidae were caught within a distance of 20 m from the river (Petersen *et al.*, 1999; Winchester *et al.*, 2002). Based on this information, a strip of 20 m around the river was taken into account. Outside this strip, the resistance value was set 10 times higher. On the other hand, migration through the air/over land is not relevant for Gammaridae because they only have an aquatic life phase. An overview of $R_{air/land}$ for *Baetis* and Limnephilidae is given in Table 8.1.

Table 8.1. Determined resistance values for migration through the air / over land ($R_{air/land}$) for *Baetis* and Limnephilidae based on the surrounding environment

Surrounding environment	**$R_{air/land}$**	
	Baetis	**Limnephilidae**
• Water surface	1	1
• Buffer strip (if present)	1	1-10
• Land use		
- Urban region	20	20-200
- Industrial area	20	20-200
- Wooded area	2	2-20
- Meadows	2	2-20
- Arable land	2	2-20
- Nature reserve	2	2-20
• Culverted river section	10-20	10-20
• Weir	2	2

The setting of the resistance values for migration upstream (R_{up}) and downstream (R_{down}) through the river was also based on expert and literature knowledge. Elliott (2003) has shown that the active migration of Gammaridae and *Baetis* in one day is maximum 5.5 - 6.0 m in the upstream and only 1.5 m in the downstream direction. The further an organism can migrate, the lower the resistance has to be. Therefore, the inverse values of these distances were taken as measure for the resistance to migrate actively through the river ($R_{up(active)}$, $R_{down(active)}$). In this way, the physical meaning of the resistance value can be seen as the number of days to migrate over one meter. Based on this and taking into account the factors affecting the active migration (e.g. presence of boulders), a resistance value was attributed to each 50 m stretch (in the vector file which has to be converted into a raster file in a next step) for the active upstream and downstream migration of Gammaridae and *Baetis*.

The calculation of the passive downstream migration (= drift) resistance ($R_{down(passive)}$) for Gammaridae was based on the flow velocity (Elliott, 2002c):

$$\bar{x} = 7.82 \times V^{0.963} \tag{9.1}$$

where $\bar{x}$ = average drift distance downstream (m) and V = flow velocity (average flow velocity of 2003 was considered) (m s^{-1}). The passive downstream migration resistance for *Baetis* was also based on flow velocity (the average flow velocity of 2003 was considered) and the fact that the drift distance is divided by two if macrophytes are present (Elliott, 2002b):

$$\bar{x} = 8.97 \times V + 0.11 \tag{9.2}$$

Resistance values between 1 and 40 for Gammaridae and between 1 and 50 for *Baetis* were obtained after rescaling and classification of the average drift distance $\bar{x}$ (Table 8.2).

According to Elliott (2002b), Limnephilidae were seldomly found in the drifting water. In this way, it can be assumed that the passive downstream migration is of less importance than the active migration. Therefore, the resistance layer for downstream migration was only based on active movements. Erman (1986) and Elliott (2003) found that the active migration of Limnephilidae species is maximally 3.5 m during 24 hours in the upstream as well as the downstream direction. The inverse value of this distance was taken as measure for the resistance to migrate actively through the river ($R_{up(active)}$, $R_{down(active)}$). Based on this and taking into account the factors affecting the active migration (presence or absence of natural river banks and macrophytes), a resistance value was attributed to each 50 m stretch. After rescaling and classification, a resistance value between 3 and 11 was obtained for the active up and downstream migration (Table 8.2).

For Gammaridae, *Baetis* and Limnephilidae the resistance values $R_{up(active)}$ and $R_{down(active)}$ for the weir were set respectively at 200 and 100, while for the culverted river sections at 50. The passive downstream migration resistance for Gammaridae and *Baetis* for the weir and the culverted river section were respectively set at 100 and 30 (Table 8.2).

Table 8.2. Determined resistance values for upstream (R_{up}) and downstream (R_{down}) migration through the river for Gammaridae, *Baetis* and Limnephilidae (n.a. = not applicable, n.s. = not significant)

	Gammaridae			***Baetis***			**Limnephilidae**		
	$R_{up(active)}$	$R_{down(active)}$	$R_{down(passive)}$	$R_{up(active)}$	$R_{down(active)}$	$R_{down(passive)}$	$R_{up(active)}$	$R_{down(active)}$	$R_{down(passive)}$
• Boulders									
- present	n.a.	n.a.	n.a.	2	7	n.a.	n.a.	n.a.	n.s.
- absent	n.a.	n.a.	n.a.	4	13	n.a.	n.a.	n.a.	n.s.
• Flow velocity	2	7	1-40	n.a.	n.a.	1-50	n.a.	n.a.	n.s.
• Macrophytes									
- present	n.a.	n.a.	n.a.	n.a.	n.a.	1-50	3-6	3-6	n.s.
- absent	n.a.	n.a.	n.a.	n.a.	n.a.	1-50	6-11	6-11	n.s.
• Natural banks									
- present	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	3-6	3-6	n.s.
- absent	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	6-11	6-11	n.s.
• Culverted river section	50	50	30	50	50	30	50	50	n.s.
• Weir	200	100	100	200	100	100	200	100	n.s.

8.3. Results

8.3.1. Source populations

Sites were considered as source populations if at least two specimens of the modelled taxon were found in the samples and if the river habitat quality (water and physical habitat quality) could be judged as suitable. Gammaridae were found at all monitored sites in the brooks Verrebeek and Dorenbosbeek (except one in the brook Verrebeek). In the Zwalm river however, Gammaridae were not found one kilometer upstream of the weir. *Baetis* was observed in the brooks Verrebeek and Dorenbosbeek and in the upstream part of the Zwalm river (Table 8.3). On the other hand, Limnephilidae were only found in the brooks Verrebeek and Dorenbosbeek (Table 8.3). In the brook Verrebeek, the latter organisms were dispersed along the whole river, while in the brook Dorenbosbeek source populations were located not further then 1.5 km downstream of that river source.

Table 8.3. Number of sampling sites and source populations of Gammaridae, *Baetis* and Limnephilidae for the brooks Verrebeek and Dorenbosbeek and the Zwalm river

River	**# of sampling sites**	**# of source populations**		
		Gammaridae	***Baetis***	**Limnephilidae**
Zwalm river	29	20	14	0
Brook Verrebeek	15	14	9	9
Brook Dorenbosbeek	16	16	7	4

Based on the source layers, including respectively the source populations of Gammaridae, *Baetis* and Limnephilidae, and the three resistance layers, the Cost Weighted Distance function could be used as a basis to set up migration models for the three studied taxa.

8.3.2. Migration model for Gammaridae

Migration costs for Gammaridae were rather low since they are widespread over the three watercourses as shown in Fig. 8.3. The maximum accumulative cost for downstream migration was 5045 units and was located at the culverted river section near Brakel city (Fig. 8.4a). To reach the river stretches upstream of the weir, a cost of 4300 units had to be realized by the Gammaridae. For upstream migration through the river, a maximum accumulative cost of 13000 units was obtained to reach the culverted river section near Brakel city (Fig. 8.4b).

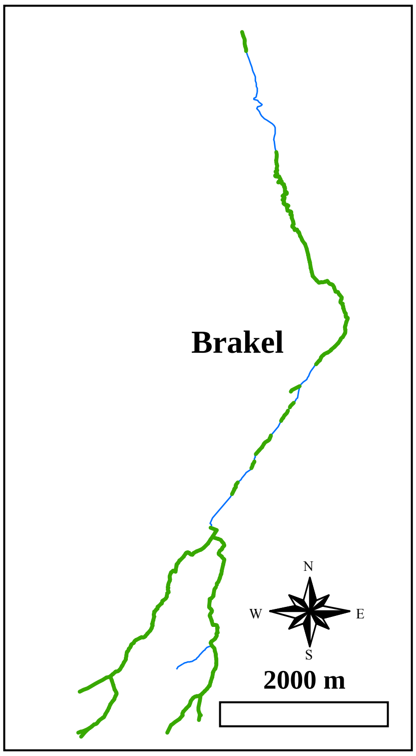


Fig. 8.3. Illustration of the source populations (indicated in green) of Gammaridae in the 'short distance' monitoring network.

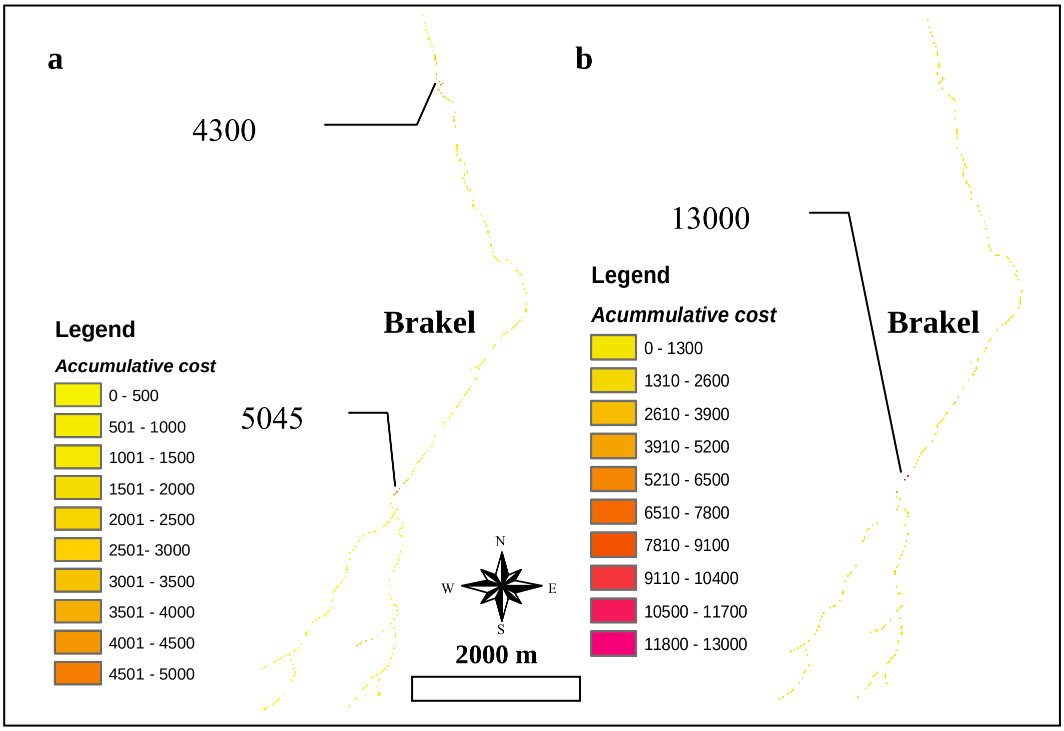


Fig. 8.4. Map of the accumulative cost for migration downstream (a) and upstream (b) through the river from the source populations of Gammaridae. The highest accumulative cost(s) are indicated (see figures and legends).

8.3.3. Migration model for *Baetis*

The maximum accumulative cost for migration through the air was more than 20000 units. However, this point was of minor importance in the present study. When only migration from a source population to a point in the river was taken into account, the accumulative cost was limited. The highest accumulative cost for migration to a point in the river (2390 units) was reached in the most downstream point of the study area (Fig. 8.5). In the centre of this figure, the influence of Brakel city on the accumulative cost value could be observed.

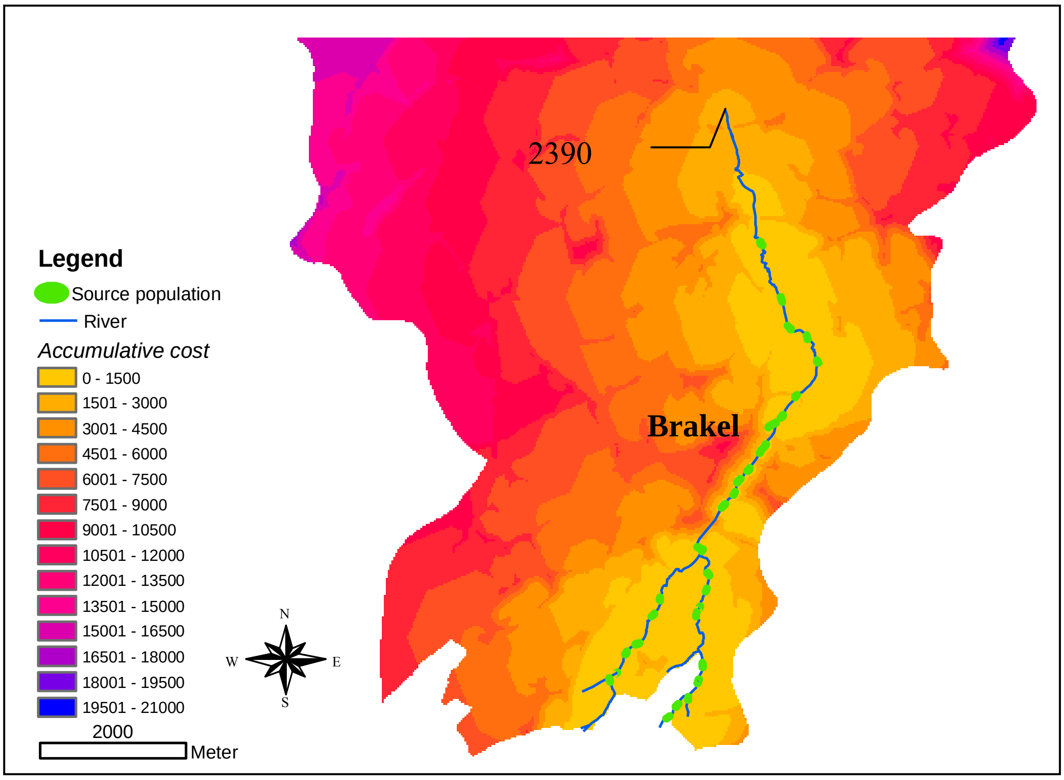


Fig. 8.5. Map of the accumulative cost for migration through the air from the source populations of *Baetis*. The highest accumulative cost along the river is indicated (see figure and legend).

The highest accumulative cost for downstream migration (9920 units) was located at the most downstream point (Fig. 8.6a). The culverted river section near Brakel city obtained also a high cost. The maximum accumulative cost for upstream migration (13300 units) was reached at this culverted river section (Fig. 8.6b).

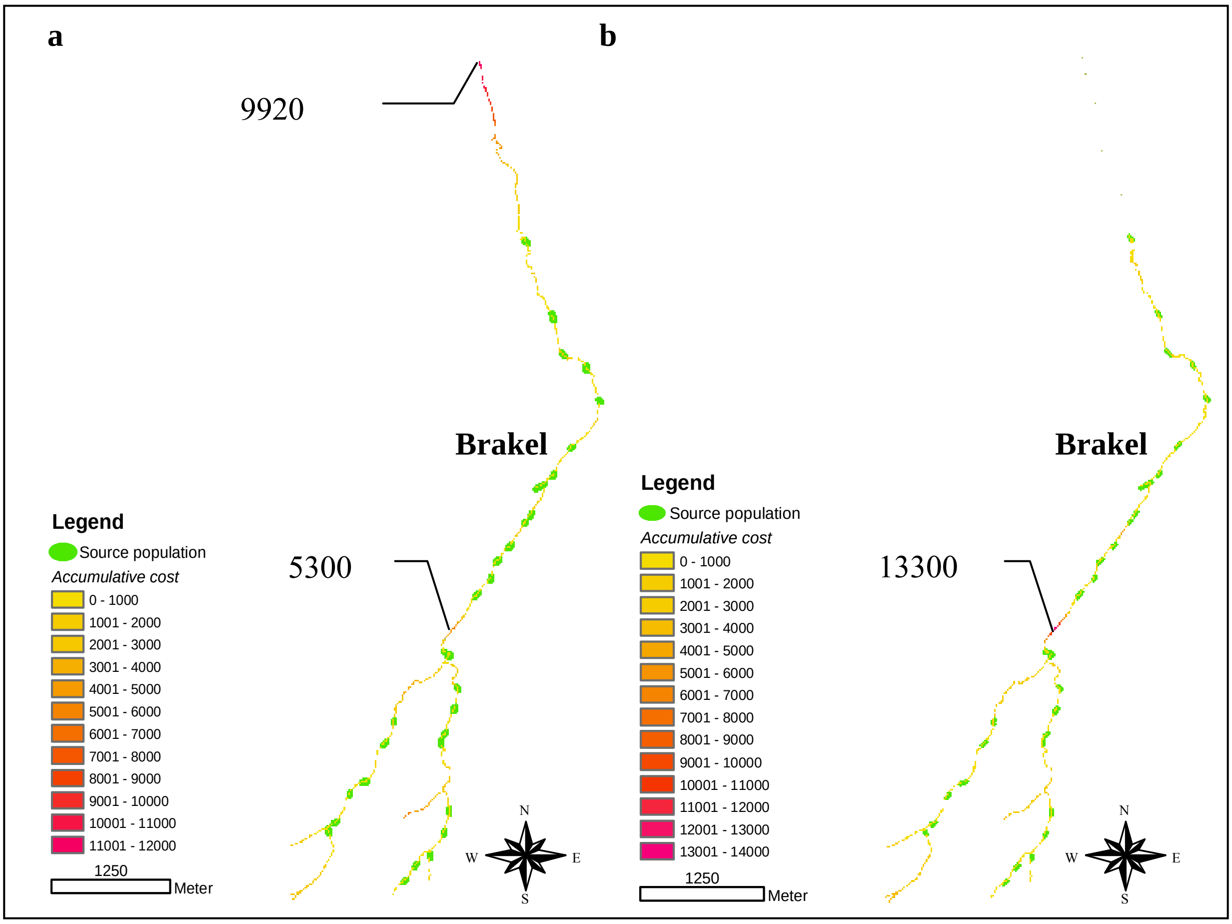


Fig. 8.6. Map of the accumulative cost for migration downstream (a) and upstream (b) through the river from the source populations of *Baetis*. The highest accumulative cost(s) are indicated (see figures and legends).

8.3.4. Migration model for Limnephilidae

For Limnephilidae, higher accumulative cost values for migration through the air were obtained in comparison with *Baetis* because a higher resistance value was attributed to points outside a strip of 20 m along the river. The maximum accumulative cost for migration through the air was about a factor ten higher than the maximum accumulative cost for *Baetis*. When only points in the vicinity of the river were taken into account, the highest accumulative cost of 115420 units was detected for the most downstream point of the Zwalm river (Fig. 8.7).

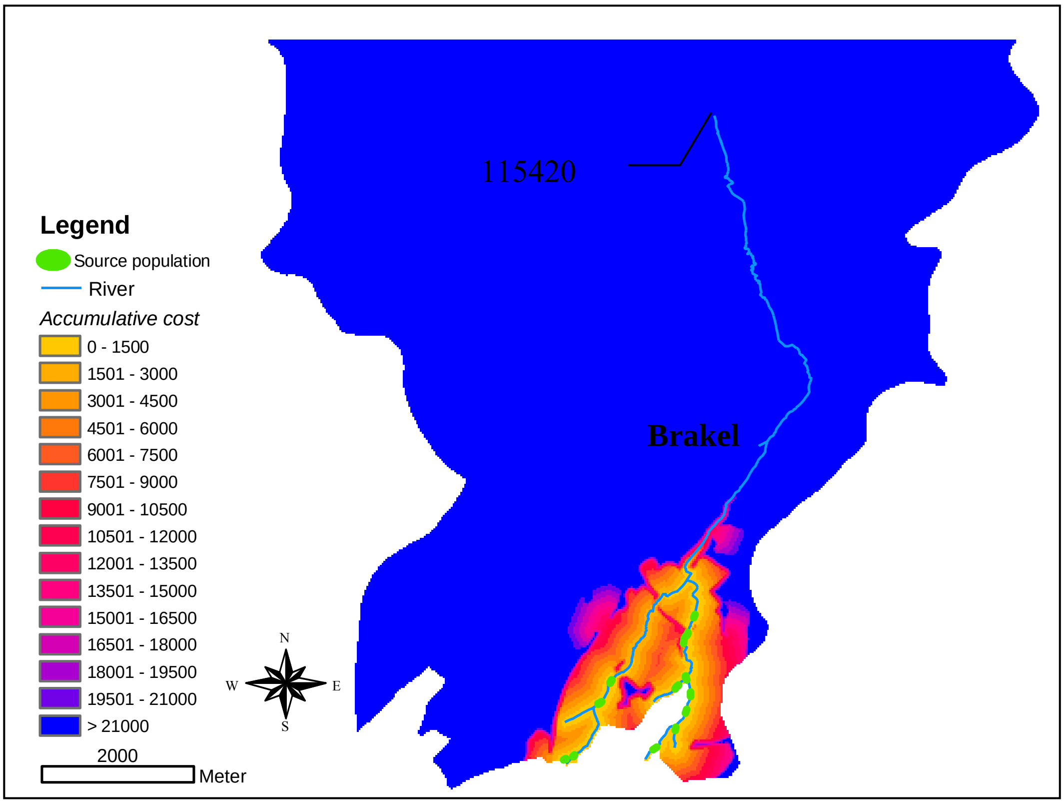


Fig. 8.7. Map of the accumulative cost for migration through the air from the source populations of Limnephilidae. The highest accumulative cost to reach a point along the river is indicated (see figure and legend).

Because the resistance values for up and downstream migration were equal (except for the weir) (cf. Table 8.2), the same map could be used for both accumulative costs. The maximum accumulative cost (101530 units) was obtained for the most downstream point of the Zwalm river (Fig. 8.8). Also the strong increase of the accumulative cost to migrate through the culverted river section could be observed. However, both accumulative costs could only be applied for the downstream migration because downstream of these points no source populations of Limnephilidae were found. In this way, upstream migration is out of the question at these points. For upstream migration, the highest accumulative cost (3685 units) was reached in a tributary of the brook Dorenbosbeek (Fig. 8.8).

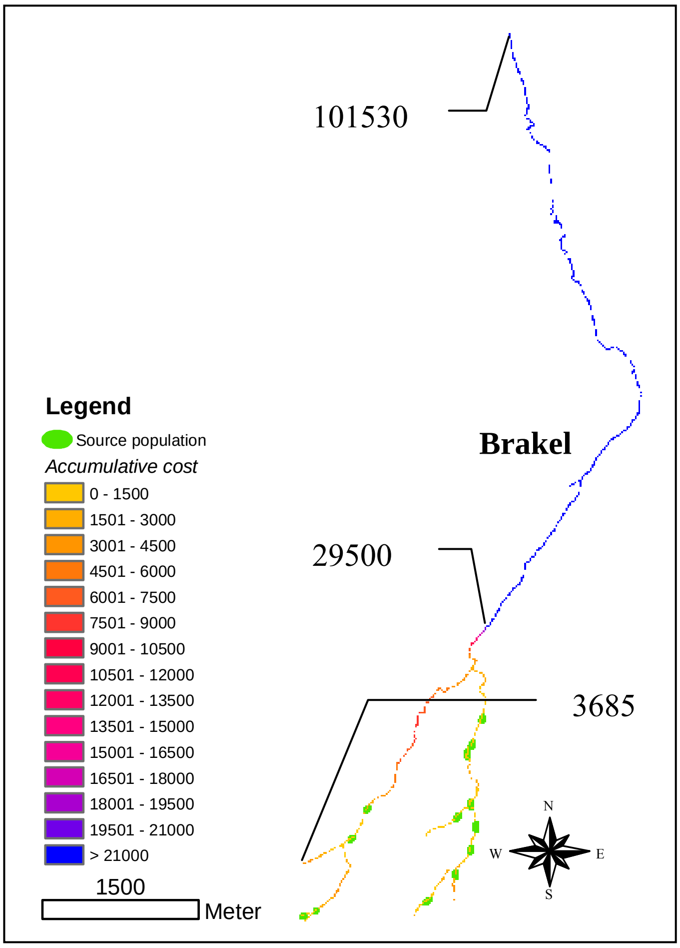


Fig. 8.8. Map of the accumulative cost for migration downstream and upstream through the river from the source populations of Limnephilidae. The highest accumulative costs are indicated (see figures and legend).

8.4. Discussion

To model the migration and recolonization possibilities of Gammaridae, *Baetis* and Limnephilidae in the Zwalm river basin the Cost Weighted Distance function was applied. This cost-distance function is an interesting and widespread tool to model movement behaviour in ecology and is proven to perform better than the Euclidean distance function (e.g. Ferreras, 2001; Chardon *et al.*, 2003; Verbeylen *et al.*, 2003). The latter function is commonly used in spatial population studies to model migration and connectivity between habitats but it is only a simple measure for the shortest distance between a suitable habitat, called a 'patch' in spatial ecology, and its nearest neighbour (Moilanen and Hanski, 2001).

The role of the environment on the connectivity of suitable habitat and the species specific migration behaviour through the environment is thereby ignored (Tischendorf and Fahrig, 2000). However, the importance of the environment for migration is increasingly being acknowledged. The importance of this topic has been illustrated in detail for terrestrial ecosystems (mammals: e.g. Bowne *et al.*, 1999; Ferreras, 2001; birds: e.g. Brooker *et al.*, 1999; butterflies; e.g. Hadded, 1999; Conradt *et al.*, 2001). Recently, its usefulness and suitability is also demonstrated for aquatic ecosystems (Michels *et al.*, 2001).

Some aspects of the method have to be emphasized before using the Cost Weighted Distance function. In the first place, the quality of the maps is highly important (Adriaensen *et al.*, 2003). All GIS packages use grid maps as input for the least cost model. This implies that a lot of GIS information available in vector format (e.g. digitalized watercourses, land use, …) has to be converted to grids before the model can be applied. Since the grid map is the only input, its quality is decisive for the quality and reliability of the resulting cost map. This has important consequences for a few aspects of the map. Relatively low resolution (large grid cells) may be used for general land cover since parcels mostly have larger dimensions. However, resolution is crucial for smaller or narrower elements in the landscape (e.g. watercourses). In this study, the optimum resolution of one grid cell was decided at 2 x 2 m. In this way, the grid size was small enough to capture the required detail, but large enough so that computer storage and analysis could be performed efficiently.

Secondly, the setting of resistance values in the resistance layer is biologically probably the most important step in the process of calculating the cost for migration (Adriaensen *et al.*, 2003). It is the link between the (non-ecological) GIS information and the ecological-behavioural aspects of the mobility of the species. Especially, when the purpose is to construct a predictive model for conservation or restoration management, it is critical that the model (and thus resistance and connectivity measures) is rigorously parameterized using empirical data and validated in independent landscapes (Moilanen and Hanski, 2001; Schadt *et al.*, 2002). For most organisms however, setting the resistance values will be a difficult process in which expert judgement and data available in literature will play an important role (see e.g. Schadt *et al.*, 2002). Also in this study, resistance values for Gammaridae, *Baetis* and Limnephilidae were mainly based on expert and literature knowledge. Only for the estimation of the resistance for passive downstream migration (= drift) of Gammaridae and *Baetis*, actual field data of flow velocity could be used respectively based on Elliott (2002c) and Elliott

(2002b). To have a better idea about the effect of the chosen resistance values for the different determining parameters on the accumulative costs, a sensitivity analysis was performed in the next chapter. Therefore, resistance values ascribed to the different environmental and river characteristics determining the migration were modified. Although in this way, a theoretical validation of the resistance values was possible, a more in depth validation of the migration models is recommended in the future. To this end, the use of artificial rivers could be helpful. On the other hand, analysing the genetic diversity or marking and tracing the macroinvertebrates could help to identify their real migration potentials and colonization routes. However, these experiments are often expensive and not always that realistic and reliable. Validation of the predictive results in the field is therefore indispensable. This field validation could be based on nets placed along the river or visual observations of migrating adults. Migration through the water could be observed based on nets placed in the water to catch drifting or actively migrating macroinvertebrates.

Although the resistance values for migration through the air/over land were very similar for *Baetis* and Limnephilidae (Table 8.1), the resulting maps of the accumulative costs were quite different. The main reason is the location of the source populations. Source populations of *Baetis* are spread over the whole study area while Limnephilidae were only found in the brooks Verrebeek and Dorenbosbeek (Table 8.3). In this way, the highest cost for migration through the air was a few orders lower for *Baetis* than for Limnephilidae. Similar conclusions could be drawn for migration through the river. Based on these high accumulative cost values, it is very unlikely that Limnephilidae will colonized the most downstream river sections very fast. Therefore, to simplify recolonization of the most downstream sections or migration in general, it is crucial that also the intervening watercourses are of good quality.

Although Gammaridae had no aerial adult phase, the migration cost remained rather low since they are very widespread over the study area. However, it is more likely that the migration behaviour is significantly more impacted by the design of weirs or culverted river sections. During upstream movement, these barriers could limit available habitat (Rawer-Jost *et al.*, 1999; Vaughan, 2002). Especially the small movements along the substrate needed to prevent extinction of the upstream reaches as discussed in the 'drift paradox' (Section 3.1.5) can be limited by these migration barriers since no compensation flight is possible for Gammaridae. Minckley (1964) reported that small dams prevented the upstream migration of Gammaridae in England. Also the Zwalm river basin is impacted by six weirs for water quantity control and several culverted river sections (Goethals and De Pauw, 2001). To improve the upstream migration in these situations, by-passes could be installed. Their effects on benthic invertebrates have been studied by Rawer-Jost *et al.* (1999), who investigated whether benthic invertebrates successfully used two types of by-passes, a boulder ramp and a concrete by-pass, for upstream movements. Their results indicated that the boulder ramp allowed for the upstream migrations, whereas the concrete by-pass was more difficult to ascend. Currently, only one weir in the Zwalm river basin is provided with a concrete by-pass, but the flow velocity and height of the steps are probably too high to allow the counter current migration of gammarids. In this way, weirs could be considered as a serious migration barrier in the Zwalm river. Also culverted river sections can be a serious obstacle for migrating Gammaridae. Especially culverts which have an outflow above the downstream water level, limit upstream passage.

8.5. Conclusions

The developed migration models using the cost-distance function have a whole range of applications. Not only the migration possibilities of observed macroinvertebrates within the study area could be modelled. Also the extension of the migration models to other species including nearly extinct as well as invasive exotic species could be of major importance in river restoration management. As such, the effects of certain interventions (e.g. weir removal or remeandering projects) in view of river management planning can be evaluated in a more reliable and integrated way compared to the local habitat suitability models. Similar strategies need to be developed for fish as well, seen the importance of migrating species in the assessment based on the Index of Biotic Integrity (IBI) for upstream brooks in Flanders. In addition, the scale of the developed models will have to be extended to the whole Zwalm river basin. Extension of the intensive monitoring campaign as done for the selected part of the Zwalm river basin, would however be very costly and time consuming. Therefore, using aerial photographs and remote sensing techniques in combination with digital maps to extract the necessary information would be recommendable. Finally, the combined use of migration and habitat suitability models would allow river managers to make a more rational selection among different restoration scenarios. In this way, they would be able to find out in advance whether and when a restoration option would have a desired effect or not.

Besides these opportunities, validation of the migration models is recommended. To this end, a first theoretical validation, based on sensitivity analysis, is performed in Chapter 9. Therefore, resistance values ascribed to the different environmental and river characteristics determining the migration are modified. However, further practical validation (e.g. based on field measurements) is advised in the future.

Chapter 9
Sensitivity analysis of the migration models

9.1. Introduction

The occurrence of a species in a habitat may depend on different factors, like habitat suitability (e.g. appropriate flow velocity, sufficient dissolved oxygen concentration) and habitat accessibility and isolation (e.g. Tischendorf and Fahrig, 2000; Verbeylen *et al.*, 2003) as mentioned in the previous chapter. Habitat accessibility and isolation is determined by the connectivity between suitable habitats, i.e. the degree to which the environment facilitates or impedes movement of organisms among these habitats (Verbeylen *et al.*, 2003). This depends not only on the specific mobility of the organism (functional connectivity, Tischendorf and Fahrig, 2000), but also on characteristics of the surrounding environment (structural connectivity), such as available habitat, distance to other suitable habitats, migration corridors or barriers, …

Previous studies have rarely considered the environment in determining accessibility of suitable aquatic habitats for a certain species. Most often, accessibility or isolation measures are solely based on distance (e.g. distance to the nearest source population). In these measures the implicit assumption is made that the environment is homogeneous with respect to animal movement and migration (Verbeylen *et al.*, 2003). Rarely migration barriers like weir and culverted river sections or elements that facilitate migration like presence of buffer strips along the river are taken into account. However, environments are seldom homogeneous. Migration barriers may hinder and change the migration path and unsuitable habitats can influence the migration and recolonization of a certain species. Due to habitat preferences, species seldom move randomly. Thanks to the new possibilities offered by GIS-systems, complexity of the environment can be included in the accessibility measure when modelling migration (e.g. Ferreras, 2001; Chardon *et al.*, 2003; Verbeylen *et al.*, 2003). Verbeylen *et al.* (2003) for example, investigated the effect of varying the resistance values on the migration cost of Red squirrel in an urban landscape to optimize their migration model performance.

The aim of the present study was to test the impact of the resistance values on the calculated migration cost of Gammaridae, *Baetis* and Limnephilidae based on the Cost Weighted Distance function (see Chapter 8). Resistance values ascribed to the different determining variables were modified. In this way, a sensitivity analysis of the migration models based on varying resistance values could be made. This sensitivity analysis can be seen as a (theoretical) validation step of the developed migration models in Chapter 8.

9.2. Material and methods

The migration models used to perform the sensitivity analyses were based on data from the 'short distance' monitoring network containing the brooks Verrebeek and Dorenbosbeek and the upstream part of the Zwalm river. As described in Chapter 8, the migration models were developed applying the Cost Weighted Distance tool, a function which is included in the Spatial Analyst extension of ArcGIS 8.3. This function finds the least accumulative cost for migration from each cell of the resistance layer to the nearest, cheapest source. Details about the Cost Weighted Distance function are provided in Chapter 8. To perform the sensitivity analyses, the resolution of one raster cell was decided at 2 x 2 m. In this way, the grid size was small enough to capture the required details, but large enough so that computer storage and analysis could be performed efficiently.

Sensitivity analyses were done for migration through the air / over land as well as for migration through the river. The performed simulations in this study are summarized below:

- 21 simulations (21 resistance sets) **for the migration of *Baetis* and Limnephilidae through the air / over land**. Therefore, the resistance values for the determining characteristics watercourse, buffer strip, land use class (urban region, industrial area, wooded area, meadows, arable land and nature reserve), weir and culverted river section were altered;
- 21 simulations (21 resistance sets) **for the downstream migration of Gammaridae and *Baetis* through the river**. Therefore, seven flow velocity values were taken into account (the minimum value, the 25 percentile value, the median value, the 75 percentile value, the maximum value, the average value of 2002 and the average value of 2003). In addition the resistance values for the weir and the culverted river section were modified. For **Limnephilidae**, 3 simulations (3 resistance sets) for the downstream migration through the river were performed, only changing the relevant characteristics weir and culverted river section;

- 3 simulations (3 resistance sets) **for the upstream migration of Gammaridae, *Baetis* and Limnephilidae through the river**. Therefore, the resistance values of the determining characteristics weir and culverted river section were altered.

Since no specific published data are available on the resistance values of the different determining surrounding environmental features and river habitat characteristics for the macroinvertebrates modelled (Gammaridae, *Baetis* and Limnephilidae), values were assigned based on expert knowledge of the species. Varying the resistance values, a sensitivity analysis of the developed migration models was performed within the Zwalm river basin. For *Baetis*, the following resistance sets were used for migration through the air / over land, taking into account that the land use classes wooded area, meadows, arable land and nature reserve and the weir element never have a higher resistance value than the land use classes urban region and industrial area and culverted river sections (Table 9.1):

- R1: all river and surrounding environmental elements have a resistance of 1, so only distance is taken into account and the effects of land use and river characteristics are eliminated;
- R2 – R3: resistance sets with 2 classes, distinguishing watercourses and buffer strips (R = 1), from the land use classes wooded area, meadows, arable land and nature reserve and the weir (with different contrasts in resistance values: R = 1, 2) and from the land use classes urban region and industrial area and culverted river sections (R = 2);
- R4 – R6: resistance set with 2 or 3 classes, distinguishing watercourses and buffer strips (R = 1), from the land use classes wooded area, meadows, arable land and nature reserve and the weir (with different contrasts in resistance values: R = 1, 2, 5) and from the classes urban region and industrial area and culverted river sections (R = 5);
- R7 – R10: resistance sets with 2 or 3 classes, distinguishing watercourses and buffer strips (R = 1), from the land use classes wooded area, meadows, arable land and nature reserve and the weir (with different contrasts in resistance values: R = 1, 2, 5, 10) and from the land use classes urban region and industrial area and culverted river sections (R = 10);
- R11 – R15: resistance sets with 2 or 3 classes, distinguishing watercourses and buffer strips (R = 1), from the land use classes wooded area, meadows, arable land and nature reserve and the weir (with different contrasts in resistance values: R = 1, 2, 5, 10, 20) and

from the land use classes urban region and industrial area and culverted river sections (R = 20);

- R16 – R21: resistance sets with 2 or 3 classes, distinguishing watercourses and buffer strips (R = 1), from the land use classes wooded area, meadows, arable land and nature reserve and the weir (with different contrasts in resistance values: R = 1, 2, 5, 10, 20, 25) and from the land use classes urban region and industrial area and culverted river sections (R = 25);

The resistance values for Limnephilidae were based on the same values. However, outside a strip of 20 m around the river, the resistance values were set 10 times higher as discussed in Section 8.2.2 (Petersen *et al.*, 1999; Winchester *et al.*, 2002).

Based on the land use map, the digitalized watercourses and these 21 resistance sets, 21 resistance maps were produced. By the Cost Weighted Distance function, the combinations of source population and resistance maps were converted to maps indicating the migration costs (CostR1 – CostR21).

Table 9.1. The 21 resistance sets with the assigned resistance values for migration through the air / over land for *Baetis* and Limnephilidae (outside a strip of 20 m around the river, the resistance values were set 10 times higher for Limnephilidae)

Resistance set	**Water surface**	**Buffer strip**	**Land use**						**Culverted river section**	**Weir**
			Urban region	**Industrial area**	**Wooded area**	**Meadows**	**Arable land**	**Nature reserve**		
R1	1	1	1	1	1	1	1	1	1	1
R2	1	1	2	2	1	1	1	1	2	1
R3	1	1	2	2	2	2	2	2	2	2
R4	1	1	5	5	1	1	1	1	5	1
R5	1	1	5	5	2	2	2	2	5	2
R6	1	1	5	5	5	5	5	5	5	5
R7	1	1	10	10	1	1	1	1	10	1
R8	1	1	10	10	2	2	2	2	10	2
R9	1	1	10	10	5	5	5	5	10	5
R10	1	1	10	10	10	10	10	10	10	10
R11	1	1	20	20	1	1	1	1	20	1
R12	1	1	20	20	2	2	2	2	20	2
R13	1	1	20	20	5	5	5	5	20	5
R14	1	1	20	20	10	10	10	10	20	10
R15	1	1	20	20	20	20	20	20	20	20
R16	1	1	25	25	1	1	1	1	25	1
R17	1	1	25	25	2	2	2	2	25	2
R18	1	1	25	25	5	5	5	5	25	5
R19	1	1	25	25	10	10	10	10	25	10
R20	1	1	25	25	20	20	20	20	25	20
R21	1	1	25	25	25	25	25	25	25	25

As discussed in Chapter 8, the resistance values for migration of Gammaridae, *Baetis* and Limnephilidae through the river were based on flow velocity, several determining river characteristics and the presence of migration barriers such as culverted river sections and weirs. During the sensitivity analysis, the flow velocity was varied between its minimum and maximum value over the two sampling periods 2002 and 2003. Seven flow velocity values were considered for each river section: the minimum value, the 25 percentile value, the median value, the 75 percentile value, the maximum value, the average value of 2002 and the average value of 2003. Also the resistance values of the weir and the culverted river sections were altered. On the other hand, the impact of the additional determining river characteristics (e.g. the presence of boulders and macrophytes for *Baetis* and the presence of macrophytes and natural banks for Limnephilidae) were supposed to be fixed since they were described as such in literature. For these characteristics, the same effects were taken into account as described in Section 8.2.2. Also for the active upstream and downstream migration through the river, the same values were assumed as described in Section 8.2.2. As a result, the resistance sets for migration through the river were defined as specified in Table 9.2 (Gammaridae and *Baetis*) and 9.3 (Limnephilidae).

Based on the digitalized watercourses and these resistance sets, resistance maps were produced. By the Cost Weighted Distance function, the combinations of source population and resistance maps were converted to maps indicating the costs for migration through the river (CostDownR and CostUpR).

Table 9.2. The resistance sets with the assigned resistance values for downstream (RDown) and upstream (RUp) migration through the river for Gammaridae and *Baetis*

Resistance set	Flow velocity	Culverted river section	Weir
RDown1	minimum value	30(1) or 50(2)	100(3) or 200(4)
RDown2	25 percentile value	30(1) or 50(2)	100(3) or 200(4)
RDown3	median value	30(1) or 50(2)	100(3) or 200(4)
RDown4	75 percentile value	30(1) or 50(2)	100(3) or 200(4)
RDown5	maximum value	30(1) or 50(2)	100(3) or 200(4)
RDown6	average value of 2002	30(1) or 50(2)	100(3) or 200(4)
RDown7	average value of 2003	30(1) or 50(2)	100(3) or 200(4)
RDown8	minimum value	45(1) or 75(2)	150(3) or 300(4)
RDown9	25 percentile value	45(1) or 75(2)	150(3) or 300(4)
RDown10	median value	45(1) or 75(2)	150(3) or 300(4)
RDown11	75 percentile value	45(1) or 75(2)	150(3) or 300(4)
RDown12	maximum value	45(1) or 75(2)	150(3) or 300(4)
RDown13	average value of 2002	45(1) or 75(2)	150(3) or 300(4)
RDown14	average value of 2003	45(1) or 75(2)	150(3) or 300(4)
RDown15	minimum value	60(1) or 100(2)	200(3) or 400(4)
RDown16	25 percentile value	60(1) or 100(2)	200(3) or 400(4)
RDown17	median value	60(1) or 100(2)	200(3) or 400(4)
RDown18	75 percentile value	60(1) or 100(2)	200(3) or 400(4)
RDown19	maximum value	60(1) or 100(2)	200(3) or 400(4)
RDown20	average value of 2002	60(1) or 100(2)	200(3) or 400(4)
RDown21	average value of 2003	60(1) or 100(2)	200(3) or 400(4)
RUp1	/	50	200
RUp2	/	75	300
RUp3	/	100	400

(1) for $R_{down(passive)}$; (2) for $R_{down(active)}$ and $R_{up(active)}$; (3) for $R_{down(active)}$ and $R_{down(passive)}$; (4) for $R_{up(active)}$.

Table 9.3. The resistance sets with the assigned resistance values for downstream and upstream migration (RDown&Up) through the river for Limnephilidae

Resistance set	Culverted river section	Weir
RDown&Up1	50	200
RDown&Up2	75	300
RDown&Up3	100	400

9.3. Results

9.3.1. Sensitivity analysis for migration through the air / over land

Characteristics of the surrounding environment are more obvious if resistance values of land use classes and migration barriers are more diverse. If a more clear distinction is made between the different land use classes, more details can be extracted from de migration models. This is illustrated in Fig. 9.1, on which the maps of the accumulative costs for migration of *Baetis* through the air / over land are presented for the resistance sets R11 – R14 (CostR11 – CostR14). On Fig. 9.1a and 9.1b (CostR11 and CostR12 where the land use class urban region has a resistance value of 20 and the land use classes wooded area, meadows, arable land and nature reserve respectively 1 and 2) for example, the location of Brakel city can clearly be observed while on Fig. 9.1d (CostR14 where the land use class urban region has a resistance value of 20 and the land use classes wooded area, meadows, arable land and nature reserve a resistance value of 10), the city can only hardly be retrieved.

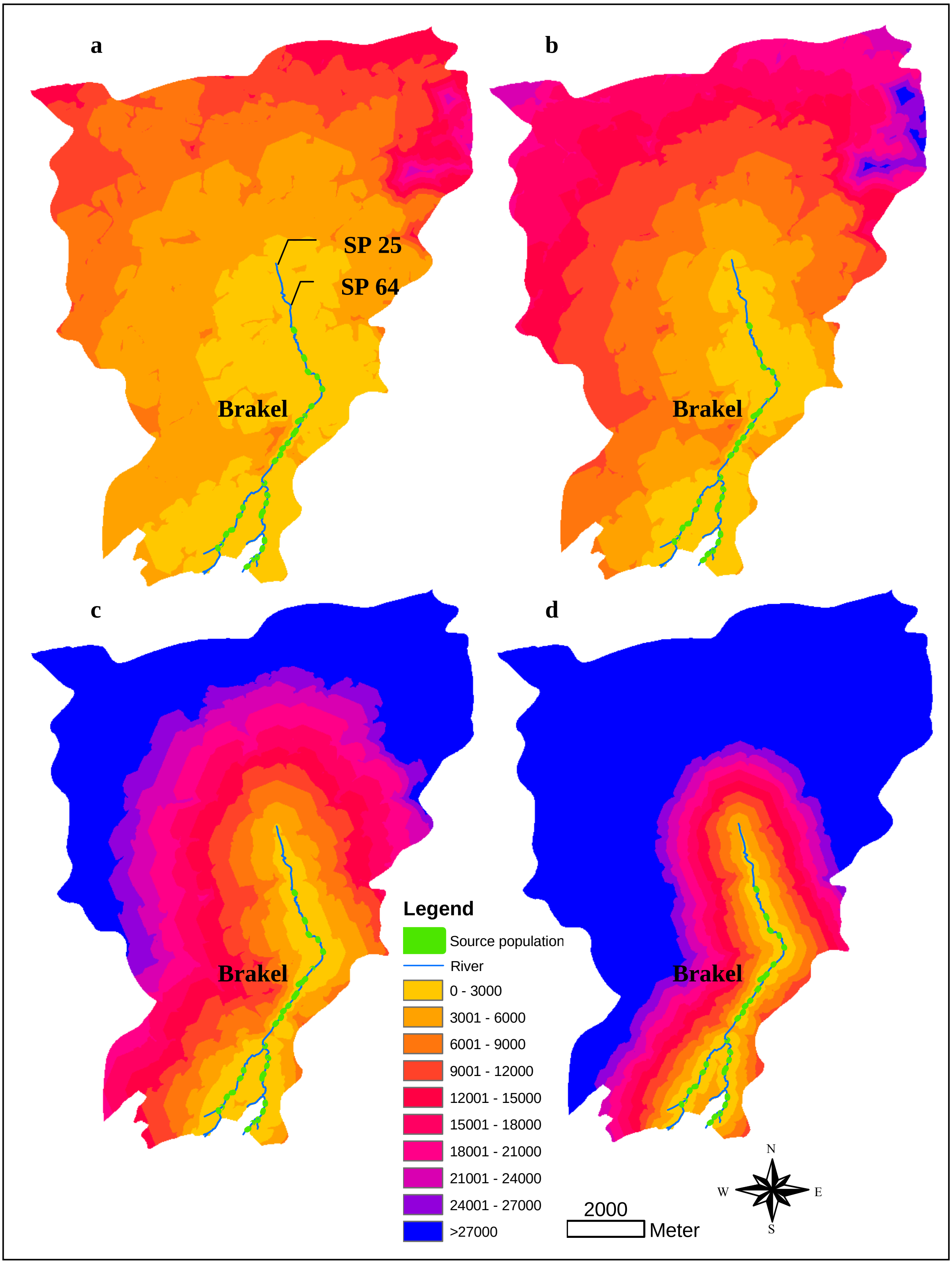


Fig. 9.1. Maps of the accumulative costs (CostR11 – CostR14) for migration through the air / over land from the source populations of *Baetis* based on the resistance sets R11 (a), R12 (b), R13 (c) and R14 (d) of *Baetis*.

Based on the sensitivity analysis, similar results were obtained for Limnephilidae. The location of Brakel city, where houses are concentrated, can easily be observed, as shown in Fig. 9.2 (CostR11 – CostR14). Downstream of the city centre however, accumulative costs increased rapidly. The latter is more obvious than for *Baetis* since no more source populations of Limnephilidae were found downstream of this town. Also for Limnephilidae, more details could be retrieved from the cost maps if the resistance values of the land use classes were more diverse (e.g. CostR12). The landscape became more homogeneous if similar resistance values were used for the different land use classes.

Table 9.3 provides more detail about the accumulative cost values for some specific sampling points. The location of these sampling points are shown on Fig. 9.1a (SP 25 is located at the most downstream point while SP 64 is situated between SP 25 and the most downstream source population) and 9.2a (SP 25 is located at the most downstream point while SP 70 and SP 80 are situated between SP 25 and the most downstream source population in the brook Verrebeek) for respectively *Baetis* and Limnephilidae.

For *Baetis*, the difference in accumulated cost based on the resistance sets R1, R2, R4, R7, R11 and R16 (Table 9.1) were rather limited. In these sets, migration resistance for water surface, buffer strip, wooded area, meadows, arable land, nature reserve and the weir were fixed at value 1 while the migration resistance for urban region, industrial area and culverted river section increased from 1 to 25. The maximum difference was only 270 units for both sampling points SP 25 and 64. Based on the tested resistance sets, the difference between the minimum (R1) and the maximum cost (R21) was a factor three for SP 64 and a factor two for SP 25. However, these differences remained rather restricted. For Limnephilidae, accumulative costs increased much faster in the downstream direction since no more source populations were found in the Zwalm river. This was even more obvious if the resistance value of the land use class urban region was increased (e.g. R1 versus R2, R4, R7, R11 and R16). Since Brakel city is located immediately downstream of the most downstream source population of Limnephilidae in the brook Verrebeek, this town boosted the accumulative cost value. In this way, if resistance values increased accumulative cost values were very high for Limnephilidae, especially in the most downstream points (e.g. SP 25).

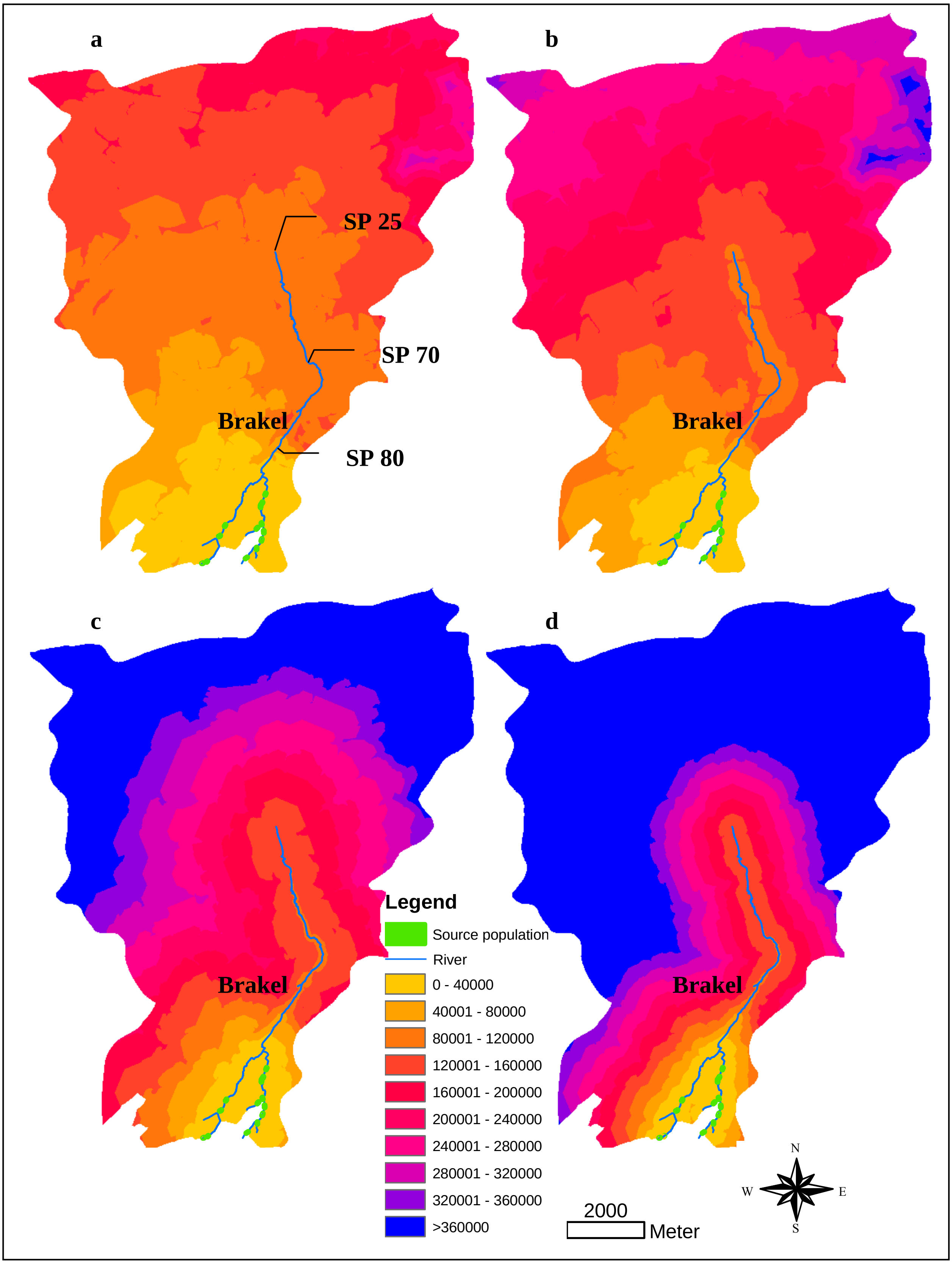


Fig. 9.2. Maps of the accumulative costs (CostR11 – CostR14) for migration through the air / over land from the source populations of Limnephilidae based on the resistance sets R11 (a), R12 (b), R13 (c) and R14 (d) of Limnephilidae.

Table 9.3. Details about the accumulative cost values for migration through the air / over land (CostR1 – CostR21) for the sampling points SP 25 and SP 64 (the locations of these sampling points are shown on Fig. 9.1a) and the sampling points SP 25, SP 70 and SP 80 (the locations of these sampling points are shown on Fig. 9.2a) for respectively *Baetis* and Limnephilidae

Cost Map	***Baetis***		**Limnephilidae**		
	SP 64	**SP 25**	**SP 80**	**SP 70**	**SP 25**
CostR1	1020	1980	4436	9631	12980
CostR2	1028	2002	7583	15187	18587
CostR3	1180	2215	8561	15854	19647
CostR4	1064	2037	15719	31685	35114
CostR5	1231	2256	17847	32353	36198
CostR6	1430	2480	19412	33765	38272
CostR7	1128	2095	34496	57432	60924
CostR8	1292	2290	35250	59800	63703
CostR9	1513	2527	35333	61163	65742
CostR10	1825	2860	39012	63466	69129
CostR11	1234	2201	60717	90316	93669
CostR12	1402	2389	64456	111442	115421
CostR13	1617	2622	66834	116095	120753
CostR14	2031	2985	69392	118248	124089
CostR15	2576	3591	73735	122597	130770
CostR16	1289	2248	70673	99064	102388
CostR17	1464	2471	74664	135487	139566
CostR18	1662	2673	78005	143481	148244
CostR19	1998	3047	83098	145699	151579
CostR20	2730	3745	88182	150155	158280
CostR21	2953	4034	88682	152263	161543

9.3.2. Sensitivity analysis for migration through the river

Sensitivity analysis of the migration models simulating the migration through the river indicated that variation in flow velocity values did not result in a big difference of the accumulative costs. This is illustrated in Table 9.4 for Gammaridae and *Baetis* for which flow velocity was an important factor determining the downstream drift. The difference in accumulative cost for downstream migration through the river between minimum (highest migration resistance) (RDown1, RDown8 and RDown15) and maximum (lowest migration resistance) (RDown5, RDown12 and RDown19) flow velocity was only 887 units for Gammaridae and 1476 units for *Baetis*. These values were based on SP 64 where no influence of the weir as migration barrier was noticeable since this weir is located more downstream. Due to the migration barrier effect of the weir (SP 25 in comparison with SP 64) the accumulative cost for downstream migration increased with approximately 5450 units in average for both taxa Gammaridae and *Baetis*. If the resistance value of the weir increased (RDown1 – RDown7 versus RDown8 – RDown14 and RDown15 – RDown21, Table 9.2), only a minor increase of ± 60 units in accumulative cost for downstream migration through the river was observed for both Gammaridae and *Baetis*. In addition, the difference in accumulative cost between river sections with minimum and maximum (CostDownR1 minus CostDownR5, CostDownR8 minus CostDownR12, CostDownR15 minus CostDownR19) flow velocity remained unchanged: 2117 units for Gammaridae and 2575 units for *Baetis*. For the downstream migration of Limnephilidae, only the variation in the resistance values of the weir and the culverted river sections were assumed to be of major importance. The values of the other determining characteristics were held constant as described in Section 9.2. Results of the sensitivity analysis for the downstream migration of Limnephilidae are shown in Table 9.5. Each increase of the resistance values of the culverted river sections with 25 (Table 9.3) resulted in an raise of the accumulative cost of 16065 and 23025 units for respectively SP 80 and SP 70 (Table 9.5). If also the weir is taken into account (SP 25), an increase of 23266 units was found for downstream migration.

Similarly, for the sensitivity analysis of the upstream migration models, only the resistance values of the weir and the culverted river sections were altered while the values of the other determining characteristics were held constant. The cost for Gammaridae to migrate upstream from the downstream source population to SP 64 was 2065, 2306 or 2547 units depending on the resistance value of the weir (Table 9.4). If the resistance of the weir increased with 100 units (Table 9.2) the accumulative cost to migrate between both sites increased with 241 units. For *Baetis* and Limnephilidae, the accumulative cost to migrate from the nearest source population to respectively SP 97 and SP 96 (upstream sites of two different tributaries of the brook Dorenbosbeek) was 1724 and 3685 units. For both taxa, there was no difference between the three resistance sets since there were no weirs or culverted river sections between the nearest source populations and these sampling points.

Table 9.4. Details about the accumulative cost values for migration through the river in the downstream (CostDownR1 – CostDownR21 for the sampling points SP 25 and SP 64 for both Gammaridae and *Baetis*) and the upstream direction (CostUp1 – CostUp3 for the sampling points SP 64 and SP 97 for respectively Gammaridae and *Baetis*) for Gammaridae and *Baetis*

Cost Map	**Gammaridae**		***Baetis***	
	SP 64	**SP 25**	**SP 64**	**SP 25**
CostDownR1	3309	9270	4777	10570
CostDownR2	3154	8840	4336	9933
CostDownR3	2900	8730	4081	9870
CostDownR4	2765	7704	3641	8543
CostDownR5	2422	7154	3301	7994
CostDownR6	3008	8347	3896	9189
CostDownR7	3050	8523	4522	9920
CostDownR8	3309	9326	4777	10630
CostDownR9	3154	8899	4336	9994
CostDownR10	2900	8794	4081	9930
CostDownR11	2765	7761	3641	8603
CostDownR12	2422	7209	3301	8055
CostDownR13	3008	8407	3896	9249
CostDownR14	3050	8592	4522	9980
CostDownR15	3309	9386	4777	10690
CostDownR16	3154	8959	4336	10054
CostDownR17	2900	8854	4081	9990
CostDownR18	2765	7821	3641	8664
CostDownR19	2422	7269	3301	8115
CostDownR20	3008	8468	3896	9309
CostDownR21	3050	8652	4522	10041
	SP 64		**SP 97**	
CostUp1	2065		1724	
CostUp2	2306		1724	
CostUp3	2547		1724	

Table 9.5. Details about the accumulative cost values for migration through the river in the downstream (CostDownR1 – CostDownR3 for the sampling points SP 80, SP 70 and SP 25) and the upstream direction (CostUp1 – CostUp3 for the sampling point SP 96) for Limnephilidae

Cost Map	**SP 80**	**SP 70**	**SP 25**
CostDownR1	40984	74586	101531
CostDownR2	57050	97611	124797
CostDownR3	73115	120636	148063
	SP 96		
CostUp1	3685		
CostUp2	3685		
CostUp3	3685		

9.4. Discussion

In the major part of the published studies using Cost Weighted Distance modelling (e.g. Bunn *et al.*, 2000; Ferreras, 2001; Chardon *et al.*, 2003), very little effort was made to test for sensitivity of the resistance values. Mostly, resistance values were chosen arbitrarily, making very little considerations on their biological meaning (especially values for hindering elements and barriers, e.g. Ricketts, 2001). Therefore, there is a need for more research (theoretical as well as practical) on the resistance values, their ecological background (e.g. Ricketts, 2001), and the sensitivity of migration models for variation in their values.

As a consequence, the sensitivity analyses performed in this chapter can be seen as a first (theoretical) step in the validation of the migration models developed in Chapter 8. Therefore, the resistance values ascribed to the different determining variables were altered. In this way, the sensitivity of the developed migration models to variation in these resistance values could be investigated.

The different determining environmental and river characteristics (e.g. the presence of a buffer strip, the type of land use, the presence of culverted river sections and weirs) were ordered in terms of increasing resistance for migrating macroinvertebrates, based on best expert judgement: (rare) literature on dispersal distance and migration behaviour and long-term experience with the taxa. The most appropriate number of resistance classes and resistance values to be included in the migration model were investigated based on sensitivity analysis by varying these factors. Although, there is no empirical evidence on the correctness of the resistance values nor the number of resistance classes (see also Ferreras, 2001; Verbeylen *et al.*, 2003), the results of this study do show sensitivity to variation in these resistance values. However, to deduct a ‘most likely’ resistance set based on this sensitivity analysis, further practical validation is recommended in the future. Field experiments can certainly support the choice and practical validation of this resistance set.

If more resistance classes were used in the Cost Weighted Distance function, better results were obtained for the models describing the migration through the air / over land. If the determining characteristics for migration were divided in three classes instead of one or two (watercourses and buffer strips; wooded area, meadows, arable land, nature reserve and weirs;

urban region, industrial area and culverted river sections), more details could be extracted from the migration models. The landscape became more homogeneous if only similar resistance values were used for the different land use classes and river characteristics. In addition, it is more realistic that crossing a meadow take less effort, is less dangerous and in this way more likely (lower resistance value) than crossing a town. In this way, more reliable results might be expected when simulating the migration behaviour of macroinvertebrates through the air / over land. So when maps are composed, sufficient and relevant details should be included. The available maps may not always allow the distinction between certain relevant elements that might be important for a certain taxon and extra digitalisation may be required. In the present study, the available land use map did not take the presence or absence of buffer strips along the river into account. However, these elements can be of paramount importance since they can serve as a corridor for migrating adults. Therefore, an extra inventory of the buffer strips along the watercourses was drawing up and digitalized afterwards. Similar results were obtained by Verbeylen *et al.* (2003). The more resistance classes they used to simulate the migration cost of Red squirrel in an urban landscape, the better the results were. In contrast to the present study, they were able to deduct the best resistance values based on collected field data (visual observations).

Additionally, to distinguish more details on the accumulative cost maps for migration through the air / over land, differences between the resistance classes has to be large enough. This was illustrated in Fig. 9.1 and 9.2. Based on this, the choice of the resistance values for migration through the air / over land of *Baetis* and Limnephilidae in Chapter 8 could be justified. However, validation based on field data is still recommended to verify the exact resistance values.

Sensitivity analysis of the migration models simulating the downstream migration of Gammaridae and *Baetis* through the river revealed that variation in flow velocity values only resulted in small differences in accumulative migration costs. In this way, it can be concluded that variation in flow velocity during the year, with exception of big flushes or droughts of course, will only have a minor impact on the behaviour of the migration models. The use of the average flow velocity of the year 2003 as basis for the passive downstream migration resistance can thus be warranted. An increase of the resistance value of the weir also resulted in a minor increase of downstream migration cost (based on the results of SP 25). If the resistance was doubled, the accumulative downstream migration cost only increased 120 units, what is rather negligible. However, the impact of variation in weir resistance is more pronounced if source populations and weir are further apart. For Limnephilidae for example, the downstream migration cost increased with approximately 45000 units if weir resistance was doubled.

For the sensitivity analysis of the upstream migration models, only the resistance values of the weir and the culverted river sections were altered while the values of the other determining environmental and river characteristics were held constant as described in Section 9.2. Conclusions could only be drawn for Gammaridae. Only for Gammaridae it was possible to simulate the upstream migration over a weir from a downstream source population towards an upstream river section (SP 64). If weir resistance was doubled, an increase in accumulative cost for upstream migration of approximately 500 units was obtained. Although this appeared to be a rather low increase in cost, 500 units represents an increase of one fourth (2547 units in stead of 2065 units), what cannot be neglected.

Additionally, all possible combinations over the different layers could be calculated. However, seen the large differences that remain based on expert knowledge, the migration routes do not alter themselves in this case. Merely a shift in velocity over this route can be expected.

9.5. Conclusions

The sensitivity analyses performed in this chapter can be seen as a theoretical validation of the migration models developed in Chapter 8. If more resistance classes were used to calculate the accumulative migration costs, better results were obtained. If the determining characteristics for migration were divided in three classes instead of one or two, more details could be extracted from the migration models. In addition, to distinguish more details on the accumulative cost maps for migration through the air / over land, differences between the resistance classes has to be large enough. The landscape became more homogeneous if only similar resistance values were used for the different land use classes and river characteristics. From the sensitivity analysis, it was also derived that variation in flow velocity values only resulted in small differences in accumulative migration costs. Variation in flow velocity during the year will thus only have a minor impact on the behaviour of the migration models. Based on these results, the choice of the resistance values in Chapter 8 for migration through the air / over land and through the river could be justified.

Although, there was no empirical evidence on the correctness of the applied resistance values nor the number of resistance classes, the results of this study showed sensitivity to variation in these resistance values. However, to deduct a 'most likely' resistance set based on this sensitivity analysis, further practical validation is recommended in the future, as reference values are very scarcely available in literature. To this end, field or lab experiments could be very useful to support the choice and practical validation of the correct resistance values.

Chapter 10

Application of Artificial Neural Networks and migration models to predict the effect of river restoration scenarios on macroinvertebrates

10.1. Introduction

In order to make the proper decisions for river restoration, one first has to understand the complex interaction between physical, chemical and biological components. The success of river restoration depends on steering the appropriate key factor(s), which differ from river to river and site to site. However, theories on river ecology are complex and not easy to use in the practice of stream management (Verdonschot and Nijboer, 2002b). For an effective restoration action at a site, river managers need a simple decision support system to handle the ecological complexity.

The main objective of this study was to illustrate and validate the practical application of the ANN habitat suitability and the migration models to support decision making in river management. In this context, it has been tried to predict the effect of four restoration projects in the Zwalm river basin on the five taxa Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae:

- project 1: remeandering of the Zwalm river at Roborst;
- project 2: construction of a collector at Elene;
- project 3: restoration of the Molenbeek at Brakel;
- project 4: removal of a weir for water quantity control in the Zwalm river.

For each restoration project, the following questions had to be answered:

- What are the restoration objectives?
- Which restoration actions are essential?
- Which sites have to be monitored to completely describe the study area of interest?
- What are the consequences of the restoration actions for the river and habitat characteristics?
- What is the effect of the river restoration on the habitat suitability of the macroinvertebrates (predicted applying ANN habitat suitability models)?
- What are the recolonization possibilities of the restored watercourse (simulation by means of migration models, only for project 3 and 4)?

10.2. Material and methods

The locations of the four river restoration project are indicated in Fig. 10.1.

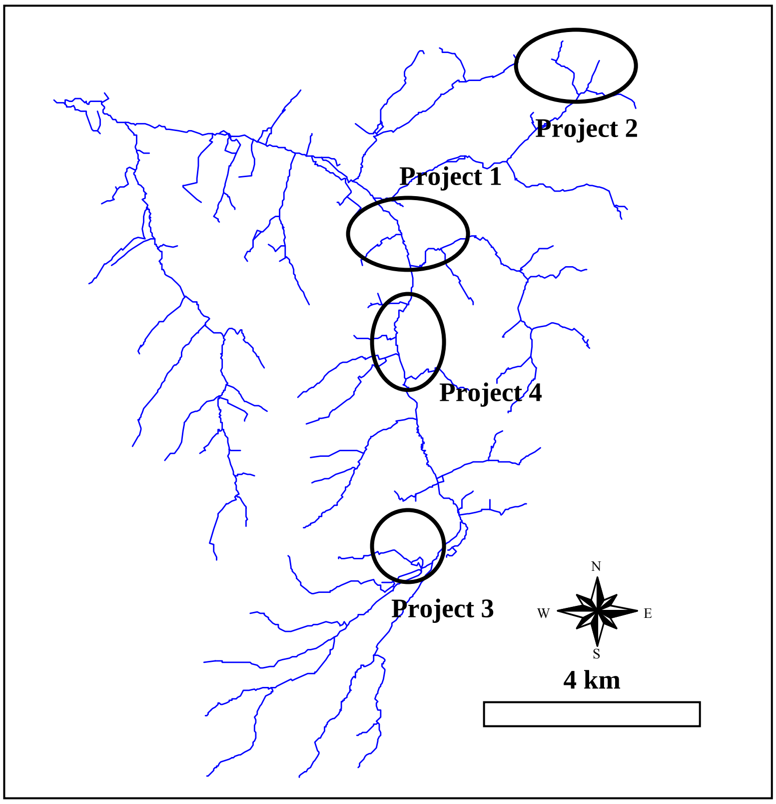


Fig. 10.1. Locations of the four restoration projects: project 1: remeandering of the Zwalm river in Roborst; project 2: construction of a collector in Elene; project 3: restoration of the Molenbeek in Brakel; project 4: removal of a weir for water quantity control in the Zwalm river.

The ANN predictions for project 1 and 2 were based on the dataset of the entire Zwalm river basin. For the predictions of project 3 and 4, the dataset of the 'short distance' monitoring network was used. For each prediction the best ANN models extracted from Chapter 6 were applied. The model characteristics and performances are summarized in Table 10.1. The migration models, as developed in Chapter 8, were only applied in project 3 and 4 because both other projects were not located in the study area of the 'short distance' monitoring network.

Table 10.1. ANN model architecture and performance of the models used for the predictions of the impact of river restoration actions on the macroinvertebrates Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae. CCI = percentage of Correctly Classified Instances; CK = Cohen's kappa

	Dataset of the entire Zwalm river basin (project 1 and 2)			**Dataset of the 'short distance' monitoring network (project 3 and 4)**		
	Model architecture	**CCI (%)**	**CK**	**Model architecture**	**CCI (%)**	**CK**
Tubificidae	24-5-1	90.3	0.12	24-15-1	94.2	0.32
Asellidae	24-10-1	79.8	0.60	24-10-1	94.2	0.88
Gammaridae	24-15-1	80.2	0.47	24-10-10-1	92.5	0.66
BAETIS	24-10-1	75.5	0.18	24-10-1	77.5	0.54
Limnephilidae	24-5-1	89.0	0.34	24-10-1	86.7	0.49

10.3. Results

10.3.1. Remeandering project of the Zwalm river in Roborst

10.3.1.1. Project definition

This project was based on a report by Belconsulting (2003). Several restoration options within an integrated water management perspective in the Zwalm river basin were described in that report. This restoration action at the Zwalm river at Roborst 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.

Near the mouth of the brook Traveinsbeek, the Zwalm river was straightened years ago. Recently there were several projects to investigate sustainable flood control measures in the Zwalm river basin. The construction of several natural flooding areas are among these, and are often combined with nature development. Also this restoration action is a combination of both (Fig. 10.2). The original project described in Belconsulting (2003) departed from reintroducing meandering without decreasing the effect of the Bostmolen weir (ca. 250 m further downstream). However, considering the construction of the flooding area, this decrease in effect of the weir could be appropriate in the future. To this end, two restoration scenarios were modelled:

- reintroducing a meandering pattern in the Zwalm river (scenario 1);
- reintroducing a meandering pattern in the Zwalm river taking into account a decrease of the effect of the Bostmolen weir (scenario 2).

The planned remeandering has an impact on several river characteristics as shown in Table 10.2. In this table, only the altered variables are presented. The expected values of the river characteristics in the studied monitoring site (sampling point 121) after remeandering (and decreasing the effect of the weir in the second scenario), were based on an upstream (sampling point 33) as well as a downstream sampling site (sampling point 25), characterized by a well-developed meandering pattern and a minimal weir effect (Fig. 10.2). Therefore, these sites could be seen as the expected situation after the remeandering works.

Table 10.2. Actual and expected altered values of the stream characteristics for sampling point 121 after reintroducing meandering (and decreasing the effect of the weir in the second scenario). These values were obtained on the basis of monitored conditions upstream and downstream of the sampling point of interest and in combination with expert knowledge

Variable	**Actual conditions**	**Expected conditions (scenario 1)**	**Expected conditions (scenario 2)**
Width (cm)	600	550	400
Meandering	5	2	2
Pools-riffles	5	3	3
Hollow banks	4	3	3
Depth (cm)	72	72	30
Flow velocity (m/s)	0.03	0.05	0.09
Fraction pebbles (%)	0.0	10.0	15.0
Fraction gravel (%)	0.0	10.0	40.0
Fraction sand (%)	29.5	40.0	5.0
Fraction loam/clay (%)	70.5	40.0	40.0

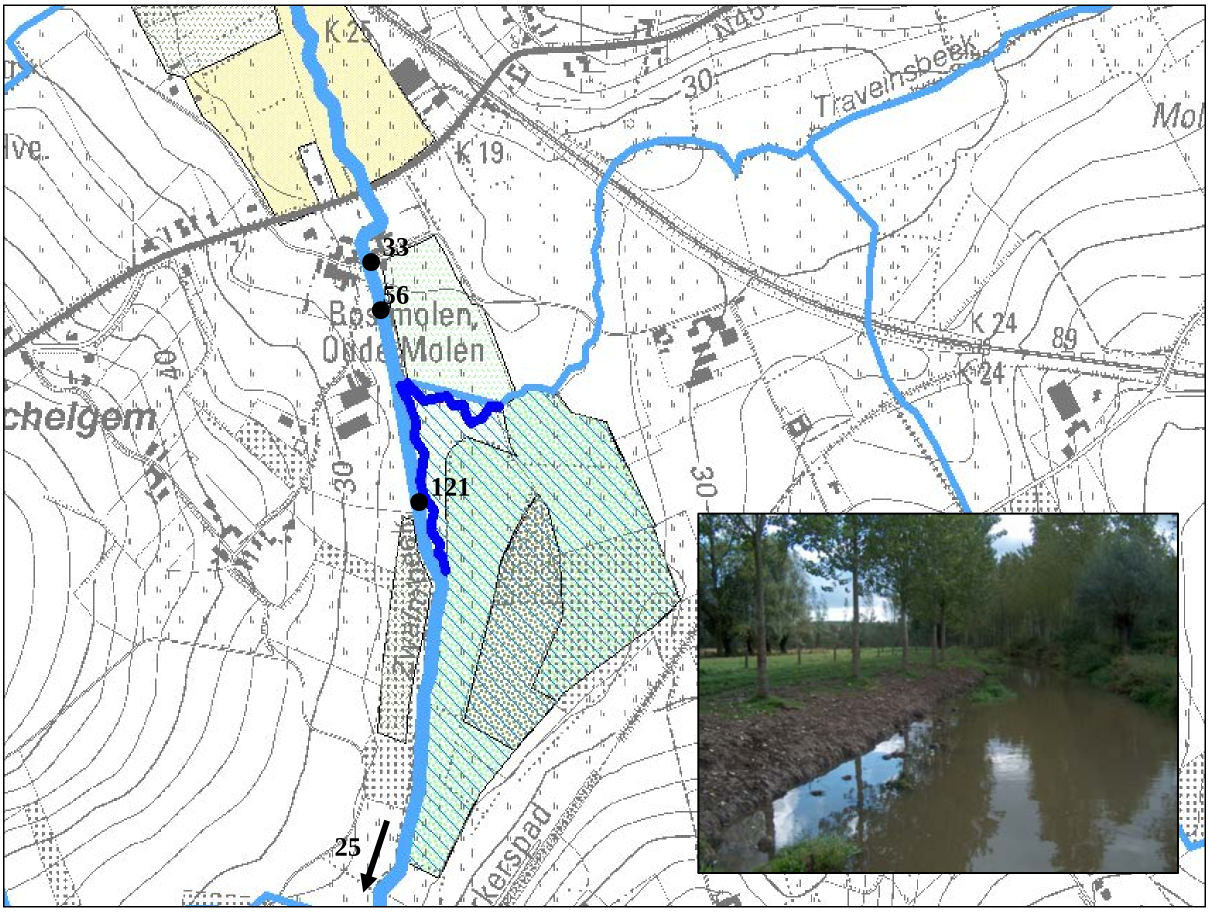


Fig. 10.2. Scheme of the planned remeandering action in the Zwalm river. Ca. 250 m upstream of the Bostmolen weir, the Zwalm river has been straightened. Belconsulting (2003) proposed to remeander the river and to install a natural flooding area (green zone) to minimize the use of the weir for water quantity control. At the right bottom, a picture of the Zwalm river (sampling point 121) in its actual condition (August 2003) is shown.

10.3.1.2. Predicted effect of the restoration actions on the habitat suitability

Based on the changes of the habitat characteristics after remeandering, with or without minimization of the weir effect, the prediction of the habitat suitability of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae was tested (Table 10.3). Table 10.3 reveals that the ANN models performed well to predict the five taxa under the actual conditions, with one major exception however, that *Baetis* was predicted absent based on the four folds. The main reason could be the bad predictive performance (Cohen's kappa = 0.18) obtained for this taxon.

For the five taxa, the following conclusions could be made: remeandering without minimizing the weir effect had no consequences on the habitat suitability of Tubificidae, Asellidae, Gammaridae and Limnephilidae, while *Baetis* was predicted absent after reintroducing meandering. When also the minimization of the weir effect was taken into account, the habitat suitability of Asellidae decreased in 50 % of the models while Gammaridae were predicted present based on the 4 models.

Table 10.3. Actual and expected taxon presence/absence values for sampling point 121 (in brackets the amount of folds out of a total of four that supports the outcome). 'Under actual conditions', 'scenario 1' and 'scenario 2' respectively means before restoration action and after remeandering without and with minimization of the weir effect

	Observed	**Predicted (under actual conditions)**	**Predicted (scenario 1)**	**Predicted (scenario 2)**
Tubificidae	present	present (4/4)	present (4/4)	present (4/4)
Asellidae	present	present (4/4)	present (4/4)	absent (2/4) present (2/4)
Gammaridae	absent	absent (4/4)	absent (4/4)	present (4/4)
Baetis	present	absent (4/4)	absent (4/4)	absent (4/4)
Limnephilidae	absent	absent (4/4)	absent (4/4)	absent (4/4)

10.3.2. Construction of a collector in Elene

10.3.2.1. Project definition

This project included several sampling points situated on the brook Molenbeek and one of its tributaries at Elene. Both streams were impacted by discharges of domestic wastewater. However, the residential area producing this wastewater, will be connected to a collector in the near future (Belconsulting, 2003). To this end, the objective of this project was to investigate the impact of physical-chemical water quality improvement in these contaminated streams.

The major part of the study area was situated upstream of Elene. Two sampling points (SP 16 and 17) were located in the brook Molenbeek and one of its tributaries upstream of Elene. Downstream of SP 17, the brook Molenbeek is dammed up over a distance of 50 m resulting in a pond for fish cultivation. More downstream, a tributary, discharging the domestic wastewater, flows out into the brook Molenbeek. Within the framework of this project, three additional sampling points were selected: one in the discharging tributary (SP 128), one upstream (SP 129) and one downstream (SP 128) of this tributary in the brook Molenbeek. At Elene the brook Molenbeek was also dammed up, resulting in another pond. A last sampling point was situated downstream of Elene (SP 36). The sampling points are shown in Fig. 10.3.

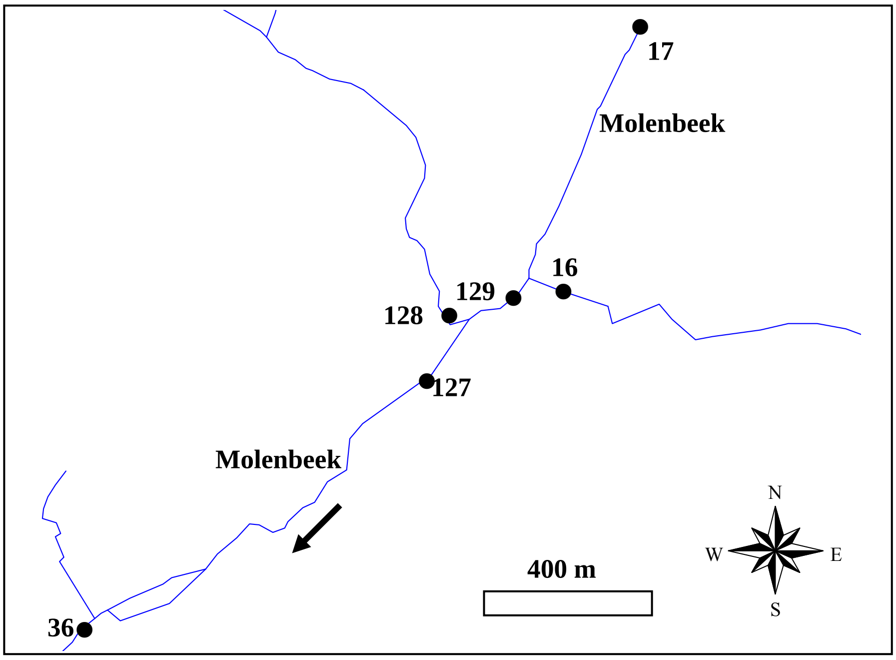


Fig. 10.3. Location of the sampling sites in project 2.

For this study, sampling sites 16 and 36 were chosen as references since they were least impacted in comparison with the other sites. Sampling sites 17 and 129 were slightly contaminated respectively by industrial and domestic wastewater, and by the eutrophicating effect of the fishpond.

This study was mainly focussing on the physical-chemical characteristics of the watercourse. The habitat characteristics were very similar and of acceptable quality over the entire study area. Due to the fishponds however, the streams were artificially dammed. Therefore, it was more realistic to include an additional restoration action in which the effect of these dams was minimized. To this end, two restoration scenarios were modelled for both sampling sites 127 and 128:

- construction of a local collector to improve the physical-chemical water quality (scenario 1);

- construction of a local collector to improve the physical-chemical water quality and minimize the dammed effect of the fishponds (scenario 2).

The planned restoration actions would have an impact on several river characteristics as shown in Table 10.4. In this table, only the altered variables are presented.

Table 10.4. Actual and expected altered values of the stream characteristics for sampling points 127 and 128 (SP 127 and 128) after construction of a local collector to improve the physical-chemical water quality (scenario 1) and decreasing the dam effect (scenario 2). These values were obtained on the basis of monitored conditions upstream (SP 16) and downstream (SP 36) the sampling points of interest and in combination with expert knowledge

Variable	**Actual condition**		**Expected conditions (scenario 1)**		**Expected conditions (scenario 2)**	
	SP 127	**SP 128**	**SP 127**	**SP 128**	**SP 127**	**SP 128**
Width (cm)	283	125	283	125	116	116
Depth (cm)	100	12	100	12	60	60
Flow velocity (m/s)	0.00	0.00	0.00	0.00	0.16	0.16
pH	7.7	7.4	7.6	7.6	7.6	7.6
Dissolved oxygen (mg/l)	7.4	4.0	8.4	8.4	8.4	8.4
Conductivity (µS/cm)	666	826	671	671	671	671
Ammonium (mg NH_4^+-N/l)	0.9	2.7	0.3	0.3	0.3	0.3
Nitrate (mg NO_3^--N/l)	3.3	2.5	1.1	1.1	1.1	1.1
Total nitrogen (mg N/l)	8.5	8.4	6.8	6.8	6.8	6.8
Ortho phosphate (mg PO_4^{3-}-P/l)	0.4	1.9	0.5	0.5	0.5	0.5
Total phosphorus (mg P/l)	0.6	2.0	0.9	0.9	0.9	0.9
COD (mg COD/l)	32	53	35	35	35	35
Fraction pebbles (%)	0.0	0.0	0.0	0.0	0.0	0.0
Fraction gravel (%)	0.0	0.0	0.0	0.0	6.9	6.9
Fraction sand (%)	24.7	37.3	24.7	37.3	45.3	45.3
Fraction loam/clay (%)	75.3	62.7	75.3	62.7	47.9	47.9

10.3.2.2. Predicted effect of the restoration actions on the habitat suitability

Based on the changes of the habitat characteristics after water quality improvement, with or without minimization of the dammed effect, the prediction of the habitat suitability of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae was tested (Table 10.5). The actual conditions were well predicted. However, Asellidae was predicted present at sampling site 127 and absent in 128 based on the four folds, while Asellidae was respectively absent and present at these sites.

Derived from the predictions of both restoration scenarios, the following conclusions could be drawn: physical-chemical water quality improvement without minimizing the dammed effect had no consequences for the habitat suitability of Tubificidae, Asellidae, *Baetis* and Limnephilidae for both sampling sites 127 and 128. Gammaridae however were predicted present at sampling point 128 on the basis of three models. When also the minimization of the dammed effect was considered, the habitat suitability of Asellidae improved in 25 % of the models at sampling point 128 while Gammaridae were predicted present based on respectively 2 and 3 models at sampling sites 127 and 128.

Table 10.5. Actual and expected taxon presence/absence values for sampling points 127 and 128 (in brackets the amount of folds out of a total of four that supports the outcome). ‘Under actual conditions’, ‘scenario 1’ and ‘scenario 2’ respectively means before restoration action and after water quality improvement without and with minimization of the dammed effect

	Observed		**Predicted (under actual conditions)**		**Predicted (scenario 1)**		**Predicted (scenario 2)**	
	SP 127	**SP 128**	**SP 127**	**SP 128**	**SP 127**	**SP 128**	**SP 127**	**SP 128**
Tubificidae	present	present	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)
Asellidae	absent	present	present (4/4)	absent (4/4)	present (4/4)	absent (4/4)	present (4/4)	absent (3/4)
Gammaridae	absent	absent	absent (4/4)	absent (4/4)	absent (4/4)	present (3/4)	present (2/4)	present (3/4)
BAETIS	absent	absent	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)
Limnephilidae	absent	absent	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)

10.3.3. Restoration of the Molenbeek in Brakel

10.3.3.1. Project definition

Before the brook Molenbeek discharges in the Zwalm river, it first flows through the city of Brakel. The habitat characteristics of this part of the watercourse are strongly modified. To this end, the objective of this project was to investigate the improvement of the habitat characteristics in this modified stream. In addition, this part of the brook Molenbeek is impacted by several discharges of domestic wastewater. Therefore, an additional improvement of the physical-chemical water quality is assumed in a second scenario. To this end, two restoration scenarios were modelled for the sampling sites 112, 116, 117, 118 and 119:

- restoration of the river habitat (scenario 1);
- restoration of the river habitat and improvement of the physical-chemical water quality (scenario 2).

Within the framework of this project, four additional samples were taken (SP 116, 117, 118 and 119). The sampling points are presented in Fig. 10.4. The actual situation of SP 112 and SP 116 is shown in Fig. 10.5. SP 119 was situated upstream of Brakel city and was still characterized by a good habitat quality. Therefore, SP 119 was selected as a reference situation for this study.

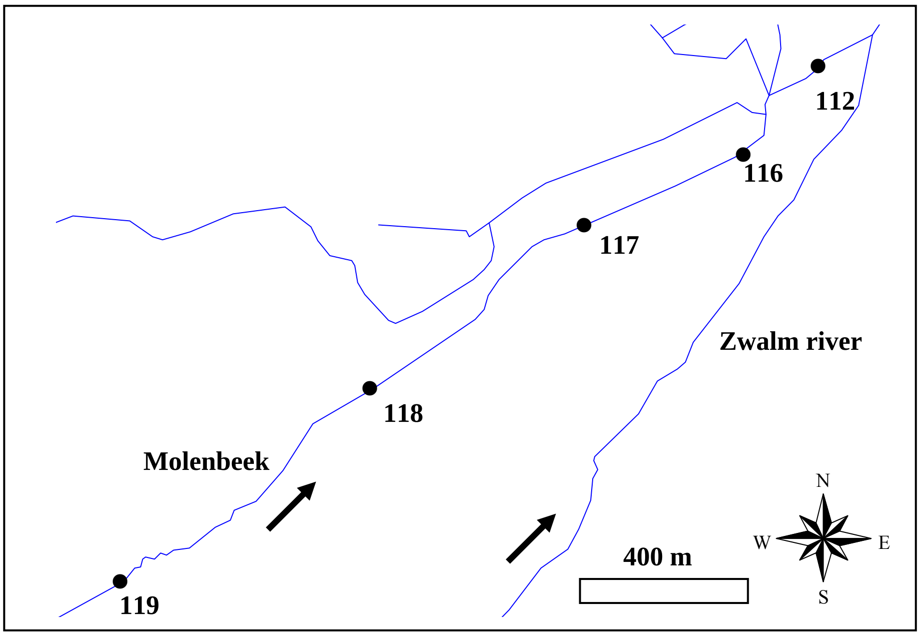


Fig. 10.4. Location of the sampling sites in project 3.

Fig. 10.5. Actual situation of the brook Molenbeek in Brakel at sampling points 112 (left) and 116 (right).

The planned restoration actions would have an impact on several river characteristics as shown in Table 10.6. Only the altered variables are presented in this table.

Table 10.6. Actual and expected altered values of the stream characteristics for sampling points 112, 116, 117, 118 and 119 after habitat restoration (scenario 1) and improvement of the physical-chemical water quality (scenario 2). These values were obtained on the basis of monitored conditions upstream (SP 119) of the sampling points of interest and in combination with expert knowledge

Variable	**Actual condition**				**Expected conditions (scenario 1)**				**Expected conditions (scenario 2)**			
	SP 112	**SP 116**	**SP 117**	**SP 118**	**SP 112**	**SP 116**	**SP 117**	**SP 118**	**SP 112**	**SP 116**	**SP 117**	**SP 118**
Width (cm)	168	270	260	205	150	150	150	150	150	150	150	150
Embankment	2	2	1	0	0	0	0	0	0	0	0	0
Meandering	4	6	3	4	3	3	3	3	3	3	3	3
Pools-riffles	5	6	5	5	5	5	5	5	5	5	5	5
Hollow banks	6	6	1	1	1	1	1	1	1	1	1	1
Depth (cm)	18	12	23	27	20	20	20	20	20	20	20	20
Flow velocity (m/s)	0.33	0.26	0.14	0.12	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
pH	8.1	8.0	7.6	7.5	8.1	8.0	7.6	7.5	7.4	7.4	7.4	7.4
Dissolved oxygen (mg/l)	7.1	7.3	8.0	8.5	7.1	7.3	8.0	8.5	8.0	8.0	8.0	8.0
Conductivity (µS/cm)	619	502	510	701	619	502	510	701	500	500	500	500
Ammonium (mg NH_4^+-N/l)	1.4	0.9	0.7	1.2	1.4	0.9	0.7	1.2	0.7	0.7	0.7	0.7
Nitrate (mg NO_3^--N/l)	3.3	4.0	3.9	4.3	3.3	4.0	3.9	4.3	3.4	3.4	3.4	3.4
Total nitrogen (mg N/l)	8.6	7.6	5.7	5.8	8.6	7.6	5.7	5.8	5.7	5.7	5.7	5.7
Ortho phosphate (mg PO_4^{3-}-P/l)	0.3	0.3	0.2	0.2	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.2
Total phosphorus (mg P/l)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
COD (mg COD/l)	21	21	9	19	21	21	9	19	9	9	9	9
Fraction pebbles (%)	99.0	83.0	55.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fraction gravel (%)	1.0	0.0	17.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fraction sand (%)	0.0	0.0	16.0	62.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0
Fraction loam/clay (%)	0.0	0.0	12.0	28.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0

10.3.3.2. Predicted effect of the restoration actions on the habitat suitability

To improve the habitat quality of the strongly modified river section in the Molenbeek brook, it was proposed to remove the artificial embankment structures along the watercourse. It would be beneficial if the infiltration of rainwater into the ground was stimulated. In addition, the discharge of domestic wastewater has to be minimized. After river restoration, hollow banks could develop. Additionally, remeandering is still possible along the major part of the impacted river section. Based on these restoration actions, the habitat characteristics are supposed to resemble these of the reference situation (SP 119).

Based on the changes of the habitat characteristics after river habitat restoration, with or without improvement of the physical-chemical water quality, the prediction of the habitat suitability of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae was tested (Table 10.7). Table 10.7 reveals that the ANN models perform well to predict the five taxa under the actual conditions, with one exception, for which a few models predicted *Baetis* absent.

For the five taxa, the following conclusions could be drawn: river habitat restoration at the Molenbeek brook had no consequences for the habitat suitability of Tubificidae, Asellidae, Gammaridae and Limnephilidae, while *Baetis* was predicted absent by 100 %, 50 %, 25 % and 50 % of the models for respectively the sites SP 112, SP 116, SP 117 and SP 118. When also the improvement of the physical-chemical water quality was considered, no changes were observed for Tubificidae, Gammaridae and Limnephilidae. On the other hand, the habitat suitability of Asellidae in SP 118 decreased on the basis of one model while the habitat suitability of *Baetis* improved in comparison with scenario 1.

Table 10.7. Actual and expected taxon presence/absence values for sampling points 112, 116, 117 and 118 (in brackets the amount of folds out of a total of four that supports the outcome). 'Under actual conditions', 'scenario 1' and 'scenario 2' respectively means before restoration actions and after restoration of the river habitat without and with improvement of the physical-chemical water quality

	Observed				**Predicted (under actual conditions)**			
	SP 112	**SP 116**	**SP 117**	**SP 118**	**SP 112**	**SP 116**	**SP 117**	**SP 118**
Tubificidae	present	present	present	present	present (4/4)	present (4/4)	present (4/4)	present (4/4)
Asellidae	present	present	present	present	present (4/4)	present (4/4)	present (4/4)	present (4/4)
Gammaridae	present	present	present	present	present (4/4)	present (4/4)	present (4/4)	present (3/4)
Baetis	present	present	present	present	present (3/4)	present (4/4)	present (3/4)	present (2/4)
Limnephilidae	absent	absent	absent	absent	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)
	Predicted (scenario 1)				**Predicted (scenario 2)**			
	SP 112	**SP 116**	**SP 117**	**SP 118**	**SP 112**	**SP 116**	**SP 117**	**SP 118**
Tubificidae	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)
Asellidae	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (3/4)
Gammaridae	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)
Baetis	present (0/4)	present (2/4)	present (3/4)	present (2/4)	present (3/4)	present (3/4)	present (3/4)	present (3/4)
Limnephilidae	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)

10.3.3.3. Simulation of the recolonization

For this restoration project, it was less useful to apply the migration models developed for Gammaridae, *Baetis* and Limnephilidae (Chapter 8) to check the recolonization possibilities of the restored sites since Gammaridae and *Baetis* were already present before the restoration actions took place. On the other hand, the habitat suitability of Limnephilidae did not improve sufficiently to get this sensitive taxon back.

10.3.4. Removal of a weir for water quantity control in the Zwalm river

10.3.4.1. Project definition

The sites on the Zwalm river near the Boembekemolen are characterized by a modification of the flow channel due to a flood control weir between sampling sites 61 and 55 (Fig. 10.6). This weir obstructs the migration of fish and other aquatic organisms, including macroinvertebrates. Upstream of the weir, the river is drastically deepened. Just in front, the depth can be nearly around two meters depending on the control level of the weir and the amount of sediments accumulated at the site. This is much deeper than under natural conditions. Also the flow velocity is reduced drastically, creating an almost stagnant water body immediately upstream of the weir (Belconsulting, 2003). This situation results in direct and indirect impacts on the river biology. A direct impact is that the shear stress is quite low, being an advantage for Asellidae for instance, but for some insect larvae like *Baetis* which 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 decreased flows can play a crucial role for the river biology as well. As a result of the reduced flows, there is a serious accumulation of sediments (due to the erosion problems in the area), containing organic materials and probably also toxic materials, such as pesticides from runoff of agricultural soils. The organic compounds are degraded by the local micro biota, reducing the amount of dissolved oxygen in the water in particular near the bottom of the deepened systems.

Removing the weir has an impact on several structural components of the river as well as presented in Table 10.8. In this table, only the altered variables are presented. In the upstream section of the removed weir, flow velocity will increase while width and depth will decrease.

Also the quality of the channel morphology will evolve positively because of an increased meandering of the river, the development of pools and riffles, and the creation of natural hollow banks. These changes are taken into account when predicting the effect of weir removal on the habitat suitability for Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae. In combination with expert knowledge, the expected values of the altered stream characteristics were obtained on the basis of monitored conditions about one kilometre upstream of the weir (sampling point 65), where the weir effect is about nihil. Therefore, it is a good site as a basis for comparison and validation also under the unaltered water quality conditions presented in this simulation exercise.

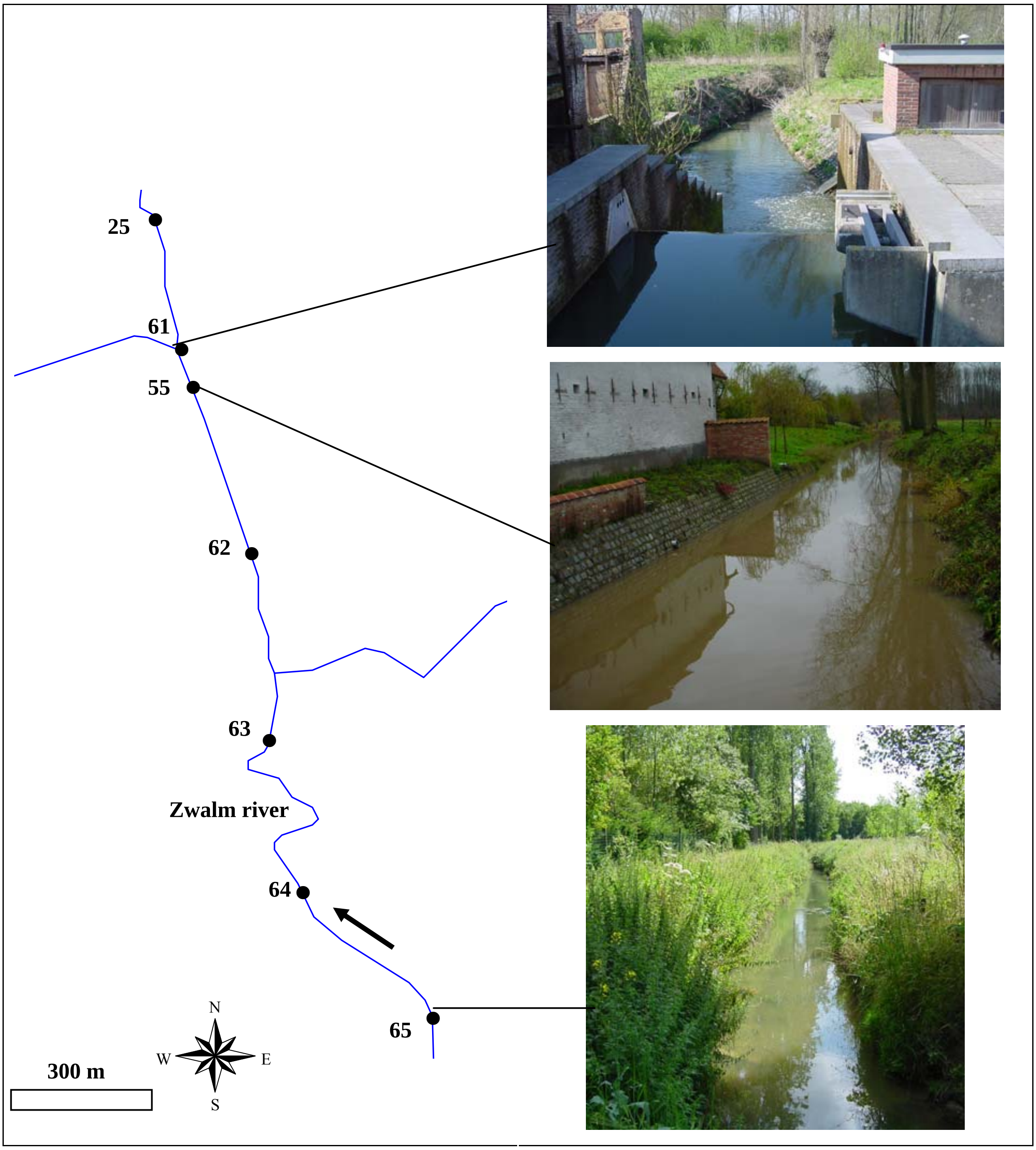


Fig. 10.6. Location of the sampling sites is in this study. At the right site, pictures of the sampling points 61, 55 and 65 are shown.

Table 10.8. Actual and expected altered values of the stream characteristics for sampling sites SP 25, 55, 61, 62, 63 and 64 after weir removal. These values were obtained on the basis of monitored conditions upstream (SP 65) of the sampling points of interest and in combination with expert knowledge

Variable	**Actual condition**						**Expected conditions**
	SP 25	**SP 61**	**SP 55**	**SP 62**	**SP 63**	**SP 64**	
Width (cm)	260	500	750	550	583	410	337
Embankment	0	1	1	0	0	0	0
Meandering	2	6	5	5	1	4	3
Pools-riffles	4	4	5	5	4	4	4
Hollow banks	4	6	5	5	1	3	4
Depth (cm)	12	27	190	130	71	62	23
Flow velocity (m/s)	0.43	0.04	0.03	0.03	0.05	0.04	0.20
Fraction pebbles (%)	98.0	74.0	0.0	0.0	0.0	11.0	38.0
Fraction gravel (%)	1.3	1.4	31.5	0.0	0.1	0.0	23.8
Fraction sand (%)	0.4	19.5	6.5	37.8	50.2	23.8	27.1
Fraction loam/clay (%)	0.3	5.1	62.0	62.2	49.7	65.2	11.1

10.3.4.2. Predicted effect of the restoration actions on the habitat suitability

In Table 10.9 the outcomes of the ANN models are presented over the four folds. In brackets the number of folds supporting this presence/absence label is indicated. For the five taxa the models were able to classify well the actual conditions (= the conditions before weir removal), based on the environmental variables. No shifts in the habitat suitability were predicted for Tubificidae, Asellidae, *Baetis* and Limnephilidae after river restoration. For Gammaridae however, sampling sites 55 and 62 were predicted suitable again based on all the models. At sampling sites 63 and 64, Gammaridae were predicted present on the basis of only one model. On the other hand, two models predicted sampling site 25 as unsuitable after weir removal.

Table 10.9. Actual and expected taxon presence/absence values for sampling points 25, 55, 61, 62, 63 and 64 (in brackets the amount of folds out of a total of four that supports the outcome). 'Under actual conditions' and 'under altered conditions' respectively means before and after weir removal

	Observed					
	SP 25	**SP 61**	**SP 55**	**SP 62**	**SP 63**	**SP 64**
Tubificidae	present	present	present	present	present	present
Asellidae	present	present	present	present	present	present
Gammaridae	present	present	absent	absent	absent	absent
Baetis	absent	absent	absent	absent	absent	absent
Limnephilidae	absent	absent	absent	absent	absent	absent
	Predicted (under actual conditions)					
	SP 25	**SP 61**	**SP 55**	**SP 62**	**SP 63**	**SP 64**
Tubificidae	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)
Asellidae	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)
Gammaridae	present (4/4)	present (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)
Baetis	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)
Limnephilidae	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)
	Predicted (under altered conditions)					
	SP 25	**SP 61**	**SP 55**	**SP 62**	**SP 63**	**SP 64**
Tubificidae	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)
Asellidae	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)	present (4/4)
Gammaridae	present (2/4)	present (4/4)	present (4/4)	present (4/4)	present (1/4)	present (1/4)
Baetis	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)
Limnephilidae	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)	absent (4/4)

10.3.4.3. Simulation of the recolonization

As predicted by the ANN models, habitat suitability for Gammaridae improved after weir removal. To check the possibility to recolonize the restored river sections, the migration models, developed in Chapter 8, could be applied. Because flow velocity was assumed to change after weir removal and an important migration barrier would disappear, corrections were made for the up and downstream migration resistances. Taking these changes into account for Gammaridae, the accumulative cost for migration downstream through the river would drop only slightly as presented in Fig. 10.7a (4270 units) compared to Fig. 8.4a (4300 units). The accumulative cost for upstream migration would remain unchanged (1760 units) (Fig. 10.7b) since no relevant parameters are altered.

To get insight into the duration of the recolonization, those distances were used that were calculated to obtain the resistances. Therefore, active and passive migration have been taken into account. For each river segment, the time to migrate through that segment has been calculated. This was based on the maximum distance which could be covered through that segment in one day. Finally, the sum from the source populations to the restored river section was considered to estimate the total recolonization time. This would result in a total migration time for Gammaridae of approximately 10 days for SP 55 and 60 days for SP 62.

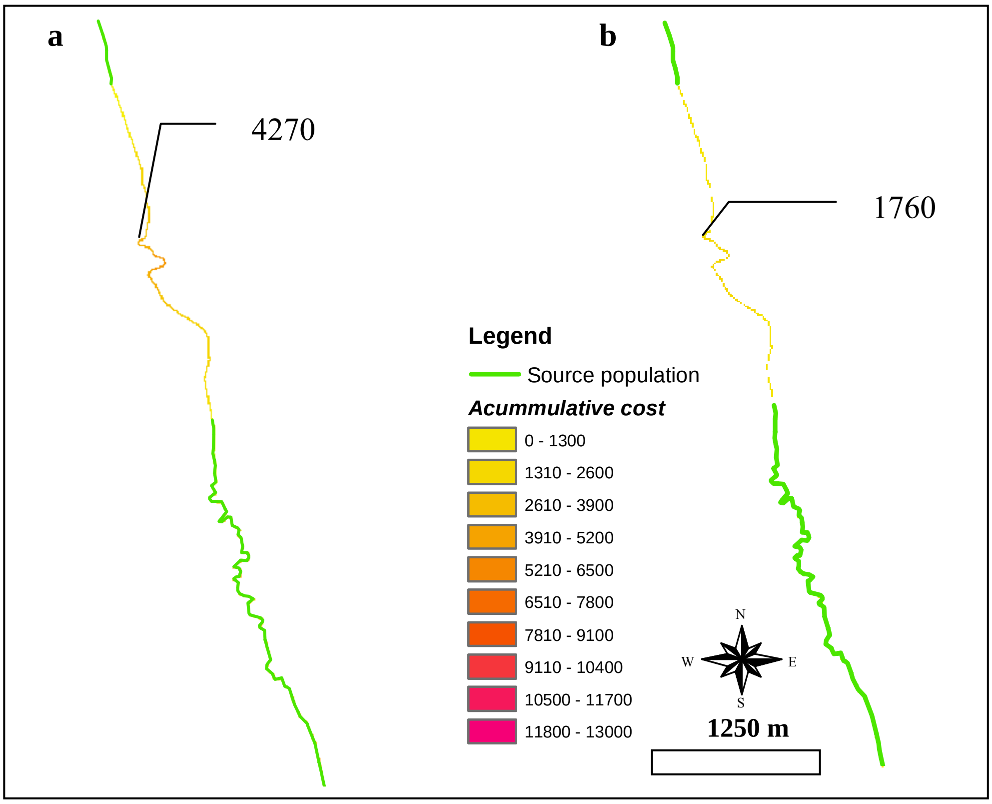


Fig. 10.7. Map of the accumulative cost for migration downstream (a) and upstream (b) through the river from the source populations of Gammaridae after weir removal. The highest accumulative costs to reach the restored habitat are indicated (see figures and legends).

Although the restored river section was evaluated as unsuitable for *Baetis* and Limnephilidae, the developed migration models still could be applied to calculate the migration possibilities from the source populations to the restored parts. This could help the river managers to make a decision on whether or not additional restoration works are necessary. For instance, if also the water quality would be improved (e.g. by the installation of local small scale wastewater treatment plants), it would be likely that *Baetis* would recolonize the restored upstream river section since, several source populations were already observed over there. For this reason, the migration model developed in Section 8.3.3 was applied to simulate the recolonization possibility of *Baetis* in the restored site. Some of the altered stream characteristics after weir removal would also have an effect on the migration model of *Baetis* (e.g. fraction of boulders and flow velocity). In addition, the migration barrier would disappear. Taking these changes into account for *Baetis*, the accumulative cost for migration downstream through the river to SP 55 would drop significantly as presented in Fig. 10.8a (5900 units) compared to Fig. 8.6a (± 9900 units). After weir removal and restoration of the river habitat, this would result in a total migration time for *Baetis* of approximately two years. On the other hand, weir removal would only have a minor effect on the accumulative cost for migration through the air as shown in Fig. 10.8b (2140 units instead of ± 2400 units). The time to recolonize the restored river section through the air would be about 275 days. In this way, one could conclude that the shortest path with the least accumulative cost from the ‘cheapest’ source population to SP 55 would be through the air (Fig. 10.8b).

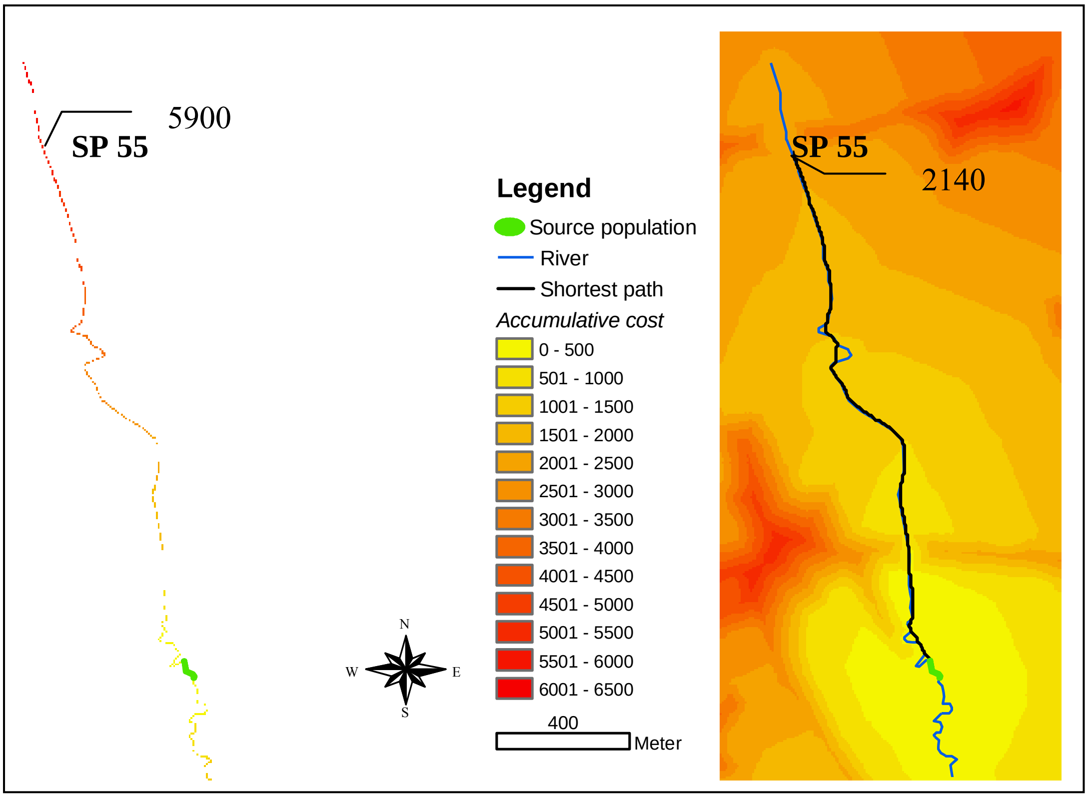


Fig. 10.8. Map of the accumulative cost for migration downstream through the river (a) and through the air (b) from the source populations of *Baetis* after weir removal. The accumulative costs to reach SP 55 are indicated (see figures and legend). The shortest path from the 'cheapest' source population to SP 55 is also shown (see black line).

Assuming that the restored habitat was suitability for Limnephilidae as well, it is less likely that Limnephilidae would recolonize these river sections since source populations were only observed in the headwaters (the brooks Verrebeek and Dorenbosbeek) of the Zwalm river basin, which are more than seven kilometre upstream of the restored river. The migration models indicate that it would take almost six years to migrate downstream through the river (assuming that also the habitat suitability of the watercourses between source population and restored river section would improve), while migration time through the air would take approximately three years.

10.4. Discussion

In the first project, the effect of remeandering without minimization of the weir effect on the five macroinvertebrates appeared to be negligible. This seems logical since remeandering does not alter the characteristics of the watercourse very drastically. Several projects considering remeandering (e.g. of the rivers Gelsa and Brede in Denmark), demonstrated however an increase of the macroinvertebrate diversity of 30 % within two years after the restoration. This is mainly caused by the increased heterogeneity of the substrate throughout the channel (Hansen, 1996). If weir minimization was not taken into account, flow velocity was still too slow to increase this heterogeneity. On the other hand, minimizing the weir effect resulted in a negative, respectively positive effect on the habitat suitability of Asellidae and Gammaridae. Based on Chapter 7, the most influencing variable for Asellidae was width. Minimizing the weir effect induced a decrease in river width which has a negative effect on the probability of presence of Asellidae (two out of four models predicted Asellidae as absent based on scenario 2). Although water quality related input variables, which were of major importance for the habitat suitability of Gammaridae (Chapter 7), were not altered in project 1, the habitat suitability of Gammaridae increased based on scenario 2 (four out of four models). However, also depth was rather important which can declare the shift in habitat suitability of Gammaridae. The habitat suitability of *Baetis* decreased. However, this was probably due to the less predictive performance of the models which tend to predict *Baetis* always as absent (Chapter 6). In addition, *Baetis* was most influenced by pollution related variables which were not altered in both scenarios (Chapter 7). Although one may assume that the self-purifying capacity of the river will increase after remeandering, this was not considered in this study. In an additional study, a negative effect was predicted on the presence of the two leeches *Erpobdella* and *Helobdella*, both occurring in more impacted streams (De Pauw and Vannevel, 1991).

For the first project, it can be concluded that only the remeandering project taking into account the minimization of the weir effect could be valuable for improving the ecological quality. Apart from the possible ecological consequences of remeandering, also the social consequences are of major importance. Although the effect of the construction of the natural flooding area was not considered in this study, it can be concluded that this, in combination with the reintroduction of meanders, enlarges the water bearing capacity of the concerning system. In this manner, this project is advantageous for the ecological value and the safety of the housings and fields against flooding (even if the weir would stow less water as supposed in the second restoration scenario), while not much agricultural area has to be sacrificed for this type of works.

Although a significant physical-chemical water quality improvement may be assumed in project 2, the impact of the installation of a domestic wastewater collector without minimizing the dam effect seemed to be very small on the five taxa. Only for Gammaridae, an improvement of the habitat suitability could be observed. However, Gammaridae was the only taxon for which pollution related variables played a key role in the prediction of the habitat suitability (Chapter 7) and for which reliable predictive performances were obtained (Chapter 6). On the other hand, minimizing the dam effect of the fishponds resulted also in a slightly positive effect on the habitat suitability of Asellidae and a further improvement for the Gammaridae. From a social point of view, it is however little realistic to assume that the dammed effect will be eliminated in the future. Nevertheless, this would be a useful restoration option. If it is not possible to remove the dams, a by-pass to avoid the fishponds could be a valuable solution.

It was less useful to investigate the river habitat restoration in project 3 since four taxa (Tubificidae, Asellidae, Gammaridae and *Baetis*) were already present before the restoration actions took place. Also after river restoration took place, the four taxa were predicted as present. Limnephilidae were however not predicted as present after restoration actions were performed, although habitat as well as water quality was improved. Although the presence/absence of Limnephilidae is mainly influenced by pollution related variables (Chapter 7), this had apparently no impact on the prediction behaviour of the models in scenario 2. Most likely, this can be explained by the sensitivity curves presented in Fig. 7.20. Not one single variable had a probability of presence higher than 0.5, the threshold to classify taxa as present. Derived from the test set however, good predictive results were obtained for Limnephilidae, i.e. the models were able to predict Limnephilidae as present based on the combination of input variables. In an additional study, a positive effect was generally predicted on the presence of the leech *Glossiphonia* and the insects *Sialis* and Tabanidae after river habitat restoration and improvement of the physical-chemical water quality. On the other hand, the restoration actions had in general a negative effect on the habitat suitability of the insect Simuliidae. It is demonstrated that a decrease in habitat quality results in a loss of biodiversity. Several restoration projects in Great-Britain and Denmark show that habitat restoration has a positive effect on the macroinvertebrate communities (e.g. Hansen, 1996; Catling, 2001). In addition to the possible ecological consequences, also the aesthetic and recreational consequences can be considered. From this point of view, the restoration actions are even more valuable.

The improved structural quality of the river in project 4 was apparently insufficient to restore the habitat for *Baetis* and Limnephilidae, two indicators of good water quality. Based on the outcomes of the contribution methods in Chapter 7 however, both taxa were most strongly affected by variables describing the stream type (stream order and distance to mouth) and pollution such as dissolved oxygen, conductivity, total phosphorus, nitrate, … which were not changed in this project. For Gammaridae, an improvement of the habitat suitability was predicted after river restoration. Two models however predicted sampling site 25 as unsuitable after weir removal. This is not so unlikely as based on the results in Chapter 7 and the expert knowledge in Chapter 4. The contribution methods revealed that depth and flow velocity had a paramount effect on the prediction of the output. After weir removal these variables were changed in an unfavourable manner (depth increased while flow velocity decreased) for Gammaridae as could be extracted from Table 10.8. This is confirmed by the Bayerisches Landesamt für Wasserwirtschaft (1996) and Verdonschot (1990) who stated that this taxon prefers small running waters with rather high flow velocity.

To get insight into the reference conditions as is necessary for the implementation of the European Water Framework Directive (EU, 2000), the removal of the weir for flood control near the mill Boembekemolen can be seen as an interesting virtual restoration action to be undertaken. By knowing what would be the ecological shifts, one is able to get insight into the reference communities, but also whether it is really worth to consider such a restoration striving to 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 rather than a nature conservation point of view.

10.5. Conclusions

To convince the water managers and stakeholders of the validity and practical usefulness of the developed models, some simulations of river restoration scenarios were performed. These scenario analyses can help to decide which restoration options to choose. In this way, these exercises can be interesting as validation instrument and can give insights into shifts of indicator organisms. Based on ANN models however, 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 four simulation exercises learned us that depending on the type of the problem, the added value of the models can differ significantly. If the values of the key environmental variables for a particular taxon are not altered during restoration (Chapter 7), the ANN models will not predict any change in habitat suitability. In this way, the restoration actions would seem to have no biological effect although water or habitat quality improved. In addition, Tubificidae, Limnephilidae and in a smaller extent *Baetis* were always predicted as respectively present and absent. One can wonder if it is worth to develop data driven models for these taxa. One could decide to use only the models of Asellidae and Gammaridae, which gave good results, or one can use knowledge based models such as Bayesian Belief Networks or fuzzy logic. The latter models are mainly based on expert knowledge. In this way, the lack of illustrative cases of presence or absence for respectively Limnephilidae and Tubificidae can be avoided. However, it is important to include as many taxa as possible to be able to describe the changes in community composition and diversity. The simulations revealed that validating the models for such exercises itself is also difficult, and the only way to really validate the models is to follow up these restorations in case they are practically realized. Perhaps the use of artificial rivers could provide an answer as well.

To be able to make a full assessment of the overall ecological effects, these types of studies need much more models concerning other biological communities (other macroinvertebrate taxa, fish, macrophytes, …). 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 into the interactions that take place as a result of changing habitat characteristics or changing pollution levels.

The results also illustrate that migration models can provide a real surplus value to the habitat suitability models since they can give an idea about the possibilities and expected time to recolonize restored river sections. This can guide water managers through the different restoration options and the decisions to be made. Nevertheless, also these models can be optimized as discussed in the previous Chapter 8.

General discussion and further research

This thesis was initiated with the objective of developing and practically exploring data driven models based on Artificial Neural Networks to predict the habitat suitability of macroinvertebrates and migration models to simulate the possibility of migration and recolonization of the taxa modelled. Both model types were developed and evaluated on the basis of data monitored in the Zwalm river basin (Flanders, Belgium). This study had a threefold purpose:

- the development of data driven habitat suitability models based on Artificial Neural Networks for macroinvertebrates in the Zwalm river basin. In addition, insight was gained into the ‘black box’ models applying three input variable contribution methods, often referred to as sensitivity analysis;
- the development of migration models for the taxa modelled in this thesis with the aim of investigating the migration and recolonization possibilities within the study area. Therefore, the Cost Weighted Distance tool of the ArcGIS Spatial Analyst extension was used;
- the practical evaluation of the developed models. To this end, the impact of four restoration projects was predicted on the macroinvertebrate habitat suitability.

Detailed results of the different approaches to the modelling problems have been covered in previous chapters. In the following paragraphs, in a more general perspective each of the above aims will be discussed in the light of the results obtained during this study. Finally, at the end of this chapter a number of thoughts and considerations are given for future research in this field.

General Discussion

Design of a monitoring strategy suitable for data driven habitat suitability modelling

Before data driven models can be applied, a reliable dataset, including relevant input and output variables, has to be available. Although a lot of data have been gathered in Flanders on the river systems, still some gaps are to be filled before these data meet the requirements of the modelling objectives. First of all, the data are spread over different institutes in Flanders using various format types, different co-ordinates, ... The existing monitoring networks of the Flemish Environment Agency (VMM) and AMINAL, both responsible for ecological river management in Flanders, are not appropriate because the monitoring approach is not adjusted to the requirements of the data needed within the aim of this thesis focussing on habitat suitability modelling.

To this end, a new dataset was set up containing 60 sampling sites equally spread over the Zwalm river basin. A broad range of all variables was measured over a four-year period. However, based on this dataset, model performances were sometimes highly variable depending on the macroinvertebrate taxa modelled, as was shown in this thesis. To get a better insight into what type of data would be needed in the future, and where these extra data could be collected in order to improve the model performance and its practical applicability, maps should be constructed indicating the ANN model residuals (observed minus predicted output value of presence/absence or abundances) (e.g. Dedecker *et al.*, 2005a; Goethals, 2005). Both studies revealed for example that the largest quantity of errors (especially underestimations of the observed occurrence) for Gammaridae were mainly encountered in the river Zwalm, while the presence/absence predictions for Asellidae contained most errors in the upstream parts scattered all over the river basin. Based on these residuals, new sampling points for particular organisms or model objectives were identified. In this manner, an even more practically directed sampling strategy could be set up. Also sites from other river basins could be added to the future data collection in order to develop more general models.

Collection of environmental variables as input for the habitat suitability models: selection of relevant input variables and data analysis

As in any prediction/forecasting model, the selection of appropriate model inputs is extremely important (Kaastra and Boyd, 1995; Faraway and Chatfield, 1998). The old saying ‘‘garbage in garbage out’’ can be applied to ANNs and a significant amount of time must be spent to perform the task of knowledge acquisition (Walczak and Cerpa, 1999). Specification of the appropriate environmental input variables is in this way an important but complex issue (Soulie, 1994). Pakath and Zaveri (1995) claim that ANNs as well as other artificial intelligence techniques are highly dependent on the specification of input variables. However, two problems may occur when selecting the input variables of the model: too many or too little environmental variables can be included to specify the relation between input (environmental river conditions) and output (habitat suitability of macroinvertebrates). Hertz *et al.* (1991) stated that the ANN performance improves as additional knowledge is provided through the input variables. This is certainly true to the extent that if a sufficient amount of information representing critical decision criteria is not given to an ANN, then the ANN (or any other modelling technique) cannot develop a correct model. The common belief is that since ANNs are learning, they will be able to determine those input variables that are important (cf. Chapter 7) and develop a corresponding model through the modification of the weights associated with the connections between the input layer and the hidden layers (Lek and Guégan, 1999). In addition, numerous variables are involved in the functioning of aquatic ecosystems and in community organisation, so that variables cannot a priori be omitted without a probable loss in information. However, Tahai *et al.* (1998) and others (e.g. D'heygere *et al.*, 2005b) claim that noise input variables produce poor generalization performance and that the presence of too many input variables can cause a decrease of the ANN model performance. In addition, the ANN models 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 datasets which are rather small, but to prove this, more research is needed). In other words, presenting a large number of inputs to the ANN models, and relying on the network to determine the critical model inputs, will usually increase the network size and after all only a limited set of variables are really used to predict the output (cf. Chapter 7). Rigorous considerations are thus needed for detecting which variables are really relevant and which ones can be neglected (Auger *et al.*, 2000).

Based on expert knowledge, literature reviews and consultation of domain experts, in this thesis 24 input variables were selected to predict the habitat suitability of macroinvertebrates as mentioned in Chapter 4. These variables include physical-chemical as well as structural river characteristics. According to Konings and Meire (2003), a conceptual relationship exists between these structural variables and the biotic quality of the river. In addition, they seemed to be crucial in the case studies of the Zwalm river basin since the water quality considerably improved due to investments in sewerage and wastewater treatment plants during the past fifteen years (VMM, 2003). Therefore, restoration projects are currently more directed towards restoration of the structural features of the river (e.g. reintroducing meanders and natural banks, …).

As discussed in Section 5.4, several theoretical reasons can be given to remove variables or instances (e.g. presence of outliers, high correlation between variables, …), but also quite some practical reasons to keep them in. To make the models however applicable in the broadest number of theoretical (e.g. model optimization) and practical (e.g. effect prediction of river restoration scenarios) cases, and to make a tryout on data that were as natural as possible, these data were kept in the dataset. As such it was possible to check whether these data driven models could deal with these kind of data as is sometimes referred to by ANN experts (e.g. Maier and Dandy, 2000).

On the other hand, more variables should be necessary to cope with all kinds of relevant ecological processes to cover different aspects in river management (e.g. what is the effect of pesticides or metals in the Zwalm river basin on the stream ecology). Besides a basic set of variables (water temperature, dissolved oxygen concentration, pH, COD, ammonium, nitrite, nitrate, ortho-phosphate, total phosphorus, chlorides and conductivity), the Flemish Environment Agency also monitors particular variables (BOD, Kjeldahl nitrogen, sulphates, total hardness, suspended solids and the heavy metals As, Ba, Cd, Cr, Cu, Fe, Hg, Mn, Pb, Se, Ni and Zn) in a selected number of sites (VMM, 2003). In addition, 67 pesticides (31 belong to the organochlor pesticides, 17 to the organophosphorus pesticides and 19 to the organonitrogen pesticides) and several other micro-organic pollutants (e.g. mono- and polycyclic aromatic compounds) are measured (VMM, 2003). Most of the last-mentioned variables were not considered in this study, except for suspended solids, which can be of importance for the habitat suitability of macroinvertebrates. Goethals (2005) included eight metal concentrations to predict the habitat suitability of Asellidae and Gammaridae in the river sediments of Flanders. None of these elements however played a major role in the prediction of the presence/absence of both taxa based on ANNs. One possible reason for this could be the fact that bio-availability of the metals was not taken into account. On the other hand, it is very important that these data cover a range broad enough 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 to the model developed. 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 simulating restoration options and predicting reference conditions. Therefore, probably more data from other river basins (e.g. international data) will be needed to improve the models described in this study.

Collection of macroinvertebrates as output for the habitat suitability models: sampling methods, sampling variability and identification levels

The collection of macroinvertebrates was based on two methods: the hand net method in the shallow river stretches (De Pauw and Vanhooren, 1983; IBN, 1984) and artificial substrates in the deeper and un-wadable river stretches (De Pauw and Vanhooren, 1983; De Pauw *et al.*, 1994). Both techniques are widely used and accepted, not only in Flanders, but also abroad. In some cases, however, also other tools like the Van Veen grab and the surber net are being applied. The first one is mainly used to take sediment samples in order to assess the river sediment quality (e.g. the Belgian Sediment Index, De Pauw and Heylen, 2001). The surber net is proposed for sampling according to the AQEM procedure (AQEM consortium, 2002). Both type of samplers provide better quantitative monitoring results than the hand net method. In this way, quantitative predictions of abundances should be more reliable. However, they were less appropriate in the present study. The Van Veen grab sampler considers indeed only the true ‘benthic’ macroinvertebrates while the surber sampler is merely developed for shallow waters with a strong current (Gabriels *et al.*, 2003). Additionally, Van de Walle (2004) revealed in a comparative study (within the framework of an inter calibration exercise for the European Water Framework Directive) that in 70 % of the sampling sites more taxa were found with the hand net method. This was confirmed by Gabriels *et al.* (2003).

Whatever method is chosen however, vital is to have some idea of the sources and effects of macroinvertebrate sampling variation. A method for the assessment of the ecological quality or an ecological model based on these data is of little value without some knowledge of the levels of ‘uncertainty’ of the sampled data. The variation of the observed fauna and the biotic index values could arise from sampling variation, sample processing, identification errors, seasonal changes or differences in weather conditions. Biologists working with benthic macroinvertebrates are since long aware of the problems of variability in what they measure and the importance of replication (Norris and Georges, 1993). Several studies have assessed sampling variation in biotic indices through replicated studies in single rivers (e.g. Armitage *et al.*, 1995; Wright, 2000; Li *et al.*, 2001). Nevertheless, the consequences of between-operator variability have rarely been addressed (Clarke *et al.*, 2002). Furse *et al.* (1981) and Mackey *et al.* (1984) indicated significant differences with respect to operators. In an exercise with students conducted by Alba-Tercedor and Sánchez-Ortega (1988) along a stream in the Sierra Nevada Mountains of southern Spain, different results in the sampling of macroinvertebrates were obtained in accordance to their difference in experience. All students detected the same trend of a decrease in water quality downstream, but with very different scores, so that clear disfunctions appeared when scores were translated to water quality class significances. Also the study described in Section 4.4 revealed small dissimilarities between the six samples taken in one site by operator A and B. Qualitative as well as quantitative differences in the macroinvertebrate fauna were detected. Not only did the abundances vary between replicas, but also the macroinvertebrate taxa composition. Nonetheless, qualitative differences in the fauna may be more important than the number of specimens of each taxon collected (Mackey *et al.*, 1984). These small dissimilarities underline the importance of adequate training of all operators involved in the monitoring programmes. The latter is confirmed by Clarke *et al.* (2002) who found between trained-operator influences on sample values that were negligible. Also Fore *et al.* (2001) detected no significant differences between field samples collected by trained volunteers and professionals.

When sorting the material and identifying the macroinvertebrates, some taxa may be missed or misidentified. Bartsch *et al.* (1998) for example subjected all samples to quality control procedures to estimate the accuracy of sorting and taxonomic identification. Sorting error varied by taxon, with small Nematoda and Oligochaeta most commonly missed. Larger, less abundant taxa such as Amphipoda, Ephemeroptera, Odonata, Pelecypoda, and Trichoptera were seldomly missed. Furse *et al.* (1981) detected significant between-operator differences in the number of taxa removed from samples during the sample processing stage. In the study performed under Section 4.4, sample processing and identification was carried out by one person to minimize this type of variation and focus on the sampling variation itself. Misidentification could be minimized if identification is performed by specialists or well trained people. However, more identification errors may be expected if identification to genus or species level is expected or if identification is done by less experienced persons. The use of species, genus or family level identification will depend on the objectives of the studies (Adriaenssens *et al.*, 2004b). If the purpose of a study is simply to detect an impact of a perturbation on macroinvertebrate communities, identification to family level may be used, but ecological interpretation remains hazardous. If the goal of the study is to know the magnitude of community changes, lower taxonomic identifications are imperatively needed. On the other hand, there is an increasing need for rapid and low-cost methods to assess and predict water quality. In this way, identification at a low taxonomic level is difficult and can be performed only by specialists who are becoming less and less numerous. In addition, if there is sufficient knowledge of the taxonomy of some major faunistic groups such as Ephemeroptera, Trichoptera and Plecoptera, other groups like Diptera and Oligochaeta for which few identification keys to species level (or genus) are available, are often neglected (Guerold, 2000). Identification to the family level may also be more appropriate when samples contain many early instar larvae, which are difficult to identify to lower taxonomic levels. Mainly based on cost-related constraints and the lack of sufficient knowledge of all taxonomic groups, it was decided to identify the macroinvertebrates to family or genus level as specified in De Pauw and Vannevel (1991).

Another factor that may lead to variation in the observed fauna could be the temporal variability and seasonal fluctuations in macroinvertebrate composition. This temporal variability refer to aquatic community changes that occur over time because of diurnal and life-cycle changes in organism behaviour or development, and seasonal or annual changes in the environment. Variation in distribution and abundance of macroinvertebrates may be caused by differences in flow-rate (Newbury, 1984), stream size and distance to the source (Minshall *et al.*, 1985), substrate (Minshall and Minshall, 1977), vegetation (Vincent, 1983), and temperature and stream discharge (Bournard *et al.*, 1987; Boulton and Lake, 1992). Seasonal variability of such factors at a site (e.g. Wade *et al.*, 1989) is one of the prominent causes of temporal variation in the macroinvertebrate community (Linke *et al.*, 1999). The phenology of species within a community (Linke *et al.*, 1999) and the move of invertebrates through their life cycle (Reece *et al.*, 2001) may also affect their presence and abundance in the aquatic community throughout the seasons. It is clear that the time scale of sampling, particularly when it is carried out over more than one season, can significantly affect results of bioassessment or an ecological model. Season should be explicitly taken into account in monitoring, assessment and modelling studies, although seasonal variation is currently most often addressed by constraining the time frame of sampling. To get the most valuable assessment for a given site, Linke *et al.* (1999) would recommend visiting and assessing the site in at least two seasons. However, annual sampling may be all that a budget will permit (Resh and McElravy, 1993). In this thesis, data were collected in the same period each year (August-September) in order to minimize temporal variation on the dataset.

Worldwide, government agencies developed quality assurance/quality control programmes (QA/QC) and audit schemes in order to ensure that operators have been trained well and have the necessary expertise to follow appropriate procedures, that methods for data collection are standardized, that these methods are of consistent and high quality, and that the quality is maintained throughout the duration of the assessment project. Quality assurance and control should pervade all aspects of an ecological study: study design, field operations, laboratory activities, data analysis and reporting. There are numerous examples of quality control programmes to refer to in the USA (Cuffney *et al.*, 1993; Barbour *et al.*, 1996; Barbour *et al.*, 1999; Moulton *et al.*, 2000), and Europe as well (e.g. EU- STAR project (Standardisation of River Classification) (EVK1-CT-2001-00089) (http://www.eu-star.at/)). The UK government agencies for example, in conjunction with CEH, have developed quality control and external audit schemes whereby very experienced biologists reanalyse a proportion of all RIVPACS samples to assess and quantify sample processing and identification errors made by agency biologists (Furse *et al.*, 1995; Clarke, 2000; Dines and Murray-Bligh, 2000). Also in Flanders, the Flemish Environment Agency developed a quality system, based on the international standard ISO 17025, to standardize sampling, sample and data processing and the determination of the Belgian Biotic Index (VMM, 2003). In addition, the VMM executes internal audits on a yearly basis in order to control the quality of data acquisition, processing and identification of the macroinvertebrates (Heylen *et al.*, 1999).

In addition, the reliability of the macroinvertebrate monitoring results is of high importance for the development and validation of habitat suitability models based on data driven techniques (e.g. Artificial Neural Networks). In this context, the predictions of reference conditions and as a consequence the prediction of indicators of a good water quality are of major importance in river management (cf. the European Water Framework Directive). However, these sensitive species (e.g. Limnephilidae) are in general less abundant. As a consequence it is more difficult to catch them, as demonstrated in Dedecker *et al.* (2005e). In this way, the monitoring efficiency can strongly influence the reliability of the data driven models. This is further demonstrated in the next paragraphs.

ANN model development

An important difficulty that ecological processes can create for presence/absence models is that some of the negative locations (sites where a taxon was not observed in the field) may be similar, and possibly identical, to positive locations (sites were a taxon was observed in the field). It is almost inevitable that species will be restricted to a few locations, so that only a small proportion of the potentially suitable sites will be occupied (Fielding and Bell, 1997). These factors will degrade the performance of the model and result in too many false positives (= commission error) (see confusion matrix Chapter 6). Apparent commission errors are derived from potentially suitable habitats correctly predicted as presence, but that are not demonstrated as such because no confirmation of the species exist there, often called ‘pseudo-absence’ (Anderson *et al.*, 2003). The lack of verification of the species may have various causes (Karl *et al.*, 2002). For example, isolated areas of potentially suitable habitats where no species were found often correspond to historical restrictions or the historical effect of speciation, for example failure of the species to disperse to a region of suitable habitat (Peterson *et al.*, 1999; Peterson and Vieglais, 2001; Anderson *et al.*, 2002a). Therefore, migration models were developed in this study to simulate the migration possibilities from a known source population to a potential habitat (unknown habitats that already exists or new possible habitats created by river restoration actions). Similarly, competition between related species likely restricts many species’ realized distributions (Peterson, 2001; Anderson *et al.*, 2002b). Other biological interactions, such as predation, may also limit some species’ distributions. In addition to historical and biotic causes, apparent commission error can also be derived from inadequate sampling: sites of real presence where the species was not found because they have not adequately been sampled by biologists (Karl *et al.*, 2002). This latter form of apparent commission error has recently been recognized in presence/absence datasets (Boone and Krohn, 1999; Karl *et al.*, 2000; Schaefer and Krohn, 2002; Stauffer *et al.*, 2002) and was also studied by Dedecker *et al.* (2005e). These authors looked at the variability of the macroinvertebrate collection in the Zwalm river basin as discussed in Section 4.4. They revealed that if taking only one sample it was likely to miss indicator organisms for a good water quality, which are in general less abundant. As a consequence, it could be concluded that the reliability of the presence/absence data driven models can be influenced by the monitoring efficiency and that commission errors are likely to occur. However, practical constraints (like lack of time and money) will be hard to overcome if multiple samples on one location have to be taken, especially when one likes to cover a large area (e.g. Flanders). Therefore, from a practical point of view, it was opted to take only one sample.

The ANN models in this thesis were evaluated on the basis of both the overall prediction success (= CCI) and the Cohen’s kappa. From an ecological point of view, the CCI is the most logic performance measure because of its easy interpretation towards water managers and policy makers. As discussed in Chapter 6 however, predictions of very common and rare taxa can be misleading if only the overall prediction success is used as performance measure. Therefore, from a mathematical point of view also the CK was recommended in the literature (e.g. Fielding and Bell, 1997; Manel *et al.*, 2001) to evaluate the performance of presence/absence models. Peterson (2001), Ponder *et al.* (2001) and Stockwell and Peterson (2002) however, stated that both CCI and Cohen's kappa are biased with datasets that lack true absence data because the pseudo-absence sites overestimate the commission error. The fact of the matter is that both performance measures include elements of omission and commission (Fielding, 2002). Nevertheless, a combination of both performance measures is very often proven to be efficient in evaluating presence/absence models (e.g. Fielding and Bell, 1997; Dedecker *et al.*, 2004; D'heygere *et al.*, 2005b; Goethals, 2005) and is as such applied to evaluate the ANN models in this study.

In general, less reliable models were constructed for very common or very rare taxa based on ANN models (respectively Tubificidae and Limnephilidae in this study). This is a well-known problem in presence/absence predictions based on data driven models. For both groups of taxa, habitat suitability models based on expert knowledge, such as fuzzy logic (Zadeh, 1965; Barros *et al.*, 2000; Adriaenssens *et al.*, 2004a) and Bayesian Belief Networks (Pearl, 1988; Adriaenssens *et al.*, 2004c) can be seen as useful alternatives. The development of these model techniques is mainly based on ecological knowledge (Jensen *et al.*, 2000). However, expert knowledge from literature is often contradictory or not available for less studied organisms. Consequently, it can be highly complicated to extract a consistent set of rules. In addition, Bayesian Belief Networks also need a large dataset to determine the conditional probabilities. In this way, Bayesian Belief Networks are difficult to apply to large, general, problems (which is often the case in ecology) because the number of conditional probabilities that must be specified can quickly become extremely large. As a consequence, many probabilities will not be well characterized (Adriaenssens, 2004). Similarly, when the input variables needed to be included in the fuzzy rule-based models become large, the number of rules required in fuzzy logic increases to an unmanageable level (Jensen *et al.*, 2000).

In comparison with predictive modelling techniques previously applied on data from the Zwalm river basin, such as fuzzy logic and Bayesian Belief Networks (Adriaenssens, 2004) and classification trees and ANNs (Goethals, 2005), the ANN models developed in this thesis show a relatively good predictive performance. In the preceding modelling studies however, only 179 instances (monitored over a three year period, 2000-2002) instead of 237 instances in the present study (over a four year period, 2000-2003) were used. In addition, both fore-mentioned authors developed merely models for the crustaceans Asellidae and Gammaridae. The highest predictive performances of these studies are summarized in Table 1. It can be concluded that the predictive performances for Asellidae were very similar for the classification trees and both ANN models, while they were much lower based on the fuzzy logic technique. On the other hand, better predictions were obtained based on the one-layered Bayesian Belief Networks. For Gammaridae, the CCIs and CKs were higher based on the ANN models developed in the present study. However, they were still lower than for the predictions based on the Bayesian Belief Networks. It has to be mentioned however that a large inherent uncertainty is present in these predictions (Adriaenssens, 2004). To this end, using a redundancy in model techniques for certain problems rather than relying on a single technique is likely to be beneficial in practice.

Table 1. The highest CCI (correctly classified instances) and Cohen's kappa (CK) values for the macroinvertebrate taxa Asellidae and Gammaridae based on the different modelling techniques applied in the present study and Adriaenssens (2004) and Goethals (2005)

Study			**CCI (%)**	**CK**
Present study	ANNs	Asellidae	79.8	0.60
	ANNs	Gammaridae	80.2	0.47
Goethals	Classification trees	Asellidae	78.6	0.57
(2005)	Classification trees	Gammaridae	75.1	0.22
	ANNs	Asellidae	81.1	0.62
	ANNs	Gammaridae	75.1	0.15
Adriaenssens	Fuzzy logic	Asellidae	25.0*	
(2004)	Fuzzy logic	Gammaridae	37.0*	
	One-layered BBN	Asellidae	86.0	0.71
	One-layered BBN	Gammaridae	85.0	0.51

* 'percentage correctness' instead of 'correctly classified instances' was used as performance measure for the fuzzy logic models. Both performance measures are highly correlated based on Adriaenssens (2004).

Although a lot of experience is already gained in ecological modelling, still many problems have to be overcome (Jorgensen, 1999). To this end several modelling techniques have been developed. A few examples are summarized here: use of fuzzy models to overcome the problem of a poor data base (Jorgensen, 1994a), use of chaos and fractal theory in modelling to improve the parameter estimation (e.g. Jorgensen, 1995), use of catastrophe theory in modelling as an attempt to model structural changes (Jorgensen, 1997), use of artificial intelligence in parameter estimation (Todorovski *et al.*, 1998), use of recently developed parameter estimation methods (Jorgensen, 1998), development of structural dynamic models by use of goal functions to account for the ecosystem properties (Jorgensen, 1986, 1988, 1999; Jorgensen and de Bernardi, 1997), development of Adaptive Agents which are capable to produce emergent behaviour and which can explore a wide range of ecosystem phenomena involving succession, adaptation and evolution (Holland and Miller, 1991; Recknagel, 2002, 2003), development of support vector machines as pattern recognition tool (Morris *et al.*, 2001).

Actually, a lot of uncertainty exists on the development of data driven models. In this study, several architectures and training methods were compared, but the outcome was not easy to summarize in a set of simple rules of thumb. 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 as presented in Chapter 2 are often not working well and trial and error is in most cases the only solution to find the most optimal model training for data driven techniques such as 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) were the major factors to develop well performing models. 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 for ANN models … A reference book for good modelling practice is published by STOWA/RIZA (1999), but this work was only concentrated on deterministic models and did not take into account data driven models. Such documents could make these ANN models 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 (Goethals, 2005). To this end, the HarmoniQuA project (Harmonising Quality Assurance in model based catchment and river basin management) (EVK2-CT2001-00097) (www.HarmoniQuA.org), a European Union-related research project contributing to the implementation of the WFD, has been launched and is aiming at providing a user friendly guidance and quality assurance framework for use in model based river basin management. This project will contribute towards enhancing the credibility of catchment and river basin modelling.

Automated development (via stochastic searches and other optimization algorithms), eventually followed by rule extraction was beyond the scope of this thesis but might look like a valuable option for the future. Algorithms based on evolutionary programming and genetic algorithms have been used successfully to determine optimal network architecture (e.g. Fang and Xi 1997; Kim and Han 2000; Zhao *et al.*, 2000; Wicker *et al.*, 2002). This is already partially worked out by D'heygere *et al.* (2003, 2005b) for classification trees and ANN models. However, so far only input variable selection was performed on the Zwalm river basin dataset. Also larger databases will certainly be necessary to be able to use these optimization techniques in a valid manner.

Three input variable contribution methods were applied to unravel the 'black box' neural networks and to retrieve in this manner the ecological meaning of the ANN model predictions. The outcomes were sometimes highly variable for the different contribution methods and taxa in the applied databases. Also the extracted ecological knowledge did often not present new insights, but in most cases however a confirmation of basic ecological expert knowledge was obtained (e.g. relations with physical habitat such as width, distance to mouth, …). On the one hand, it is confirmed that these techniques can extract thorough ecological knowledge, but on the other hand the added value seems rather limited if a lot of studies are performed yet and expert knowledge is available. However, for taxa for which no or limited knowledge is available, this can already mean an important step forward. In addition, these contribution methods are very useful to reveal the predictive behaviour of the ANN models. It could indeed be concluded that the combination of the PaD method with the Profile method gives a fairly correct idea of the ecological meaning of the models and their practical relevance for decision support in river management.

Nevertheless, Olden *et al.* (2004) stated that a comparison of the different contribution methodologies is not valid if empirical datasets (e.g. relating the presence/absence of macroinvertebrates to a set of local habitat variables) are used. Using empirical datasets precludes the ability to establish generalizations regarding the 'true accuracy and precision' of the different approaches because the 'true importance of the variables' is unknown. The fact that the dataset of the Zwalm river basin has been analysed in a number of other studies (e.g. Adriaenssens, 2004; Goethals, 2005) does not change the fact that the true correlative characteristics of the entire population (i.e. all macroinvertebrates and their associated habitat conditions in the entire river basin) remain unknown. As such, these data are only representative of a sample of the entire population. To this end, Olden *et al.* (2004) compared these methodologies using a Monte Carlo simulation experiment with data exhibiting defined (and consequently known) numerical 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. The results of Olden *et al.* (2004) show that a Connection Weight approach that uses raw input-hidden and hidden-output connection weights in the neural network provides the best methodology for accurately quantifying variable importance and should be favoured over other approaches commonly used in the ecological literature.

Migration model development

Since water quality in Flanders gradually improves (VMM, 2003), recolonization of new suitable habitats gets more and more of importance. However, accessibility of these restored habitats can be restricted due to several migration barriers along the river (Monden *et al.*, 2005). In this way, source populations of certain species can be isolated. As a result, migration to (new) potential habitats can be limited or even impossible. Because ANN models only predict the effect of (altered) environmental river characteristics on the biology (i.e. the habitat suitability), the development of migration models could be useful. These models are able to check the migration possibilities from known source populations to potential habitats. In this way, they can be of important additional value for river restoration management. In addition, they can direct the attention on the conservation of valuable source populations which are of major importance as starting point for the ecological recovery of the watercourses.

To model the migration and recolonization possibilities of Gammaridae, *Baetis* and Limnephilidae in the Zwalm river basin, the Cost Weighted Distance function was applied. The setting of resistance values in the resistance layer is biologically probably the most important step in the process of calculating the cost of migration (Adriaensen *et al.*, 2003). As mentioned before, setting the resistance values was a difficult process in which expert judgement and data available in literature played an important role. Although a theoretical validation of the resistance values was done based on a sensitivity analysis (the resistance values ascribed to the different determining variables were altered to investigate the changes in the resulting accumulative costs and the sensitivity of the developed migration model on variation of the applied resistance values), a more in depth validation of the migration models is recommended in the future. To this end, the use of artificial rivers could be helpful. On the other hand, analysing the genetic diversity or marking and tracing the macroinvertebrates could help to identify their real migration potentials and colonization routes. However, these experiments are often expensive and not always that realistic and reliable. Validation of the predictive results in the field is therefore indispensable. This field validation could be based on nets placed along the river or visual observations of migrating adults. Migration through the water could be observed based on nets placed in the water. Unfortunately, it appeared quasi-impossible to include these validation steps in the framework and time frame of the present thesis. Nevertheless, these steps have to be considered during future research.

The migration model using the cost-distance function has a whole range of applications. Not only the migration possibilities of observed macroinvertebrates within the study area can be modelled but also the extension to other species including nearly extinct as well as invasive exotic species (e.g. *Dikerogammarus villosus* (Crustacea, Amphipoda), *Corbicula fluminea* (Mollusca, Bivalvia)) will prove to be of major importance in river restoration management. Also effects of certain interventions in the river (like weir removal or remeandering projects) in view of river management planning can be judged in a more reliable and integrated manner compared to the local habitat suitability models. Similar strategies will also have to be developed for fish, seen the importance of migrating species (Monden *et al.*, 2005). In addition, the scale of the developed models will have to be extended to the whole Zwalm river basin and maybe beyond. Extension of the intensive monitoring campaign as done for the selected parts of the Zwalm river basin, would however be very costly and time consuming. Therefore, using aerial photographs and remote sensing techniques in combination with digital maps to extract the necessary information would be recommendable.

Model applications for decision support in water management

The first steps towards a reliable model application for decision support in water management have been mentioned and discussed before. These include among others a clear description of the objectives, appropriate data collection, analysis of the available data, examination of the model structure and parameters, model validation, … These can be summarized as the 'credibility' of the models (van der Molen, 1999). The term credibility is defined and used in several ways as:

- 'a property which depends upon the success in all phases of the model-building procedure' (Young, 1983);
- 'a sufficient degree of belief in the validity of the model to justify its use for research and decision making' (Rykiel, 1996). The latter author relates credibility to the amount of knowledge available, the purpose of the model and the consequences of any decision based on it;
- 'a model becomes credible if the model went through the procedure of systems analysis and if the uncertainties involved in the predictions have been considered' (van der Molen, 1999).

Beside this certification of 'good modelling practice' however, the analysis of the requirements from the managers' and stakeholders' point of view is also 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). There is however no standard tendency to use models for river management, in particular in Flanders (Goethals, 2005). 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. To this end, van der Molen (1999) made a distinction between 'credibility' and 'acceptability'. After the credibility of the model has been specified, the model may be either accepted or rejected. So, credibility is the more technical appropriateness of the model while acceptability is the perception of the managers of its practical value. Mostly modellers perform model analyses, but managers also use model results. They have to judge if model results are acceptable to be used in decision-making. However, involving users (river managers) in the model development process is not easy as was demonstrated in the COST 626 project 'European Aquatic Modelling Network'. The aim of this project was to stimulate the integrated use of models for decision support in water management. To determine and assess model credibility and to specify the acceptability of the model van der Molen (1999) set up some criteria. These criteria are listed in Table 2. By extension, they have been applied on the developed ANN and migration models in order to test the credibility and acceptability of the models in this thesis (Table 2).

Table 2. Criteria to determine model credibility and to specify the acceptability (this part is normally completed by the manager, stakeholder, …) of a model (after van der Molen (1999)). The criteria are applied on the developed ANN and migration models

Criteria for credibility	**ANN model**	**Migration model**
Objectives of the model are specified and the choice of state variables is in agreement with these objectives	Prediction of the habitat suitability of macroinvertebrates in rivers based on environmental river characteristics. The single state variable is the macroinvertebrate presence/absence	Simulation of the migration and recolonization possibilities of macroinvertebrates. The single state variable is the cost to migrate from one point to another
Dimensions of the modelled system and *aggregations* in time and space meet the objectives and the availability of data	Models were applied to the entire Zwalm river basin and the 'short distance' monitoring network	Models were applied to the 'short distance' monitoring network
The available *data* are utilized sufficiently and the system identification is not hampered by lack of input data and observations; uncertainties in the data are considered	Input variables are summarized in Chapter 4 and include physical-chemical as well as structural river characteristics. Data analysis is performed as described in Chapter 5	Input variables are summarized in Chapter 8 and include environmental and river characteristics influencing migration of the considered taxa. Uncertainty could partly be retrieved from the sensitivity analysis in Chapter 9
During sensitivity analysis, parameter estimation and validation, the appropriateness of the *model structure* is examined	Model architecture is examined based on comprehensive parameter estimation (Chapter 6) en sensitivity analysis (Chapter 7)	The model structure of the migration model was fixed and was not examined in this way
Model *parameters* are fixed at well documented values or are properly estimated	The optimal model parameters (e.g. learning rate, number of hidden neurons, …) were fixed or estimated in Chapter 6	Model parameters (e.g. grid size, …) were fixed as specified in Chapter 8
Model *validation* is based on an independent set of observations and the results are quantified and related to the objectives	Model validation was based on *k*-fold cross-validation as specified in Chapter 6	Theoretical model validation was based on sensitivity analysis (Chapter 9)

Continuation of Table 2. Criteria to determine model credibility and to specify the acceptability (this part is normally completed by the manager, stakeholder, …) of a model (after van der Molen (1999)). The criteria are applied on the developed ANN and migration models

Criteria for credibility	**ANN model**	**Migration model**
The *uncertainties* in model structure, model parameters and model predictions are addressed and quantified to a certain extend	Uncertainties are encountered using standard deviation measures on model predictions and sensitivity analysis	The uncertainty in model output is acceptable for speculations and gross conclusions on the migration possibilities of the considered taxa and was addressed based on the sensitivity analysis in Chapter 9
Criteria for acceptability (specified by the river managers)	**ANN model**	**Migration model**
The *motivation* for the initiation of a modelling project is known	The motivation to develop the habitat suitability models is to fill this gap in river management in Flanders so far. In this manner, the ecological effects of river restoration projects can be supported in a more scientific way	Since ANN models do not take accessibility of the restored river sections into account, migration models can fill this gap. They are able to investigate the connectivity between populations or the possibility to migrate from a source population to a restored river section
Constraints in time and money for model development and applications are specified	No constraints in time and money were specified for the development of the models	No constraints in time and money were specified for the development of the models
Arguments and *consequences* for approval or rejection of the model (results) are discussed	Arguments for acceptance/rejection of the model were not specified in advance, but acceptability of the results may be positively affected by the credibility of the modelling approach. In addition, the models are practically evaluated based on the simulation of four river restoration scenarios (Chapter 10)	Arguments for acceptance/rejection of the model were not specified in advance. However, the practical usefulness of the models was demonstrated based on the simulation of river restoration scenarios (Chapter 10)

To convince the river managers and stakeholders of the validity and practical usefulness of the developed models in this thesis, some simulations of river restoration scenarios were performed in the last chapter. These scenario analyses can help to decide which restoration options to choose. In this way, these exercises can be useful as practical validation instruments and give insights into shifts of indicator organisms. However, there are several needs for improvement to make the models practically more reliable, as was clearly shown in Chapter 10. For example, one can wonder if it is useful to include Tubificidae and Limnephilidae in the predictions since the ANN models were not able to predict shifts in habitat suitability of both taxa. On the other hand, there is a need to make simulations of all taxa to detect the shift of the whole community. Also other models can be included to predict the input variables of these habitat suitability models, such as land use, water quantity and quality models, … These practical case studies also deliver insights into how to improve the data collection (e.g. where to take additional samples) to develop and validate models that are of practical use for decision support in river management. Currently, in the Zwalm river basin river restoration project which are in progress are followed up. However, results are not available yet.

As mentioned before, to enhance the acceptability of the developed models in decision support, it would be beneficial to include and combine the outcomes of other models. Recently, several practical concepts and software systems were developed related to decision support in environmental and water management (e.g. Booty *et al.*, 2001; Caminiti, 2004; Cuddy and Gandolfi, 2004; Mysiak *et al.*, 2005; Holmes *et al.*, 2005; Bazzani, 2005; Pallottino *et al.*, 2005).

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). In addition, it is of major importance to carefully consider the general and specific application areas of the applied models to avoid a wrong extrapolation or coupling of existing knowledge or models (Ceccaroni *et al.*, 2004). However, when a lot of models are already available, it is probably the best option.

Particularly in Flanders, it would be beneficial to couple the existing hydraulic and river quality models applied by the different governmental administrations and institutions to the ecological habitat suitability models (and migration models) developed in this study. These hydraulic and water quality models could provide useful insights into the effect of river restoration scenarios on the environmental river characteristics used as input variables for the habitat suitability models. The model ISIS (applied by AWZ, AMINAL Water, VLM and several provinces in Flanders) for example could be of major additional value to simulate the hydraulic effects of weir removal or remeandering on the river width, depth, … In addition, this model is able to make a more realistic estimation of the (intension and distance of the) effects in upstream and downstream direction (e.g. the risks for flooding downstream). On the other hand, the water quality models SIMCAT, SENTWA and SEPTWA (applied by VMM) could be used to simulate the effect of reduction of domestic, agricultural and industrial discharges on the river water quality. SIMCAT is a mathematical model used to describe the river water quality throughout a catchment, while SENTWA and SEPTWA are models to evaluate the transport of nutrients and pesticides to the river. In this way, a more reliable and scientific based estimation of the impacts of river restoration actions on the environmental input variables of the habitat suitability models could be made.

In addition, Goethals (2005) described the potential link between ecological models and socio-economical models and stakeholder information needs, since economic valuation can play an important role to analyse the costs and benefits of river restoration options within the policy area of water management.

Main conclusions based on the present thesis

As stated at the beginning of this chapter, the three major objectives of this PhD research were 1) the development of habitat suitability models for macroinvertebrates in the Zwalm river basin based on ANNs, 2) the development of migration models for macroinvertebrates as extension of the habitat suitability models and 3) the prediction of river restoration scenarios as practical evaluation of the developed models. The main conclusion based on the present thesis are summarized below:

1) <u>ANN model development</u>

- the combination of two performance measures (the percentage of Correctly Classified Instances, CCI and the Cohen's kappa) was proven to be efficient evaluating presence/absence ANN models. From an ecological point of view, the CCI is the most logic performance measure because of its easy interpretation towards water managers and policy makers while the Cohen's kappa can be recommended from a mathematical point of view;
- for the macroinvertebrate taxa Asellidae and Gammaridae, good predictive results were obtained based on the ANN models in comparison with other habitat suitability models (classification trees, fuzzy logic and Bayesian Belief Networks). The predictions for the very common taxon Tubificidae and the very rare taxon Limnephilidae were less satisfactory since the ANN models tend to 'learn' that very rare taxa are always absent and very common taxa are always present. On the one hand, supplementary data comprising sites where Tubificidae are absent and Limnephilidae are present, can be included. On the other hand, habitat suitability models based on expert knowledge can be applied to overcome this problem in the future;
- because the size of the training and test set influences respectively the generalization capability of the model and accuracy of the model performance, the optimal size of training and test set was searched for based on the cross-validation procedure. It could be decided that the optimal number of folds was four (three fourth of the data is used for training and one fourth is used for testing) for all five taxa. To this end, one may conclude that 4-fold cross-validation is a good consideration between model generalization capacity and accuracy;
- network architecture is generally known to be highly problem dependent. However, comparing the different model architectures for each taxon in the present study, predictive performances were very similar. In this way, no straightforward conclusions towards the five taxa could be drawn according to the optimal number of hidden layers and neurons;
- the input variable contribution methods applied to the ANN models were useful to select ecologically essential variables to describe the species' habitat(s) and to include these in monitoring campaigns for river assessment. In particular the insight into the sensitivity curves was useful, by showing how the environmental variables affect the biological communities. The input variables ammonium, COD, conductivity and total phosphorus were of major importance in predicting the habitat suitability of Gammaridae, while for

Asellidae, the environmental characteristics describing the stream type (such as width, stream order and distance to mouth) were the most contributing variables. For Tubificidae, it was very difficult to extract the major environmental variables and to make conclusions about environmental preferences. To a smaller extend, the variables nitrate, distance to mouth, total phosphorus, sand and stream order were indicated as key variables. For *Baetis*, beside pollution related variables (ammonium, nitrate, conductivity and dissolved oxygen), also habitat characteristics (suspended solids, temperature, pools-riffles, depth and stream order) seemed to be important to explain the presence/absence. Stream order, dissolved oxygen, total nitrogen and phosphate were the highest contributing variables for Limnephilidae. However, it was often difficult to find major trends for the five taxa in the Zwalm river basin, the three contribution methods and the different folds. The first two can be explained by different ecological preferences of the taxa and by the different aspects the three contribution methods deal with. The instability over the different folds is perhaps related to the relative small size of the datasets in combination with a high variability of the sites, the high number of input variables or outliers in the measurements. This will therefore need further research based on larger datasets and use of sub-samplings. In spite of this, it could be concluded from this study that 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 of river management.

2) <u>Migration model development</u>

- because ANN models only predict the effect of (altered) environmental river characteristics on the biology (i.e. the habitat suitability), the development of the migration models was very useful. These models were able to check the migration possibilities from known source populations to potential habitats (created after river restoration or naturally existing). In addition, they gave an indication of the timing of the expected effects (expressed in days);
- setting the resistance values was a difficult process in which expert judgement and data available in literature played an important role. However, data was very scarce and often incomplete. Therefore, it was difficult to validate the model results in a proper way. However, a first step in the theoretical validation process was made performing a sensitivity analysis varying the resistance values ascribed to the different determining variables. If more resistance classes were used to calculate the accumulative migration

costs for migration through the air / over land and differences between the resistance classes were large enough, more details could be extracted from the migration models. It was also derived that variation in flow velocity values only resulted in small differences in accumulative migration costs for migration through the river. Variation in flow velocity during the year would thus only have a minor impact on the behaviour of the migration models. However, practical field validation studies would still be appropriate but it was quasi-impossible to include these validation studies in the time frame of the present thesis. Nevertheless, these steps have to be considered during future research.

3) Prediction of river restoration scenarios

- the simulations of river restoration scenarios have shown to be useful to convince river managers and stakeholders of the validity and practical usefulness of the developed models;
- however, no shifts in habitat suitability could be predicted for Tubificidae and Limnephilidae. One can wonder if these taxa have to be included for the evaluation of the river restoration actions. One could decide to use only the models of Asellidae and Gammaridae, which gave good results, or one can use knowledge based models such as Bayesian Belief Networks or fuzzy logic. In this way, the lack of illustrative cases of presence or absence for respectively Limnephilidae and Tubificidae can be avoided. However, it is important to include as many taxa as possible to be able to describe the changes in community composition and diversity;
- the four simulation exercises learned us that depending on the type of the problem, the added value of the models can differ significantly. If the values of the key environmental variables for a particular taxon were not altered during restoration, the ANN models did not predict any change in habitat suitability. In this way, the restoration actions would seem to have no biological effect although water or habitat quality improved;
- based on the migration models, an approximation of the migration route and the migration time (expressed in days) to the restored river sections could be given for Gammaridae, *Baetis* and Limnephilidae.

Further research

This PhD research illustrated that habitat suitability models based on Artificial Neural Network models can be an interesting tool to get insight into the relation between environmental river characteristics and macroinvertebrates. On the other hand, the development of migration models for macroinvertebrates has shown to be useful as extension of the habitat suitability models. The combined use of both models can be helpful 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 facing daily .

Several difficulties and research questions remain however unsolved. Therefore, a number of recommendations and considerations for further research are being put forward here.

The need for good datasets was one of the major problems that was encountered during this study. Although a lot of data have been gathered on the river systems in Flanders, there are still some gaps to fill before these data will meet the requirements of our modelling objectives. Therefore, a new monitoring network (in the Zwalm river basin) has been constructed in order to develop reliable and useful predictive habitat suitability models for macroinvertebrates. Nevertheless, the collected dataset was far from optimal. On the one hand, inclusion of new, essential variables, for example pesticides, heavy metals and micro pollutants could be beneficial to explain the habitat suitability of the macroinvertebrates. On the other hand, 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 to the model developed. Concerning the range of the data, in particular in Flanders, there is a major lack of river ecosystems of good ecological status. This makes it difficult to develop well performing models for testing restoration options and predicting reference conditions. Therefore, more data from other river basins maybe abroad (e.g. international data) will be needed to improve the models described in this study. However, also financial and time constraints can play a major role in deciding whether certain data are to be included in the database or not. Rigorous methods are therefore needed to decide which explanatory variables or combinations of variables should be entered in the model. To this end, genetic algorithms have proven to be very efficient to automatically select the relevant input variables for Artificial Neural Networks (e.g. D'heygere *et al.*, 2005b).

Besides the optimization of the dataset, a further optimization and extension of the ANN models and migration models (especially further validation) is required. More macroinvertebrate taxa have to be considered. From a practical point of view, it might be justified 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. Because river managers have extensive experience with these indices, they also will have better understanding of the use of models predicting directly an ecological index. Also the development of habitat suitability models for fish, macrophytes, … will be beneficial for the implementation of the European Water Framework Directive. On the other hand, extension of both migration and habitat suitability models to other river basins in Flanders needs to be taken into account in the future. However, sufficient attention should be paid to the place (chosen sampling sites should include different levels of water and habitat quality, different types of land use, all river types, all river basins and should be easy to reach) and number (on the one hand, enough sites should be included to develop reliable models, on the other hand, the number of sampling sites has to be as low as possible to reduce costs) of sampling sites. Probably, measurements of the Flemish Environment Agency can be integrated in the future.

Nevertheless, the amount of knowledge and data to develop, train and validate these models will probably remain a major bottleneck for their successful application, at least in the near future. To this end, also the coupling of the habitat suitability models based on ANN models to hydraulic (e.g. ISIS) and river water quality models (e.g. SIMCAT, SENTWA and SEPTWA) will be necessary to improve the reliability and the practical applicability of the habitat suitability models. Coupling of terrestrial modelling modules to aquatic habitat suitability models is another future prospect, because many of the considered macroinvertebrate taxa, in particular the insect larvae that have a terrestrial adult stage. In this context, the development of migration models is an important step forwards. Because of the very limited instances in which rare species were present (e.g. Limnephilidae in this study), it is most likely that other habitat suitability models such as fuzzy logic and Bayesian Belief Networks will be more appropriate because the development of these model techniques is mainly based on ecological knowledge. Add to this fact that the availability of proper and reliable expert knowledge is of crucial importance, as well as a good validation set. Also embedding these models in a GIS environment will make it possible to deduce inputs from GIS maps as well as producing outputs on GIS maps. Moreover, the habitat factors can be considered at different scales and an interactive system can be provided for. In this way, decision makers will have the possibility to quickly modify parameters and visualize the results of simulations, which can be part of a more participatory oriented river basin management (Welp, 2001).

Taking the above recommendations and considerations into account, an important step towards the collaboration between modellers and river managers and stakeholders can be made in the near future.

References

Adriaensen, F., Chardon, P., Gulinck, H., De Blust, G., Verhagen, R. and Matthysen, E. (2000). Quantitative evaluation of the connectivity function of landscape elements based on connectivity models. Final report of the Flemish Impuls Program for Nature conservation project (VLINA project 97/01). Brussels, 138 p. (in Dutch).

Adriaensen, F., Chardon, J.P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H. and Matthysen, E. (2003). The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning*, **64**, 233-247.

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

Adriaenssens, V., Dedecker, A., D'heygere, T., Goethals, P.L.M and De Pauw, N. (2003). Relations between structural characteristics and macro-invertebrate communities in the Zwalm river basin at different spatial scales. In: J.-J., Symoens and K., Wouters (Eds.), *Biological evaluation and monitoring of the quality of surface waters*, Cercle Hydrobiologique de Bruxelles, Brussels, 81-86 p.

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

Adriaenssens, V., Goethals, P.L.M., Trigg, D. and De Pauw, N. (2004c). 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.

Adriaenssens, V., Goethals, P.L.M. and De Pauw, N. (2005). Fuzzy knowledge-based models for prediction of *Asellus aquaticus*, *Asellus meridianus* and *Gammarus pulex* in watercourses. *Ecological Modellling*. (accepted).

Alba-Tercedor, J. and Sánchez-Ortega, A. (1988). A rapid and simple method to assess the water quality of running waters based on Hellawell (1978). *Limnética*, **4**, 51-56.

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

Allan, J.D. and Malmquist, B. (1989). Diel activity of *Gammarus pulex* (Crustacea) in a South Swedish stream: comparison of drift catches vs baited traps. *Hydrobiologia*, **179**, 73-80.

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.

Anderson, R.P., Gómez-Laverde, M. and Peterson, A.T. (2002a). Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography*, **11**, 131-141.

Anderson, R.P., Peterson, A.T. and Gómez-Laverde, M. (2002b). Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, **98**, 3-16.

Anderson, R.P., Lew, D. and Peterson, A.T. (2003). Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, **162**, 211-232.

Andrén, H. (1994). Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat. *Oikos*, **71**, 355-366.

Angeline, P.J., Sauders, G.M. and Pollack, J.B. (1994). An evolutionary algorithm that constructs recurrent neural networks. *IEEE Transactions on Neural Networks*, **5**, 54–65.

AQEM consortium (2002). Manual for the application of the AQEM (The Development and Testing of an Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates) system. A comprehensive method to assess European streams using benthic macroinvertebrates, developed for the purpose of the Water Framework Directive. Version 1.0, February 2002.

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

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.

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.

Armitage, P.D., Pardo, I. and Brown, A. (1995). Temporal constancy of faunal assemblages in mesohabitats. Application to management. *Archiv für Hydrobiologie*, **133**, 367–387.

Auger, P., Charles, S., Viala, M. and Poggiale, J. (2000). Aggregation and emergence in ecological modelling: integration of ecological levels. *Ecological Modelling*, **127**, 11-20.

Austin, M.P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.

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.

Baguette, M., Mennechez, G., Petit, S. and Schtickzelle, N. (2003). Effect of habitat fragmentation on dispersal in the butterfly Proclossiana eunomia. *Comtes Rendus Biologies*, **326**, 200-209.

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., Stribling, J.B., Gerritsen, J. and Karr, J.R. (1996). Biological Criteria: Technical Guidance for Streams and Small Rivers, Revised Edition. In: G.R., Gibson (Ed.), EPA 822-B-96-001. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.

Barbour, M.T., Gerritsen, J., Snyder, B.D. and Stribling, J.B. (1999). Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.

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.

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

Bartsch, L.A., Richardson, W.B. and Naimo, T.J. (1998). Sampling benthic macroinvertebrates in a large floodplain river: considerations of study design, sample size, and cost. *Environmental Monitoring and Assessment*, **52**, 425–439.

Bayerisches Landesamt für Wasserwirtschaft (1996). Ökologische Typisierung der aquatischen Makrofauna. Informationsberichte des Bayerischen Landesamtes für Wasserwirtschaft. Heft 4/96.

Bazzani, G.M. (2005). An integrated decision support system for irrigation and water policy design: DSIRR. *Environmental Modelling & Software*, **20(2)**, 153-163.

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.

Belconsulting (2003). Ecological inventory and vision forming within the scope of integrated water management, The Zwalm. Under the authority of AMINAL Department Water, 136 p. (in Dutch).

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

Berthold, M. and Hand, D. (1999). Intelligent Data Analysis, an Introduction. Springer-Verlag, Berlin-Heidelberg, 400 p.

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.

Biswas, A.K. (1991). Water for sustainable development in the 21st century. *Water International*, **16**, 219–224.

Blöch, H. (1999). The European Union Water Framework Directive: taking European water policy into the next millennium. *Water Science and Technology*, **40(10)**, 67-71.

Bohle, H.W. (1978). Relation between food supply, drift and microdistribution of larvae of *Bäetis rhodani*. Investigations in a stream model. *Archiv für Hydrobiologie*, **84**, 500-525.

Boone, R.B. and Krohn, W.B. (1999). Modeling the occurrence of bird species: are the errors predictable? *Ecological Applications*, **9**, 835-848.

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.

Boulton, A.J. and Lake, P.S. (1992). The ecology of two intermittent streams in Victoria, Australia. III. Temporal changes in faunal composition. *Freshwater Biology*, **27**, 123-138.

Bournaud, M. and Cogerino, L. (1986). Les microhabitats aquatiques des rives d'un grand cours d'eau: approche faunistique. *Annales de Limnologie - International Journal of Limnology*, **23(3)**, 285-294.

Bournard, M., Tachet, H., Roux, A.L. and Auda, Y. (1987). The effects of seasonal and hydrological influences on the macroinvertebrates of the Rhone River, France 1. Methodological aspects. *Archives of Hydrobiology*, **109**, 287-304.

Bowden, G.J., Dandy, G.C. and Maier, H.R. (2002). An evaluation of methods for the selection of inputs for an artificial neural network based river model. In: F., Recknagel (Ed.), *Ecological Informatics, Understanding Ecology by Biologically-Inspired Computation*, Springer-Verlag, Berlin-Heidelberg, 215-232 p.

Bowne, D.R., Peles, J.D. and Barrett, G.W. (1999). Effects of landscape spatial structure on movement patterns of the hispid cotton rat (*Sigmodon hispidus*). *Landscape Ecology*, **14**, 53-65.

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.

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.

Breiman, L., Friedman, J.H., Olsen, A. and Stone, C.G. (1984). Classification and Regression Trees. Wadsworth International Group, Belmont, CA, USA.

Breine, J., Simoens, I., Goethals, P., Quataert, P., Ercken, D., Van Liefferinghe, C. and Belpaire, C. (2004). A fish-based index of biotic integrity for upstream brooks in Flanders (Belgium). *Hydrobiologia*, **522(1-3)**, 133-148.

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

Brooker, L., Brooker, M. and Cale, P. (1999). Animal dispersal in fragmented habitat: measuring habitat connectivity, corridor use and dispersal mortality. *Conservation Ecology*, **3(1)**, 4.

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.

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.

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.

Bullock, J.M., Kenward, R. and Hails, R. (2002). Dispersal Ecology. Blackwell Science, Oxford, 458 p.

Bunn, A.G., Urban, D.L. and Keitt, T.H. (2000). Landscape connectivity: A conservation application of graph theory. *Journal of Environmental Management*, **59(4)**, 265-278.

Cairns, Jr.J. and Pratt, J.R. (1993). A history of biological monitoring using benthic macroinvertebrates. In: D.M., Rosenberg and V.H., Resh (Eds.), *Freshwater Biomonitoring and Benthic Macroinvertebrates*, Chapman & Hall, New York, 10-27 p.

Caminiti, J.E. (2004). Catchment modelling - a resource manager's perspective. *Environmental Modelling & Software*, **19(11)**, 991-997.

Catling, C. (2001)**. Site Visit Notes. In: Proceedings of the River Restoration Centre Siminar, River Restoration and Chalk Streams, University of Hertfordshire, Hatfield, 29-34 p.**

Cauwenberghs, K. (2003). Modellering voor watersystemen. *Tijdschrift Water*, **7**, 1-33.

Ceccaroni, L., Cortés, U. and Sànchez-Marrè, M. (2004). OntoWEDSS: augmenting environmental decision-support systems with ontologies. *Environmental Modelling & Software*, **19(9)**, 785-797.

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.

Chardon, J.P., Adriaensen, F. and Matthysen, E. (2003). Incorporating landscape elements into a connectivity measure: a case study for the Speckled wood butterfly (*Pararge aegeria* L.). *Landscape Ecology*, **18**, 561-573.

Chave, P. (2001). The EU Water Framework Directive, an introduction. IWA Publishing, London, 208 p.

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: F., Recknagel (Ed.), *Ecological Informatics, Understanding Ecology by Biologically-Inspired Computation*, Springer-Verlag, Berlin-Heidelberg, 127-178 p.

Ciborowski, J.J.H. (1983). Influence of current velocity, density, and detritus on drift of two mayfly species (Ephemeroptera). *Canadian Journal of Zoology*, **61**, 119-125.

Clarke, R.T. (2000). Uncertainty in estimates of river quality based on RIVPACS. In: J.F., Wright, D.W., Sutcliffe and M.T., Furse (Eds.), *Assessing the Biological Quality of Freshwaters. RIVPACS and Other Techniques*, Freshwater Biological Association, Ambleside, 39-54 p.

Clarke, R.T., Furse, M.T., Gunn, R.J.M., Winder, J.M. and Wright, J.F. (2002). Sampling variation in macroinvertebrate data and implications for river quality indices. *Freshwater Biology*, **47**, 1735-1751

Clarke, R.T., Wright, J.F. and Furse, M.T. (2003). RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling*, **160**, 219-233.

Clobert, J., Danchin, E., Dhondt, A.A. and Nichols, J.D. (2001). Dispersal. Oxford University Press, Oxford, 480 p.
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20(1)**, 37-46.

Conradt, L., Roper, T.J. and Thomas, C.D. (2001). Dispersal behaviour of individuals in metapopulations of two British butterflies. *Oikos*, **95**, 416-424.

Cook, D.F., Ragsdale, C.T. and Major, R.L. (2000). Combining a neural network with a genetic algorithm for process parameter optimization. *Engineering Applications of Artificial Intelligence*, **13**, 391-396.

Corkum, L.D., Pointing, P.J. and Ciborowski, J.J.H. (1977). The influence of current velocity and substrate on the distribution and drift of two species of mayflies (Ephemeroptera). *Canadian Journal of Zoology*, **55**, 1970-77.

Corkum, L.D. and Clifford, H.F. (1979). The importance of species associations and substrate types to behavioural drift. In: J.F., Flannagan and K.E., Marshall (Eds), *Advances in Ephemeroptera Biology*, Plenum Publishing Corporation, New York, 331-341 p.

Cuddy, S.M. and Gandolfi, C. (2004). Integrated catchment modelling and decision support. *Environmental Modelling & Software*, **19(11)**, 989-990.

Cuffney, T.F., Gurtz, M.E. and Meador, M.R. (1993). Guidelines for processing and quality assurance of benthic invertebrate samples collected as part of the national water quality assessment program. United States Geological Survey, Report No. 93-407. 80 p.

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

Cuppen, H.P.J.J. (1980). De macrofauna in een aantal droogvallende en permanente stilstaande wateren in het ruilverkavelingsgebied Brummen-Voorst. Regionale milieu-raad Oost-Veluwe, Nederland.

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.

Davies, P.E. (2000). Development of a national river bioassessment system (AUSRIVAS). In: J.F., Wright, D.W., Sutcliffe and M.T., Furse (Eds.), *Assessing the biological quality of freshwaters: RIVPACS and other techniques*, Freshwater Biological Association, Ambleside, 79-82 p.

Davis, W.S. and Simon, T.P. (1995). Biological Assessment and Criteria Tools for Water Resource Planning and Decision Making. Lewis Publisher, Boca Raton.

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. *TheScientificWorldJOURNAL*, **2**, 96-104.

Dedecker, A., Goethals, P.L.M., Gabriels, W. and De Pauw, N. (2004). 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., D'heygere, T. and De Pauw, N. (2005a). Development of an in-stream migration model for *Gammarus pulex* L. (Crustacea, Amphipoda) as a tool in river restoration management. *Aquatic Ecology*. (submitted).

Dedecker, A.P., Goethals, P.L.M., D'heygere, T., Gevrey, M., Lek, S. and De Pauw, N. (2005b). Selecting variables of habitat suitability 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. (2005c). Application of Artificial Neural Network models to analyse the relationships between *Gammarus pulex* L. (Crustacea, Amphipoda) and river characteristics. *Environmental Monitoring and Assessment*. (in press).

Dedecker, A.P., Goethals, P.L.M. and De Pauw, N. (2005d). Sensitivity and robustness of stream model based on artificial neural networks for the simulation of different management scenarios.. In: S., Lek, M., Scardi, P.F.M., Verdonschot, J.-P., Descy and Y.S., Park (Eds.), *Modelling community structure in freshwater ecosystems*. Springer-Verlag, 133-146 p.

Dedecker, A.P., Janssen, K., Wouters, H., Thas, O., Goethals, P.L.M. and De Pauw, N. (2005e). Explaining the variability in data collection for river assessment purposes. *Freshwater Biology*. (in preparation).

De Loose, L., Verheyen, R.F., Bervoets, L., Weiss, L., Schneiders, A., Wils, C., Van Maele, T.M., De Pauw, N., Baillieul, M., Blust, R. and Vanderborght, O. (1995). Ecologische kwaliteitsdoelstellingen voor oppervlaktewater. Studie uitgevoerd in opdracht van de Vlaamse Milieumaatschappij. Groep voor Toegepaste Ecologie (GTE). Antwerpen, België, 185 p. + bijlagen.

De Pauw, N. (2000). Using RIVPACS as a modelling tool to predict the impacts of environmental changes. In: J.F., Wright, D.W., Sutcliffe and M.T., Furse (Eds.), *Assessing the Biological Quality of Freshwaters. RIVPACS and Other Techniques*, Freshwater Biological Association, Ambleside, 39-54 p.

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

De Pauw, N. and Vannevel, R. (1991). Macroinvertebrates and water quality. Stichting Leefmilieu. Dossier N° 11, Antwerp, 316 p. (in Dutch).

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

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

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 Heylen, S. (2001). Biotic index for sediment quality assessment of watercourses in Flanders, Belgium. *Aquatic Ecology*, **35(2)**, 121-133.

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. *TheScientificWorldJournal*, **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. (2005a). Use of genetic algorithms to select input variables in neural network models for the prediction of benthic macroinvertebrates. *Aquatic Ecology*. (submitted).

D'heygere, T., Goethals, P.L.M. and De Pauw, N. (2005b). Genetic algorithms for optimisation of predictive ecosystems models based on decision trees and neural networks. *Ecological Modelling*. (in press).

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.

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.

Dines, R.A. and Murray-Bligh, J.A.D. (2000). Quality assurance and RIVPACS. In: J.F., Wright, D.W., Sutcliffe and M.T., Furse (Eds.), *Assessing the Biological Quality of Freshwaters. RIVPACS and Other Techniques*, Freshwater Biological Association, Ambleside, 71-78 p.

Efron, B. (1983). Estimating the error rate of a prediction rule: some improvements on cross-validation. *Journal of the American Statistical Association*, **78**, 316–331.

Elbersen, J.W.H., Verdonschot, P.F.M., Roels, B. and Hartholt, J.G. (2003). Definitiestudie -Kaderrichtlijn Water (KRW). I. Typologie Nederlandse Oppervlaktewateren. Alterra, Wageningen, 61 p.

Ellien, C., Thiebaut, E., Dumas, F., Salomon, J.C. and Nival, P. (2004). A modelling study of the respective role of hydrodynamic processes and larval mortality on larval dispersal and recruitment of benthic invertebrates: example of *Pectinaria koreni* (Annelida : Polychaeta) in the Bay of Seine (English Channel). *Journal of Plankton Research*, **26(2)**, 117-132.

Elliott, J.M. (1971a). Upstream movements of benthic invertebrates in a Lake District stream. *Journal of Animal Ecology*, **40**, 235-252.

Elliott, J.M. (1971b). The distances travelled by drifting invertebrates in a Lake District stream. *Oecologia*, **6**, 350-379.

Elliott, J.M. (2002a). A continuous study of the total drift of freshwater shrimps, *Gammarus pulex*, in a small stony stream in the English Lake District. *Freshwater Biology*, **47**, 75-86.

Elliott, J.M. (2002b). Time spent in the drift by down-stream-dispersing invertebrates in a Lake District stream. *Freshwater Biology*, **47**, 97-106.

Elliott, J.M. (2002c). The drift distances and time spent in the drift by freshwater shrimps, *Gammarus pulex*, in a small stony stream , and their implications for the interpretation of downstream dispersal. *Freshwater biology*, **47**, 1403-1417.

Elliott, J.M. (2003). A comparative study of the dispersal of 10 species of stream invertebrates. *Freshwater Biology*, **48**, 1652-1668.

Elliott, J.M., Humpesch, U.H. and Macan, T.T. (1988). Larvae of the British Ephemeroptera: A key with ecological notes. Freshwater Biological Association, Scientific Publication N°49, 147 p.

Englund, G. and Hambäck, P.A. (2004a). Scale-dependence of movement rates in stream invertebrates. *Oikos*, **105(1)**, 31-40.

Englund, G. and Hambäck, P.A. (2004b). The scale-dependence of emigration rates. *Ecology*, **85(2)**, 320-327.

Erman, N.E. (1986). Movements of elf-marked caddisfly lavae, *Chyrandra centralis* (Trichoptera: Limnephilidae), in a sierran spring stream, California, USA. *Freshwater Biology*, **16**, 455-464.

Erni, B., Liechti, F. and Bruderer, B. (2003). How does a first year passerine migrant find its way? Simulating migration mechanisms and behavioural adaptations. *Oikos*, **103(2)**, 333-340.

ESRI (2001). ArcGIS Desktop. Environmental Systems Research Institute. Redlands, California.

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. European Union, The European Parliament, The Council, PE-CONS 3639/1/00 REV 1 EN, 62 p. + annexes.

EU (2001). Strategic document: Common strategy on the implementation of the Water Framework Directive. 18 p. + annexes.

Falkenmark, M. (1997). Society's interaction with the water cycle: a conceptual framework for a more holistic approach. *Hydrological Science Journal*, **42**, 451-466.

Fang, J. and Xi, Y. (1997). Neural network design based on evolutionary programming. *Artificial Intelligence in Engineering*, **11**, 155-161.

Faraway, J. and Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using airline data. *Applied Statistics*, **47(2)**, 231-250.

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

Ferreras, P. (2001). Landscape structure and asymmetrical inter-patch connectivity in a metapopulation of the endangered Iberian Lynx. *Biological Conservation*, **100**, 125-136.

Fielding, A.H. (2002). What are the appropriate characteristics of an accuracy measure? In: J.M., Scott, P.J., Heglund, M.L., Morrison, J.B., Haufler, M.G., Raphael, W.A., Wall and F.B., Samson (Eds.), *Predicting Species Occurrences: Issues of Accuracy and Scale*, Island Press, Washington, DC, 271-280 p.

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.

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.

Fore, L.S., Paulsen, K. and O'Laughlin, K. (2001). Assessing the performance of volunteers in monitoring streams. *Freshwater Biology*, **46**, 109-123.

Furse, M.T., Wright, J.F., Armitage, P.D. and Moss, D. (1981). An appraisal of pond-net samples for biological monitoring of lotic macro-invertebrates. *Water Research*, **15**, 679–689.

Furse, M.T., Clarke, R.T., Winder, J.M., Symes, K.L., Blackburn, J.H., Grieve, N.J. and Gunn, R.J.M. (1995). Biological Assessment Methods: Controlling the Quality of Biological Data. Package 1: the Variability of Data Used for Assessing the Biological Condition of Rivers. R&D Note 412. National Rivers Authority, Bristol.

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., Goethals, P., Adriaenssens, V., Heylen, S. and De Pauw, N. (2003). Development of a score or index for macroinvertebrates for the Flemish rivers and lakes according to the European Water Framework Directive and testing of REFCOND. Final report. Ghent, Belgium, Ghent University, Laboratory of Environmental Toxicology and Aquatic Ecology, Research Unit Aquatic Ecology**,** 72 p. + appendices.

Gabriels, W., Dedecker, A.P., Goethals, P.L.M., Lek, S. and De Pauw, N. (2005). Analysing and predicting the effect of river pollution on macrobenthos communities in Flanders (Belgium) using a stepwise input variable selection procedure in combination with artificial neural networks. *Aquatic Ecology*. (submitted).

Garcia, L.A. and Armbruster, M. (1997). A decision support system for evaluation of wildlife habitat. *Ecological Modelling,* **102**, 287-300.

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

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.

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

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, 71 p.

Gledhill, T., Sutcliffe, D.W. and Williams, W.D. (1993). British freshwater Crustacea Malacostraca: A key with ecological notes. Freshwater Biological Association, 173 p.

Glozier, N.E. and Culp, J.M. (1989). Experimental investigations of diel vertical movements by lothic mayflies over substrate surfaces. *Freshwater Biology*, **21**, 253-260.

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. (2005). Data driven development of predictive ecological models for benthic macroinvertebrates in rivers. PhD thesis, Ghent University, Belgium, 377 p.

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

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. In: F., Recknagel (Ed.), *Ecological Informatics, Understanding Ecology by Biologically-Inspired Computation*, Springer-Verlag, Berlin-Heidelberg, 91-108 p.

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, 412 p.

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, Nederland.

Graham, C.H. (2001). Factors influencing movement patterns of Keel-billed Toucans in a fragmented tropical landscape in Southern Mexico. *Conservation Biology*, **15(6)**, 1789-1798.

Green, M. and Alerstam, T. (2002). The problem of estimating wind drift in migrating birds. *Journal of Theoretical Biology*, **218(4)**, 485-496.

Greenberg, L.A. and Giller, P.S. (2000). The potential of flat-bed passive integrated transponder antennae for studying habitat use by stream fishes. *Ecology of Freshwater Fish*, **9**, 74-80.

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

Guerold, F. (2000). Influence of taxonomic determination level on several community indices. *Water Research*, **34(2)**, 487-492.

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

Gulinck, H., Dufourmont, H., Brunovsky, M., Andries, A. and Wouters, P. (1993). Satellite images for detection of changes in rural landscapes: a landscape ecological perspective. *EARSeL Adv. Remote Sensing*, **2**, 84-90.

Gyselman, E.C. (1979). The mechanisms that maintain population stability of selected species of Ephemeroptera in a temperate. Stream. In: J.F., Flannagan and K.E., Marshall (Eds), *Advances in Ephemeroptera Biology*, Plenum Publishing Corporation, New York, 309-319 p.

Haddad, N.M. (1999). Corridor use predicted from behaviours at habitat boundaries. *American Naturalist*, **153**, 215-227.

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

Haimes, Y.Y. (1992). Sustainable development: A holistic approach to natural resources management. *Water International*, **17**, 187–192.

Halse, S.A. (1999). AusRivAS: using macroinvertebrates to assess ecological condition of rivers in Western Australia. *Freshwater Biology*, **41**, 269-282.

Hansen, H.O. (1996). River restoration – Danish experience and examples. National Environmental Research Institute, Silkeborg, Denmark, 99 p.

Harby, A., Baptist, M., Dunbar, M.J. and Schmutz, S. (2004). State-of-the-art in data sampling, modelling analysis and applications of river habitat modelling. European Aquatic Ecological Modelling Network, EU COST action 626, 313 p.

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

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

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.

Hayes, J.W., Hughes, N.F. and Kelly, L.H. (2003). Overview of a process-based model relating stream discharge to the quantity and quality of brown trout habitat. Proceedings of the International IFIM Workshop. Fort Collins, Colorado.

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

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

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

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.

Hertz, J., Krogh, A. and Palmer, R. (1991). Introduction to the Theory of Neural Computation. Addison-Wesley, Reading, MA.

Heylen, S., Goethals, P.L.M. and De Pauw, N. (1999). Audit present water quality determination of the most important watercourses with the Belgian Biotic Index and proposal for an improved measurement strategy. Ghent, Belgium, Group for Applied Ecology; Ghent University, 58 p. + appendices. (In Dutch).

Hildrew, A.G. (1996). Whole river ecology: spatial scale and heterogeneity in the ecology of running waters. *Archiv für Hydrobiologie*, **10**, 323-343.

Hill, M.O. (1979). Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology,* **61,** 237-249.

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*, **195**, 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. In: F., Recknagel (Ed.), *Ecological Informatics, Understanding Ecology by Biologically-Inspired Computation*, Springer-Verlag, Berlin-Heidelberg, 179-192 p.

Holland, D.G. (1976). The distribution of the freshwater Malacostraca in the area of the Mersey and Weaver River Authority. *Freshwater Biology*, **6**, 265–76.

Holland, J.H. and Miller, J.H. (1991). Artificial adaptive agents in economic theory. *American Economic Review*, **81(1)**, 365-370.

Holmes, M.G.R., Young, A.R., Goodwin, T.H. and Grew, R. (2005). A catchment-based water resource decision-support tool for the United Kingdom. *Environmental Modelling & Software*, **20(2)**, 197-202.

Holthuis, L.B. (1956). Notities betreffende limburgse Crustacea III. De Amphipoda (Vlokreeftjes) van Limburg. *Nat. historisch maandblad*, **45(7-8)**.

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

Huang, G. H. and Xia, J. (2001). Barriers to sustainable water-quality management. *Journal of Environmental Management* , **61(1)**, 1-23.

Hubbard, S., Babak, P., Sigurdsson, S.T. and Magnusson, K.G. (2004). A model of the formation of fish schools and migrations of fish. *Ecological Modelling*, **174(4)**, 359-374.

Hughes, D.A. (1970). Some factors affecting drift and upstream movements of *Gammarus pulex*. *Ecology*, **51**, 301-305.

Hultin, L., Svensson, B. and Ulfstrand, S. (1969). Upstream movements of insects in a south Swedish small stream. *Oikos*, **20**, 553-557.

Humphries, S. (2002). Dispersal in drift-prone macroinvertebrates: a case for density-independence. *Freshwater Biology*, **47**, 921-929.

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

Humston, R., Olson, D.B. and Ault, J.S. (2004). Behavioral assumptions in models of fish movement and their influence on population dynamics. *Transactions of the American Fisheries Society*, **133(6)**, 1304-1328.

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.

Ibarra, A.A., Gevrey, M., Park, Y.S., Lim, P. and Lek, S. (2003). Modelling the factors that influence fish guilds composition using a back-propagation network: Assessment of metrics for indices of biotic integrity. *Ecological Modelling*, **160(3)**, 281-290.

IBN (1984). Biological water quality: determination of the biotic index based on aquatic macroinvertebrates, NBN T92-402. Institut Belge de Normalisation (IBN).

Illies, J. (1961). Versuch einer allgemeinen biozönotische Gliedering der Fliesgewässer. *International Review of Hydrobiology*, **46**, 205-213.

IN (2002). Ontwikkeling van een hydro-ecologisch model voor vallei-ecosystemen in Vlaanderen, ITORS-Vl. VLINA 00/16, 129 p. + annexes.

Jensen, M.E., Reynolds, K., Andreasen, J. and Goodman, I.A. (2000). A knowledge-based approach to the assessment of watershed condition. *Environmental Monitoring & Assessment*, **64**, 271-283.

Jochems, H., Schneiders, A., Denys, L. and Van den Bergh, E. (2003). Typologie van de oppervlaktewateren in Vlaanderen. Eindverslag van het project VMM. KRLW-typologie 2001. Instituut voor Natuurbehoud, Brussel, België. 51 p.

Jongman, R.H.G., ter Braak, C.J.F. and van Tongeren, O.F.R. (1995). Data analysis in community and landscape ecology. Cambridge, Cambridge University Press.

Jorgensen, S.E. (1986). Structural Dynamics model. *Ecological Modelling*, **31**, 1-9.

Jorgensen, S.E. (1988). Use of models as an experimental tool to show the structural changes are accompanied by increased exergy. *Ecological Modelling*, **41**, 117-126.

Jorgensen, S.E. (1994). Fundamentals of Ecological Modelling, 2nd edition. Elsevier, Amsterdam, 630 p.

Jorgensen, S.E. (1995). The growth rate of zooplankton at the edge of chaos. *Journal of Theoretical Biology*, **175**, 13-21.

Jorgensen, S.E. (1997). Integration of Ecosystem Theories: A Pattern, 2nd edition. Kluwer, Dordrecht, 400 p.

Jorgensen, S.E. (1998). An improved parameter estimation procedure in lake modelling. *Lakes and Reservoirs: Research Management*, **3**, 139-142.

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

Jorgensen, S.E. (2002a). Recent trends in the development of ecological models applied on aquatic ecosystems. *The ScientificWorldJOURNAL*, **2**, 387-406.

Jorgensen, S.E. (2002b). Explanation of ecological rules and observation by application of ecosystem theory and ecological models. *Ecological Modelling*, **158**, 241-248.

Jorgensen, S.E. and de Bernardi, R. (1997). The application of a model with dynamic structure to simulate the effect of mass fish mortality on zooplankton structure in Lago de Annone. *Hydrbiologia*, **356**, 87-96.

Jorgensen, S.E., Verdonschot, P. and Lek, S. (2002). Explanation of the observed structure of functional feeding groups of aquatic macro-invertebrates by an ecological model and the maximum exergy principle. *Ecological Modelling*, **158(3)**, 223-231.

Jung, N.S., Lee, J.J., Heinemann, P.H., Kim, D.S. and Kim, H.J. (2004). Development of an areal elderly migration model considering spatial interaction. *Journal of Urban Planning and Development*, **130(4)**, 175-183.

Kaastra, I. and Boyd, M.S. (1995). Forecasting futures trading volume using neural networks. *Journal of Futures Markets*, **15(8)**, 953-970.

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-Ampipoda); part 1. Gammarus pulex-group and related species. *Bijdrage tot de dierkunde*, **47(1)**.

Kareiva, P.M. (1983). Local movement in herbivorous insects: applying a passive diffusion model to mark-recapture field experiments. *Oecologia*, **57**, 322-327.

Karl, J.W., Heglund, P.J., Garton, E.O., Scott, J.M., Wright, N.M. and Hutto, R.L. (2000). Sensitivity of species–habitat relationship model performance to factors of scale. *Ecological Applications*, **10**, 1690-1705.

Karl, J.W., Svancara, L.K., Heglund, P.J., Wright, N.M. and Scott, J.M. (2002). Species commonness and the accuracy of habitat relationship models. In: J.M., Scott, P.J., Heglund, M.L., Morrison, J.B., Haufler, M.G., Raphael, W.A., Wall and F.B., Samson (Eds.), *Predicting Species Occurrences: Issues of Accuracy and Scale*, Island Press, Washington, DC, 573-580 p.

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, Covelo, CA .

Kim, K. and Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, **19**, 125-132.

Klemm, D.J., Lewis, P.A., Fulk, F. and Lazorchak, J.M. (1990). Macroinvertebrate Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters. EPA.600/4-90/030, Environmental Monitoring Systems Laboratory, Office of Modeling, Monitoring Systems, and Quality Assurance, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH.

Klemm, D.J., Blocksom, K.A., Thoeny, W.T., Fulk, F.A., Herlihy, A.T., Kaufmann, P.R. and Cormier, S.M. (2002). Methods development and use of macroinvertebrates as indicators of ecological conditions for streams in the Mid-Alantic Highllands region. *Environmental Monitoring and Assessment*, **78**, 169-212.

Knaapen, J.P., Scheffer, M. and Harms, B. (1992). Estimating habitat isolation in landscape planning. *Landscape and Urban Planning*, **23**, 1-16.

Kohler, S.L. (1985). Identification of stream drift mechanisms: an experimental and observational approach. *Ecology*, **66**, 79-89.

Kohler, S.L. and McPeek, M.A. (1989). Predation risk and the foraging behaviour of competing stream insects. *Ecology*, **70**, 1811-1825.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43,** 59-69.

Konings, P. and Meire, P. (2003). Ecologische inventarisatie en visievorming. Eindrapport, Technical 2618 report, Universiteit Antwerpen, Onderzoeksgroep Ecosysteembeheer, Antwerpen.

Kovats, Z.E., Ciborowski, J.J.H. and Corkum, L.D. (1996). Inland dispersal of adult aquatic insects. *Freshwater Biology*, **47**, 265-276.

Kundzewicz, Z.W. (1997). Water resources for sustainable development. *Hydrological Science Journal*, **42**, 467–480.

Kwok, T.Y. and Yeung, D.Y. (1997a). Constructive algorithms for structure learning in feedforward neural networks for regression problems. *IEEE Transactions on Neural Networks*, **8(3)**, 630–645.

Kwok, T.Y. and Yeung, D.Y. (1997b). Objective functions for training new hidden units in constructive neural networks. *IEEE Transactions on Neural Networks*, **8(5)**, 1131–1148.

Laë, R., Lek, S., and Moreau, J. (1999). Predicting fish yield of African lakes using neural networks. *Ecological Modelling*, **120**, 325–335.

Lam, D. and Swayne, D. (2001). Issues of EIS software design: some lessons learned in the past decade. *Environmental Modelling & Software*, **16**, 419-425.

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 & Software*, **19**, 317-324.

Landis, J.R. and Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-174.

Largier, J.L. (2003). Considerations in estimating larval dispersal distances from oceanographic data. *Ecological Applications*, **13(1)**, 71-89.

Lee, J.H.W., Huang, Y., Dickman, M. and Jayawardena, A.W. (2003). Neural network modelling of coastal algal blooms. *Ecological Modelling*, **159(2-3)**, 179-201.

Lek, S., Belaud, A., Dimopoulos, I., Lauga, J. and Moreau, J. (1995). Improved estimation, using neural networks, of the food consumption of fish populations. *Marine 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.

Lek, S. and Guégan, J.F. (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, **120**, 65-73.

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.

Li, J., Herlihy, A., Gerth, W., Kaufmann, P., Gregory, S., Urquhart, S. and Larsen, D. (2001). Variability in stream macroinvertebrates at multiple spatial scales. *Freshwater Biology,* **46**, 87-97.

Liano, K. (1996). Robust error measure for supervised neural network learning with outliers. *IEEE Transactions on Neural Networks*, **7(1)**, 246-250.

Linke, S., Bailey, R.C. and Schwindt, J. (1999). Temporal variability of stream bioassessments using benthic macroinvertebrates. *Freshwater Biology*, **42(3)**, 575-584.

Macan, T.T. (1961). Factors that limit the range of freshwater animal communities. *Biological Revue*, **52**, 45-70.

Mackey, A.P., Cooling, D.A. and Berrie, A.D. (1984). An evaluation of sampling strategies for qualitative surveys of macroinvertebrates in rivers, using pond nets. *Journal of Applied Ecology*, **21**, 515–534.

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.

Maddock, I. (1999). The importance of physical habitat assessment for evaluating river health. *Freshwater Biology*, **41**, 373-391.

Magnusson, K.G., Sigurdsson, S.T., Babak, P., Gudmundsson, S.F. and Dereksdottir, E.H. (2004). A continuous density Kolmogorov type model for a migrating fish stock. *Discrete and Continuous Dynamical Systems – Series*, **4(3)**, 695-704.

Maier, H.R., Dandy, G.C. and Burch, M.D. (1998). Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray, South Australia. *Ecological Modelling*, **105**, 257-272.

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 & Software*, **15**, 101-124.

Maier, H.R. and Dandy, G.C. (2001). Neural network based modelling of environmental variables: a systematic approach. *Mathematical and Computer Modelling*, **33**, 669-682.

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 Ecology*, **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.

Maren, A., Harston, C. and Pap, R. (1990). Handbook of neural computing applications. Academic Press, San Diego, CA.

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.

Martin, T.R. and Holdich, D.M. (1986). The acute lethal toxicity of heavy metals to peracarid crustaceans (with particular reference to freshwater Asellids and Gammarids). *Water Research*, **20**, 1137-1147.

Masters, T. (1993). *Practical Neural Network Recipes in C++*. Academic Press, San Diego.

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.

Mastrorillo, S., Dauba, F., Oberdorff, T., Guégan, J.F. and Lek, S. (1998). Predicting local fish species richness in the Garonne river basin. *Ecology*, **321**, 423-428.

Matthiopoulos, J. (2003). The use of space by animals as a function of accessibility and preference. *Ecological Modelling*, **159**, 239-268.

Meijering, M.P.D. (1972). Experimentelle Untersuchungen zur Drift und Aufwanderung von Gammariden in Fliessgewässern. *Archiv für Hydrobiologie*, **70**, 133-205.

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: P., Calow and G.E., Petts (Eds.), *The rivers handbook*, Vol. 2., Blackwell Scientific Publications, Oxford, 144-170 p.

Meyer, J.L. (1997). Stream health: Incorporating the human dimension to advance stream ecology. *Journal of the North American Benthological Society*, **16**, 439–447.

Michels, E., Cottenie, K., Neys, L., De Gelas, K., Coppin, P. and De Meester, L. (2001). Geographical and genetic distances among zooplankton populations in a set of interconnected ponds : a plea for using GIS modelling of the effective geographical distance. *Molecular Ecology*, **10**, 1929-1938.

Minckley, W.L. (1964). Upstream movements of *Gammarus* (Amphipoda) in Doe Run, Meade county, Kentucky. *Ecology*, **45**, 195-197.

Minshall, G.W. and Minshall, J.N. (1977). Microdistribution of benthic invertebrates in a rocky mountain stream. *Hydrobiologia*, **55**, 231-249.

Minshall, G.W., Petersen, R.C. and Nimz, C.F. (1985). Species richness in streams of different size from the same drainage basin. *American Naturalist*, **125**, 16-38.

MIRA-T (2004). Milieu- en natuurrapport Vlaanderen thema's, Kwaliteit Oppervlaktewater. Vlaamse Milieumaatschappij.

Moilanen, A. and Hanski, I. (2001). On the use of connectivity measures in spatial ecology. *Oikos*, **95**, 147-151.

Mommsen, T.P. (2004). Salmon spawning migration and muscle protein meetabolism: the August Krogh principle at work. *Comparative Biochemistry and Physiology B – Biochemistry & Molecular Biology*, **139(3)**, 383-400.

Monden, S., Van Liefferinge, C., Vandenauweele, I., Simoens, I., Beyens, J., Denayer, B., Yseboodt, R., Meire, P. and De Charleroy, D. (2005). Databank vismigratieknelpunten op prioritaire waterlopen in het Vlaamse Gewest. IBW-UIA databank, http://vismigratie.instnat.be.

Montgomery, D.C. and Peck, E.A. (1982). Introduction to Linear Regression Analysis. John Wiley, New York.

Moog, O. (1995). Fauna Aquatica Austriaca, A comprehensive species inventory of Austrian aquatic organisms with ecological notes. Wasserwirtschafts – Kataster, Bundesministerium für Land- und Forstwirtshaft, Wien.

Morris, C. W., Autret, A. and Boddy, L. (2001). Support vector machines for identifying organisms - a comparison with strongly partitioned radial basis function networks. *Ecological Modelling*, **146**, 57-67.

Moulton, S.R., Carter, J.M., Grotheer, S.A., Cuffney, T.F. and Short, T.M. (2000). Methods of analysis by the U.S. Geological Survey National water laboratory processing, taxonomy, and quality control of benthic macroinvertebrate samples. U.S. Geological Survey. Open-File Report 00-212. Denver, Colorado.

Müller, K. (1966). Die tagesperiodik von Flieβwasserorganismen. *Zeitschrift für Morphologie und Ökologie der Tiere*, **56**, 93-142.

MVG, LIN and AMINAL department Water (2000). Digital version of the Flemish Hydrological Atlas, scale 1/10000 (OC GIS-Flanders).

Mysiak, J., Giupponi, C. and Rosato, P. (2005). Towards the development of a decision support system for water resource management. *Environmental Modelling & Software*, **20(2)**, 203-214.

Nabhan, T.M. and Zomaya, A.Y. (1994). Toward generating neural network structures for function approximation. *Neural Networks*, **7(1)**, 89-99.

Nilsson, C., Pizzuto, J.E., Moglen, G.E., Palmer, M.E., Stanley, E.H., Bockstael, N.E. and Thompson, L.C. (2003). Ecological forecasting and the urbanization of stream ecosystems: challenges for economists, hydrologists, geomorphologists, and ecologists. *Ecosystems,* **6**, 659-674.

Norris, R.H. and Georges, A. (1993). Analysis and interpretation of benthic macroinvertebrate surveys. In: D.M., Rosenberg and V.H., Resh (Eds.), *Freshwater biomonitoring and benthic macroinvertebrates*, Chapman and Hall, New York.

Norris, R.H. and Norris, K.R. (1995). The need for biological assessment of water quality: Australian perspective. *Australian Journal of Ecology,* **20**, 1-6.

Nykanen, M., Huusko, A. and Lahti, M. (2004). Movements and habitat preferences of adult grayling (*Thymallus thymallus* L.) from late winter to summer in a boreal river. *Archiv für Hydrobiologie*, **161(3)**, 417-432.

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.

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.

Pakath, R. and Zaveri, J.S. (1995). Specifying critical inputs in a genetic algorithm-driven decision support system: an automated facility. *Decision Sciences*, **26(6)**, 749-779.

Pallottino, S., Sechi, G.M. and Zuddas P. (2005). A DSS for water resources management under uncertainty by scenario analysis. *Environmental Modelling & Software*, **20(8)**, 1031-1042.

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., 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(3)**, 265-280.

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.

Patuwo, E., Hu, M.Y. and Hung, M.S. (1993). Two-group classification using neural networks. *Decision Sciences*, **24(4)**, 825–845.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems. Morgan Kauffman Publishers, San Fransisco, CA.

Peckarsky, B.L., Fraissinet, P.R., Penton, M.A. and Conklin, D.J. (1989). Freshwater macroinvertebrates of north-eastern North America. Cornell University Press, Cornell, 442 p.

Peckarsky, B.L., Cowan, C.A. and Anderson, C. (1994). Consequences and plasticity of specialized predatory behaviour of stream-dwelling stonefly larvae. *Ecology*, **75**, 166-181.

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, 168 p.

Petersen, I., Winterbottom, J.H., Orton, S., Friberg, N., Hildrew, A.G., Speirs, D.C. and Gurney, W.S.C. (1999). Emergence and lateral dispesal of adult Plecoptera and Trichoptera from Broadstone Stream, UK. *Freshwater Biology*, **42**, 401-416.

Peterson, A.T. (2001). Predicting species' geographic distributions based on ecological niche modeling. *Condor*, **103**, 599-605.

Peterson, A.T., Soberón, J. and Sánchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265-1267.

Peterson, A.T. and Vieglais, D.A. (2001). Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *Bioscience*, **51**, 363-371.

Piramuthu, S., Shaw, M. and Gentry, J. (1994). A classification approach using multi-layered neural networks. *Decision Support Systems*, **11(5)**, 509-525.

Pitcher, J. and Taylor, P.D. (1998). An experimental assessment of landscape connectivity. *Oikos*, **83**, 166-174.

Plate, E.J. (1993). Sustainable development of water resources: A challenge to science and engineering. *Water International*, **18**, 84–93.

Poch, M., Comas, J., Rodríguez-Roda, I., Sànchez-Marrè, M. and Cortés, U. (2004). Designing and building real environmental decision support systems. *Environmental Modelling & Software* (in press).

Ponder, W.F., Carter, G.A., Flemons, P. and Chapman, R.R. (2001). Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology*, **15**, 648-657.

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 tress. *Machine Learning*, **1(1)**, 81-106.

Rabeni, C.F. and Minshall, G.W. (1977). Factors affecting microdistribution of stream benthic insects. *Oikos*, **29**, 33-43.

Rawer-Jost, C., Kappus, B., Böhmer, J., Jansen, W. and Rahmann, H. (1999). Upstream movements of benthic macroinvertebrates in two different types of fishways in southwestern Germany. *Hydrobiologia*, **391**, 47-61.

Rech, V.H., Brown, A.V., Covich, A.P., Gurtz, M.E., Li, H.W., Minschall, G.W., Reice, S.R., Sheldon, A.L., Wallace, J.B. and Wissmar, R. (1988). The role of disturbance in stream ecology. *Journal of the North American Benthological Society*, **7**, 433-455.

Recknagel, F. (2001). Applications of machine learning to ecological modeling. *Ecological Modelling*, **146**, 303-310.

Recknagel, F. (2002). Ecological informatics: understanding ecology by biologically-inspired computation. Springer-Verlag, Berlin-Heidelberg, 398 p.

Recknagel, F. (2003). Simulation of food web and species interactions by adaptive agents embodied with evolutionary computation: a conceptual framework. *Ecological Modelling*, **170**, 291-302.

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

Reece, P.F., Reynoldson, T.B., Richardson, J.S. and Rosenberg, D.M. (2001). Implications of seasonal variation for biomonitoring with predictive models in the Fraser River catchment, British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 1411-1418.

Reed, R. (1993). Pruning algorithms - A review. *IEEE Transactions on Neural Networks*, **4(5)**, 740-747.

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.

Reice, S.R., Wissmar, R.C. and Naiman, R.J. (1990). Disturbance regimes, resilience, and recovery of animal communities and habitats in lotic ecosystems. *Environmental Management*, **14**, 647-659.

Reisen, W.K. and Prins, R. (1972). Some ecological relationships of the invertebrate drift in Praters Creek, Pickens County, South Carolina. *Ecology*, **53**, 876-884.

Resh, V.H. and McElravy, E.P. (1993). Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In: D.M., Rosenberg and V.H., Resh (Eds.), *Freshwater Biomonitoring and Benthic Macroinvertebrates*, Chapman & Hall, New York, 159-194 p.

Reyjol, Y., Lim, P., Belaud, A. and Lek, S. (2001). Modelling of microhabitat used by fish in natural and regulated flows in the river Garonne (France). *Ecological Modelling*, **146**, 131-142.

Reynoldson, T.B., Bailey, R.C., Day, K.E. and Norris, R.H. (1995). Biological guidelines for freshwater sediment based on benthic assessment of sediment (the BEAST) using a multivariate approach for predicting biological state. *Australian Journal of Ecology,* **20,** 198-219.

Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E. and Rosenberg, D.M. (1997). The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society*, **16(4)**, 833-852.

Ricketts, T.H. (2001). The matrix matters: effective isolation in fragmented landscapes. *The American Naturalist*, **158**, 87-99.

RIZA (2000). Database ter beschikking gesteld door RIZA (Rijksinstituut voor Integraal Zoetwaterbeheer en Afvalwaterbehandeling (Institute for Inland Water Management and Waste Water Treatment)).

Rizzoli, A.E. and Young, W.J. (1997). Delivering environmental decision support systems: software tools and techniques. *Environmental Modelling & Software*, **12**, 237-249.

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.

Rojas, R. (1996). Neural networks: A systematic introduction. Springer-Verlag, Berlin.

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.

Rykiel, E.J.R. (1996). Testing ecological models: the meaning of validation. *Ecological Modelling*, **90**, 229-244.

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, 43 p.

Scardi, M. (2001). Advances in neural network modeling of phytoplankton primary production. *Ecological Modelling*, **146**, 33-45.

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.

Schadt, S., Knauer, F., Kaczensky, P., Revilla, E., Wiegand, T. and Trepl, L. (2002). Rule-based assessment of suitable habitat and patch connectivity for Eurasian Lynx in Germany. *Ecological Applications*, **12**, 1469-1483.

Schaefer, S.M. and Krohn, W.B. (2002). Predicting vertebrate occurrences from species habitat associations: improving the interpretation of commission error rates. In: J.M., Scott, P.J., Heglund, M.L., Morrison, J.B., Haufler, M.G., Raphael, W.A., Wall and F.B., Samson (Eds.), *Predicting Species Occurrences: Issues of Accuracy and Scale*, Island Press, Washington, DC, 419-427 p.

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.

Schneiders, A., Wils, C. and Verheyen, R. (1999). The use of ecological information in the selection of quality objectives for river conservation and restoration in Flanders (Belgium). *Aquatic Ecosystem Health and Management*, **2,** 137-154.

Schonfisch, B. and Kinder, M. (2002). A fish migration model. *Lecture Notes in Computer Science*, **2493**, 210-219.

Scott, J.M., Heglund, P.J., Samson, F., Haufler, J., Morrison, M., Raphael, M. and Wall, B. (2002). Predicting species occurrences: issues of accuracy and scale. Island Press, Covelo, CA, 868 p.

Simonovic, S.P. (1996a). Decision support system for sustainable management of water resources: 1 general principles. *Water International*, **21**, 223–232.

Simonovic, S.P. (1996b). Decision support system for sustainable management of water resources: 2 Case studies. *Water* International, **21**, 233–244.

Skidmore, A.K., Gauld, A. and Walker, P. (1996). Classification of Kangaroo habitat distribution using three GIS models. *International Journal of Geographical Information Systems*, **10**, 441–454.

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., Richarson, 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 Söderström, O. (1987). Upstream movements of invertebrates in running waters – a review. *Archiv für Hydrobiologie*, **111**, 197-208.

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. *Det. Kong. Danske Vidensk. Selsk. Biol. Slr*, **5(4)**, 1-34.

Soulie, F. (1994). Integrating neural networks for real world applications. In: A., Zurada, A., Marks and A., Robinson (Eds.), *Computational Intelligence: Imitating Life*, IEEE Press, New York, 396 p.

Speirs, D.C. and Gurney, W.S.C. (2001). Population persistence in rivers and estuaries. *Ecology*, **82**, 1219-1237.

Statzner, B. and Mogel, R. (1985). An example showing that drift net catches of stream mayflies (*Baetis spp*., Ephemeroptera, Insecta) do not increase during periods of higher substrate surface densities of the larvae. *Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie*, **22**, 3238-3243.

Stauffer, H.B., Ralph, C.J. and Miller, S.L. (2002). Incorporating detection uncertainty into presence-absence surveys for marbled murrelet. In: J.M., Scott, P.J., Heglund, M.L., Morrison, J.B., Haufler, M.G., Raphael, W.A., Wall and F.B., Samson (Eds.), *Predicting*

Species Occurrences: Issues of Accuracy and Scale, Island Press, Washington, DC, 357-365 p.

Steel, E.A., Feist, B.E., Jensen D.W., Pess, G.R., Sheer, M.B., Brauner, J.B. and Bilby, R.E. (2004). Landscape models to understand steelhead (*Oncorhynchus mykiss*) distribution and help prioritise barrier removals in the Willamette basin, Oregon, USA. *Canadian Journal of Fisheries and Aquatic Sciences*, **61(6)**, 999-1011.

Steenbergen (1993). Macrofauna-atlas of North Holland: Distribution maps and responses to environmental factors of aquatic invertebrates. Haarlem, The Netherlands, 650 p.

Stockwell, D.R.B. and Peterson, A.T. (2002). Controlling bias in biodiversity data. In: J.M., Scott, P.J., Heglund, M.L., Morrison, J.B., Haufler, M.G., Raphael, W.A., Wall and F.B., Samson (Eds.), *Predicting Species Occurrences: Issues of Accuracy and Scale*, Island Press, Washington, DC, 537-546 p.

STOWA/RIZA (1999). Vloeiend Modelleren in het Waterbeheer, Handboek Good Modelling Practice; STOWA-rapport 99-05, Rijkswaterstaat-RIZA-rapport 99.036, ISBN-nr. 90-5773-056-1.

Strahler, A. (1957). Quantative analyses of watershed geomorphology. *American Geophysical Union transactions*, **38**, 913-920.

Tachet, H., Richoux, P., Bournaud, M. and Usseglio-Polaterra, P. (2002). Invertébrés d'eau douce, systématique, biologie, écologie. CNRS Editions, Paris, 587 p.

Tahai, A., Walczak, S. and Rigsby, J.T. (1998). Improving artificial neural network performance through input variable selection. In: P., Siegel, K., Omer, A., deKorvin and A., Zebda (Eds.), *Applications of Fuzzy Sets and The Theory of Evidence to Accounting II*, JAI Press, Stamford, Connecticut, 277 p.

Tankersley, R. and Orvis, K. (2003). Modeling the geography of migratory pathways and stopover habitats for neotropical migratory birds. *Conservation Ecology*, **7(1)**.

ter Braak, C.J.F. and Smilauer, P. (1998). CANOCO reference manual and user's guide to Canoco for Windows: software for canonical community ordination (version 4). Microcomputer Power (Ithaca, NY USA), 352 p.

Tikkanen, P., Muotka, T. and Hahta, A. (1994). Predator detection and avoidance by lotic mayfly nymphs of different size. *Oecologia*, **99**, 252-259.

Tischendorf, L. and Fahrig, L. (2000). On the use and measurement of landscape connectivity. *Oikos*, **90**, 7-19.

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

Tolkamp, H.H. (1980). Organism-substrate relationships in lowland streams. PhD thesis, Wageningen University, The Netherlands.

Tolkamp. H.H. (1982). Microdistribution of macroinvertebrates in lowland streams. *Hydrobiological Bulletin*, **16(2-3)**, 133-148.

Townsend, C.R. and Hildrew, A.G. (1976). Field experiments on the drifting, colonization en continuous redistribution of stream benthos. *Journal of Animal Ecology*, **45**, 759-772.

Townsend, C.R., Arbuckle, C.J., Crowl, T.A. and Scarsbrook, M.R. (1997a). The relationship between land use and physicochemistry, food resources and macroinvertebrate communities in tributaries of the Taieri River, New Zealand: a hierarchically scaled approach. *Freshwater Biology,* **37,** 177-191.

Townsend, C.R., Scarsbrook, M.R. and Dolédec, S. (1997b). Quantifying disturbance in streams: alternative measures of disturbance in relation to macroinvertebrate species traits and species richness. *Journal of the North American Benthological Society*, **16**, 531-544.

Van Bockstal, P. (2003). Toelichting bij het Decreet Integraal Waterbeleid. *Tijdschrift Water,* **10,** 1-6.

Vandenberghe, V., van Griensven, A., Bauwens, W., Goethals, P., De Pauw, N., Meirlaen, J., Van Vooren, L. and Vanrolleghem, P. (2000). The importance of on-line water quality measurements, application on the river Dender. *@WEL,* **5**, 1-6. (in Dutch).

van der Molen, D. (1999). The role of eutrophication models in water management. PhD thesis, Wageningen University, the Netherlands, 167 p.

Van de Walle, A. (2004). Uitwerking vaneen methode voor het beoordelen van macro-invertebraten in Vlaanderen volgens de Europese Kaderrichtlijn Water. Scriptie voorgedragen tot het behalen van de graad van Bio-ingenieur in de Cel- en Genbiotechnologie, Universiteit Gent, 110 p.

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.

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.

Van Tongeren, O. (1986). Flexclus, an interactive program for classification and tabulation of ecological data. *Acta Botanica Neerlandica*, **35(3),** 137-142.

Vaughan, D.M. (2002). Potential impact of road-stream crossings (culverts) on the upstream passage of aquatic macroinvertebrates. US Forest Service Report, 15 p.

Verbeylen, G., De Bruyn, L. Adriaensen, F. and Matthysen, E. (2003). Does matrix resistance influence Red squirrel (*Sciurus vulgaris* L. 1758) distribution in an urban landscape? *Landscape Ecology*, **18**, 791-805.

Verdonschot, P.F.M. (1990). Ecological characterization of surface waters in the province of Overijssel (the Netherlands). PhD thesis, Wageningen Agricultural University.

Verdonschot, P.F.M. (2000a). Natuurlijke levensgemeenschappen van de Nederlandse binnenwateren, Deel 1, Bronnen. Alterra, Wageningen, 88 p.

Verdonschot, P.F.M. (2000b). Natuurlijke levensgemeenschappen van de Nederlandse binnenwateren, Deel 2, Beken. Alterra, Wageningen, 128 p.

Verdonschot, P.F.M. and Nijboer, R.C. (2002a). A review of model design. Wageningen, Alterra, Green World Research, 92 p.

Verdonschot, P.F.M. and Nijboer, R.C. (2002b). Towards a decision support system for stream restoration in the Netherlands: an overview of restoration projects and future needs. *Hydrobiologia*, **478**, 131-148.

Vincent, B. (1983). Variations spatiales et saisonnieres de la structure de groupements macrobenthiques lithoraux en climat froid. *Hydrobiologia*, **102**, 175-186.

VLAREM (1999). VLAREM II. Conclusions of the Flemish Government of 01/06/95 on general and sectoral guidelines on environmental hygiene. (in Dutch).

VMM (2003). Water quality – water discharges 2002. Flemish Environment Agency, VMM, Aalst, 162 p. (in Dutch).

Voinov, A., Fitz, C., Boumans, R. and Costanza, R. (2004). Modular ecosystem modeling. *Environmental Modelling and Software*, **19**, 285-304.

Wade, K.R., Ormerod, S.J. and Gee, A.S. (1989). Classification and ordination of macroinvertebrate assemblages to predict stream acidity in upland Wales. *Hydrobiologia*, **171**, 59-78.

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. (1995). Developing neural nets currency trading. *Artificial Intelligence in Finance*, **2(1)**, 27-34.

Walczak, S. and Cerpa, N. (1999). Heuristic principles for the design of artificial neural networks. *Information and Software Technology*, **41(2)**, 107-117.

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. and Fontama, V.N. (2000). New approaches to river quality classification based upon Artificial Intelligence. In: J.F., Wright, D.W., Sutcliffe and M.T., Furse (Eds.), *Assessing the Biological Quality of Freshwaters: RIVPACS and Other Techniques*, Freshwater Biological Association, Ambleside, 263-280 p.

Wallin, M., Wiederholm, T. and Johnson, R.K. (2003). Guidance on establishing reference conditions and ecological status class boundaries for inland surface waters. Produced by CIS Working Group 2.3 – REFCOND. Final version 7.0., 73 p. + annexes.

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.

Waters, T.F. (1962). Diurnal periodicity in the drift of stream invertebrates. *Ecology*, **43**, 316-320.

Waters, T.F. (1965). Interpretation of invertebrate drift in streams. *Ecology*, **46**, 327-334.

Waters, T.F. (1972). The drift of stream insects. *Annual Review of Entomology*, **17**, 253-272.

Weber, N.S., Booker, D.J., Dunbar, M.J., Ibbotson, A.T. and Wheater, H.S. (2004). Modelling stream invertebrate drift using particle tracking. In: D., García de Jalón Lastra and P.V., Martínez (Eds.), *Aquatic Habitats: Analysis & Restoration*, Fifth International Symposium on Ecohydraulics, Madrid, 1129-1134 p.

Wei, B., Sugiura, N. and Maekawa, T. (2001). Use of artificial neural network in the prediction of algal blooms. *Water Research*, **35(8)**, 2022-2028.

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.

Welp, M. (2001). The use of decision support tools in participatory river basin management. *Physics and Chemistry of the Earth*, **26(7-8)**, 535-539.

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 analysis, 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. (1991a). The Gammarus-Asellus ratio as an index of organic pollution. *Water Research*, **25(3)**, 333-339.

Whitehurst, I.T. (1991b). The effects of sampling techniques on the Gammarus, Asellus ratio. *Water Research*, **25(6)**, 745-748.

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.

Wicker, D., Rizki, M.M. and Tamburino, L.A. (2002). E-Net: Evolutionary neural network synthesis. *Neurocomputing*, **42**, 171-196.

Wiederholm, T. (1980). Use of benthos in lake monitoring. *Journal of the Water Pollution Federation*, **52**, 537-547.

Williams, D.D. and Feltmate, B.W. (1992). Aquatic insects. C.A.B. International, Wallingford, Oxford, 360 p.

Williams, D.D. and Williams, N.E. (1993). The upstream/downstream movement paradox of lotic invertebrates: quantitative evidence from a Welsh mountain stream. *Freshwater Biology*, **30**, 199-218.

Wils, C., Schneiders, A., Bervoets, L., Nagels, A., Weiss, L. and Verheyen, R.F. (1994). Assessment of the ecological value of rivers in Flanders (Belgium). *Water Science and Technology*, **30(10)**, 37-47.

Wilson, H. and Recknagel, F. (2001). Towards a generic artifical neural network model for dynamic predictions of algal abundance in freshwater lakes. *Ecological Modelling*, **146**, 69-84.

Winchester, N.N., Nimmo, A.P. and Cross, S.F. (2002). Riparian reserve zones: comparison of adult Trichoptera communities in area of ancient Sitka spruce rain forest and adjacent growth. Proceedings of the 10th International Symposium on Trichoptera, 625-640 p.

Wissinger, S.A., Sparks, G.B., Rouse, G.L., Brown, W.S. and Steltzer, R.H. (1996). Intraguild predation and cannibalism among larvae of detritivorous caddisflies in subalpine wetlands. *Ecology*, **77**, 2421-2430.

Wissinger, S.A., Whiteman, H.H., Sparks, G.B., Rouse, G.L. and Brown, W.S. (1999). Tradeoffs between competitive superiority and vulnerability to predation in caddisflies along a permanence gradient in subalpine wetlands. *Ecology*, **80**, 2102-2116.

Wissinger, S.A., Brown, W.S. and Jannot, J.E. (2003). Caddisfly life histories along permanence gradients in high-altitude wetlands in Colorado (U.S.A.). *Freshwater Biology*, **48**, 255-270.

Witten, I.H. and Frank, E. (2000). Data Mining: practical machine learning tools and techniques with Java implementations. Morgan Kaufmann Publishers, San Francisco. 369 p.

Woodiwiss, F.S. (1980). Biological monitoring of surface water quality. Summary Report. Commission of the European Communities. ENV/787/80-EN, 45 p.

Wooster, D. and Sih, A. (1995). A review of the drift and activity responses of stream prey to predator presence. *Oikos*, **73**, 3-8.

Wright, J.F. (1995). Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology*, **20**, 181-197.

Wright, J.F. (2000). An introduction to RIVPACS. In: J.F., Wright, D.W., Sutcliffe and M.T., Furse (Eds.), *Assessing the Biological Quality of Freshwaters. RIVPACS and Other Techniques*, Freshwater Biological Association, Ambleside, 1-24 p.

Wright, J.F., Moss, D., Armitage, P.D. and Furse, M.T. (1984). A preliminary classification of running-water sites in Great Britain based on macroinvertebrate species and the prediction of community type using environmental data. *Freshwater Biology,* **14,** 221-256.

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. Freshwater Biological Association, Ambleside, 373 p.

Xia, J. and Takeuchi, K. (1999). Barriers to sustainable management of water quality and quantity. *Hydrological Science Journal*, **44**, 462–474.

Yamada, K., Elith, J., McCarthy, M. and Zerger, A. (2003). Eliciting and integrating expert knowledge for wildlife habitat modelling. *Ecological Modelling*, **165**, 251-264.

Yao, X. (1993). A review of evolutionary artificial neural networks. *International Journal of Intelligent Systems*, **8(4)**, 539–567.

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. (1983). The validity and credibility of models for badly defined systems. In: M.B., Beck and G. Van Straten (Eds.), *Uncertainty and Forecasting of Water Quality*, Springer-Verlag, Berlin, 69-98 p.

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 and Software*, **15**, 257-265.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, **8**, 338-353.

Zalewski, M. and Roberts, R. (2003). Ecohydrology – a new paradigm for integrated resources management. *SIL News*, **40**, 1-5.

Zhao, W., Chen, D. and Hu, S. (2000). Optimizing operating conditions based on ANN and modified Gas. *Computers and Chemical Engineering*, **24**, 61–65.

Summary

In river management in Flanders, nowadays, besides the use of ecological indicators for the assessment of water quality, no use is made of other decision support techniques that enable the linkage between abiotic and biotic river characteristics. However, the major aim of the European Water Framework Directive (WFD) is to reach 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. To assess the rivers by comparing the actual status to a reference status, 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 into the relation between the aquatic communities and the human activities affecting these water systems. Insights into these relations will also be valuable for detection of causes of particular river conditions as well as for decision-making in river restoration and protection management to meet and sustain the requirements set by the WFD.

In spite of the ecological objectives of the WFD, ecological models have been rarely used so far to support river management and water policy. Models have however several interesting applications in this context. In particular, ‘habitat suitability models’ that can predict the habitat requirements of organisms based on environmental river characteristics might be very useful. This type of models has only very recently been recognized as a significant component of conservation planning.

The overall aim of the present thesis was to determine the appropriate variables and ecosystem processes by using a data driven modelling technique, based on Artificial Neural Networks (ANNs), to predict the habitat suitability of biological communities present in rivers. The ANN models were developed and applied in a MATLAB environment. The research mainly focused on macroinvertebrates in brooks and small rivers in the Zwalm river basin, a sub-basin of the Upper-Scheldt river basin (Flanders, Belgium). The selected sampling sites were characterized by a gradient ranging from nearly natural situations to severely impacted (water pollution, physical habitat degradation) ones.

The ANN model results illustrated the convenience of using two performance measures to evaluate the model predictions. From an ecological point of view, the overall prediction success (= percentage of Correctly Classified Instances, CCI) is the most logic performance measure because of its easy interpretation towards water managers and policy makers. There is clear evidence however, that the CCI is influenced by the frequency of occurrence of the organism being modelled. Therefore, a second performance measure (the Cohen's kappa, CK) was used, since the effect of prevalence on the CK appeared to be negligible. Because the size of the training and test set influences respectively the generalization capability of the model and accuracy of the model performance, the optimal size of training and test set was searched for. Based on both the CCI and CK, the best option was using three fourth for training and one fourth for testing (cf. 4-fold cross-validation). Network architecture is generally known to be highly problem dependent. However, comparing the different model architectures for each organism, predictive performances were very similar. In this way, no straightforward conclusions towards the taxa could be drawn according to the optimal number of hidden layers and neurons.

In addition and combined with the ANN models, three methods (the Weights, PaD and Profile method) were used to analyse the contribution of environmental variables to predict the presence/absence of the macroinvertebrates in a reliable manner and to detect the major river characteristics to describe the habitat suitability of the different taxa. It was often difficult to find major trends over the five taxa, the three contribution methods and the different folds. The first two can be explained by the different ecological preferences of the taxa and by the different aspects the three contribution methods deal with. The instability over the different folds is perhaps related to the relative small size of the datasets in combination with the high variability of the sites, 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. Nevertheless, it could be concluded that the combined use of the PaD method and 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.

Ecosystem models, such as ANNs, can act as interesting tools to support decision-making in river restoration management. In general however, these habitat suitability models do not include spatial and temporal relationships. Migration dynamics of the predicted organisms and migration barriers along the river may therefore deliver important additional information on the effectiveness of the restoration plans. In this context, migration models were developed for the Zwalm river basin as extension of the habitat suitability models. These models are able to examine the connectivity between population patches or the possibility to migrate from a source population to recolonize a restored river section. The migration models were based on the Cost Weighted Distance function, a powerful tool within the ArcGIS Spatial Analyst extension of ArcGIS 8.3. A (theoretical) validation of the resistance values was done based on a sensitivity analysis. Therefore, the resistance values ascribed to the different determining variables were altered to investigate the changes in the resulting accumulative costs and the sensitivity of the developed migration model on variation of the applied resistance values.

To convince water managers and stakeholders of the validity and practical usefulness of the developed models, four simulations of river restoration scenarios were performed (a remeandering project, the construction of collector, a river restoration project, and the removal of a weir for water quantity control). The scenario analyses can help to select the most valuable restoration options. In this way, these exercises can be interesting as validation instrument of the developed models and can give insights into shifts of indicator organisms.

It could be concluded that the combined use of both models (habitat suitability and migration model) can be very beneficial to make simulations of the potential ecological effects of river restoration actions (as well as the effects of river deterioration) and can as such support the decision making process that river managers are daily facing.

However, a number of recommendations and considerations for further research can be put forward in order to enhance their value for decision support in river management. In the first place, optimization of the monitoring networks (e.g. tuning the selection of sampling sites to the prediction of river restoration actions) and the dataset used (e.g. inclusion of new, essential variables on the one hand and remove less relevant input variables on the other hand) can be beneficial. In addition, further optimization and extension of the ANN models and migration models (especially the practical validation) is requested. The development of habitat suitability models for other macroinvertebrates, as well as for fish, macrophytes, … can also be beneficial based on the needs of the European Water Framework Directive. On the other hand, extension of both models to other river basins in Flanders (in order to enlarge the predictive range of the models) needs to be incorporated in the future. Also the coupling of the habitat suitability models to hydraulic and river water quality models will be required and improve the reliability and the practical applicability of the habitat suitability models.

Samenvatting

In het rivierbeheer in Vlaanderen wordt momenteel naast het toepassen van ecologische indicatoren voor de beoordeling van de waterkwaliteit, geen gebruik gemaakt van andere beslissingsondersteunende middelen die het verband kunnen leggen tussen de abiotiek en de biotiek van een waterloop. De Europese Kaderrichtlijn Water (KRW) stelt echter voor alle lidstaten van de Europese Unie een goede ecologische toestand van alle waterlichamen voorop tegen 2015. Een groot deel van deze waterlichamen kan als stromende wateren of rivieren worden beschouwd. 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 ecologische rivierbeoordeling te kunnen ontwikkelen. Kennis over deze relaties kan eveneens nuttig zijn bij zowel het opsporen van oorzaken van bepaalde riviercondities als bij beslissingsondersteuning inzake rivierherstel en beheer om zo aan de eisen van de KRW te voldoen.

Ondanks de ecologische doelstellingen van de KRW worden ecologische modellen tot nog toe zelden gebruikt om het rivierbeheer te ondersteunen. Modellen kennen nochtans verschillende interessante toepassingen in deze context. Vooral habitatgeschiktheidsmodellen die de habitatvereisten van organismen kunnen voorspellen op basis van rivierkarakteristieken blijken zeer nuttig te zijn. Dit type modellen is echter nog maar zeer recent aanvaard binnen het milieubeheer.

Een eerste doelstelling van deze thesis is het bepalen van geschikte variabelen en ecosysteemprocessen door het toepassen van artificiële neurale netwerken (ANN), een gegevensgebaseerde techniek, bij het voorspellen van de habitatgeschiktheid van biologische gemeenschappen in rivieren. Deze ANN modellen werden ontwikkeld en toegepast in een MATLAB omgeving. Het onderzoek richt zich vooral op macro-invertebraten in beken en smalle rivieren in het Zwalmbekken, een deelbekken van het Boven-Scheldebekken, in Vlaanderen (België). De geselecteerde staalnamepunten worden gekenmerkt door een gradiënt gaande van een bijna natuurlijke tot een sterk verstoorde toestand (waterverontreiniging, fysische habitatverstoring).

De resultaten van de modellen illustreerden vooreerst het nut om twee evaluatiecriteria te gebruiken bij het beoordelen van ANN voorspellingen. Door de gemakkelijke interpretatie naar watermanagers en beleidsmensen toe is vanuit een ecologisch standpunt het aantal juiste voorspellingen (de CCI) het meest geschikt. Er is echter duidelijk aangetoond dat de CCI wordt beïnvloed door de frequentie van voorkomen van het gemodelleerde organisme. Daarom werd een tweede evaluatiecriterium (de Cohens kappa, CK) gebruikt. De invloed van de frequentie van voorkomen op de CK bleek immers verwaarloosbaar te zijn. Omdat de grootte van de training- en testdataset enerzijds de mogelijkheid tot veralgemenen van de modellen en anderzijds de accuraatheid van de evaluatiecriteria kan beïnvloeden werd gezocht naar de optimale grote van training- en testdataset. Op basis van de CCI en de CK bleek het gebruik van drie vierden van de dataset voor training en één vierde voor het testen optimaal te zijn (cf. 4-fold cross-validatie). De opbouw van het neuraal netwerk is in het algemeen zeer probleem afhankelijk. Bij het vergelijken van de verschillende modelconstructies voor elk organisme bleken de verschillen in performantie echter zeer klein te zijn. Zo kon naar modelopbouw (aantal verborgen lagen en neuronen) geen eenduidige conclusie per organisme getrokken worden.

Er werden eveneens drie methoden (de Weights, de PaD en de Profile methode) toegepast om het aandeel van de verschillende milieuvariabelen op de voorspelling van de aan- en afwezigheid van de macro-invertebraten aan te geven. Op deze manier konden de belangrijkste riviereigenschappen per organisme bepaald worden. Het was niet altijd evident om trends over de vijf organismen, de drie methodes en de verschillende ‘folds’ te achterhalen. De eerste twee kunnen toegeschreven worden aan de verschillende ecologische voorkeur van elk organisme en de verschillende aspecten die de drie methoden behandelen. De instabiliteit over de verschillende ‘folds’ kan waarschijnlijk verklaard worden door de relatief kleine datasets waarmee wordt gewerkt. Daarbij komt nog de grote variabiliteit binnen de staalnameplaatsen, het groot aantal invoervariabelen en de uitbijters in de metingen. Verder onderzoek op basis van grotere datasets is daarom aangewezen. Desondanks kan besloten worden dat een combinatie van de PaD en de Profile methode een zeer goed inzicht verschaft in de ecologische betekenis van de modellen en hun praktische relevantie voor beslissingsondersteuning in het rivierbeheer.

Ecosysteemmodellen, zoals ANN, houden meestal geen rekening met ruimtelijke en temporele relaties. De migratiemogelijkheden van de voorspelde organismen en migratieknelpunten langs de rivier kunnen daarom interessante bijkomende informatie leveren bij het beoordelen van herstelmaatregelen. Daarom werden als uitbreiding van de habitatgeschiktheidsmodellen migratiemodellen ontwikkeld voor een deel van het Zwalmbekken. Deze modellen beschikken over de mogelijkheid om de verbondenheid van populaties of de mogelijkheid tot migratie vanuit een bronpopulatie naar een hersteld riviersegment te onderzoeken. Deze migratiemodellen zijn gebaseerd op de 'Cost Weighted Distance' functie, een krachtig instrument binnen de ArcGis Spatial Analyst extensie van ArcGIS 8.3. Om de migratiemodellen (theoretisch) te valideren werd een sensitiviteitsanalyse uitgevoerd. Daarbij werden de waarden, toegekend aan de weerstanden, gevarieerd om zo de gevoeligheid van de ontwikkelde migratiemodellen en de wijzigingen in voorspelde uitvoer te onderzoeken.

Om de watermanagers en beheerders te overtuigen van de mogelijkheden en het praktisch nut van de ontwikkelde modellen, werden vier herstelscenario's voor waterlopen in het Zwalmbekken gesimuleerd, waaronder een hermeanderingsproject, de aanleg van een collector, de aanleg van een natuurvriendelijke oever en het verwijderen van een stuw voor waterbeheersing. Deze oefeningen bewijzen het nut van de ontwikkelde modellen en geven tevens inzicht in de verschuiving van indicatororganismen bij het doorvoeren van herstelmaatregelen.

Er kan besloten worden dat de combinatie van habitatgeschiktheids- en migratiemodellen zeer nuttig is bij het voorspellen van de mogelijke ecologische effecten van rivierherstel, evenals de effecten van rivierverstoring. Op deze manier kunnen ze als beslissingsondersteunend instrument dienen voor rivierbeheerders.

Desondanks kunnen een aantal bedenkingen en aanbevelingen voor verder onderzoek aangehaald worden. In de eerste plaats zou een optimalisatie van de monitoringsnetwerken (bijvoorbeeld de selectie van staalnamepunten afstemmen op de voorspelling van herstelmaatregelen) en de gebruikte datasets (bijvoorbeeld het inpassen van nieuwe, essentiële variabelen enerzijds en het verwijderen van minder relevante invoervariabelen anderzijds) een verdere verbetering van de voorspellingen kunnen opleveren. Daarenboven zou een verdere optimalisatie en uitbreiding van de ANN en migratiemodellen (vooral de praktische validatie) zeer nuttig kunnen zijn. Anderzijds zou, met de eisen van de KRW in het achterhoofd, de ontwikkeling van habitatgeschiktheidsmodellen voor andere macro-invertebraten, vissen, waterplanten, … de waarde van de modellen kunnen verhogen. Tenslotte is de uitbreiding van beide modellen naar andere stroombekkens in Vlaanderen (om de voorspellingsrange van de modellen te vergroten) alsook de koppeling van de habitatgeschiktheidsmodellen met hydraulische en waterkwaliteitsmodellen in de toekomst noodzakelijk om de betrouwbaarheid en praktische toepasbaarheid van deze modellen te verbeteren.

Curriculum vitae

Personalia

Name:	Andy Pierre Dedecker
Date and place of birth:	13/02/1978, Torhout (Belgium)
Married	Eefje Verschaeve (25/04/1978, Brugge (Belgium))
Address (private):	Groendreef 137, B-9000 Gent
Address (work)	Department Applied Ecology and Environmental Biology Ghent University J. Plateaustraat 22, B-9000 Gent E-mail: andy.dedecker@UGent.be Tel.: 0032 (0)9 264 37 08 Fax.: 0032 (0)9 264 41 99

Education

2001: **Bioscience Engineer, option Environmental Technology**, Ghent University, Faculty of Bioscience Engineering, Coupure Links 653, B-9000 Gent. Script 'Ecosystem monitoring and modelling of the Zwalm river basin based on Artificial Neural Networks', Promoter: Prof. Dr. N. De Pauw.

Publications

- **International journal publications with peer-review**

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. In Proceedings of the 2nd Symposium on European Freshwater Systems. *TheScientificWorldJOURNAL*, **2**, 96-104.

Dedecker, A.P., Goethals, P.L.M., Gabriels, W. and De Pauw, N. (2004). Optimization 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., D'heygere, T., Gevrey, M., Lek, S. and 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*.

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

Dedecker, A.P., Goethals, P.L.M., D’heygere, T., Gevrey, M., Lek, S. and De Pauw, N. (submitted). Selecting variables of habitat suitability of *Asellus* (Crustacea, Isopoda) by applying input variable contribution methods to Artificial Neural Network models. *Environmental Modeling & Assessment*.

Dedecker, A.P., Gabriels, W., Lek, S., Goethals, P.L.M. and De Pauw, N. (submitted). Applications of artificial neural networks predicting macroinvertebrates in freshwaters. *Journal of Applied Ecology Review*.

Dedecker, A.P., Van Melckebeke, K., Goethals, P.L.M. and De Pauw, N. (submitted). Development of migration models for macroinvertebrates in the Zwalm River basin as a tool in river assessment and restoration management. *Ecological Modelling*.

Dedecker, A.P., Janssen, K., Wouters, H., Thas, O., Goethals, P.L.M. and De Pauw, N. (in preparation). Explaining the variability in data collection for river assessment purposes. *Freshwater Biology*.

Dakou, E., D’heygere, T., Dedecker, A.P., Goehals, P.L.M., Lazaridou-Dimitriadou, M. and De Pauw, N. (submitted). Use of rule induction techniques for prediction of macroinvertebrate taxa in the river Axios (Northern Greece). *Aquatic Ecology*.

Dakou, E., D’heygere, T., Dedecker, A.P., Goethals, P.L.M., De Pauw, N. and Lazaridou-Dimitriadou, M. (submitted). Development of artificial neural network models predicting macroinvertebrate taxa in the river Axios (Northern Greece). *Annales de Limnologie – International Journal of Limnology*.

Goethals, P.L.M., Dedecker, A.P., Bouma, J.J., François, D., Verstraete, A., D’heygere, T., Mouton, A., De Clercq, M. and De Pauw, N. (submitted). Development of the Water Ecology Decision Support System (WAECO-DSS) to enable integrated cost-benefit analyses in river restoration management: case study on the Zwalm river basin (Belgium). *Journal of Environmental Management*.

Gabriels, W., Dedecker, A.P., Goethals, P.L.M., Lek, S. and De Pauw, N. (submitted). Analysing and predicting the effect of river pollution on macrobenthos communities in Flanders (Belgium) using a stepwise input variable selection procedure in combination with artificial neural networks. *Aquatic Ecology*.

- **Publications in international proceedings with peer-review**

Dedecker, A.P., Goethals, P.L.M., Gabriels, W. and De Pauw, N. (2002). Optimisation of Artificial Neural Network (ANN) model design for prediction of macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium). In: A.E., Rizzoli and A.J., Jakeman (Eds.), *Integrated Assessment and Decision Support, proceedings of the 1st biennial meeting of the International Environmental Modelling and Software Society*, Vol.2, 142-147 p.

D'heygere, T., Goethals, P., Dedecker, A. and De Pauw, N. (2003). Development of a monitoring network to model the habitat suitability of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). In: D.A., Post (Ed.), MODSIM 2003: *Integrative Modelling of*

Biophysical, Social and Economic Systems for Resource Management Solutions, Townsville, Australia, 753-758 p.

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). In: D.A., Post (Ed.), MODSIM 2003: *Integrative Modelling of Biophysical, Social and Economic Systems for Resource Management Solutions*, Townsville, Australia, 971-976 p.

Adriaenssens, V., Dedecker, A., D'heygere, T., Goethals, P.L.M and De Pauw, N. (2003). Relations between structural characteristics and macro-invertebrate communities in the Zwalm river basin at different spatial scales. In: J.-J., Symoens and K., Wouters (Eds.), *Biological evaluation and monitoring of the quality of surface waters*, Cercle Hydrobiologique de Bruxelles, Brussels, 81-86 p.

Dedecker, A.P., Van Melckebeke, K., D'heygere, T., Goethals, P.L.M. and De Pauw, N. (2004). Impact study of physical habitat restoration on aquatic macroinvertebrates in rivers by means of neural networks and migration models: case study of the mayfly *Baetis* (Insecta – Ephemeroptera). In: D.G.J., Lastra and P.V., Martínez (Eds.), *proceedings of the 5th International Symposium on Ecohydraulics, Aquatic Habitats: Analysis & Restoration*, Madrid, Spain, 1157-1161 p.

- **Other journal publications**

Dedecker, A., Goethals, P., Gabriels, W. and De Pauw, N. (2001). River management applications of ecosystem models predicting aquatic macroinvertebrate communities based on artificial neural networks (ANN's). *Communications in Agricultural and Applied Biological Sciences*, **66(4)**, 207-211.

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. *Communications in Agricultural and Applied Biological Sciences*, **66(1)**, 71-86.

Dedecker, A., Goethals, P.L.M., D'heygere, T. and 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). *Communications in Agricultural and Applied Biological Sciences*, **67(4)**, 105-108.

Adriaenssens, V., D'heygere, T., Dedecker, A., Goethals, P.L.M. and De Pauw, N. (2002). Relations between structural characteristics and macroinvertebrate communities in the Zwalm river basin at different spatial scales. *Communications in Agricultural and Applied Biological Sciences*, **67(4)**, 129-132.

D'heygere, T., Adriaenssens, V., Dedecker, A., Gabriels, W., Goethals, P.L.M. and De Pauw, N. (2002). Development of a decision support system for integrated water management in the Zwalm river basin, Belgium. *Communications in Agricultural and Applied Biological Sciences*, **67(4)**, 159-162.

Dedecker, A.P., Goethals, P.L.M. and De Pauw, N. (2003). Overview and quantification of the factors affecting the upstream and downstream movements of *Gammarus pulex* (Amphipoda). *Communications in Agricultural and Applied Biological Sciences*, **68(1)**, 25-32.

Dedecker, A.P., Van Melckebeke, K., Goethals, P.L.M. and De Pauw, N. (2004). Development of migration models for macroinvertebrates in the Zwalm river basin as a tool in river assessment and restoration management. *Communications in Agricultural and Applied Biological Sciences*, **69(2)**, 81-84.

D'heygere, T., Mouton, A., Dedecker, A.P., Adriaenssens, V., Goethals, P.L.M. and De Pauw N. (2004). Development of predictive ecological models for river restoration. *Communications in Agricultural and Applied Biological Sciences*, **69(2)**, 15-18.

• Chapters in books

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. In: F., Recknagel (Ed.), *Ecological Informatics. Understanding ecology by biologically-inspired computation*, Springer, Berlin**,** 91-108 p.

Dedecker, A.P., Goethals, P.L.M. and De Pauw, N. (2005). Sensitivity and robustness of stream model based on artificial neural networks for the simulation of different management scenarios.. In: S., Lek, M., Scardi, P.F.M., Verdonschot, J.-P., Descy and Y.S., Park (Eds.), *Modelling community structure in freshwater ecosystems*. Springer-Verlag, 133-146 p.

• Proceedings without peer-review

Dedecker, A., Goethals, P. Gabriels, W. and De Pauw, N. (in press). Prediction of macroinvertebrate communities in the Zwalm catchment by means of artificial neural networks. In: Scheldt catchment. In: K. Buis (Ed.), KVAB, Antwerpen, Belgium (in dutch).

Meetings and symposia

Workshop on Ecological research in the Scheldt catchment, Brussels, Belgium, 29-30 March 2001
Contribution: Dedecker, A., Goethals, P., Gabriels, W. and De Pauw, N. Prediction of macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium) based on artificial neural networks: poster presentation.

2nd Symposium for European Freshwater Sciences (SEFS2), Toulouse, France, 7-12 July 2001
Contribution: Dedecker, A., Goethals, P., Gabriels, W. and De Pauw, N. Development and application of ecosystem models based on artificial neural networks for integrated river management: poster presentation.

Contribution as co-author: Goethals, P., Dzeroski, S., Dedecker, A., Raes, N., Adriaenssens, V., Gabriels, W. and De Pauw, N. Comparing classification trees, artificial neural networks and fuzzy logic models to predict macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium): oral presentation.

Workshop on Parameter selection in modelling aquatic community structure, Namur, Belgium, 15-16 September 2001.

7th FLTBW PhD Symposium, Ghent, Belgium, 10 October 2001
Contribution: Dedecker, A., Goethals, P., Gabriels, W. and De Pauw, N. Development and application of ecosystem models based on artificial neural networks for integrated river management: poster presentation.

B-IWA Happy hour, Brussels, Belgium, 22 October 2001
Contribution as co-author: Goethals, P., D'heygere, T., Raes, N., Dedecker, A., Adriaenssens, V., Gabriels, W., Heylen, S., Detemmerman, L., Houzet, H., Decraene, E., De Clercq, K., Simons, F. and De Pauw, N. Integrating monitoring modelling and assessment in ecological river management: oral presentation.

Workgroup meeting COST Action 626, Vigo, Spain, 7-10 November 2001.

NECO**V** Wintermeeting on New developments in ecology, Antwerp, Belgium, 12-13 December 2001
Contribution: Dedecker, A., Goethals, P., Gabriels, W. and De Pauw, N. Development and application of ecosystem models based on artificial neural networks for integrated river management: poster presentation.

Symposium on the Status and trends of Belgian fauna with the focus on exotic species, Brussels, Belgium, 14 December 2001.

Workgroup meeting COST Action 626, Oulu, Finland, 30-31 May 2002.

B-IWA Happy hour, Brussel, Belgium, 10 June 2002
Contribution as co-author: D'heygere, T., Adriaenssens, V., Dedecker, A., Gabriels, W., Goethals, P.L.M. and De Pauw, N. Development of a Decision Support System for integrated water management in the Zwalm river basin, Belgium: poster presentation.

Symposium of the International Environmental Modelling and Software Society (iEMSs) on Integrated Assessment and Decision Support, Lugano, Switzerland, 24-27 June 2002
Contribution: Dedecker, A., Goethals, P., Gabriels, W. and De Pauw, N. Optimisation of Artificial Neural Network (ANN) model design for prediction of macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium): oral presentation.

3rd Conference of the International Society for Ecological Informatics (ISEI3), Rome, Italy, 26-30 August 2002
Contribution: Dedecker, A., Goethals, P.L.M. and De Pauw, N. Sensitivity and robustness of predictive neural network ecosystem models for simulation of different management scenarios: oral presentation.
Contribution as co-author: Goethals, P.L.M., Dedecker, A., D'heygere, T., Adriaenssens, V. and De Pauw, N. River quality assessment based on fuzzy logic: oral presentation.

8th FLTBW PhD Symposium, Ghent, Belgium, 9 October 2002.
Contribution: Dedecker, A., Goethals, P.L.M., D'heygere, T. and De Pauw, N. Use of Artificial Neural Networks (ANNs) and Geographical Information Systems (GIS) to simulate the migration of macroinvertebrates in the Zwalm river basin (Flanders, Belgium): poster presentation.
Contribution as co-author: Adriaenssens, V., Dedecker, A., D'heygere, T., Goethals, P.L.M. and De Pauw, N. Relations between structural characteristics and macroinvertebrate communities in the Zwalm river basin: poster presentation.

Conference on Ecological Informatics Applications in Water Management, Ghent, Belgium, 6-7 November 2002
Contribution as co-author: Adriaenssens, V., Goethals, P.L.M., Dedecker, A., D'heygere, T. and De Pauw, N. Data collection in the Zwalm river basin for the prediction of the effect of restoration actions on macroinvertebrate communities: oral presentation.
Contribution: Dedecker, A., Goethals, P.L.M., D'heygere, T. and De Pauw, N. Use of Artificial Neural Networks (ANNs) and Geographical Information Systems (GIS) to simulate the migration of macroinvertebrates in the Zwalm river basin (Flanders, Belgium): poster presentation.

Symposium on Current Themes in Ecology: Spatial Ecology, Wageningen, The Netherlands, 22 November 2002.

COST Action 626: European Aquatic Modelling Network, Scaling Subgroup Meeting, Ghent, Belgium, 11-14 December 2002
Contribution as co-author: Adriaenssens, V., D'heygere, T., Dedecker, A., Goethals, P.L.M. and De Pauw, N. Relations between structural characteristics and macroinvertebrate communities in the Zwalm river basin at different spatial scales: oral presentation.

NECOV Wintermeeting 2003, Leiden, The Netherlands, 15-16 January 2003.

3rd Symposium for European Freshwater Sciences (SEFS3), Edinburgh, UK, 13-18 July 2003
Contribution: Dedecker, A., Goethals, P.L.M., D'heygere, T. and De Pauw, N. Combining ANNs and migration models to simulate the habitat suitability and accessibility to *Gammarus pulex* in the Zwalm river basin (Belgium): oral presentation.

9th FLTBW PhD Symposium, Leuven, Belgium, 16 October 2003.

NECOV Wintermeeting 2004, Ghent, Belgium, 14-15 Januari 2004.

B-IWA Happy hour, Brussels, Belgium, 7 June 2004
Contribution as co-author: Goethals, P., Dedecker, A., D'heygere, T., Adriaenssens, V., Gabriels, W., Depestele, J., Mouton, A., Dominguez, L., Zarkami, R., Ambelu, A. and De Pauw, N. Ecotechnological solutions for river management: poster presentation.

10th FLTBW PhD Symposium, Ghent, Belgium, 29 September 2004
Contribution: Dedecker, A., Van Melckebeke, K., D'heygere, T., Goethals, P.L.M. and De Pauw, N. Development of migration models for macroinvertebrates in the Zwalm river basin as a tool in river assessment and restoration management: poster presentation.

4th Conference of the International Society for Ecological Informatics (ISEI4), Busan (Pusan), South Korea, 24-28 October 2004
Contribution: Dedecker, A., Van Melckebeke, K., Goethals, P.L.M. and De Pauw, N. Development of migration models for macroinvertebrates in the Zwalm river basin as a tool in river assessment and restoration management: oral presentation.

NECO**V** Wintermeeting 2005, 's Hertogenbosch, The Netherlands, 26-27 januari 2005
Contribution: Dedecker, A., Van Melckebeke, K., Goethals, P.L.M. and De Pauw, N. Development of migration models for macroinvertebrates in rivers: oral presentation.

Scientific awards

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 and 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 and N. De Pauw.

Foreign Research Visits

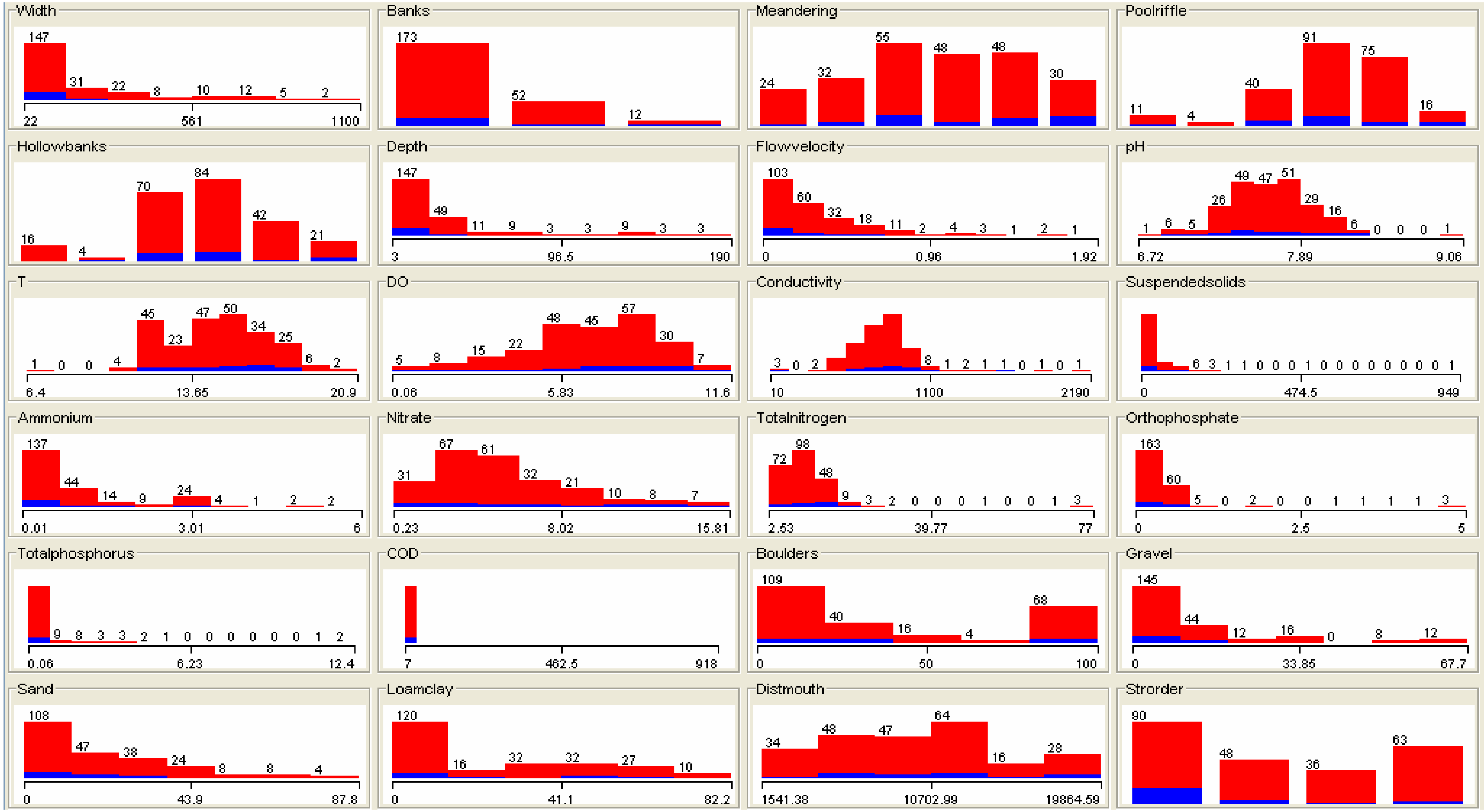
12-13 December 2002: LADYBIO UMR 5172, CNRS-University Paul Sabatier, Toulouse, France.

16-20 November 2003: LADYBIO UMR 5172, CNRS-University Paul Sabatier, Toulouse, France, related to the Tournesol project 'Development of predictive river ecosystem models for decision support in integrated water management'.

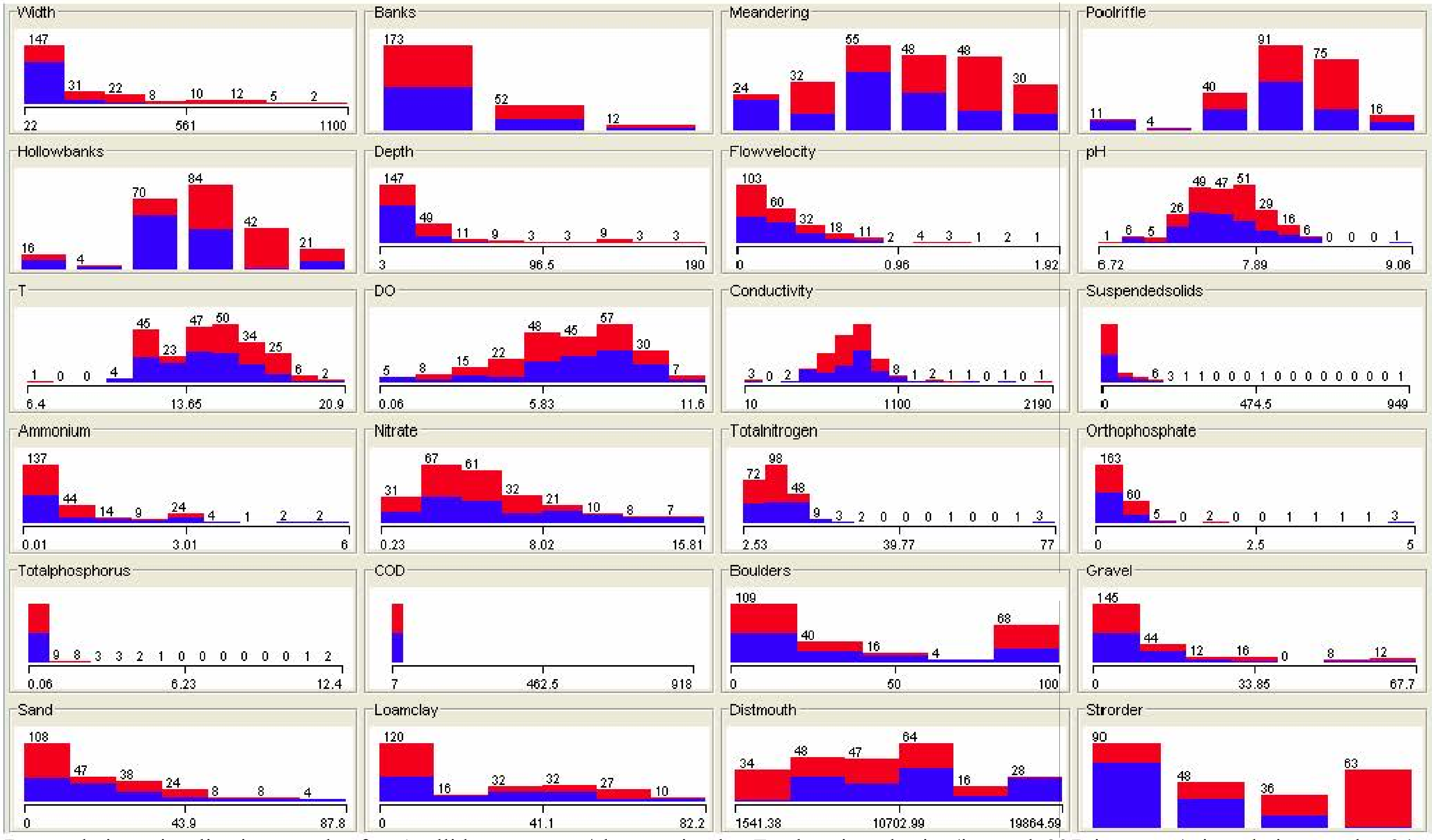
27 March - 1 May 2005: VLIR 'South initiative 2005', Indonesia, 'Evaluation, optimisation and implementation of biological monitoring techniques to improve the Way Besai river quality in Indonesia'.

Appendices

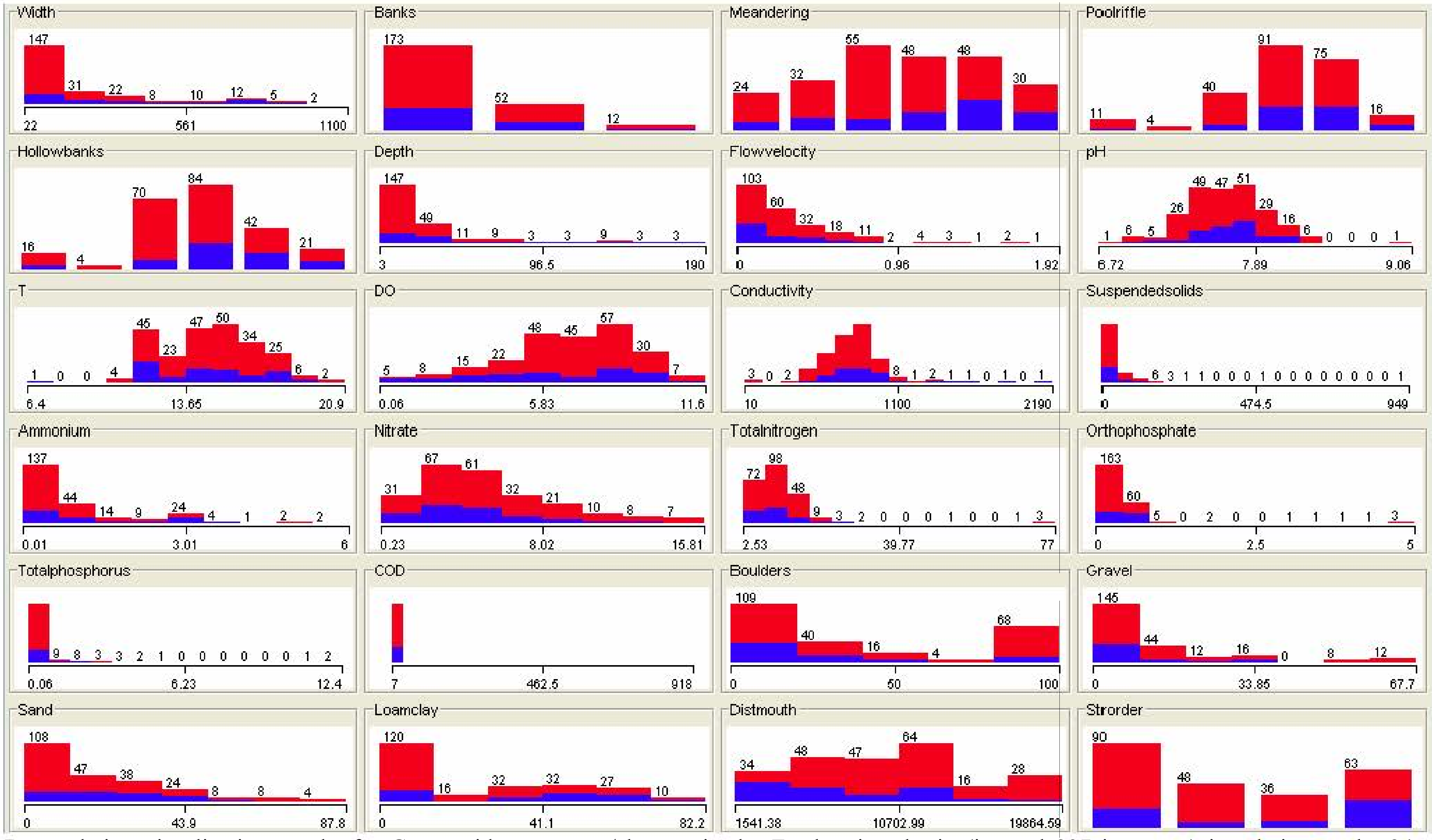
Appendix 1: Data relation visualization graphs for the presence/absence of Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae in the Zwalm river basin (in total 237 instances) and the 'short distance' monitoring network (in total 120 instances) in relation to the 24 environmental variables (absent = blue; present = red).



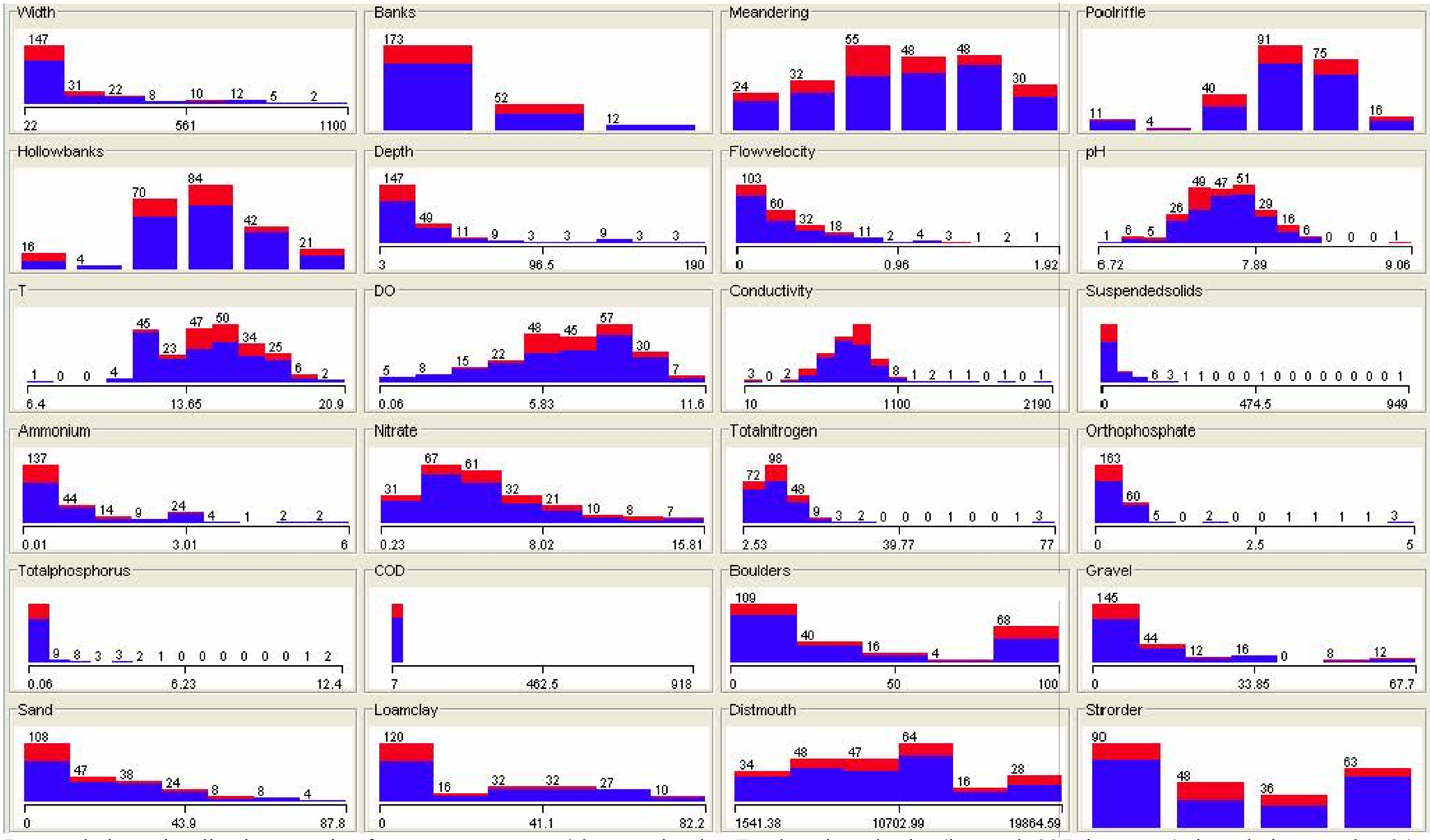
Data relation visualization graphs for Tubificidae presence/absence in the Zwalm river basin (in total 237 instances) in relation to the 24 environmental variables (Tubificidae absent in 25 instances (blue), Tubificidae present in 212 instances (red)).



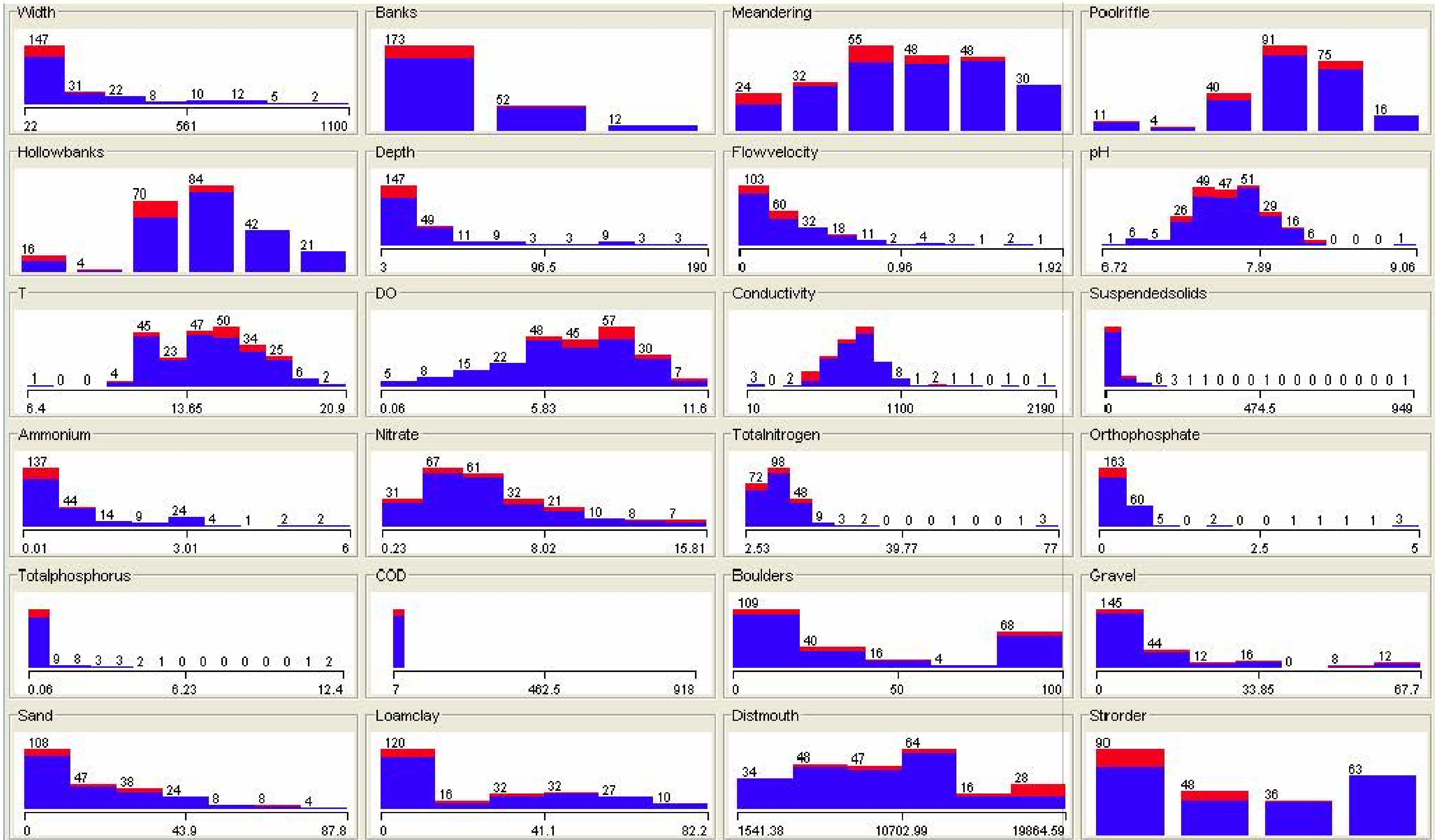
Data relation visualization graphs for Asellidae presence/absence in the Zwalm river basin (in total 237 instances) in relation to the 24 environmental variables (Asellidae absent in 117 instances (blue), Asellidae present in 120 instances (red)).



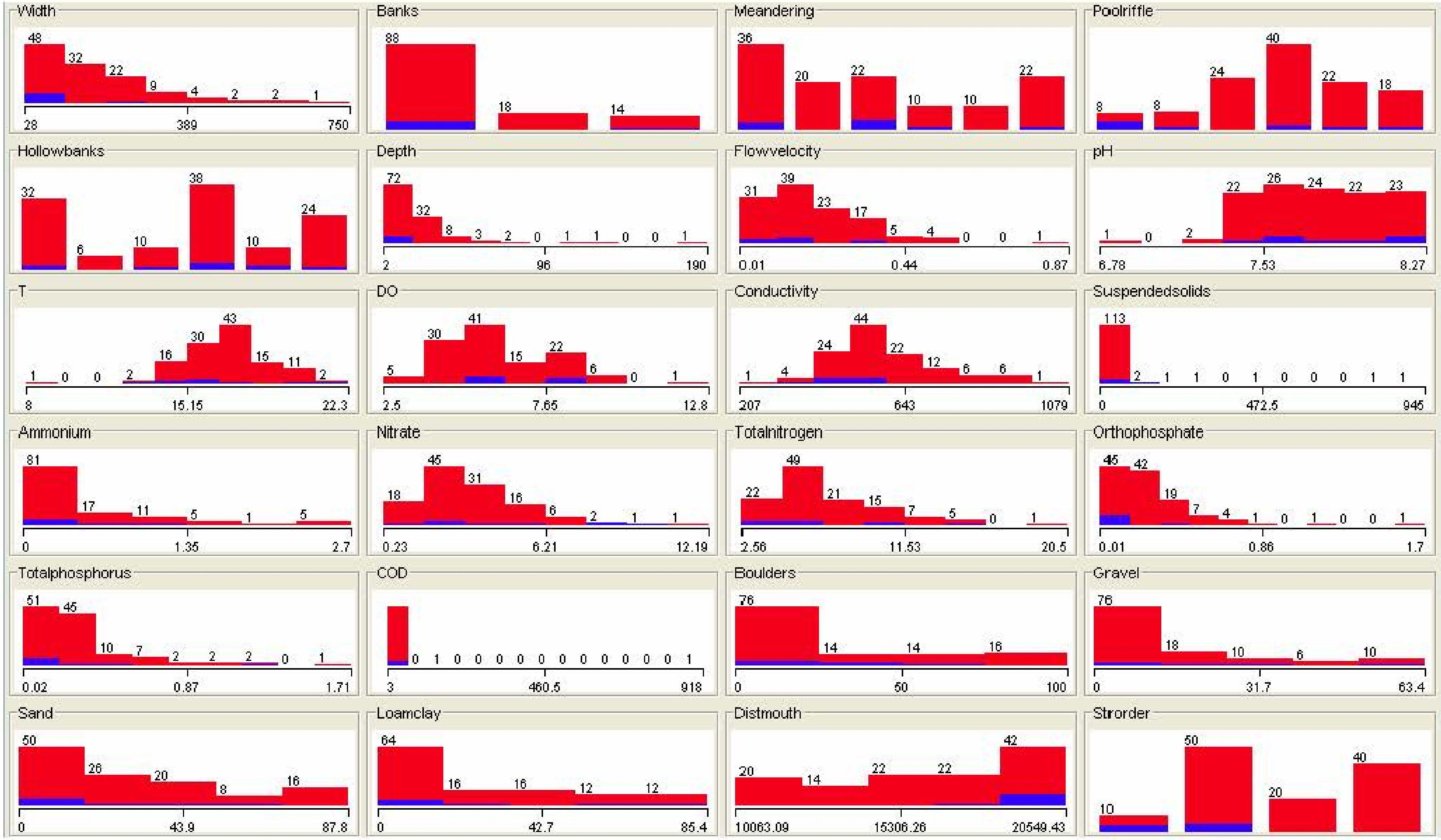
Data relation visualization graphs for Gammaridae presence/absence in the Zwalm river basin (in total 237 instances) in relation to the 24 environmental variables (Gammaridae absent in 64 instances (blue), Gammaridae present in 173 instances (red)).



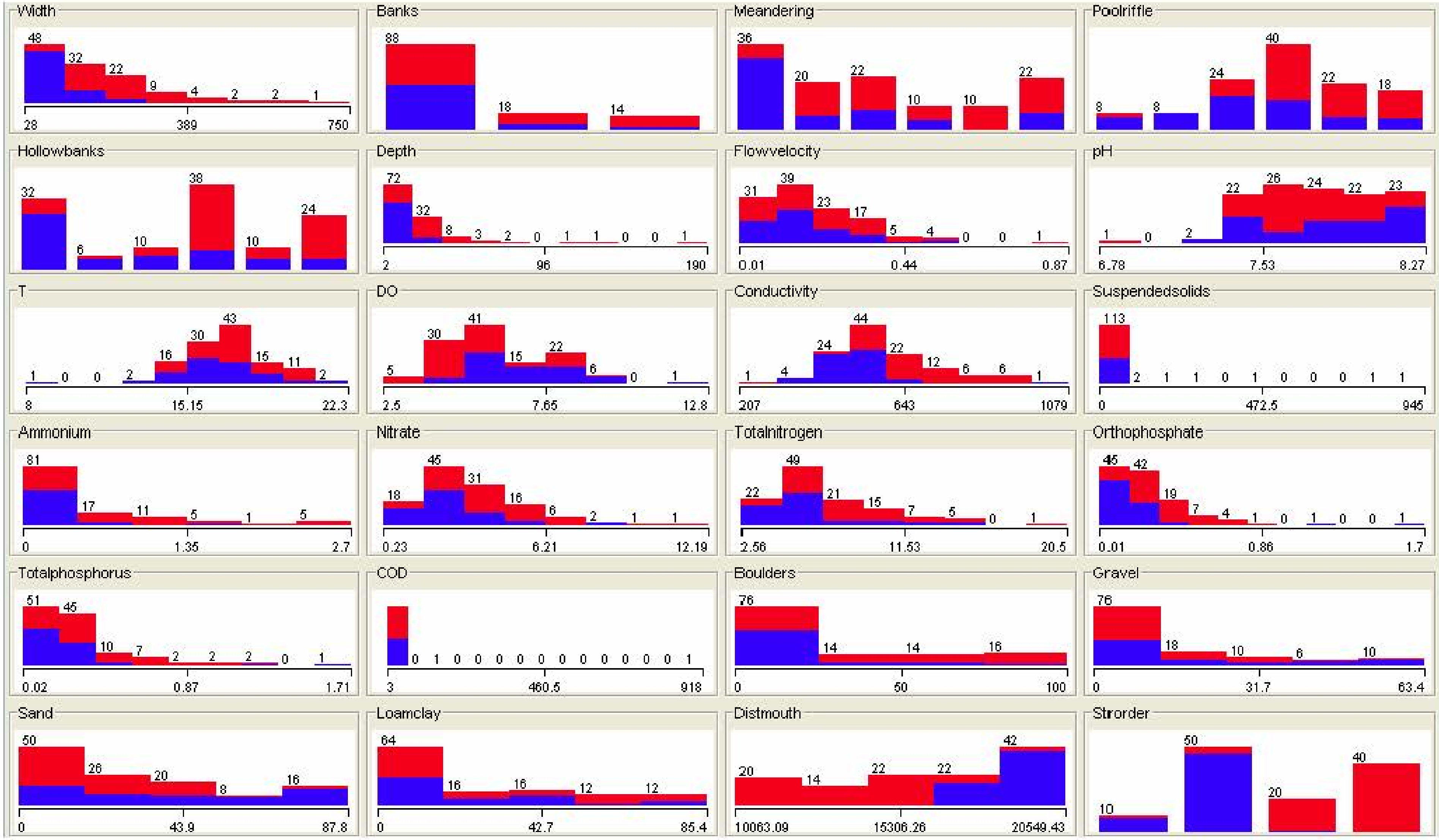
Data relation visualization graphs for *Baetis* presence/absence in the Zwalm river basin (in total 237 instances) in relation to the 24 environmental variables (*Baetis* absent in 179 instances (blue), *Baetis* present in 58 instances (red)).



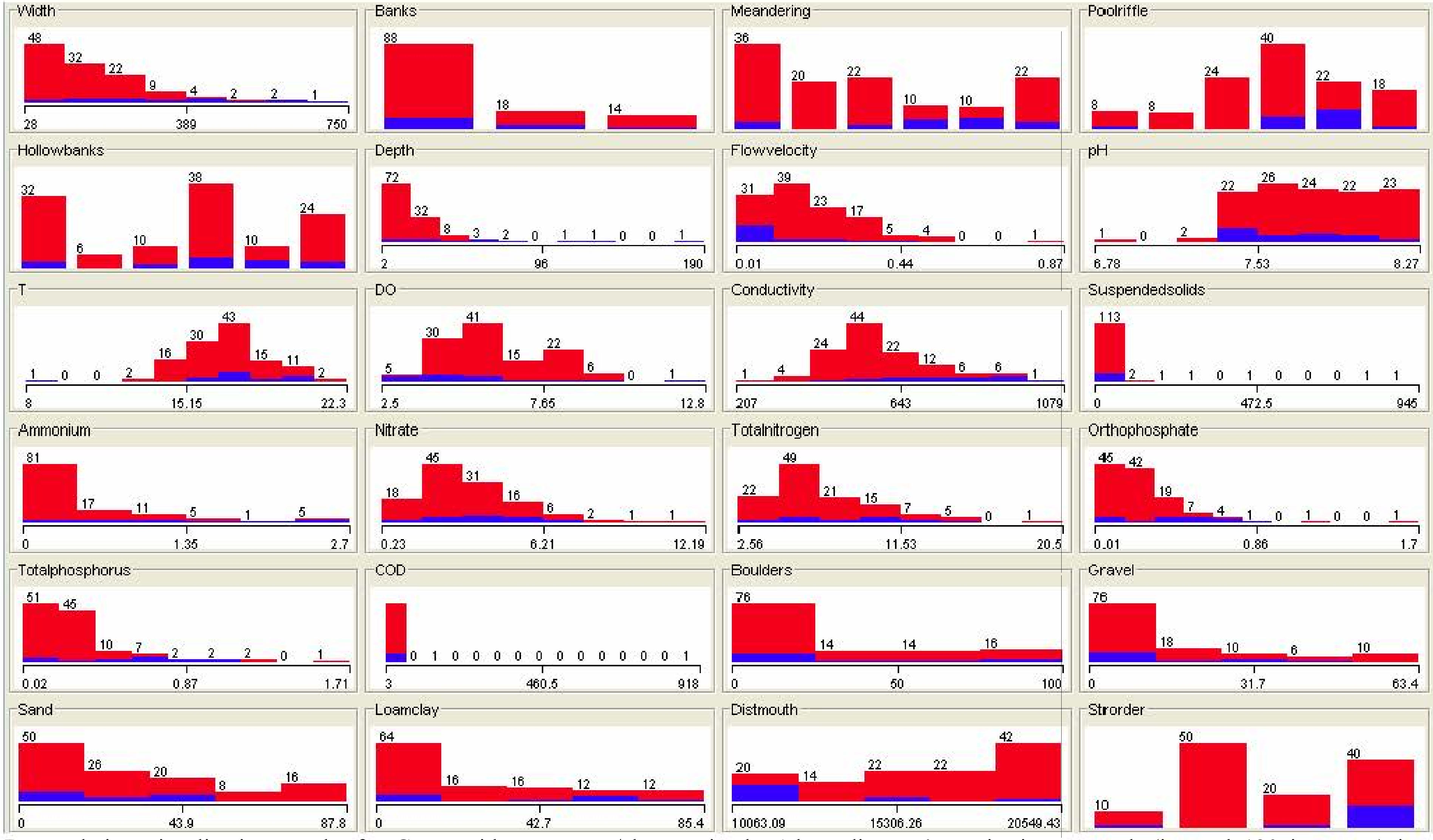
Data relation visualization graphs for Limnephilidae presence/absence in the Zwalm river basin (in total 237 instances) in relation to the 24 environmental variables (Limnephilidae absent in 208 instances (blue), Limnephilidae present in 29 instances (red)).



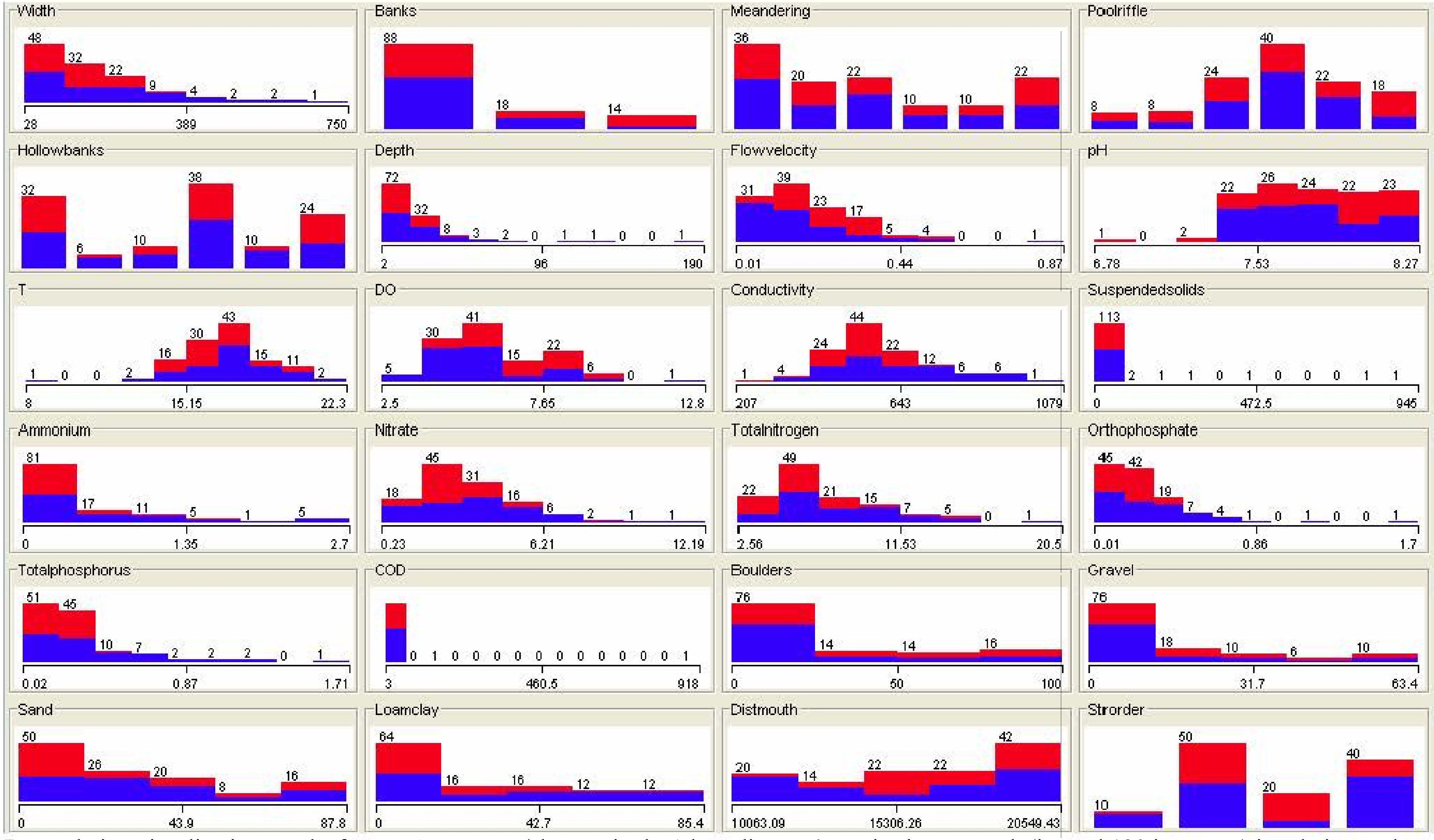
Data relation visualization graphs for Tubificidae presence/absence in the ‘short distance’ monitoring network (in total 120 instances) in relation to the 24 environmental variables (Tubificidae absent in 9 instances (blue), Tubificidae present in 111 instances (red)).



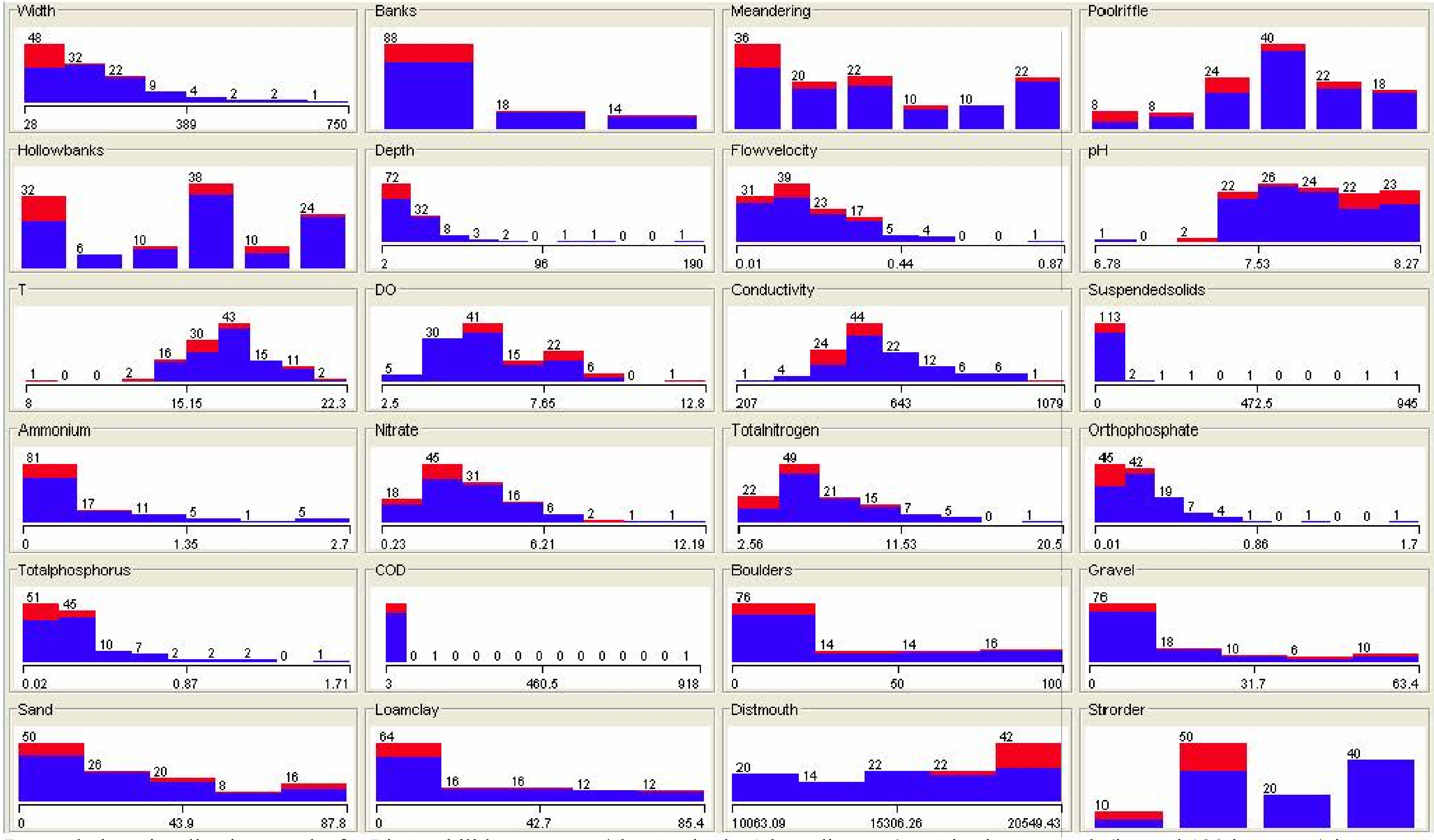
Data relation visualization graphs for Asellidae presence/absence in the ‘short distance’ monitoring network (in total 120 instances) in relation to the 24 environmental variables (Asellidae absent in 55 instances (blue), Asellidae present in 65 instances (red)).



Data relation visualization graphs for Gammaridae presence/absence in the 'short distance' monitoring network (in total 120 instances) in relation to the 24 environmental variables (Gammaridae absent in 17 instances (blue), Gammaridae present in 103 instances (red)).



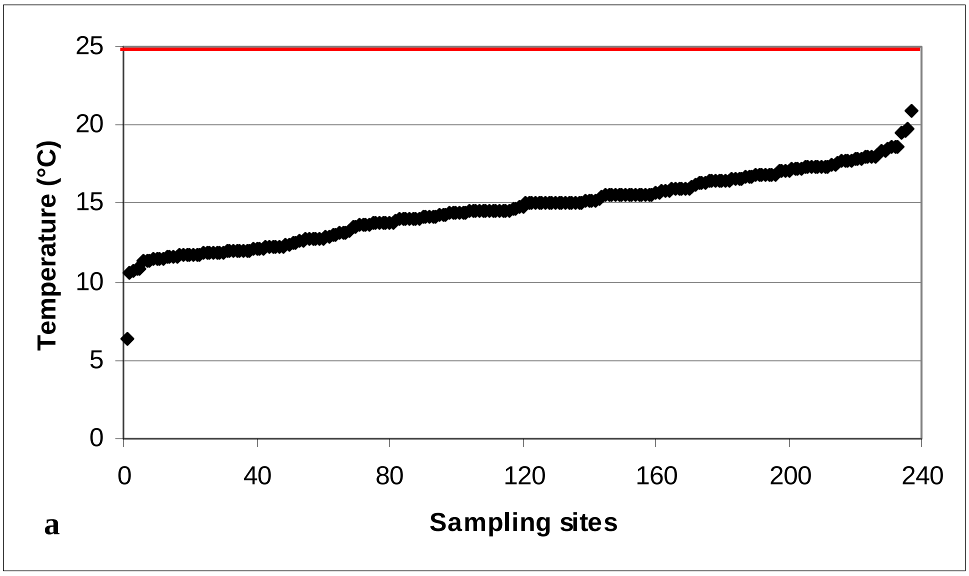
Data relation visualization graphs for *Baetis* presence/absence in the ‘short distance’ monitoring network (in total 120 instances) in relation to the 24 environmental variables (*Baetis* absent in 68 instances (blue), *Baetis* present in 52 instances (red)).

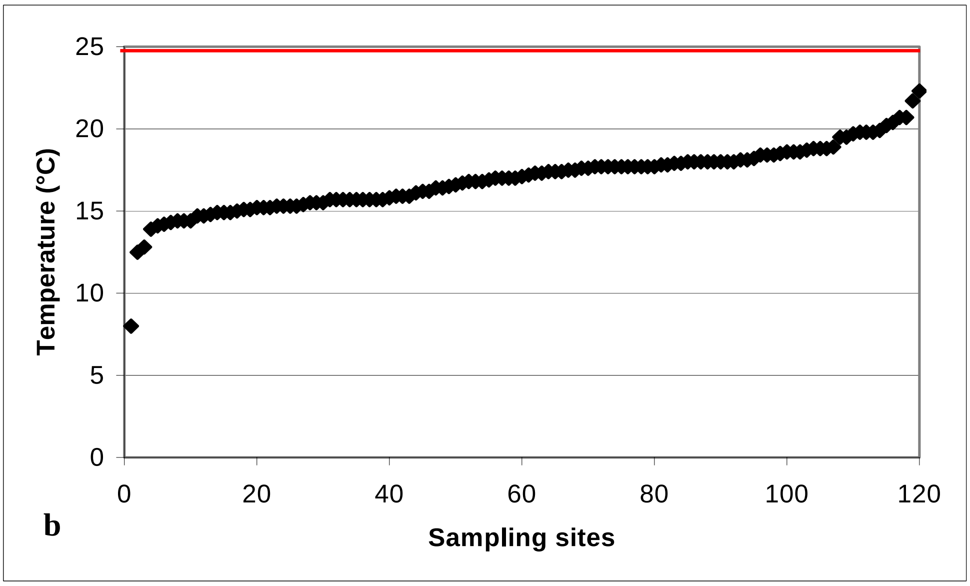


Data relation visualization graphs for Limnephilidae presence/absence in the ‘short distance’ monitoring network (in total 120 instances) in relation to the 24 environmental variables (Limnephilidae absent in 99 instances (blue), Limnephilidae present in 21 instances (red))

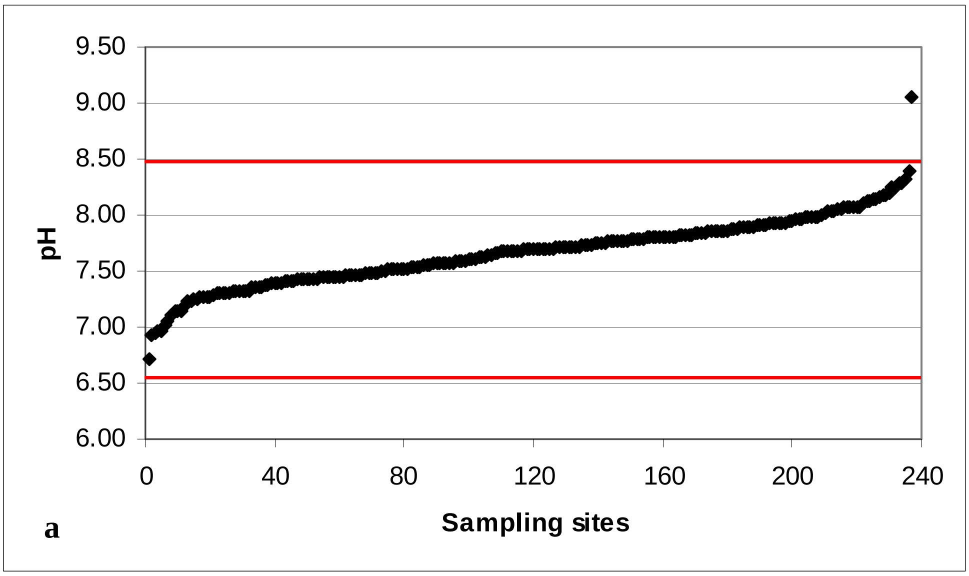
Appendix 2: Distribution of the environmental input variable, with indication of the quality standard where appropriate in red, over the 237 (a, Zwalm river basin) and 120 (b, 'short monitoring' monitoring network) sampling sites.

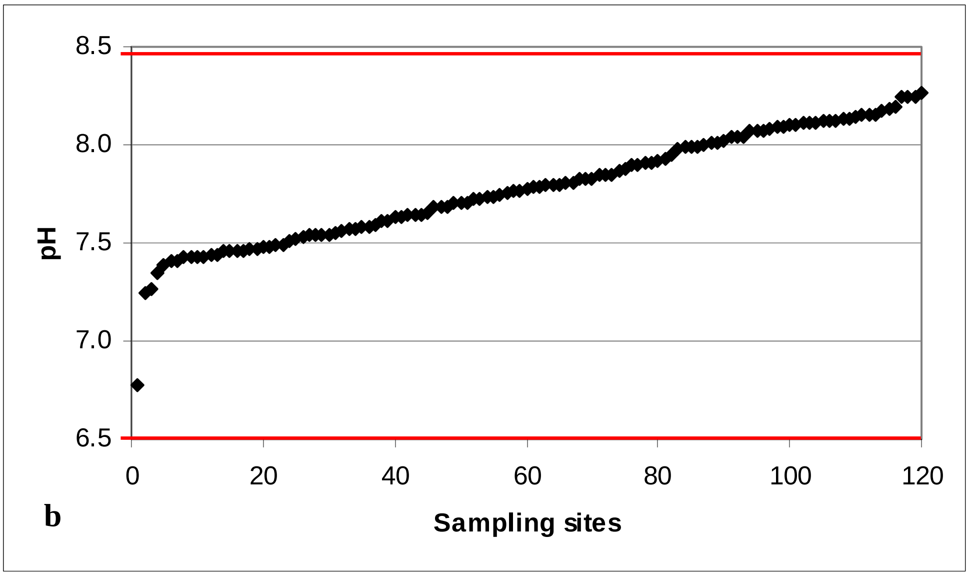
Temperature:



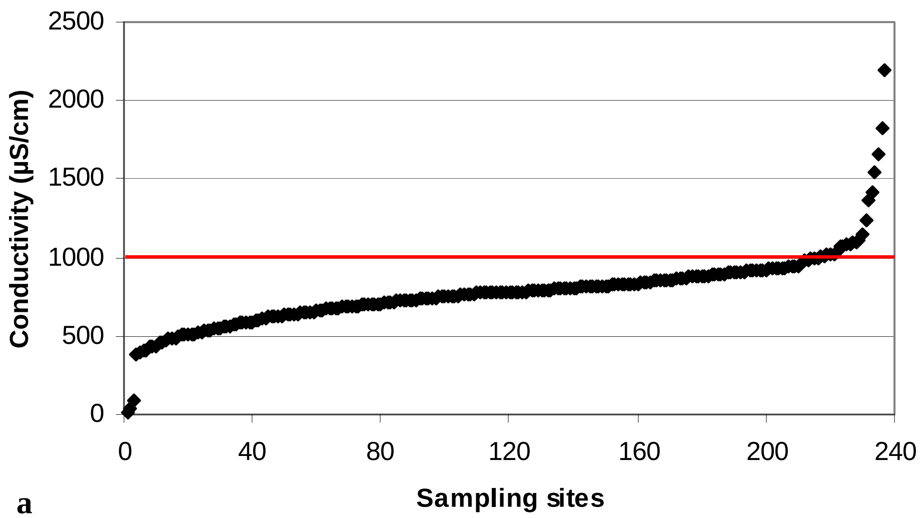


pH:

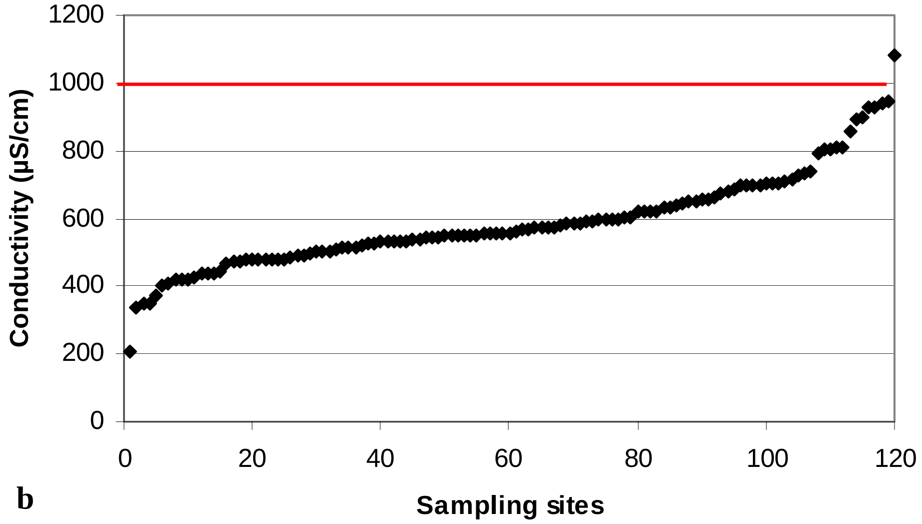

9.50
9.00
8.50
8.00
7.50
7.00
6.50
6.00
pH
0
40
80
120
160
200
240
Sampling sites
a


8.5
8.0
7.5
7.0
6.5
pH
0
20
40
60
80
100
120
Sampling sites
b

Conductivity:

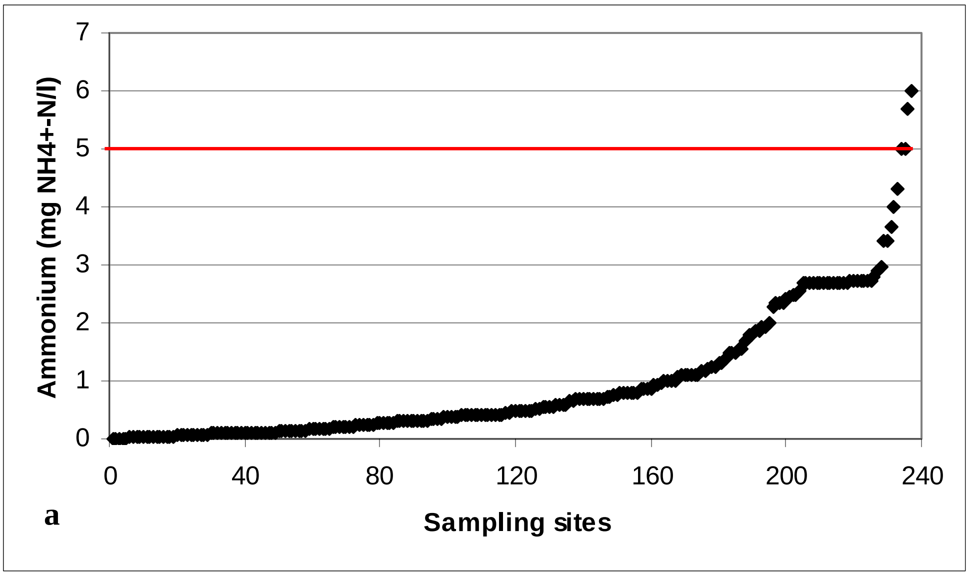


a

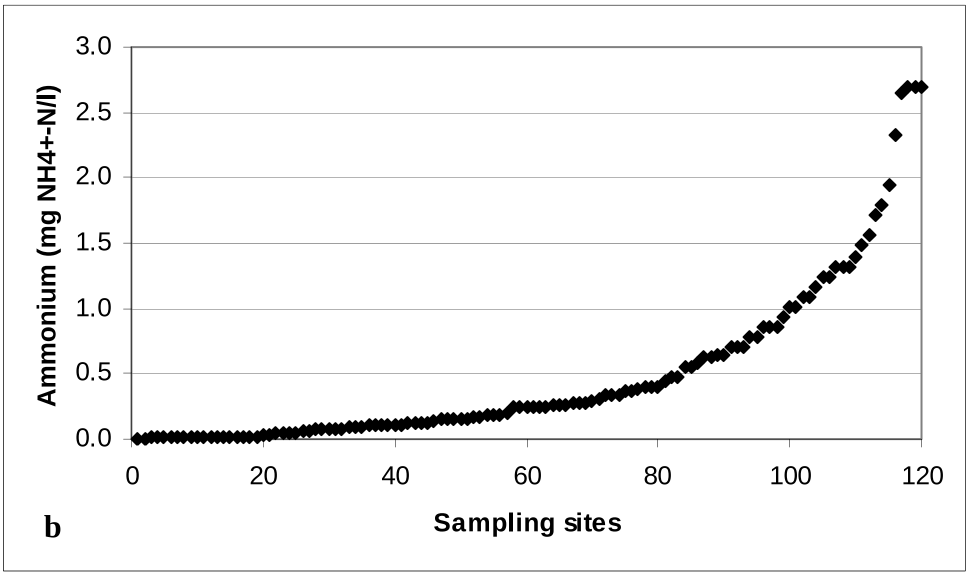


b

Ammonium:

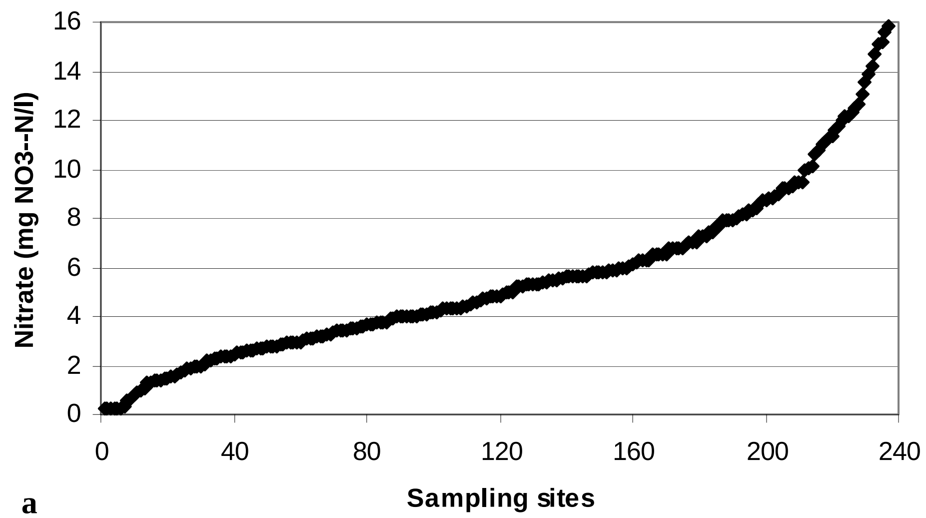


Ammonium (mg NH4+-N/l)
7
6
5
4
3
2
1
0
0
40
80
120
160
200
240
a
Sampling sites

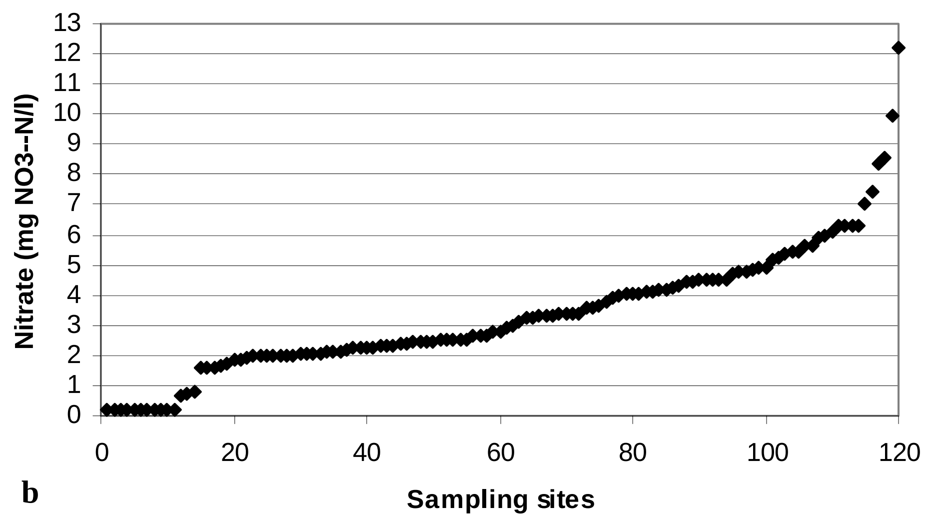


Ammonium (mg NH4+-N/l)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
20
40
60
80
100
120
b
Sampling sites

Nitrate:

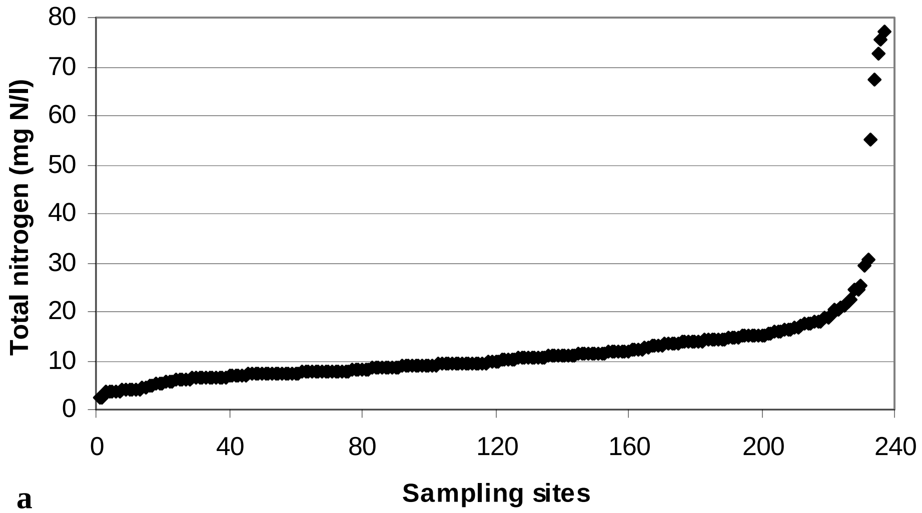

Nitrate (mg NO3--N/l)
Sampling sites
0
2
4
6
8
10
12
14
16
0
40
80
120
160
200
240

a

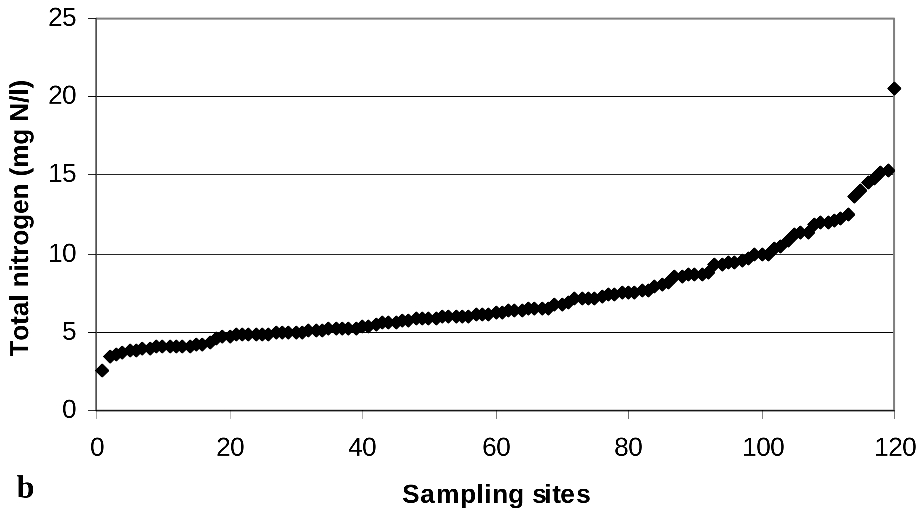

Nitrate (mg NO3--N/l)
Sampling sites
0
1
2
3
4
5
6
7
8
9
10
11
12
13
0
20
40
60
80
100
120

b

Total nitrogen:

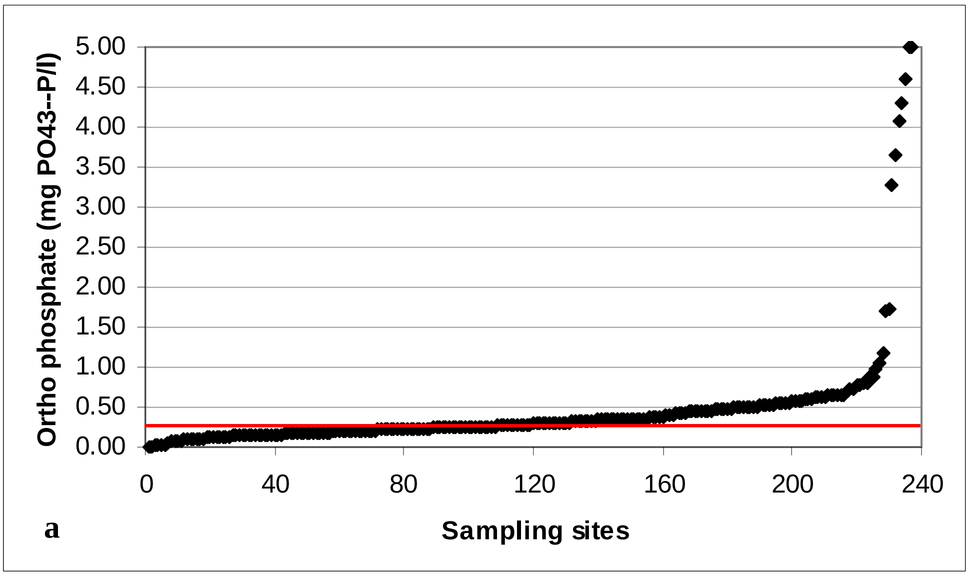


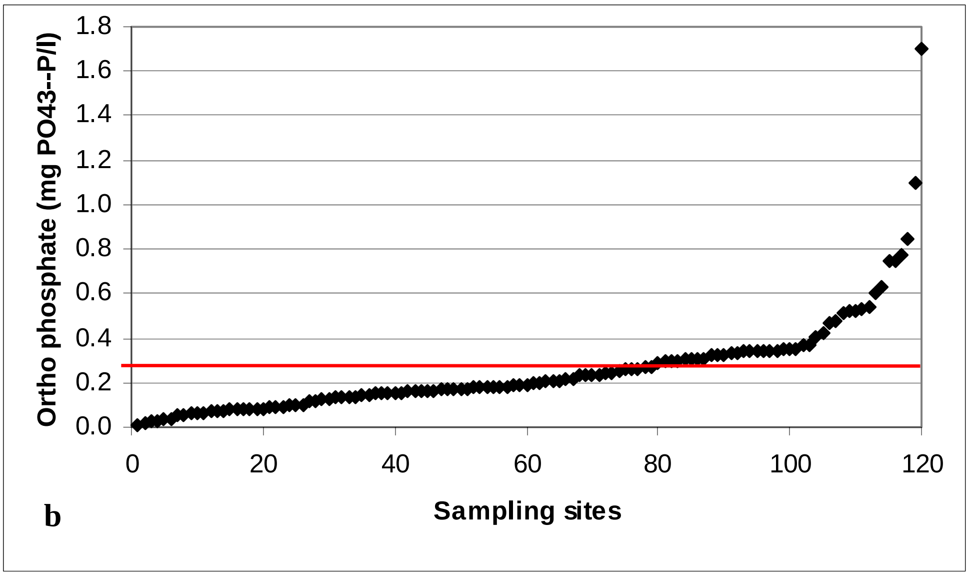
a



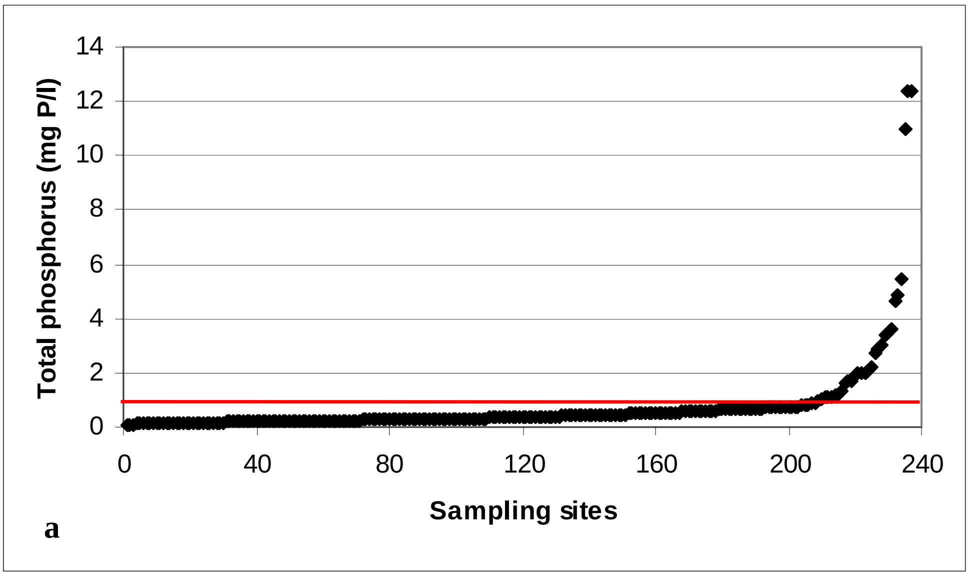
b

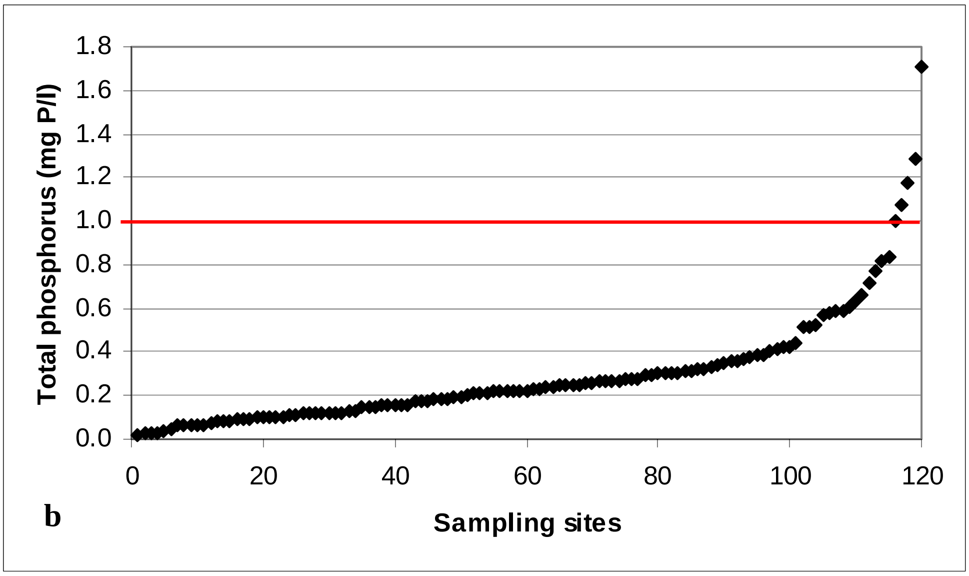
Ortho phosphate:



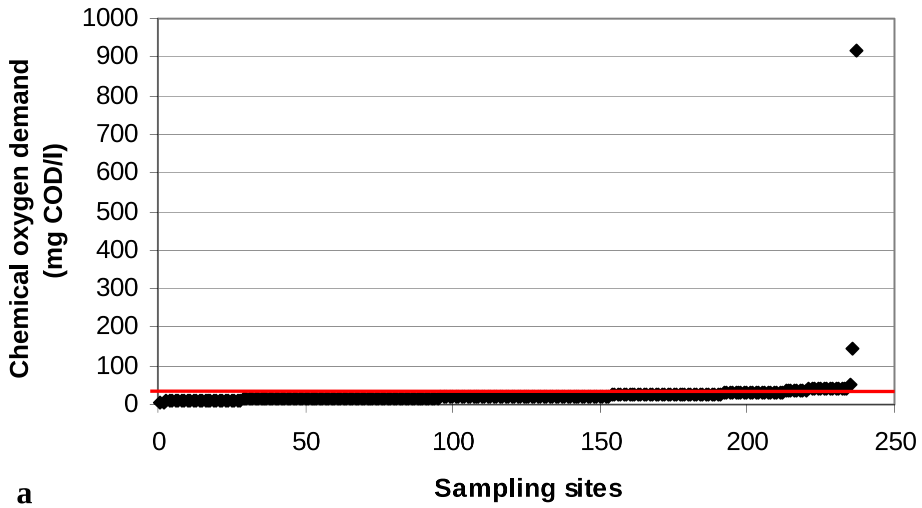


Total phosphorus:

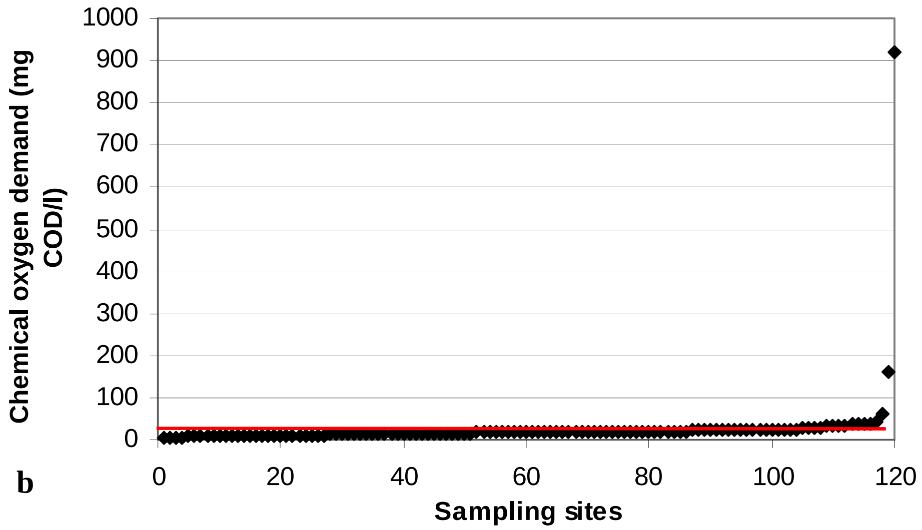




Chemical oxygen demand:

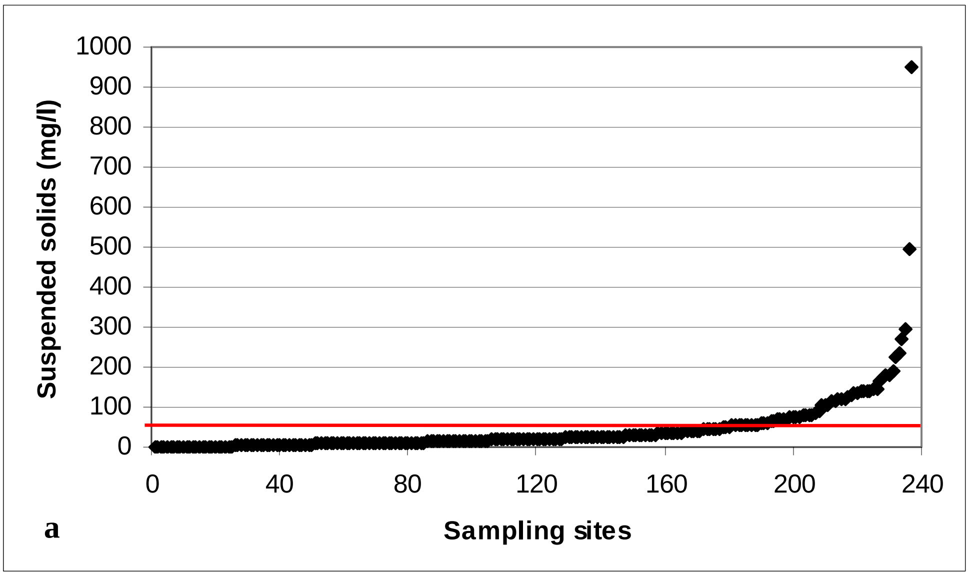


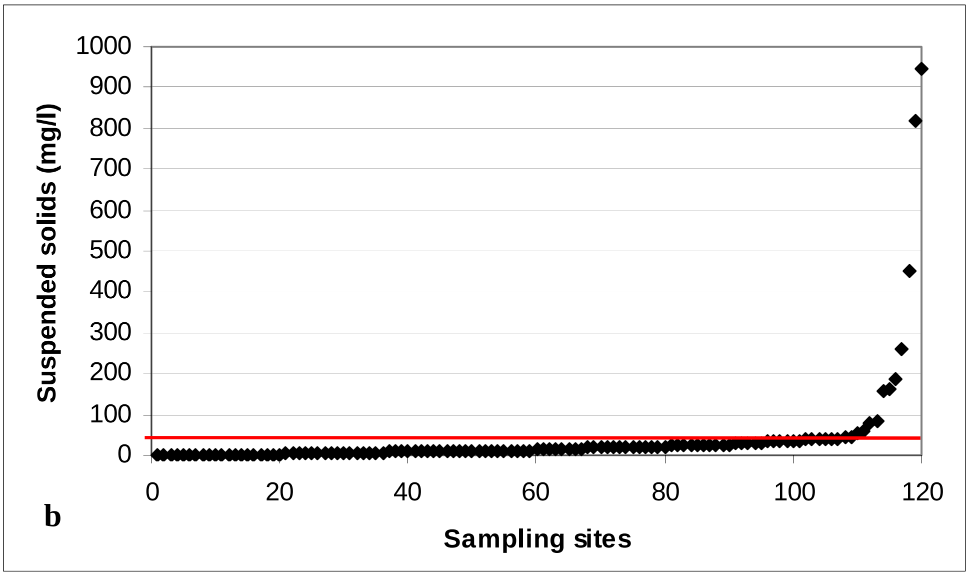
a



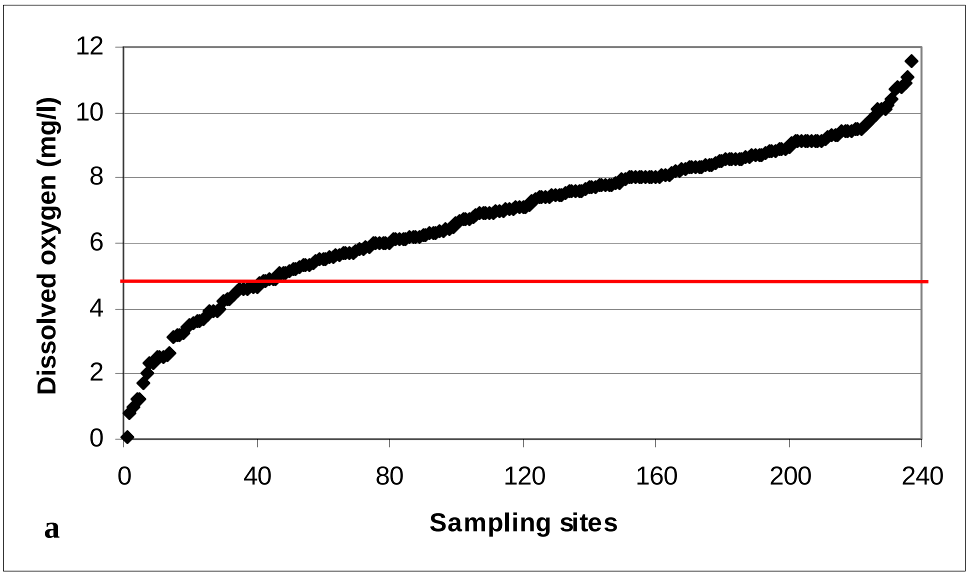
b

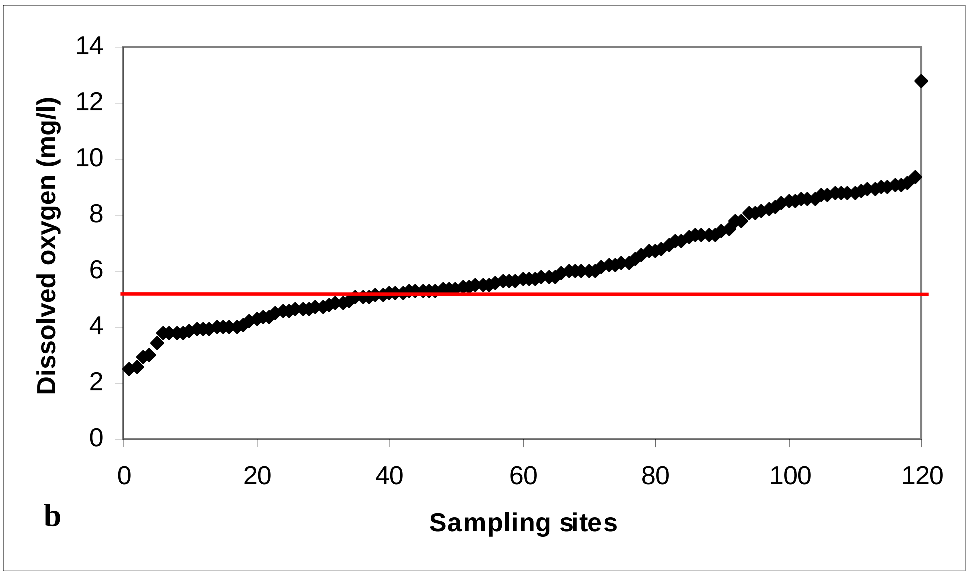
Suspended solids:



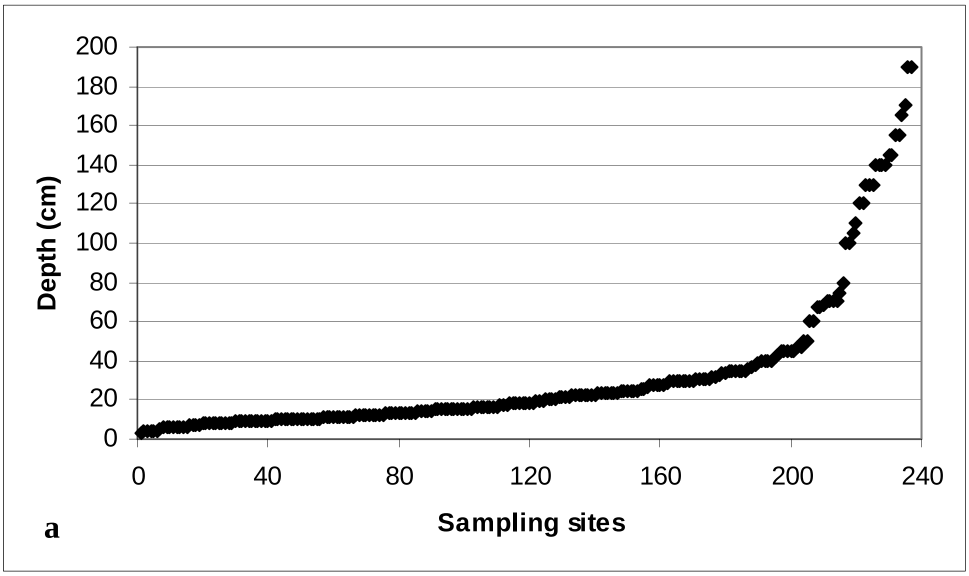


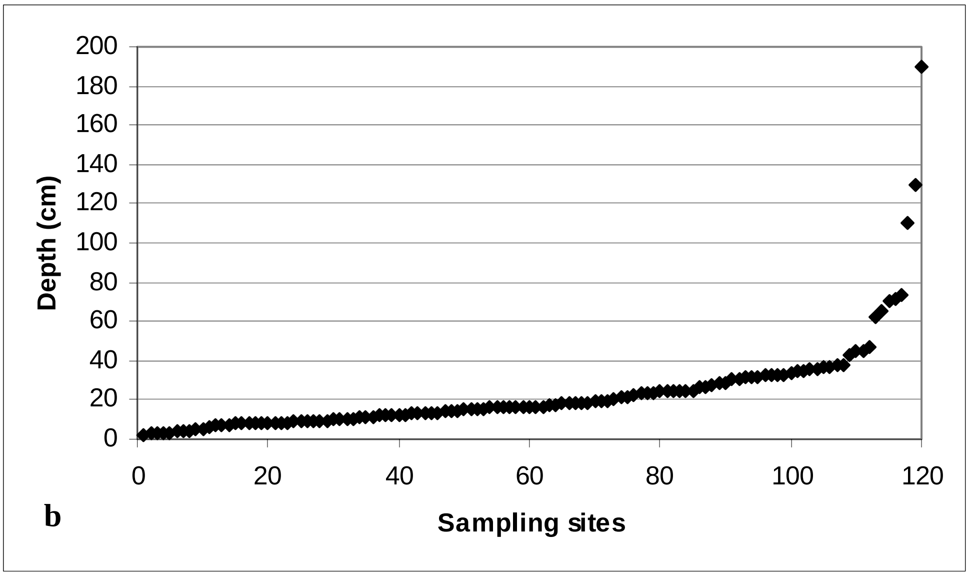
Dissolved oxygen:



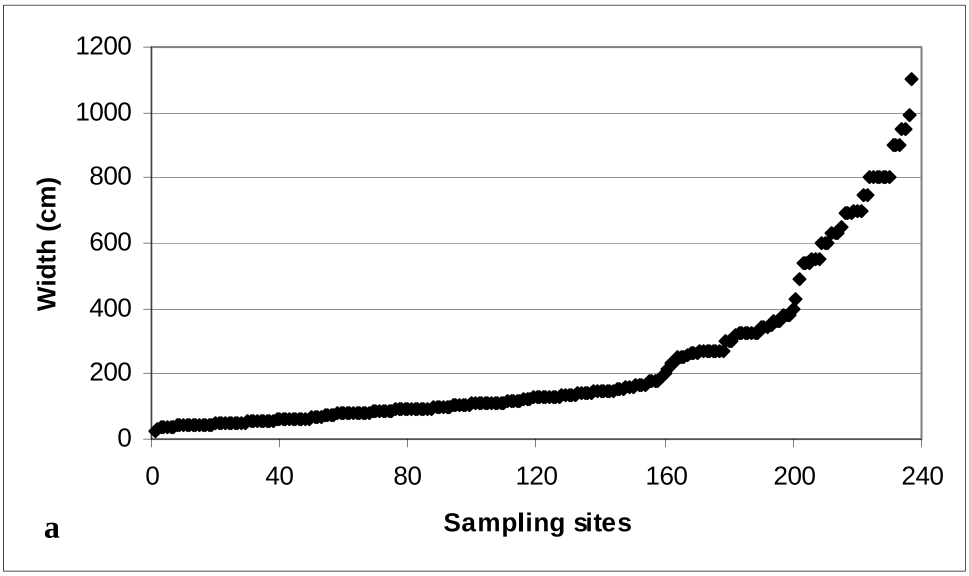


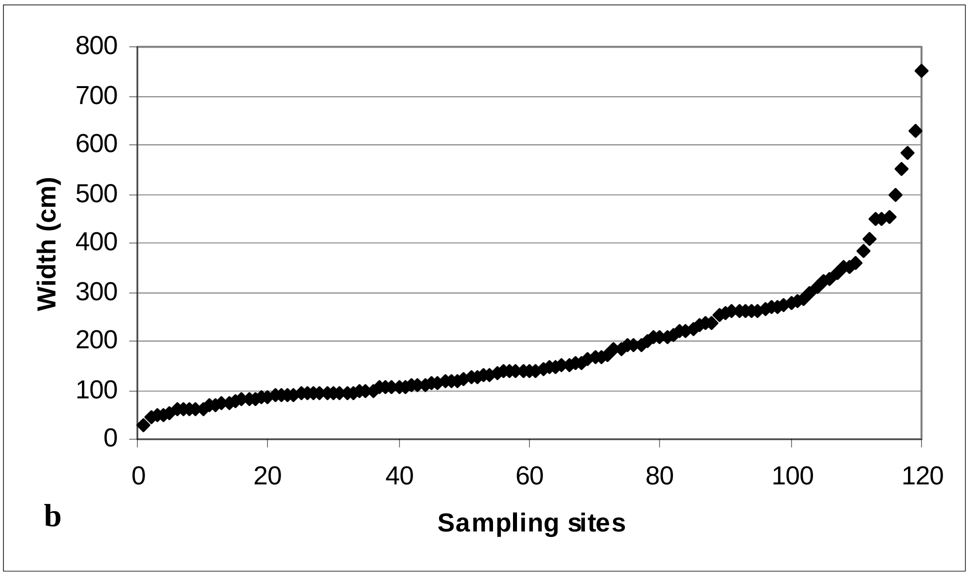
Depth:


200
180
160
140
120
100
80
60
40
20
0
Depth (cm)
0
40
80
120
160
200
240
Sampling sites
a

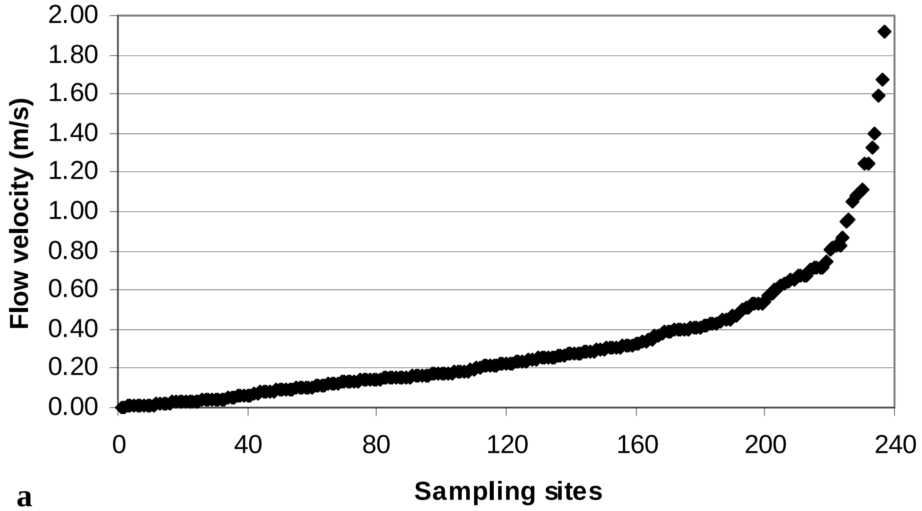

200
180
160
140
120
100
80
60
40
20
0
Depth (cm)
0
20
40
60
80
100
120
Sampling sites
b

Width:

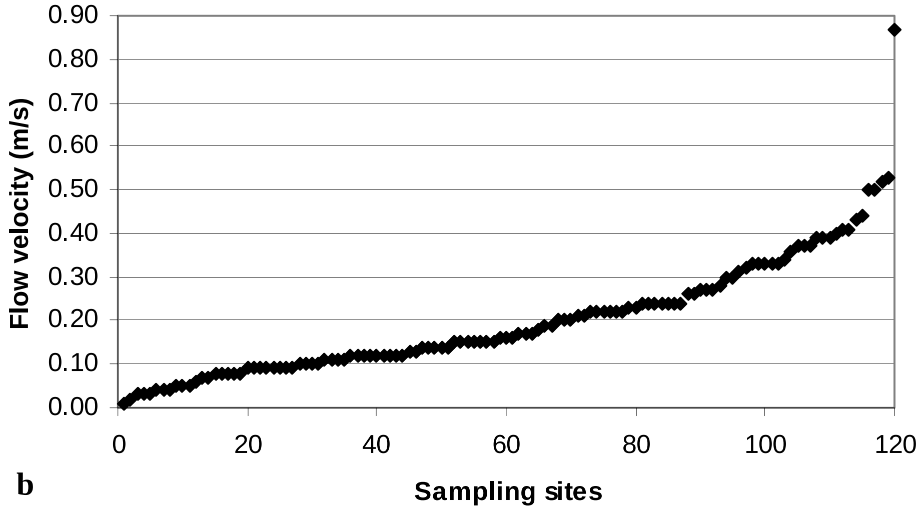

1200
1000
800
600
400
200
0
Width (cm)
0
40
80
120
160
200
240
Sampling sites
a


800
700
600
500
400
300
200
100
0
Width (cm)
0
20
40
60
80
100
120
Sampling sites
b

Flow velocity:

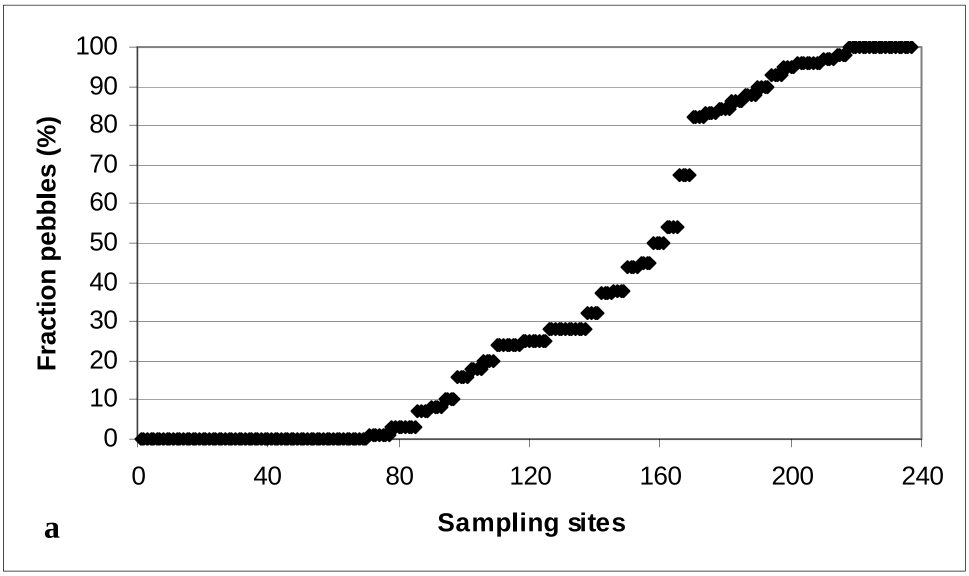


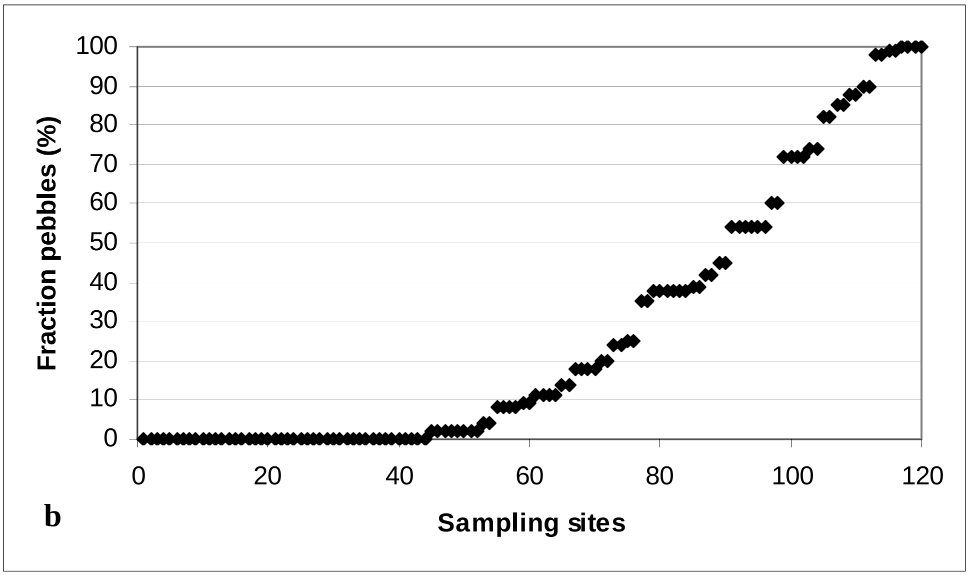
a



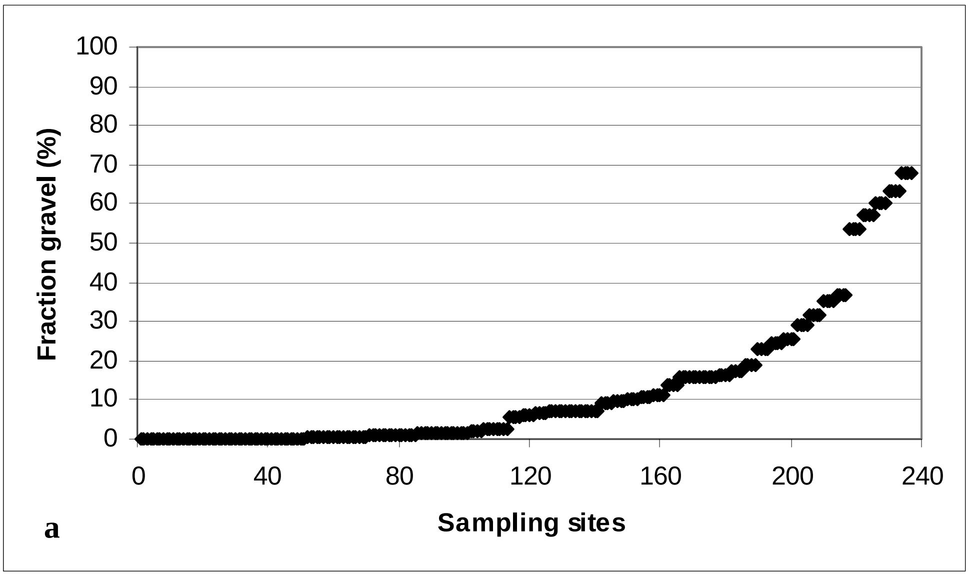
b

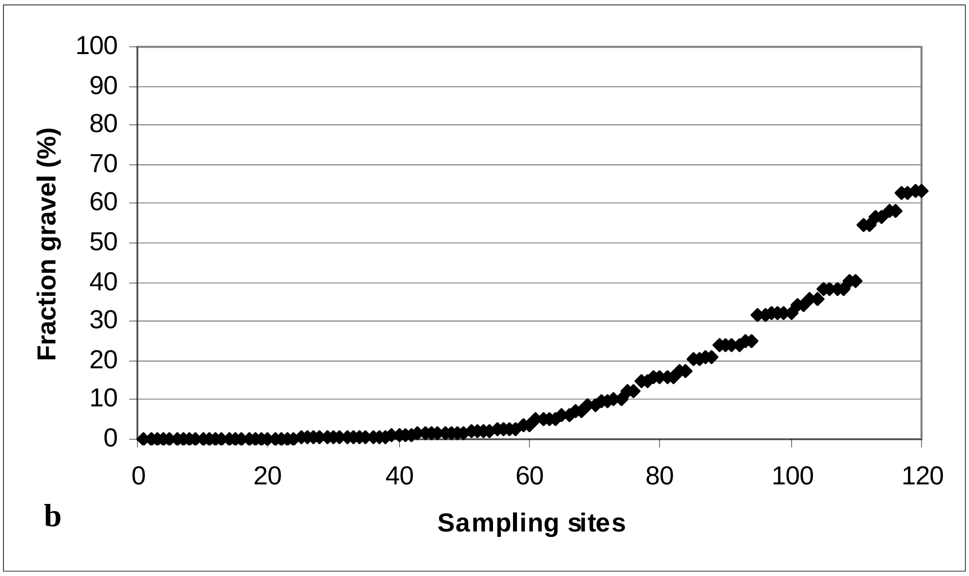
Fraction pebbles:



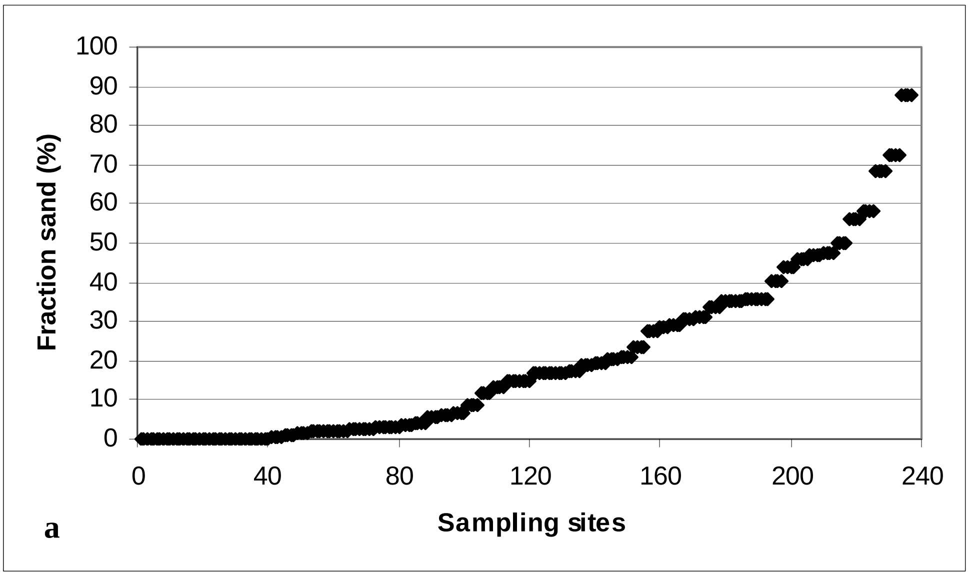


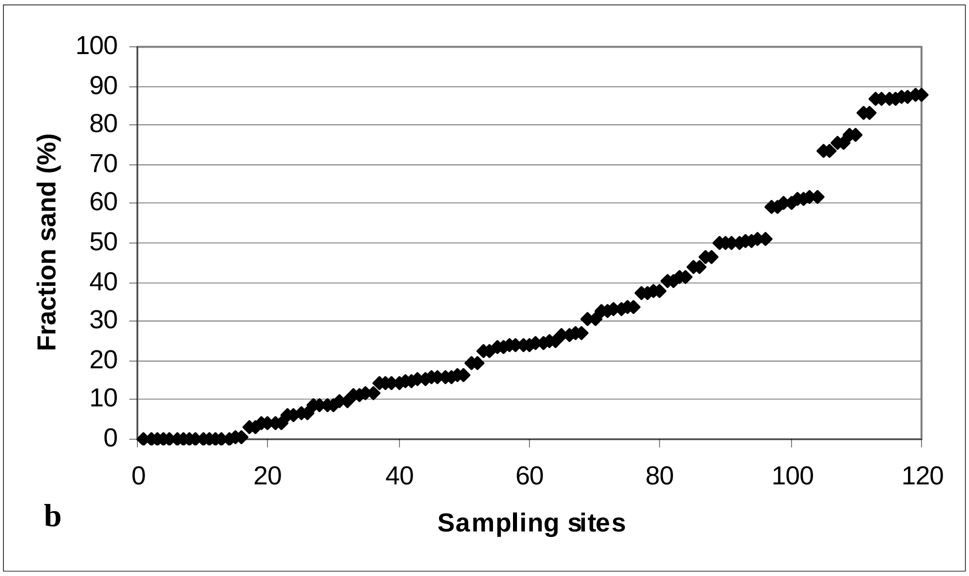
Fraction gravel:



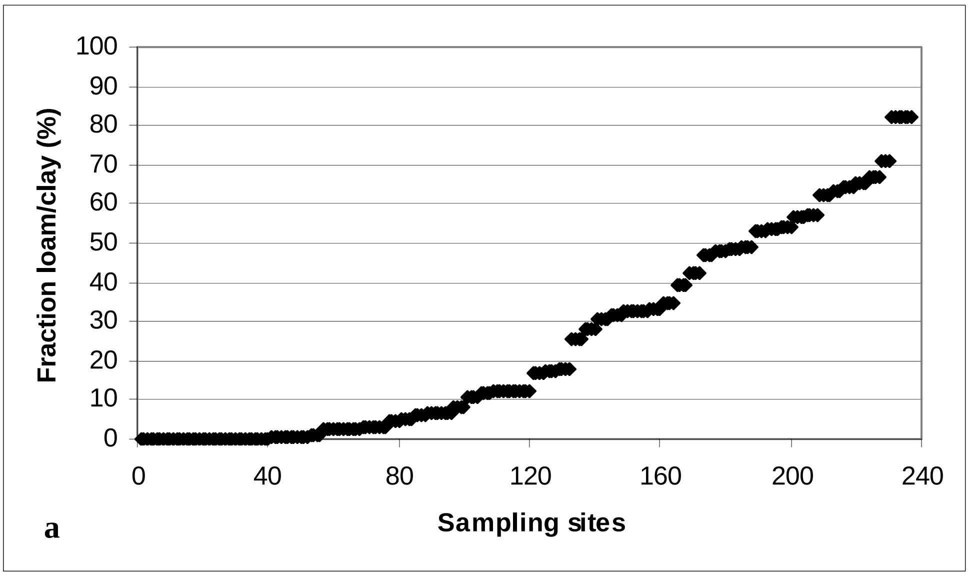


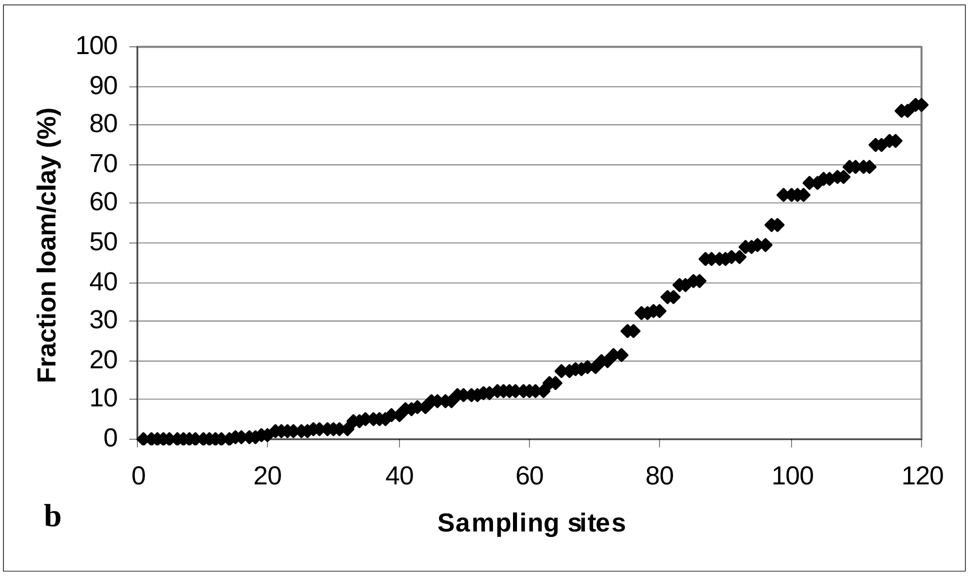
Fraction sand:



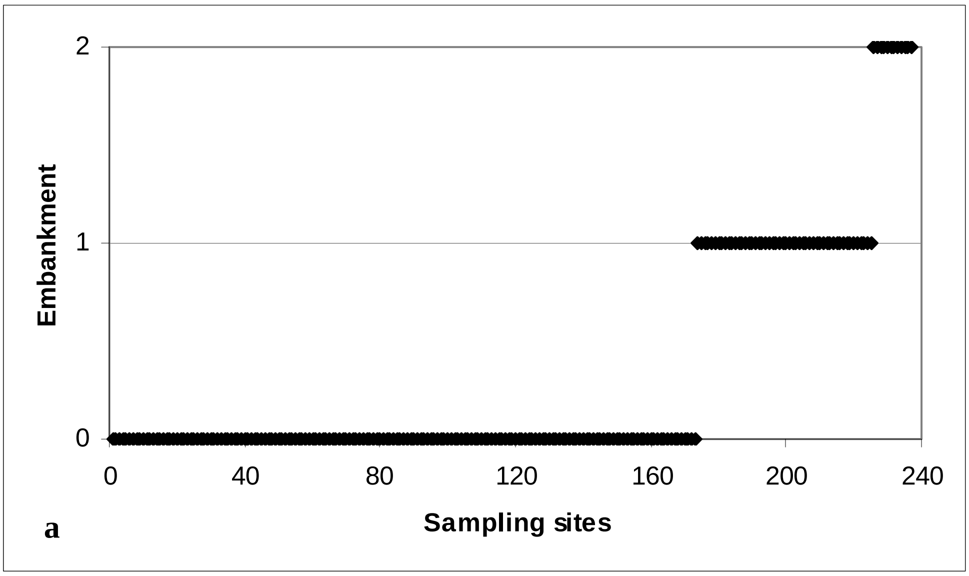


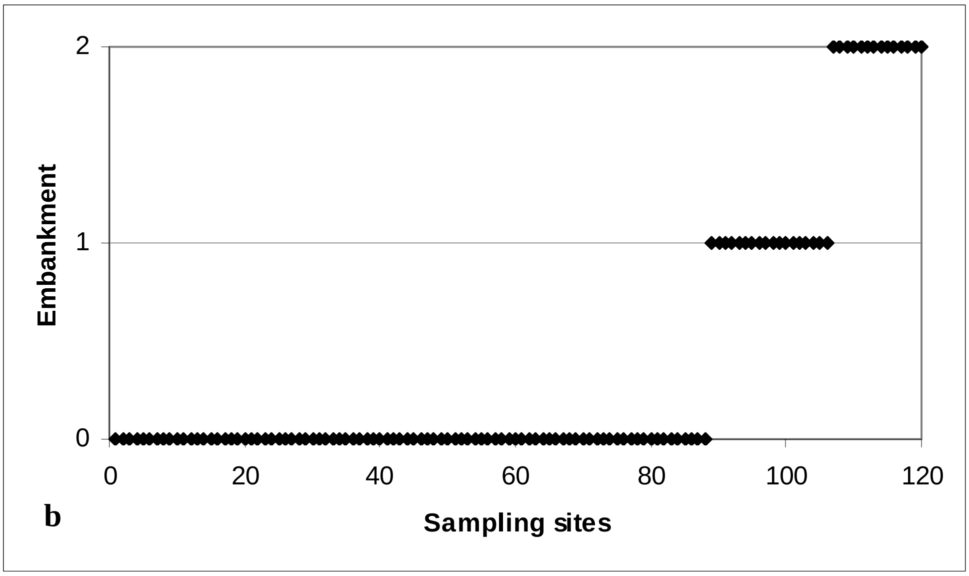
Fraction loam/clay:



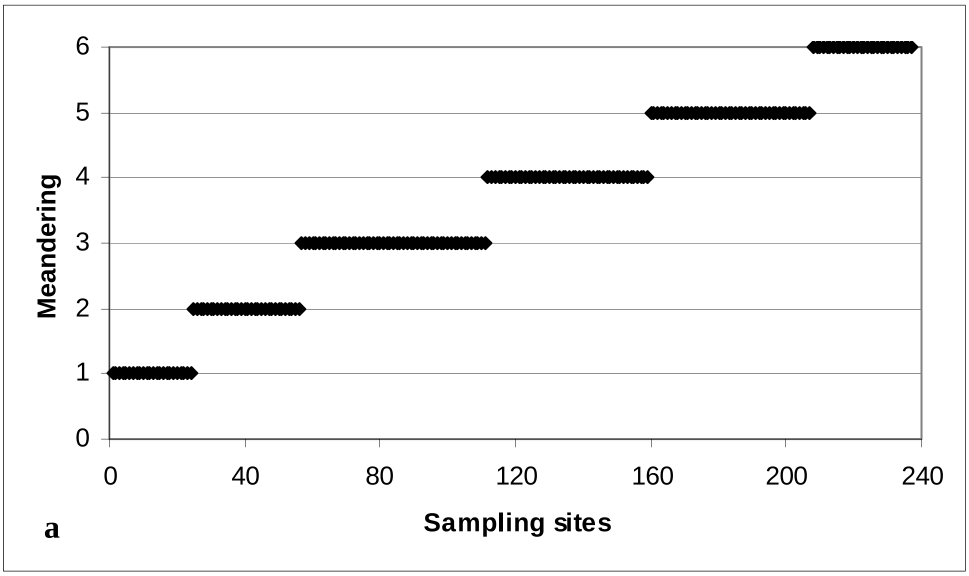


Embankment:

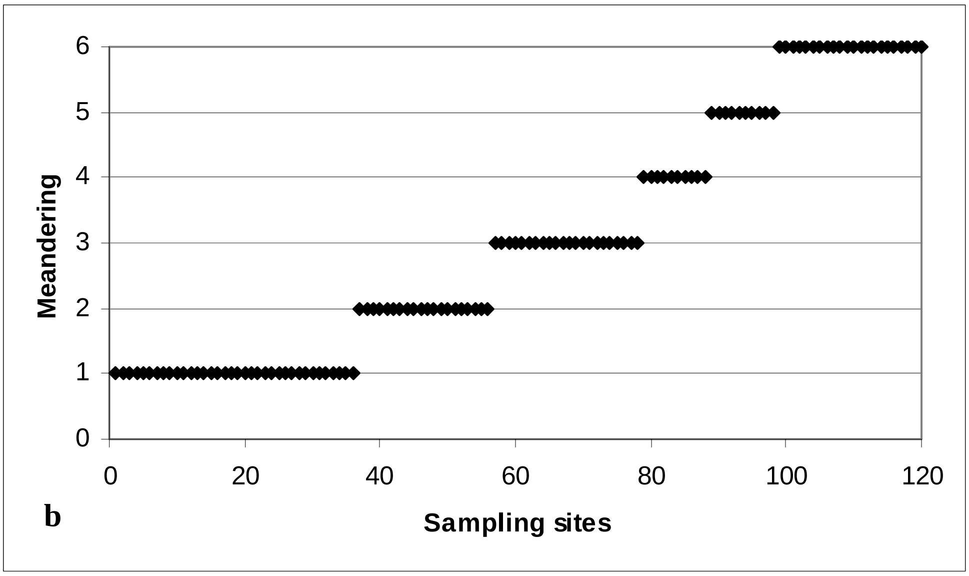

2
1
0
Embankment
0
40
80
120
160
200
240
Sampling sites
a


2
1
0
Embankment
0
20
40
60
80
100
120
Sampling sites
b

Meandering:

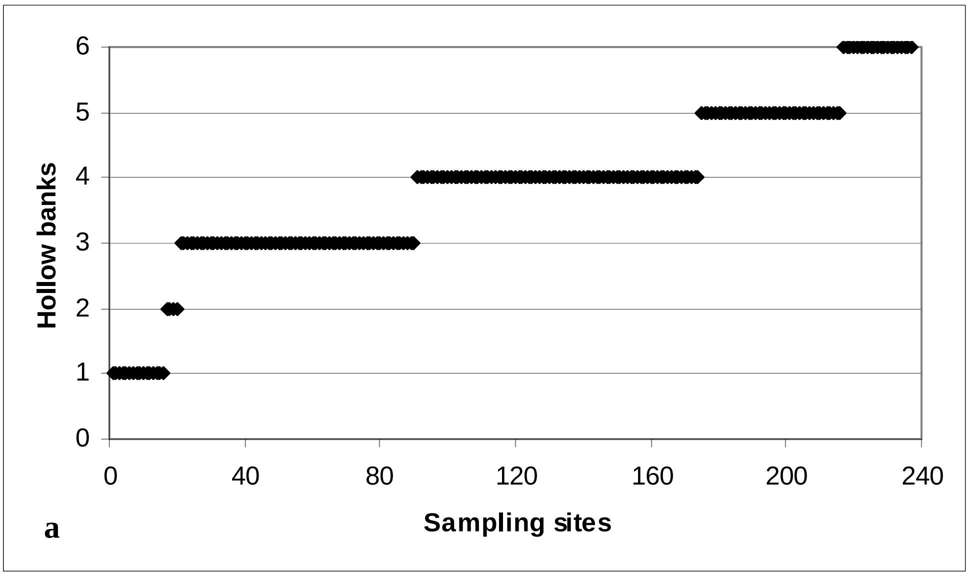


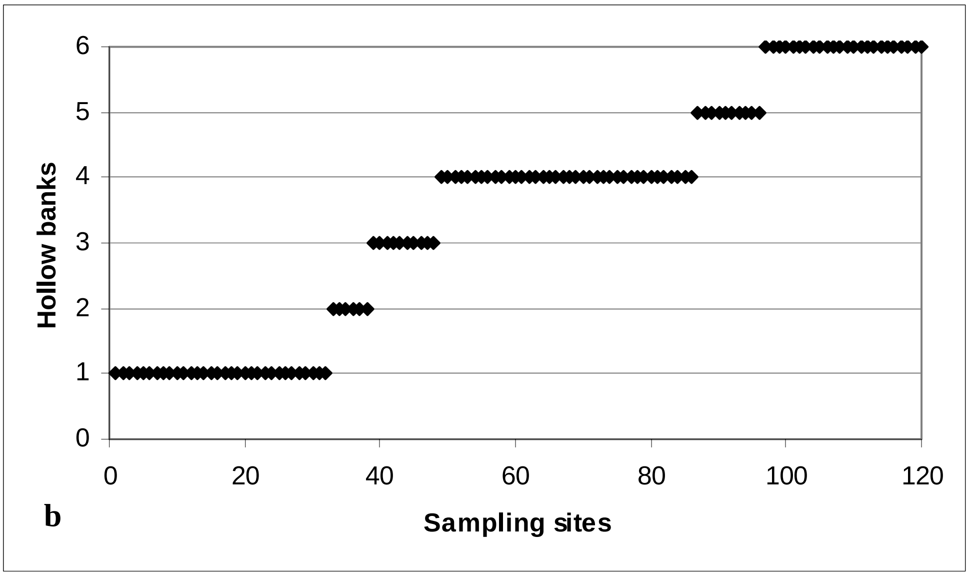
a



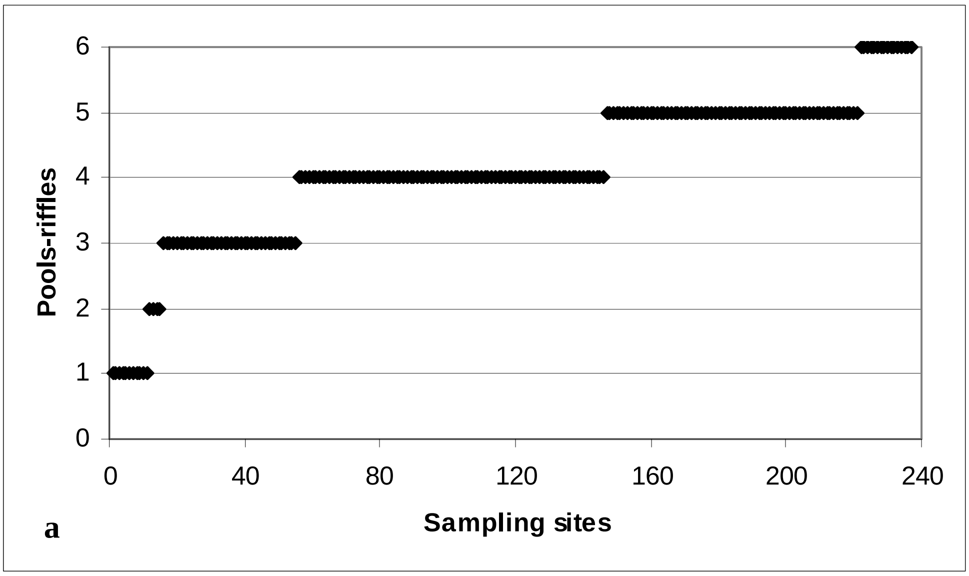
b

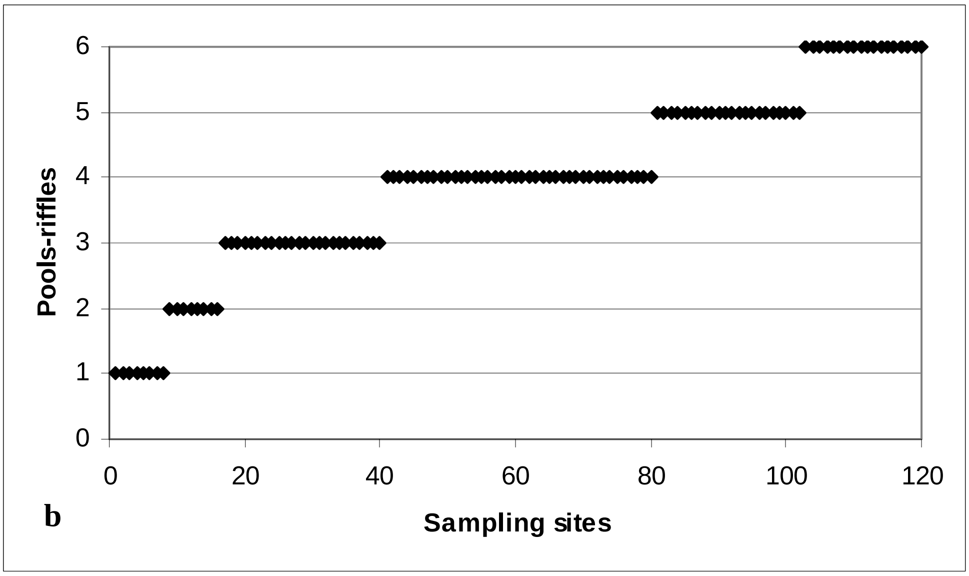
Hollow banks:



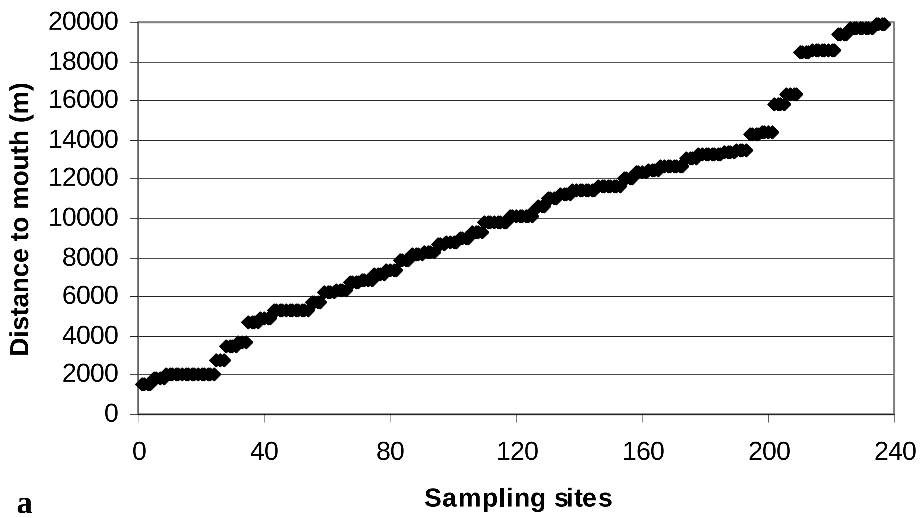


Pools-riffles:

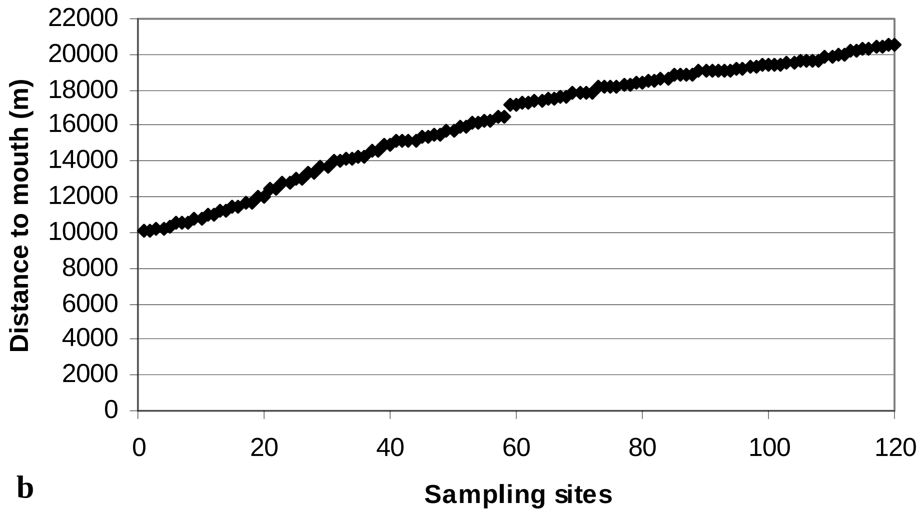




Distance to mouth:

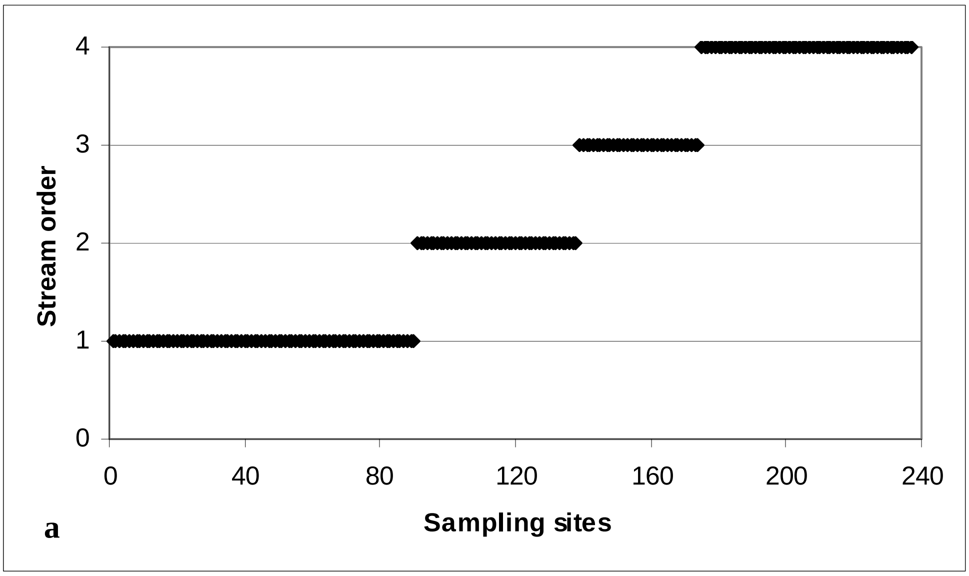


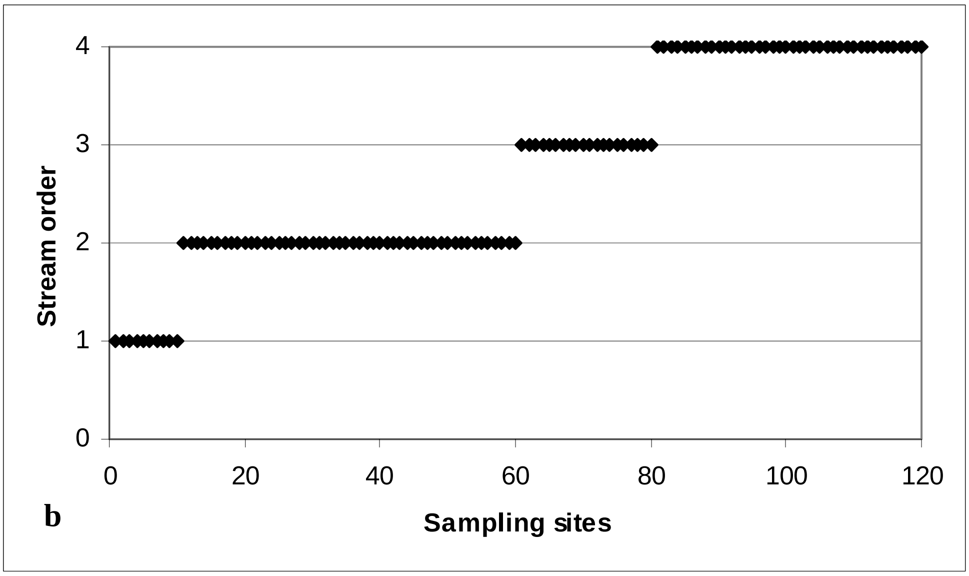
a



b

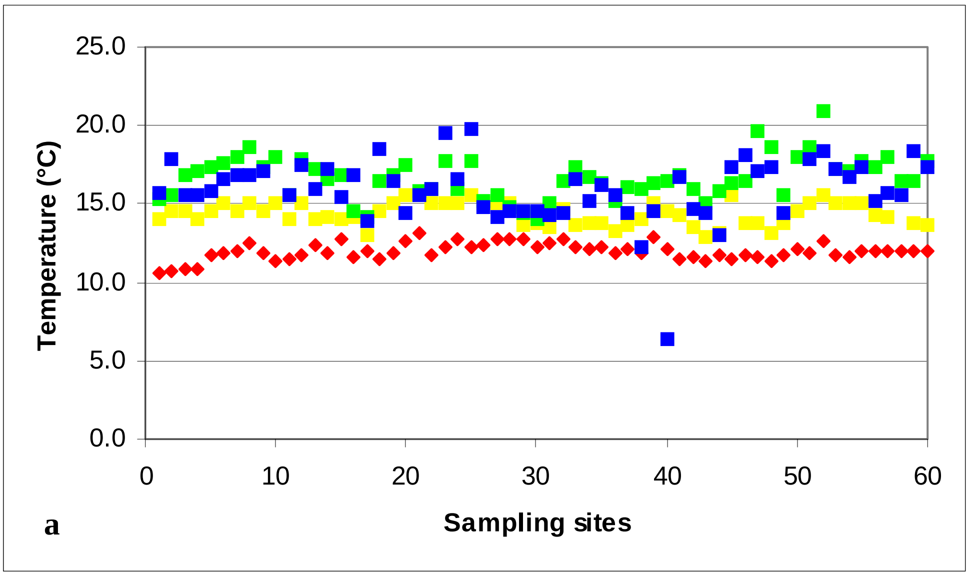
Stream order:

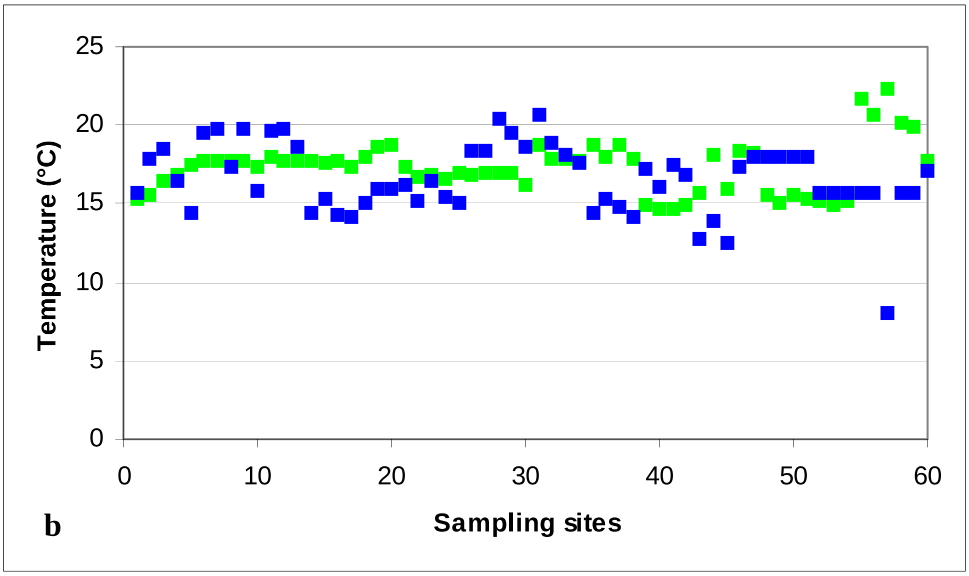




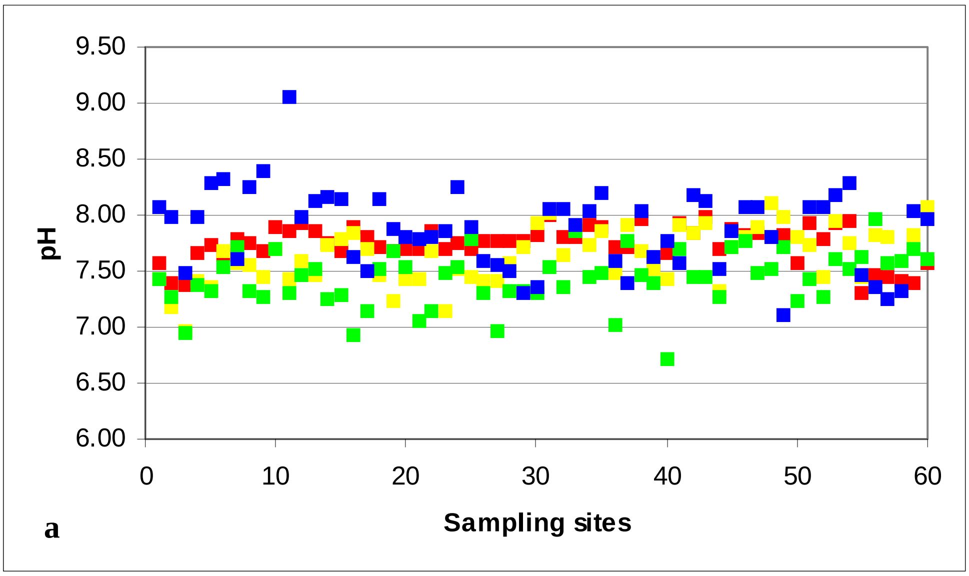
Appendix 3: Spatial variation of the appropriate environmental input variables over the respectively 4 years (a, Zwalm river basin, 237 sampling sites) and 2 years (b, 'short distance' monitoring network, 120 sampling sites) (red = 2000; yellow = 2001; green = 2002; blue = 2003).

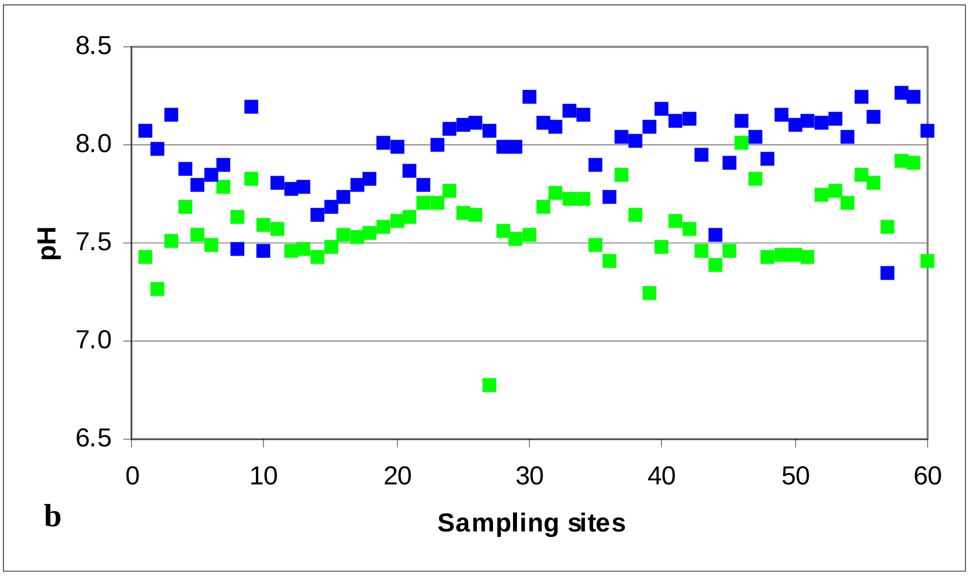
Temperature:



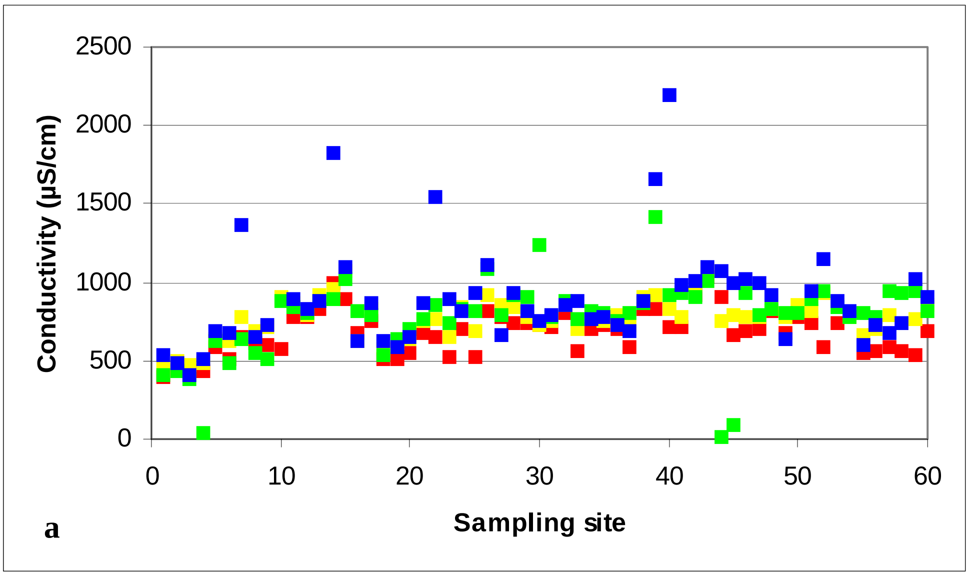


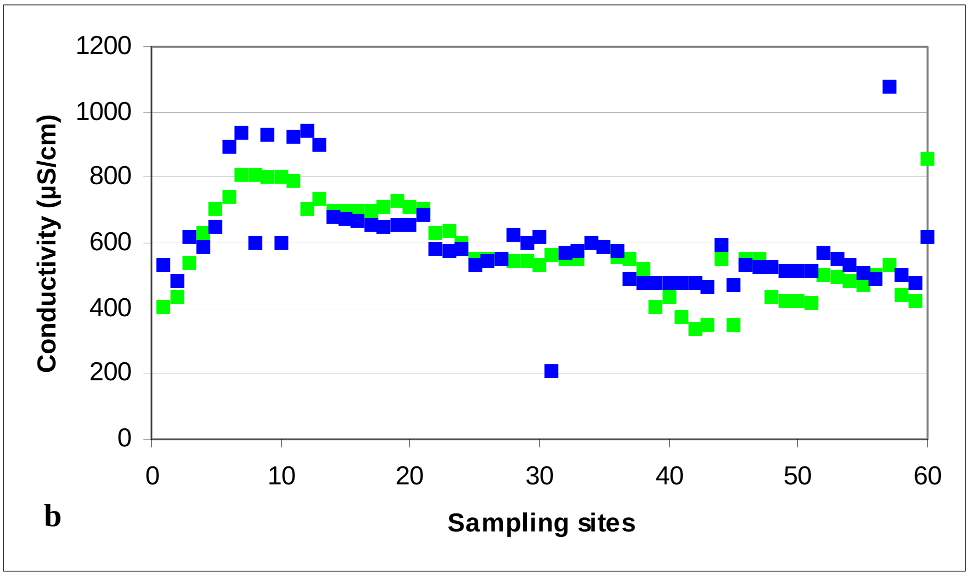
pH:


9.50
9.00
8.50
8.00
7.50
7.00
6.50
6.00
pH
0
10
20
30
40
50
60
Sampling sites
a

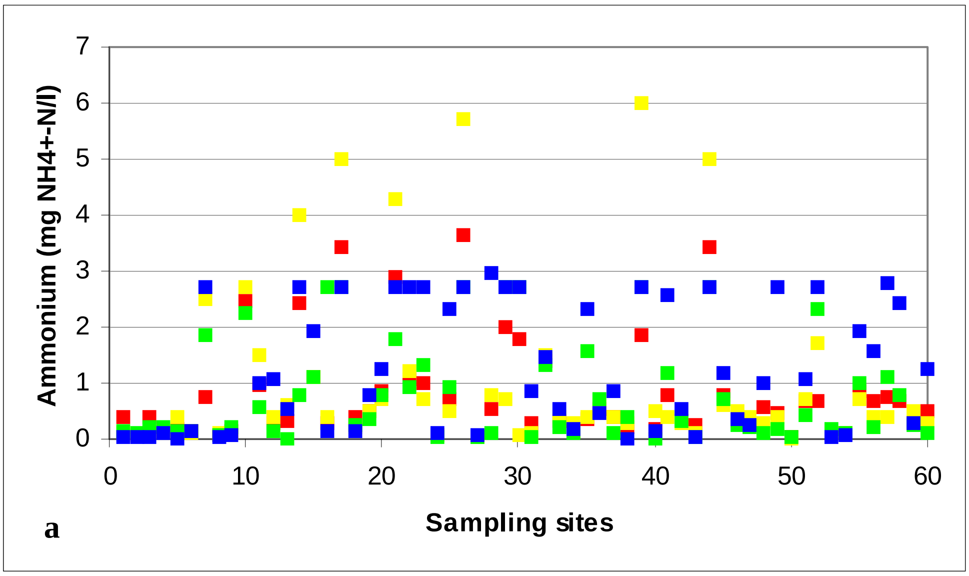

8.5
8.0
7.5
7.0
6.5
pH
0
10
20
30
40
50
60
Sampling sites
b

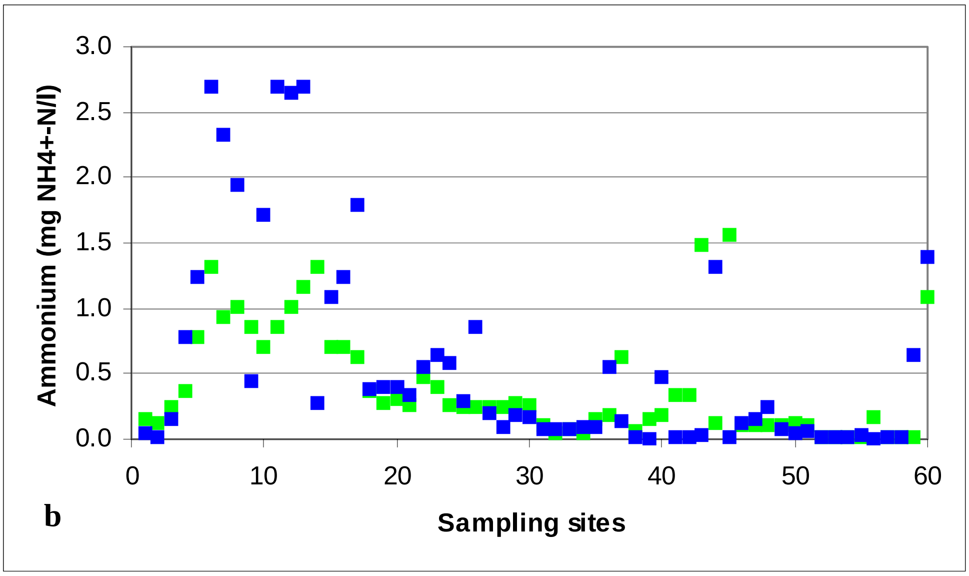
Conductivity:



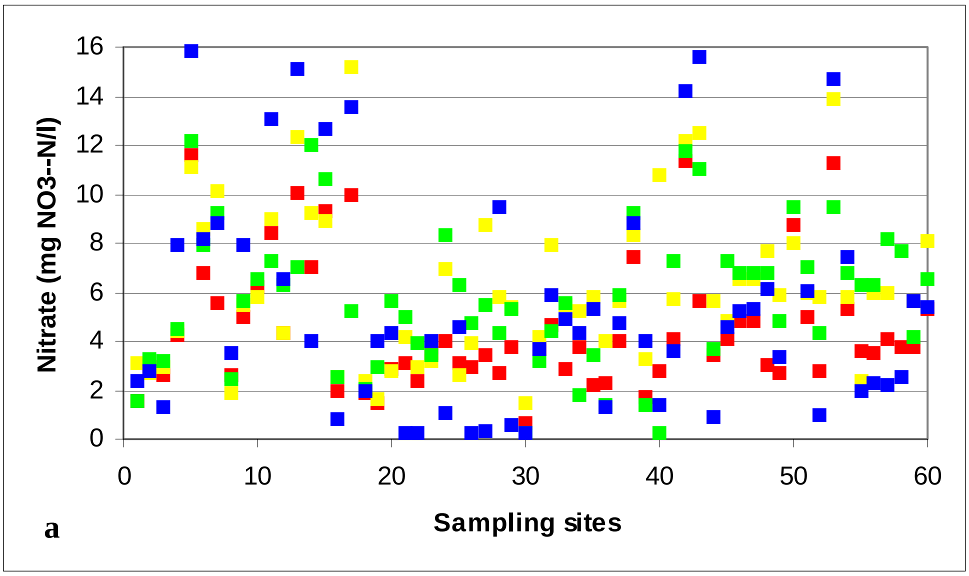


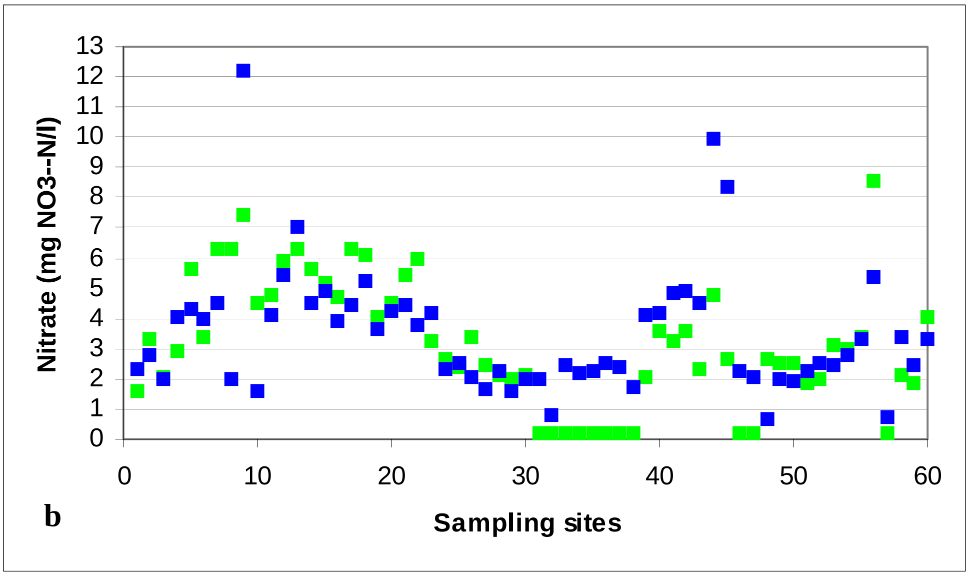
Ammonium:


Ammonium (mg NH4+-N/l)
Sampling sites
a

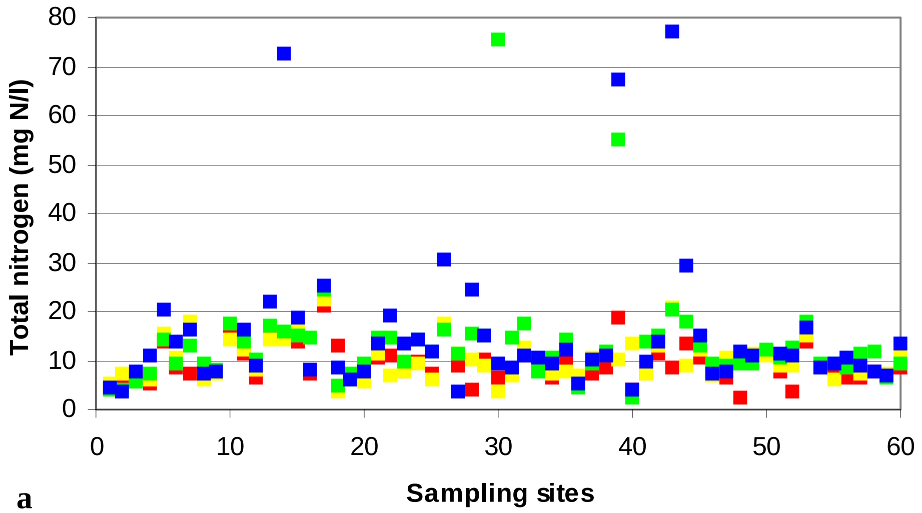

Ammonium (mg NH4+-N/l)
Sampling sites
b

Nitrate:

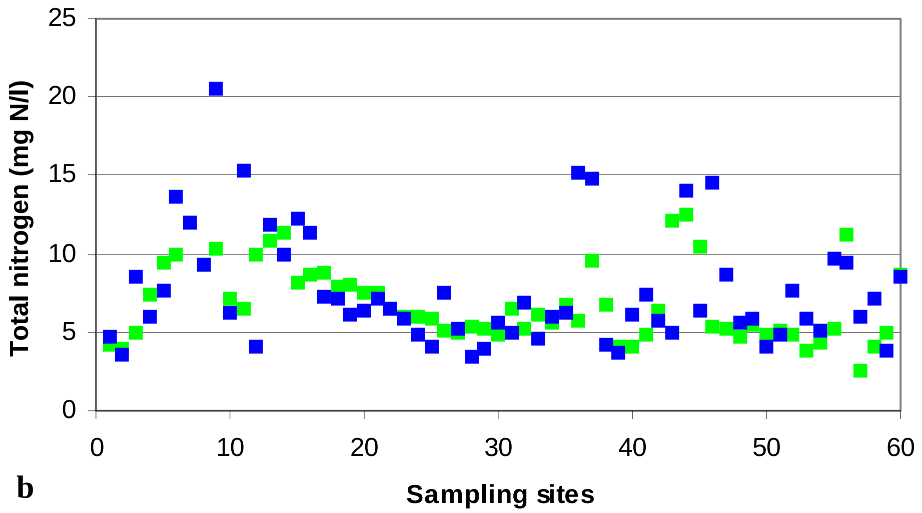

Nitrate (mg NO3--N/l)
Sampling sites
a


Nitrate (mg NO3--N/l)
Sampling sites
b

Total nitrogen:

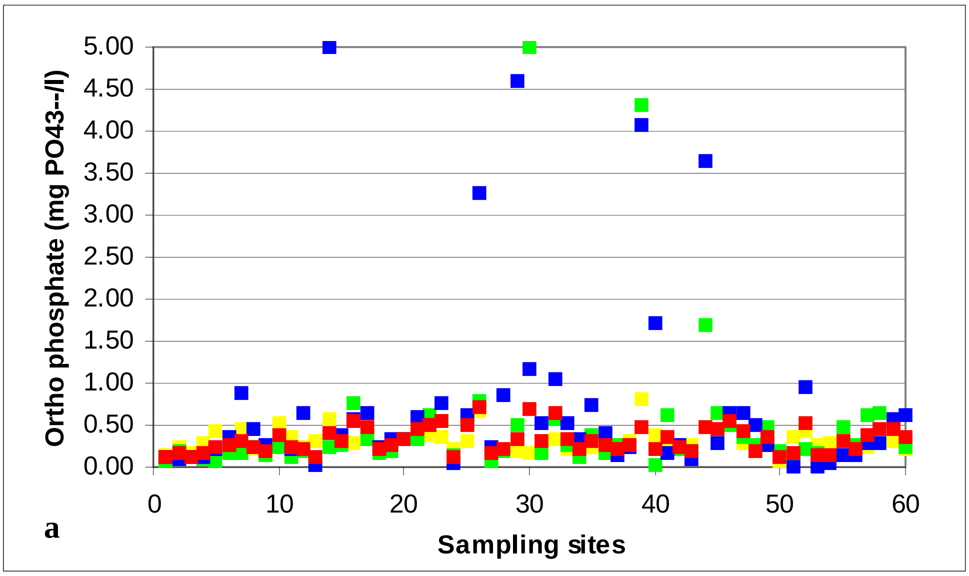


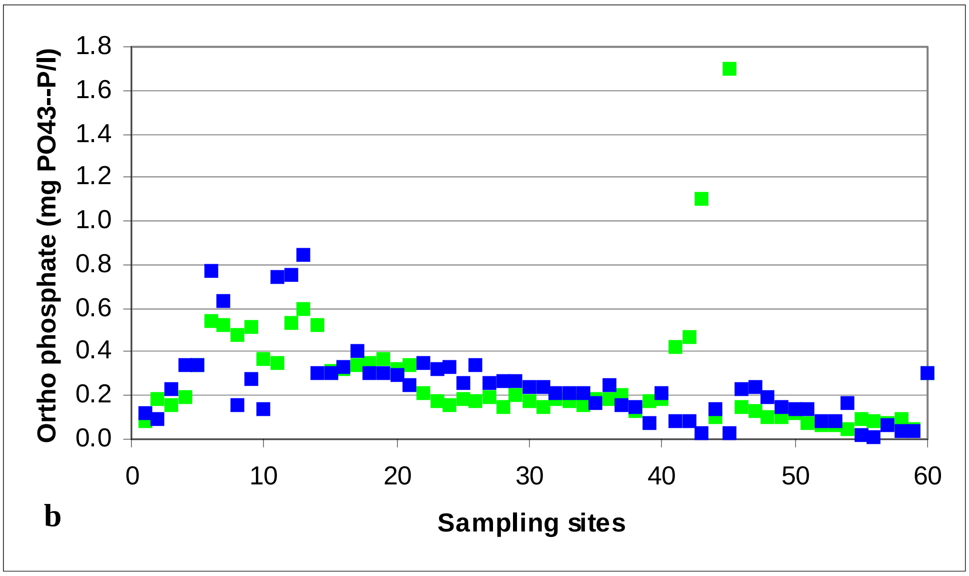
a



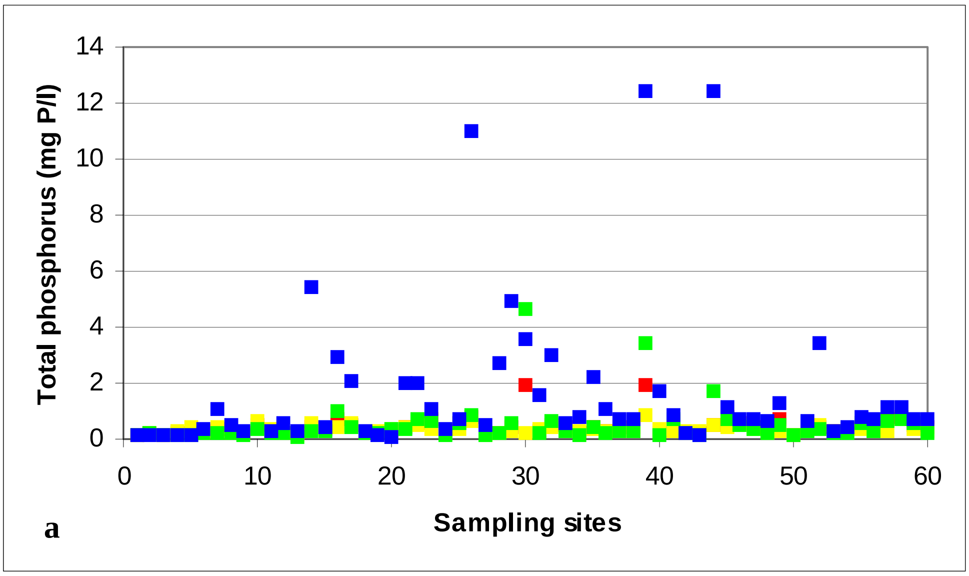
b

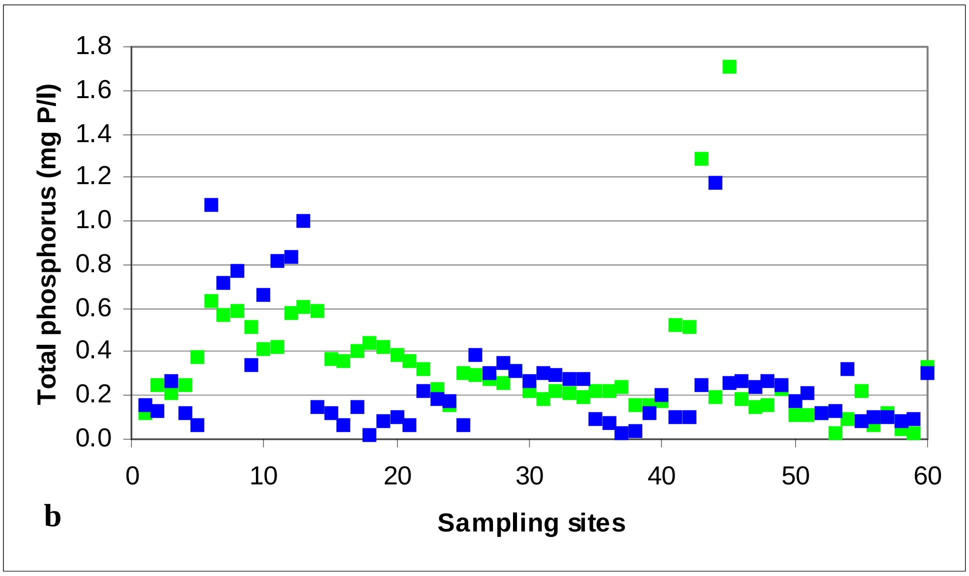
Ortho phosphate:



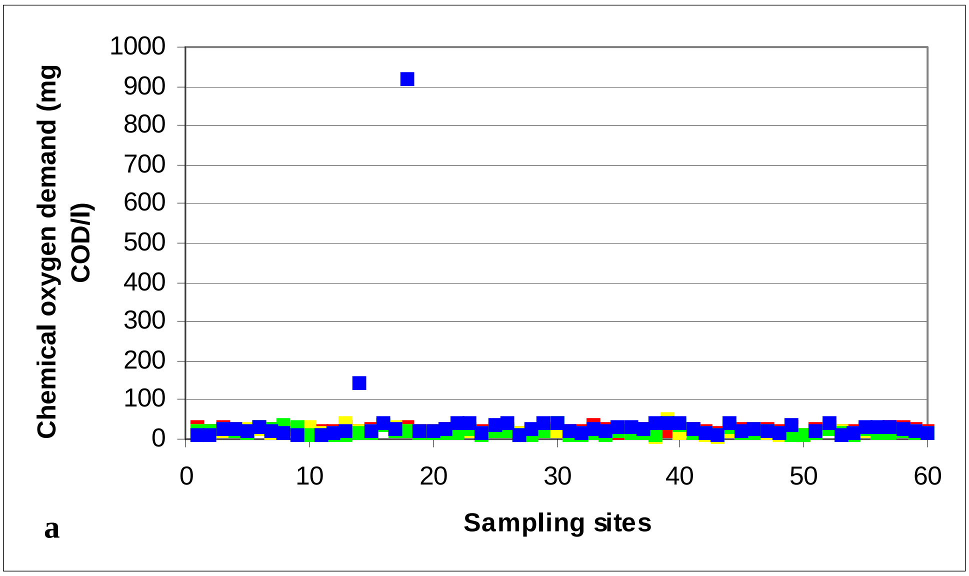


Total phosphorus:

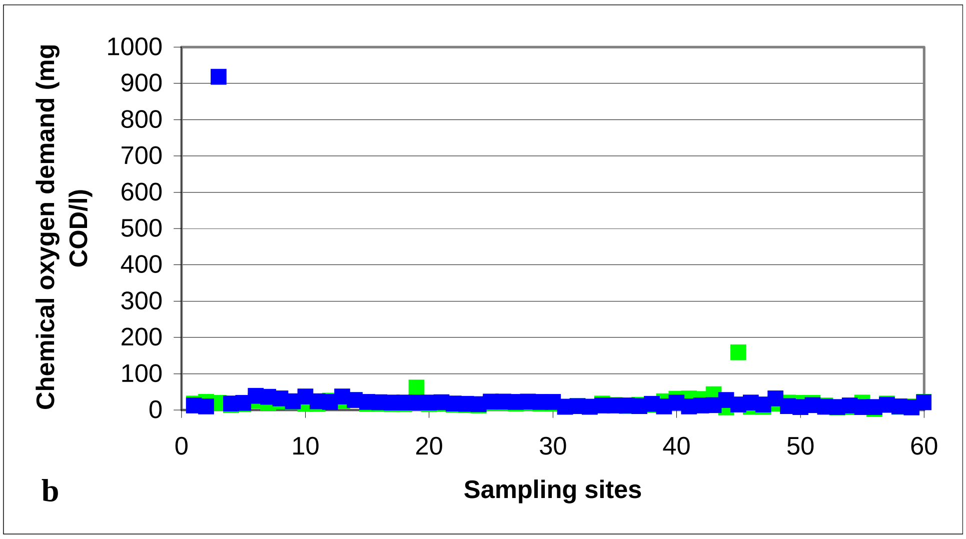




Chemical oxygen demand:

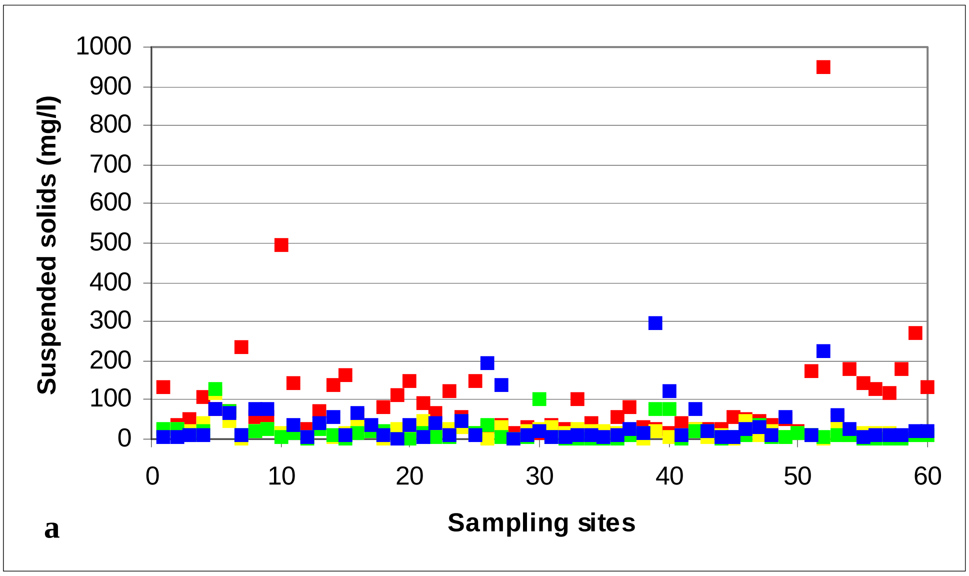


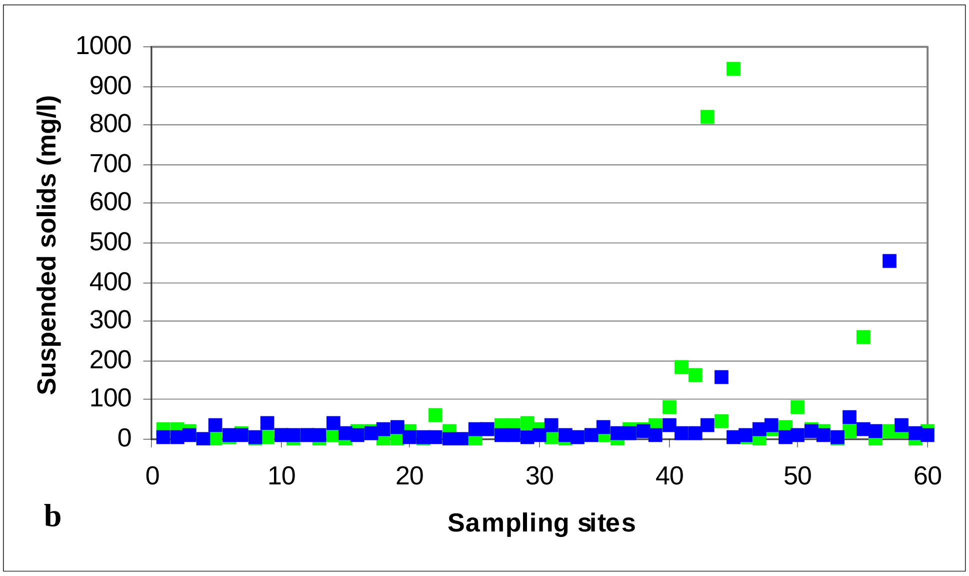
a



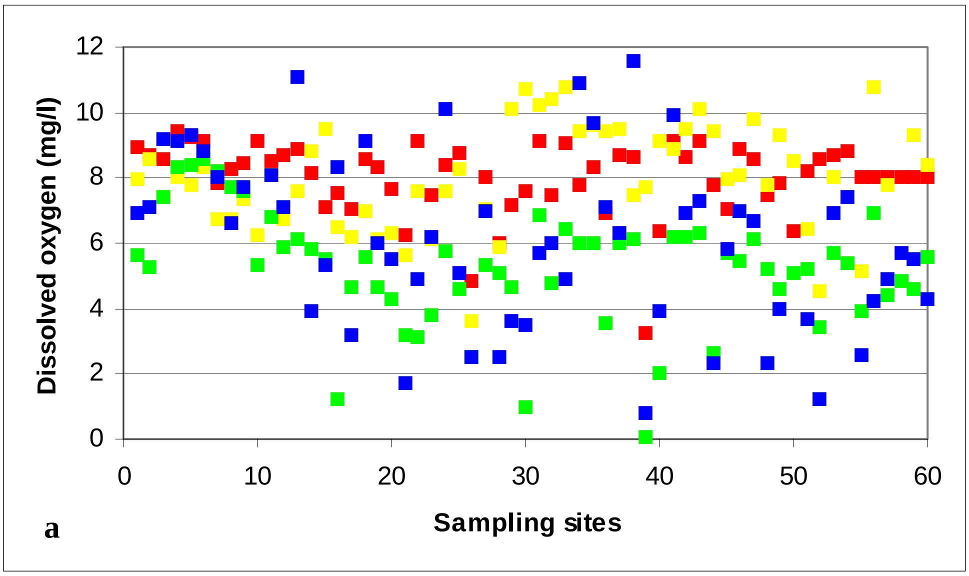
b

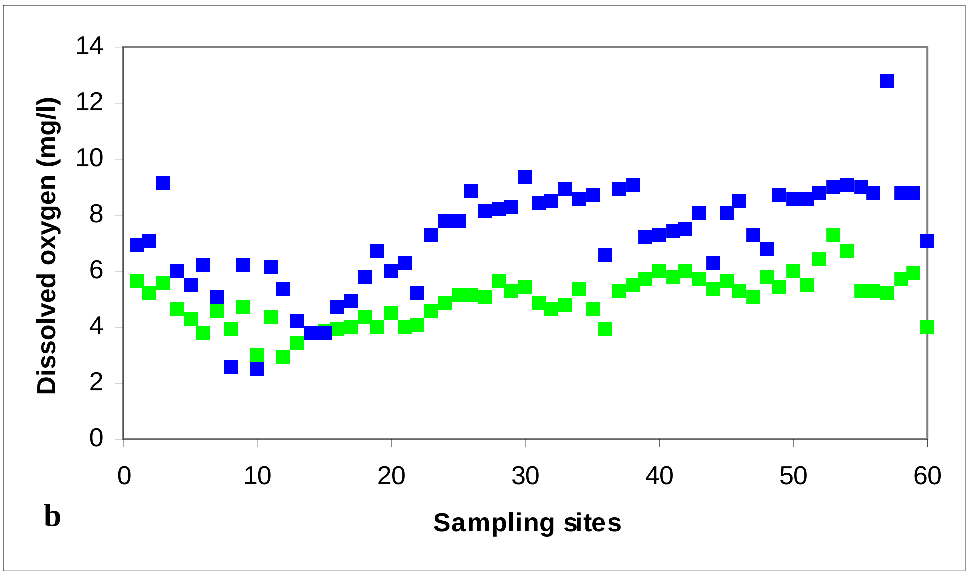
Suspended solids:



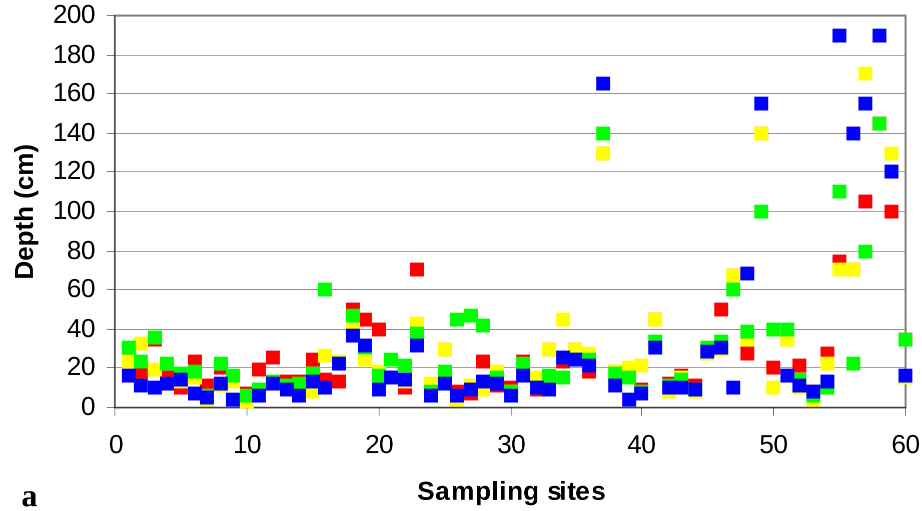


Dissolved oxygen:

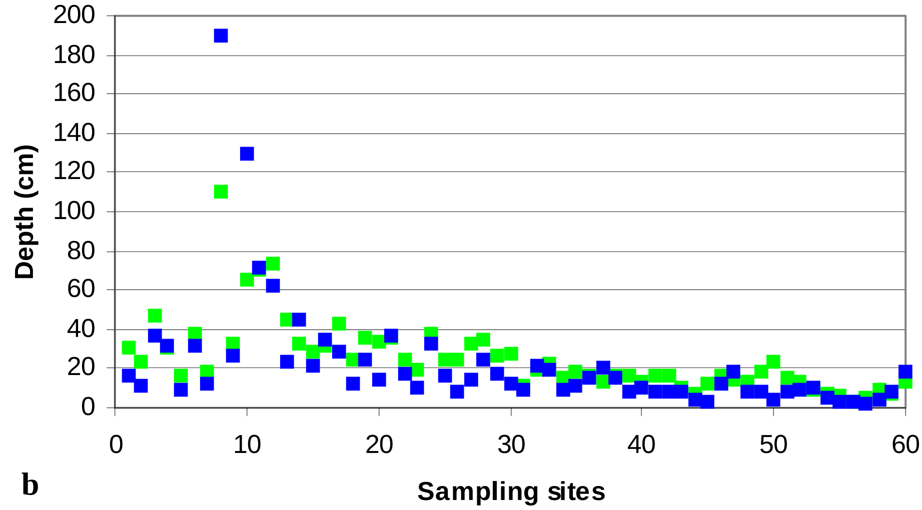




Depth:

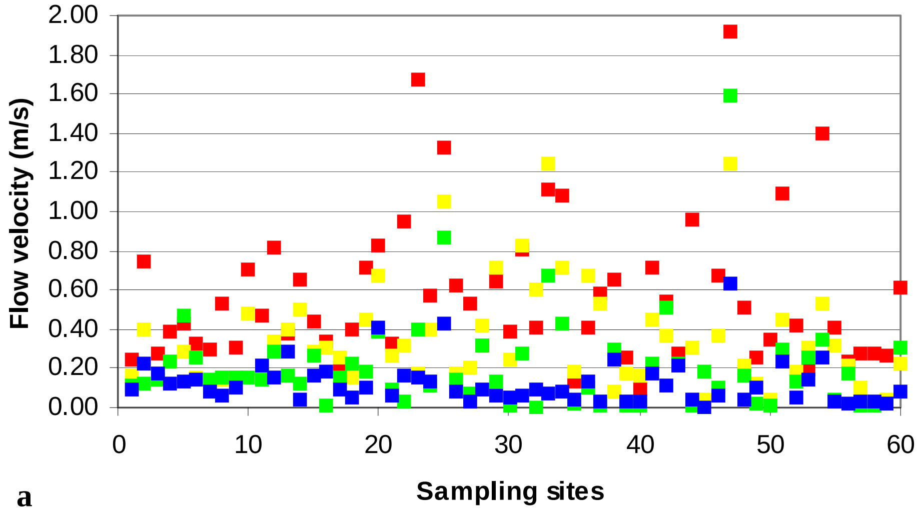

Depth (cm)
200
180
160
140
120
100
80
60
40
20
0
0
10
20
30
40
50
60
Sampling sites

a

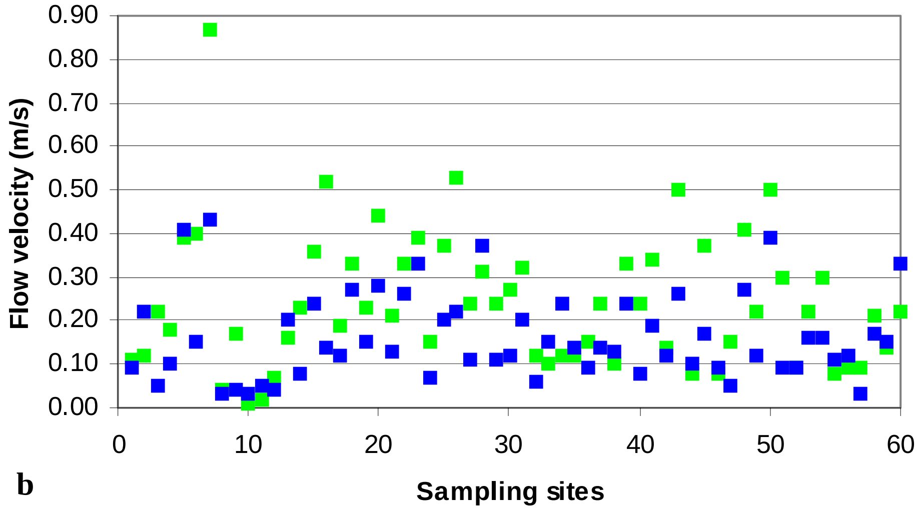

Depth (cm)
200
180
160
140
120
100
80
60
40
20
0
0
10
20
30
40
50
60
Sampling sites

b

Flow velocity:



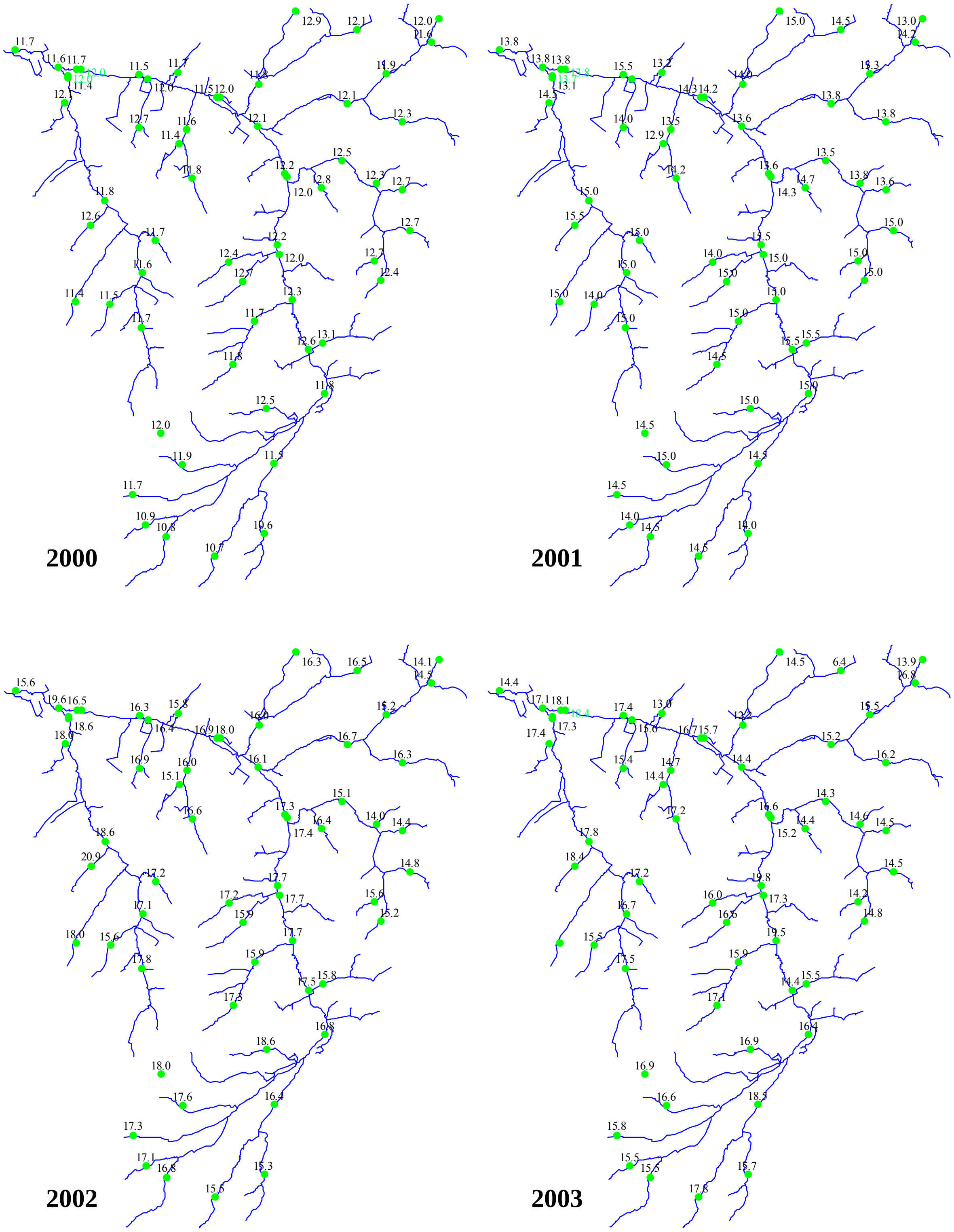
a



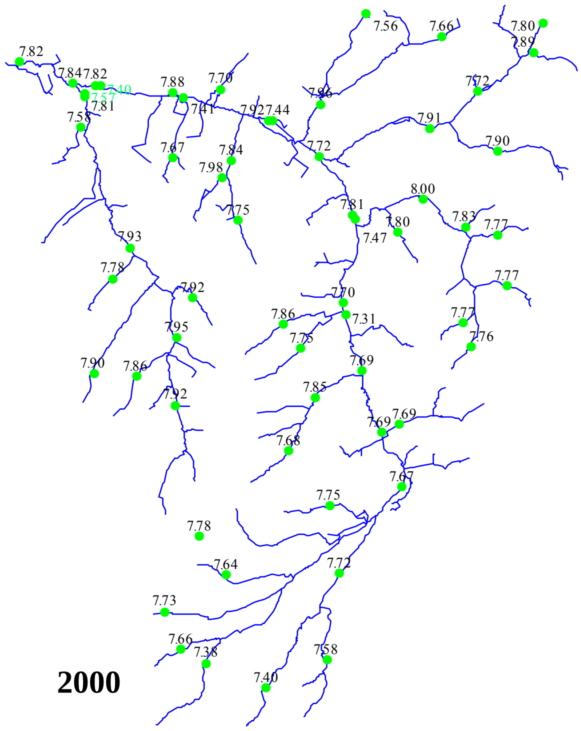
b

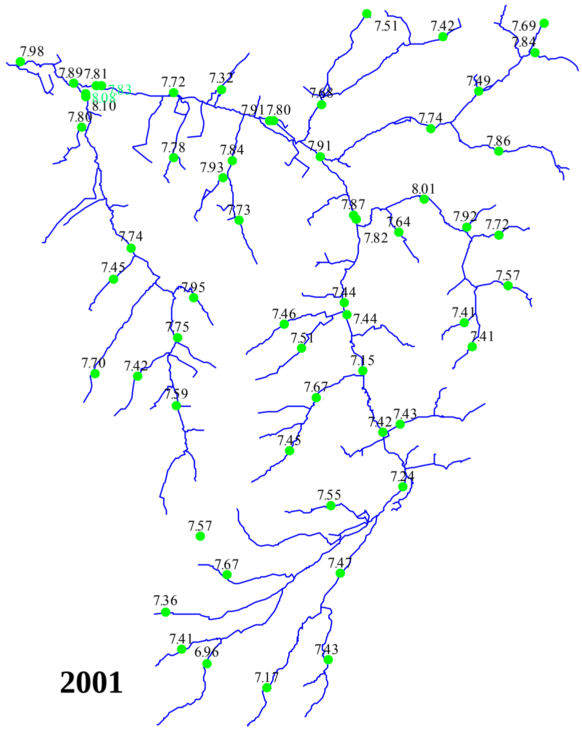
Appendix 4: Spatial distribution of the input variable (green dot complies with the quality standard, red dot does not comply with the quality standard) and the macroinvertebrates Tubificidae, Asellidae, Gammaridae, *Baetis* and Limnephilidae (green dot = present, red dot = absent) for the databases of the Zwalm river basin and the 'short distance' monitoring network.

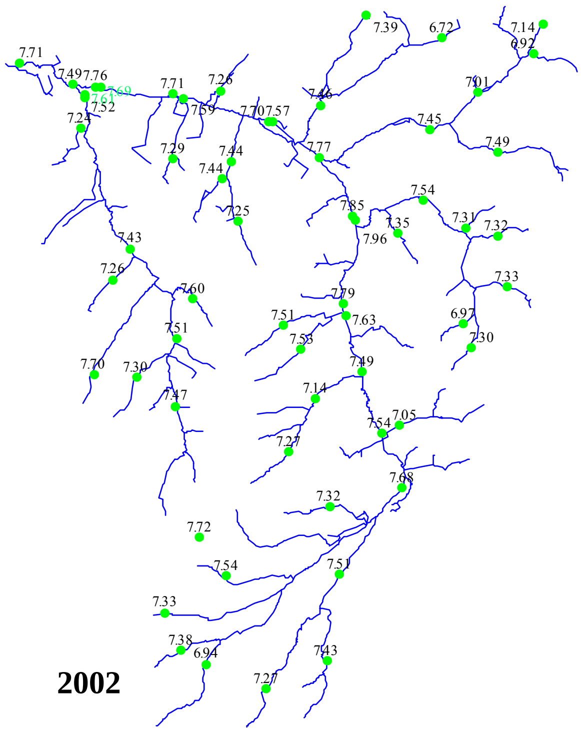
Temperature (the Zwalm river basin):

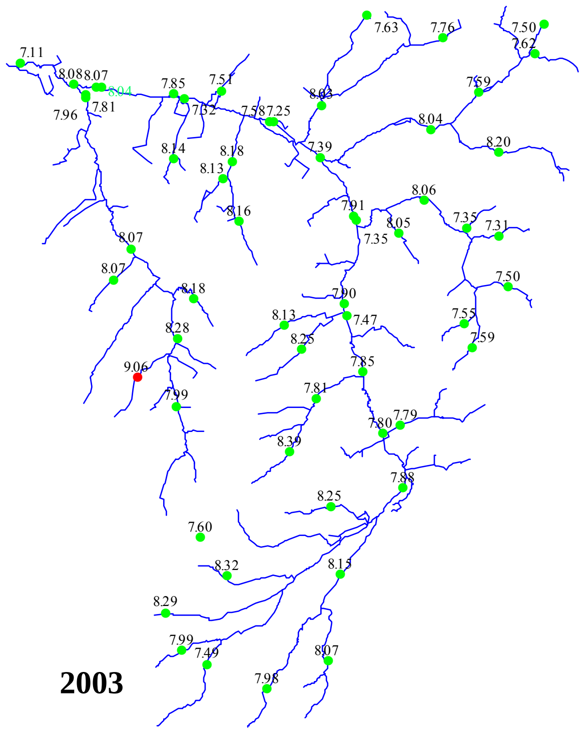


pH (the Zwalm river basin):

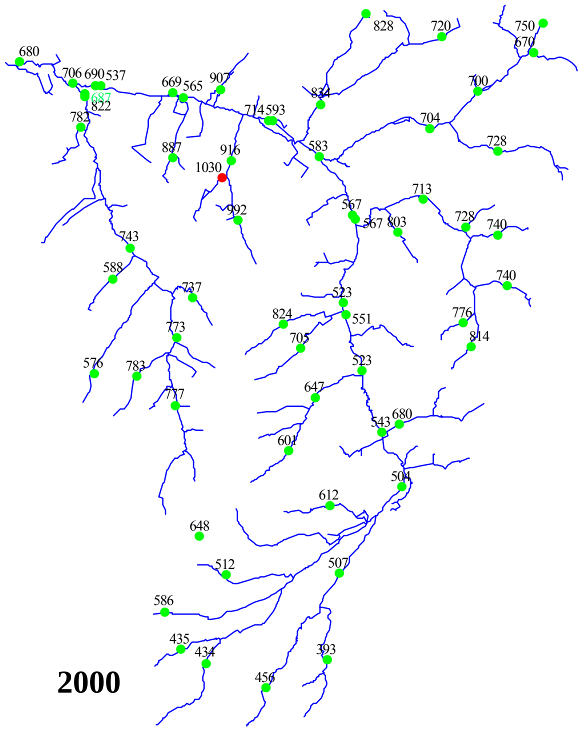


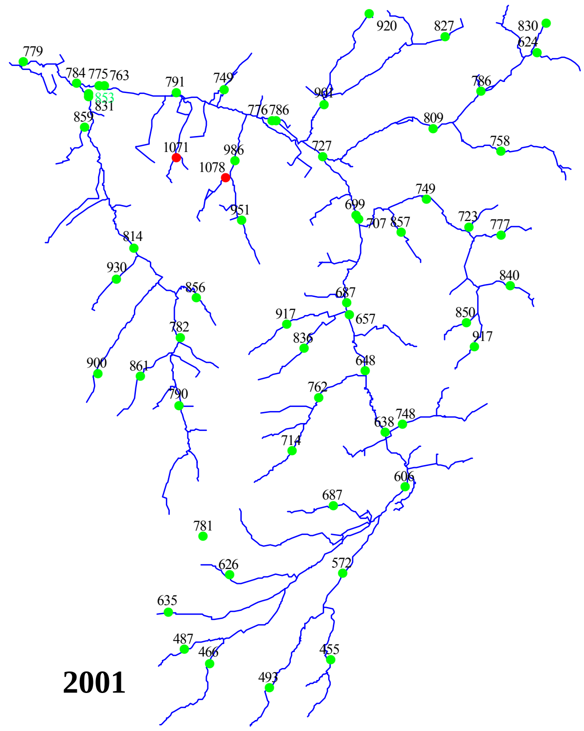


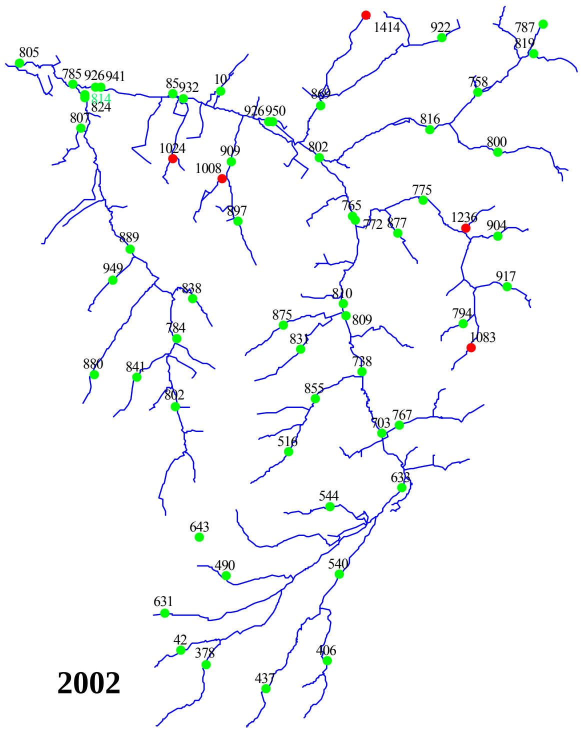


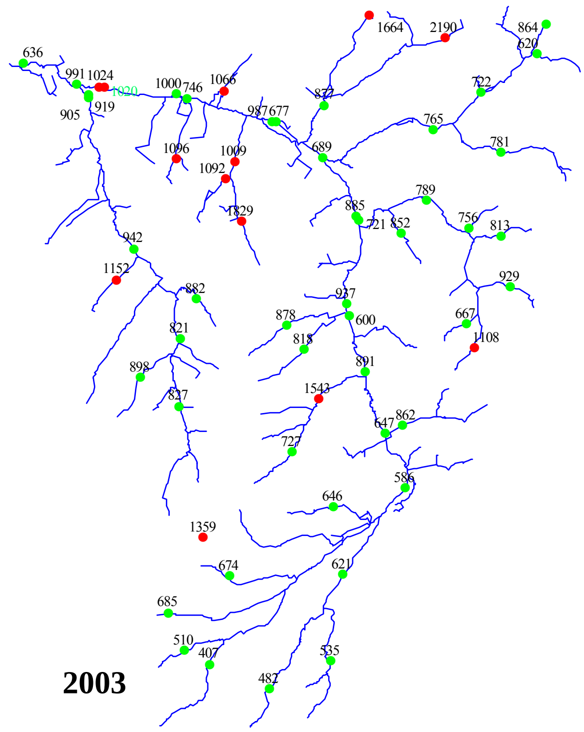


Conductivity (the Zwalm river basin):

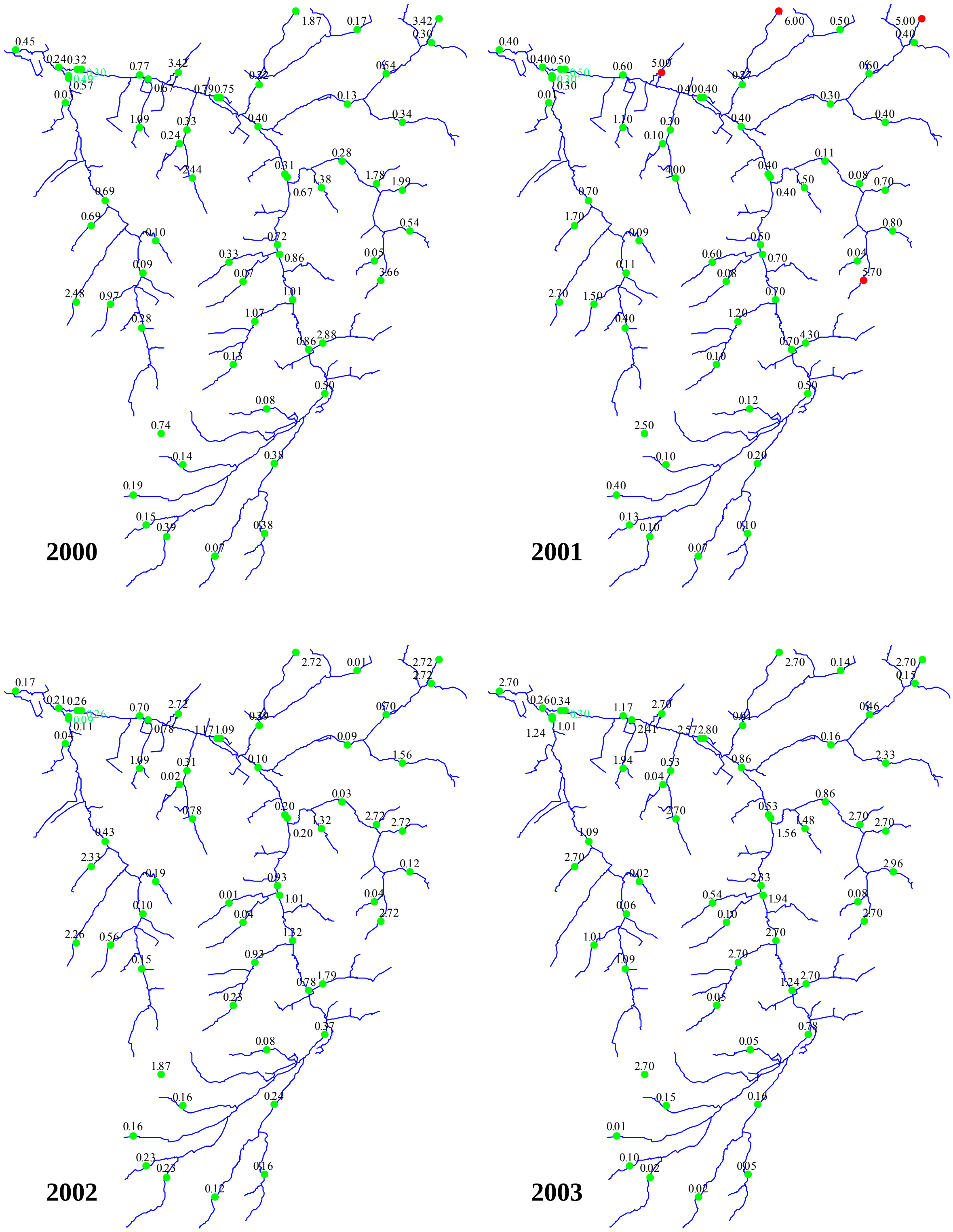




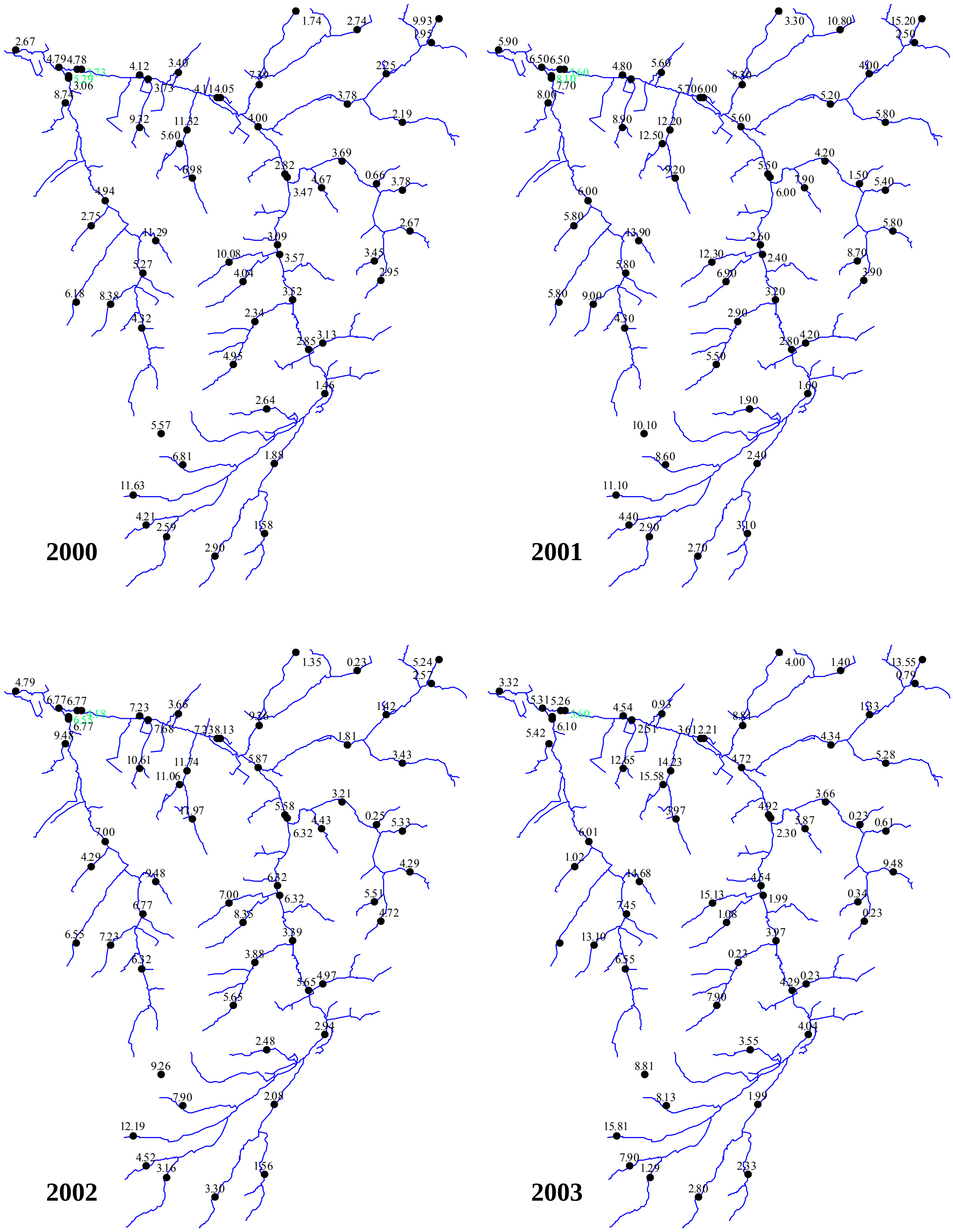




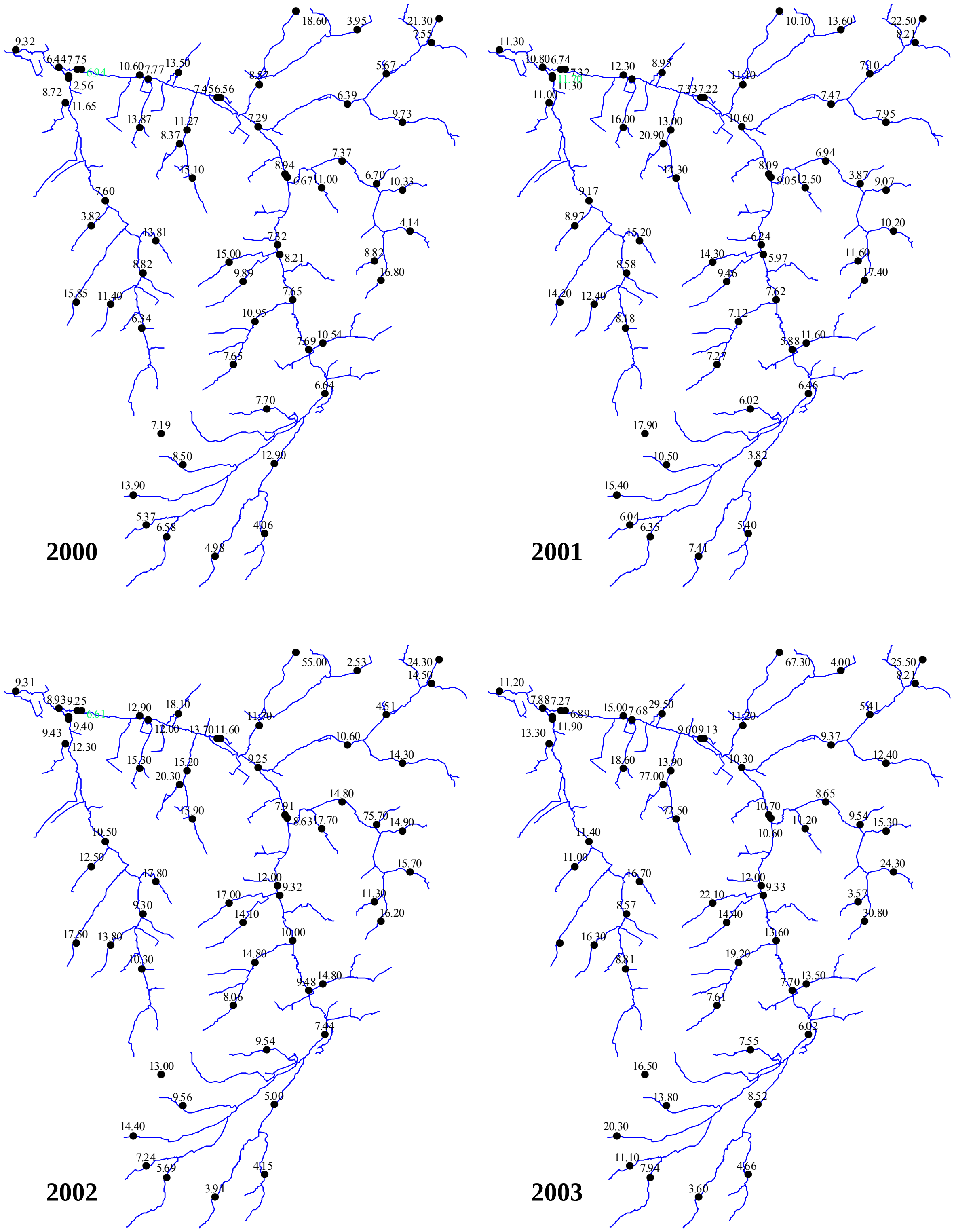
Ammonium (the Zwalm river basin):



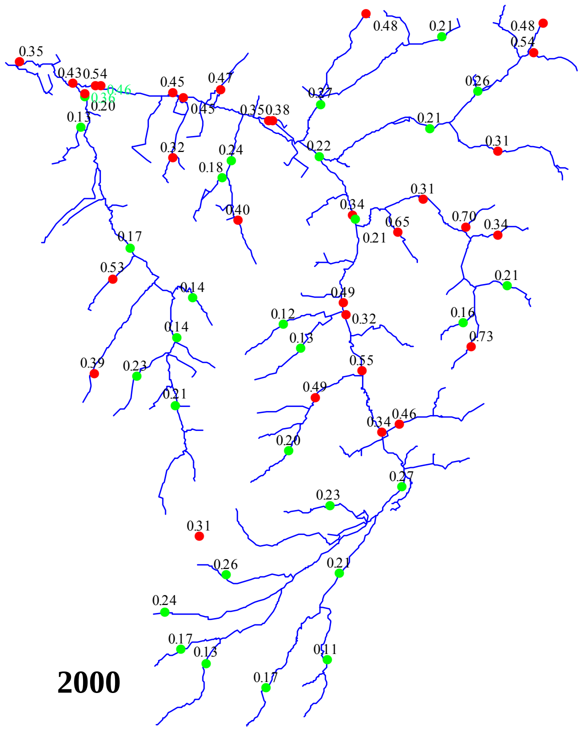
Nitrate (the Zwalm river basin):

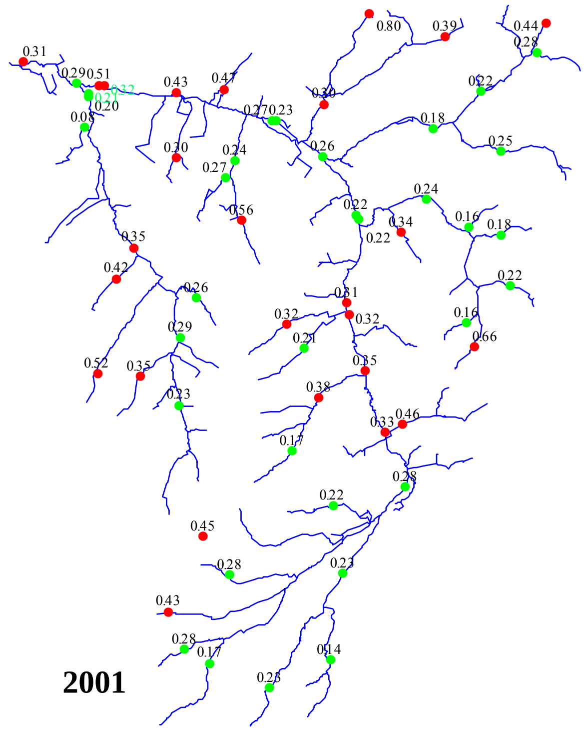


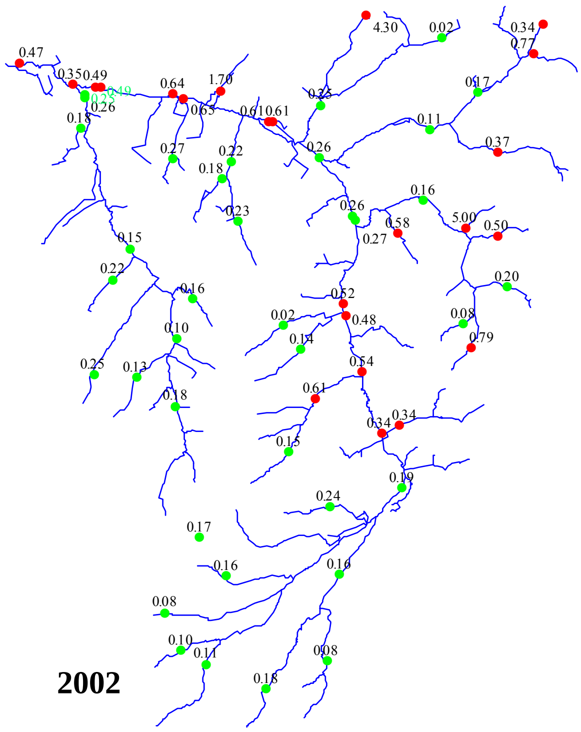
Total nitrogen (the Zwalm river basin):

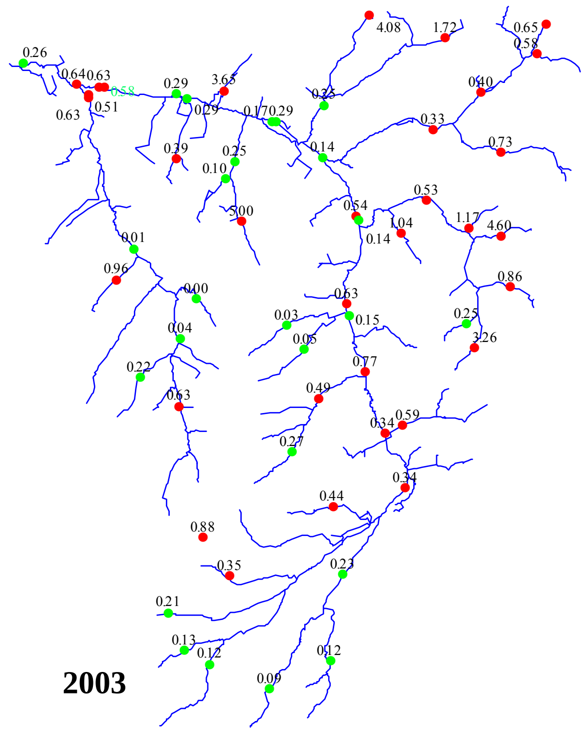


Ortho phosphate (the Zwalm river basin):

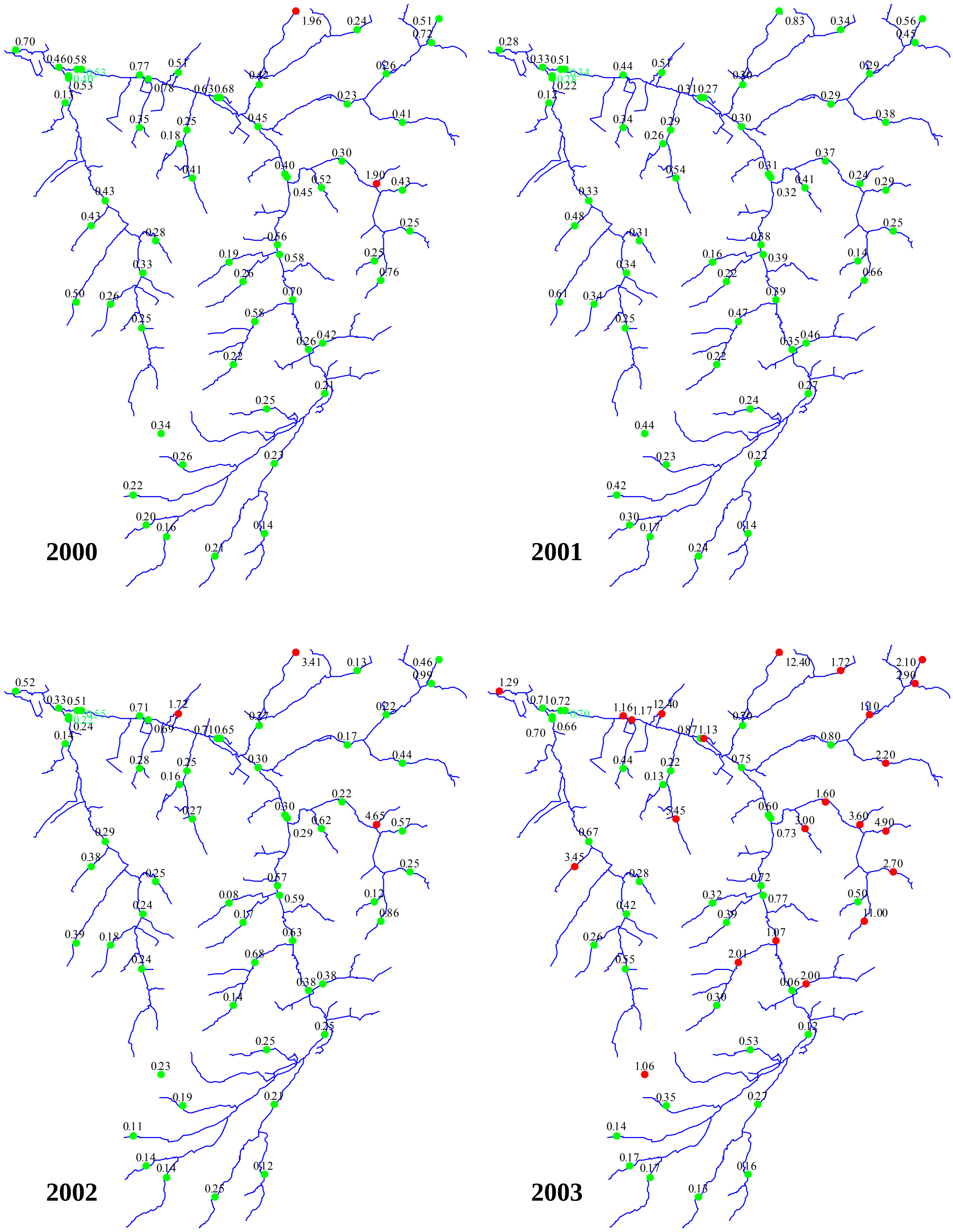




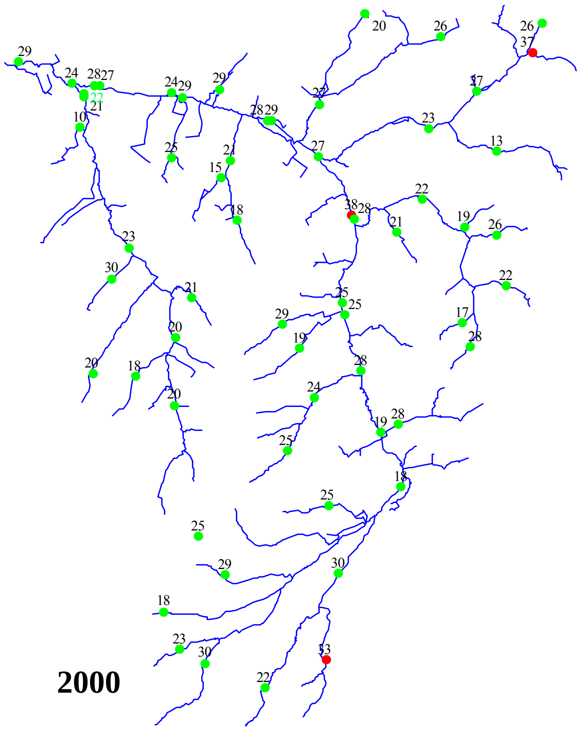


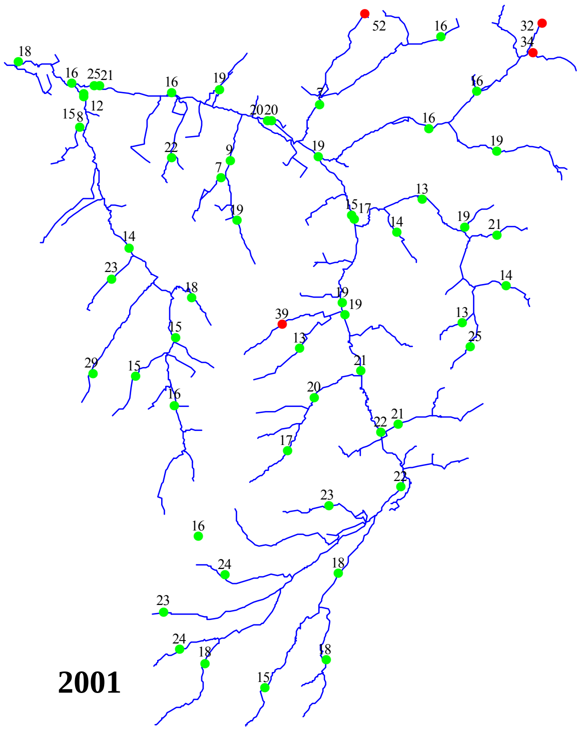


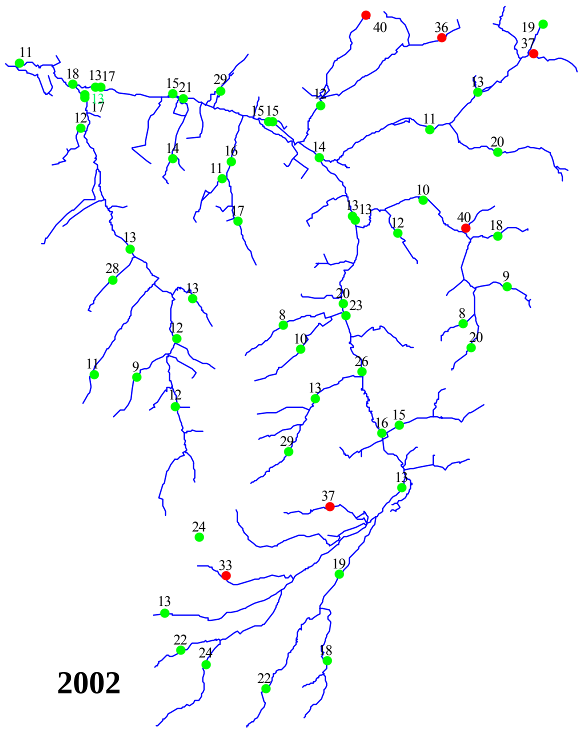
Total phosphorus (the Zwalm river basin):

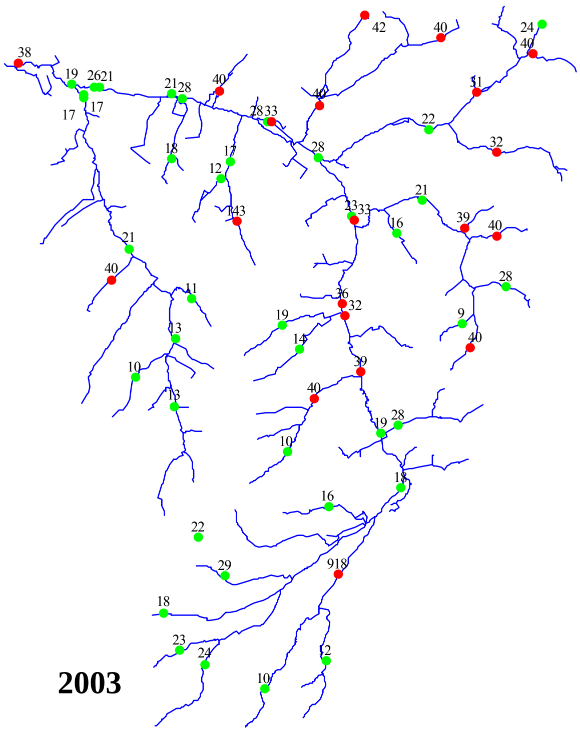


Chemical oxygen demand (the Zwalm river basin):

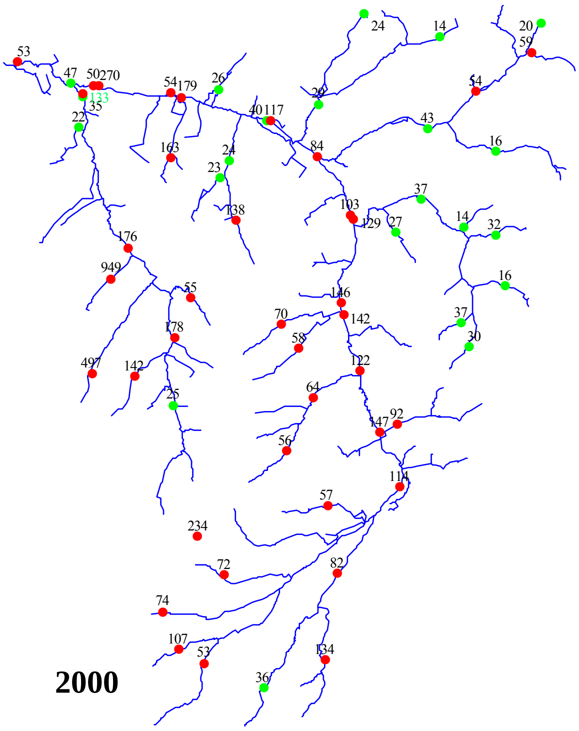




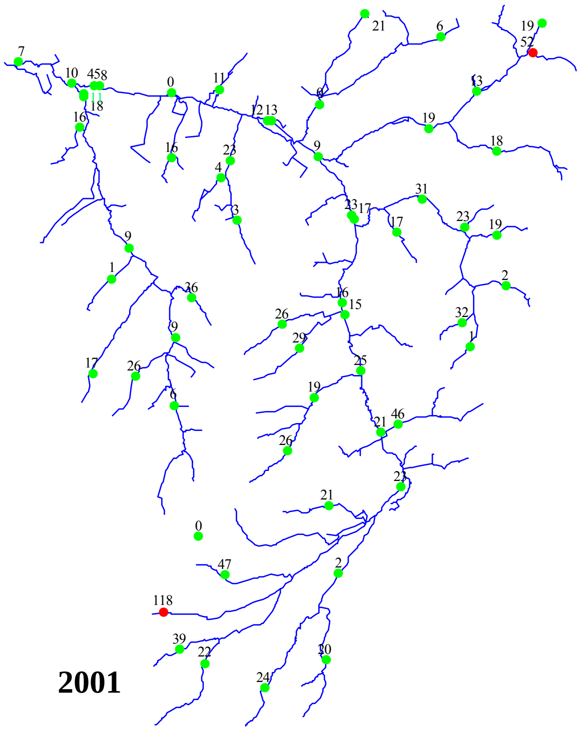




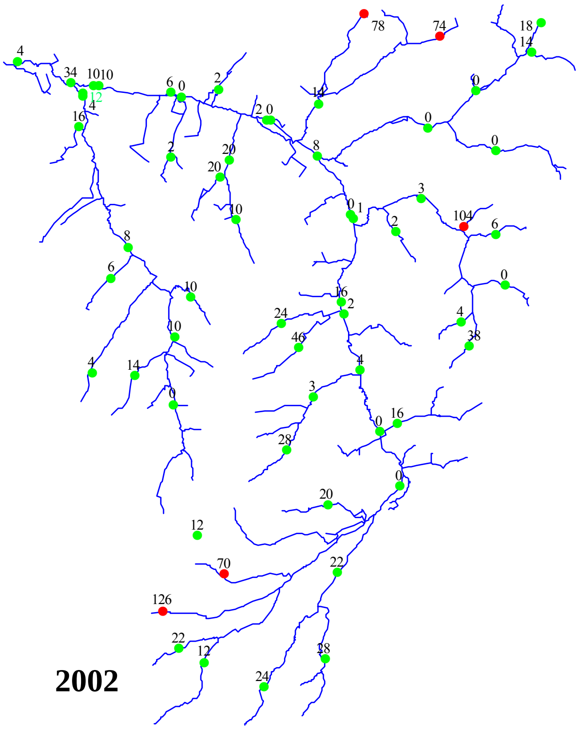
Suspended solids (the Zwalm river basin):



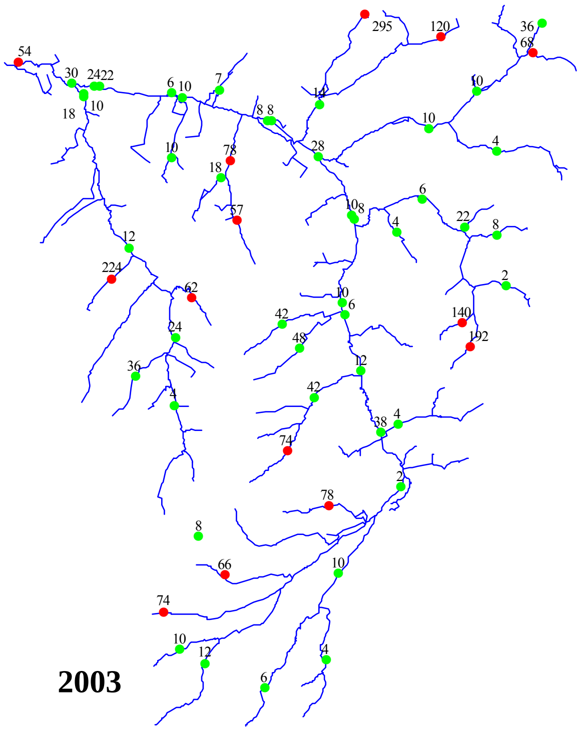
2000



2001

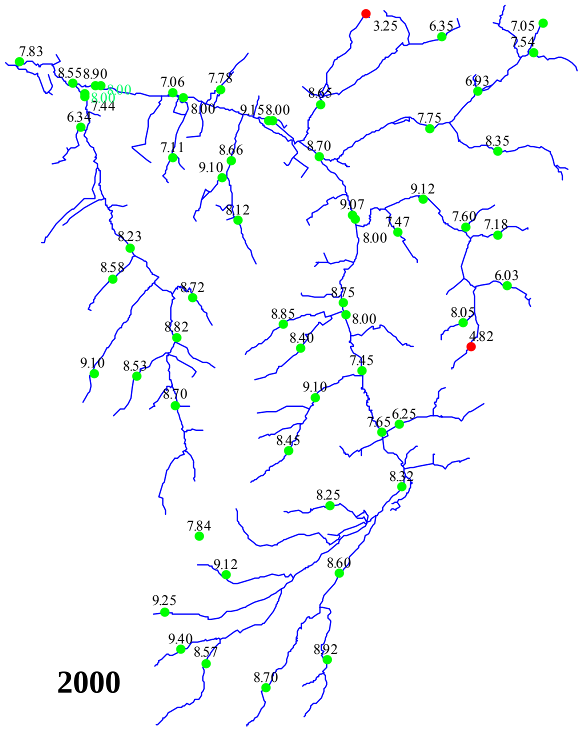


2002

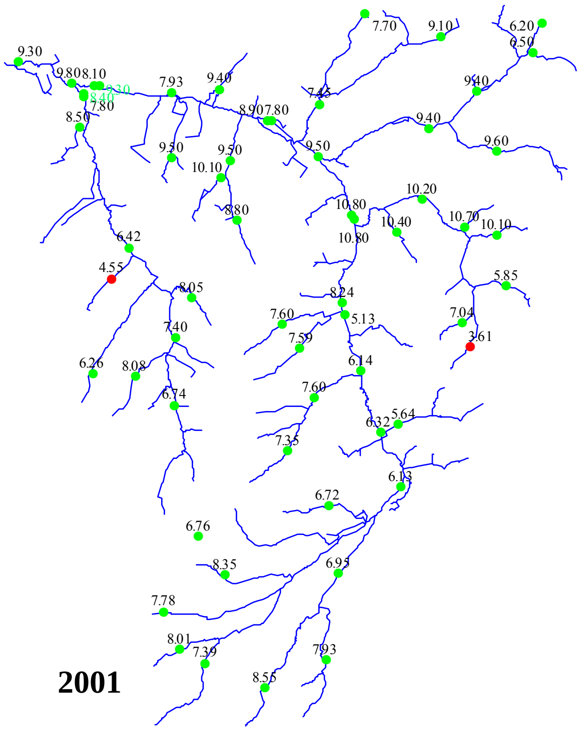


2003

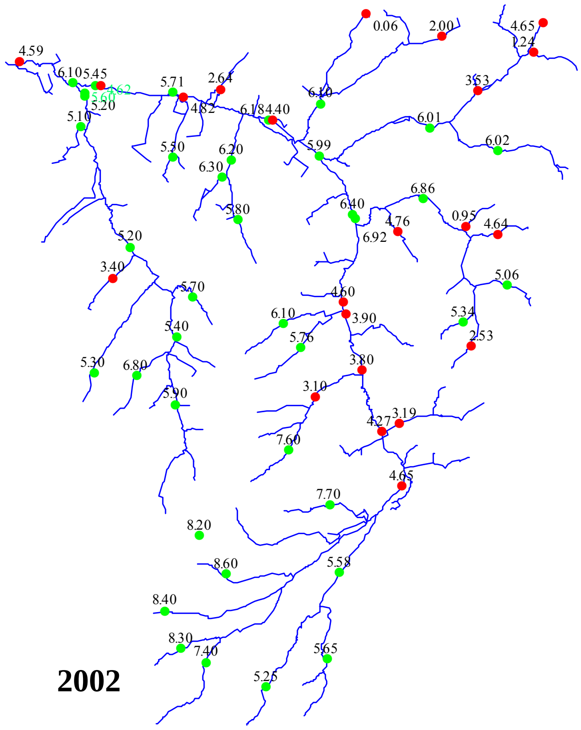
Dissolved oxygen (the Zwalm river basin):



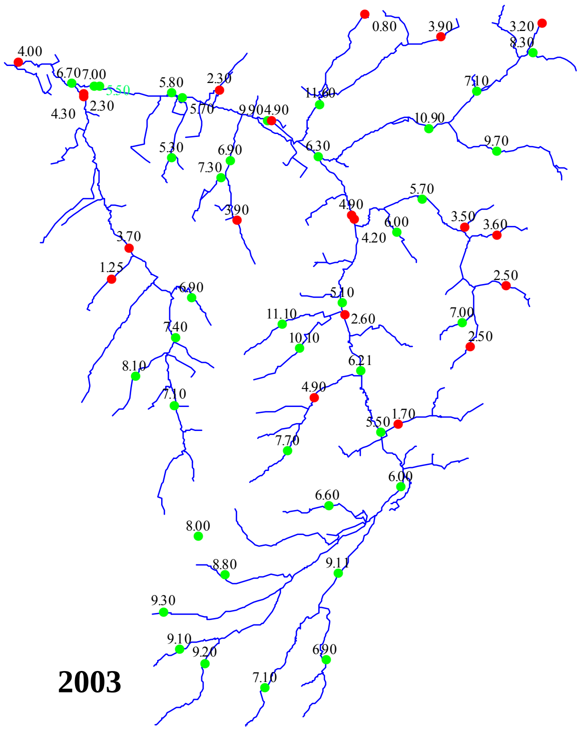
2000



2001

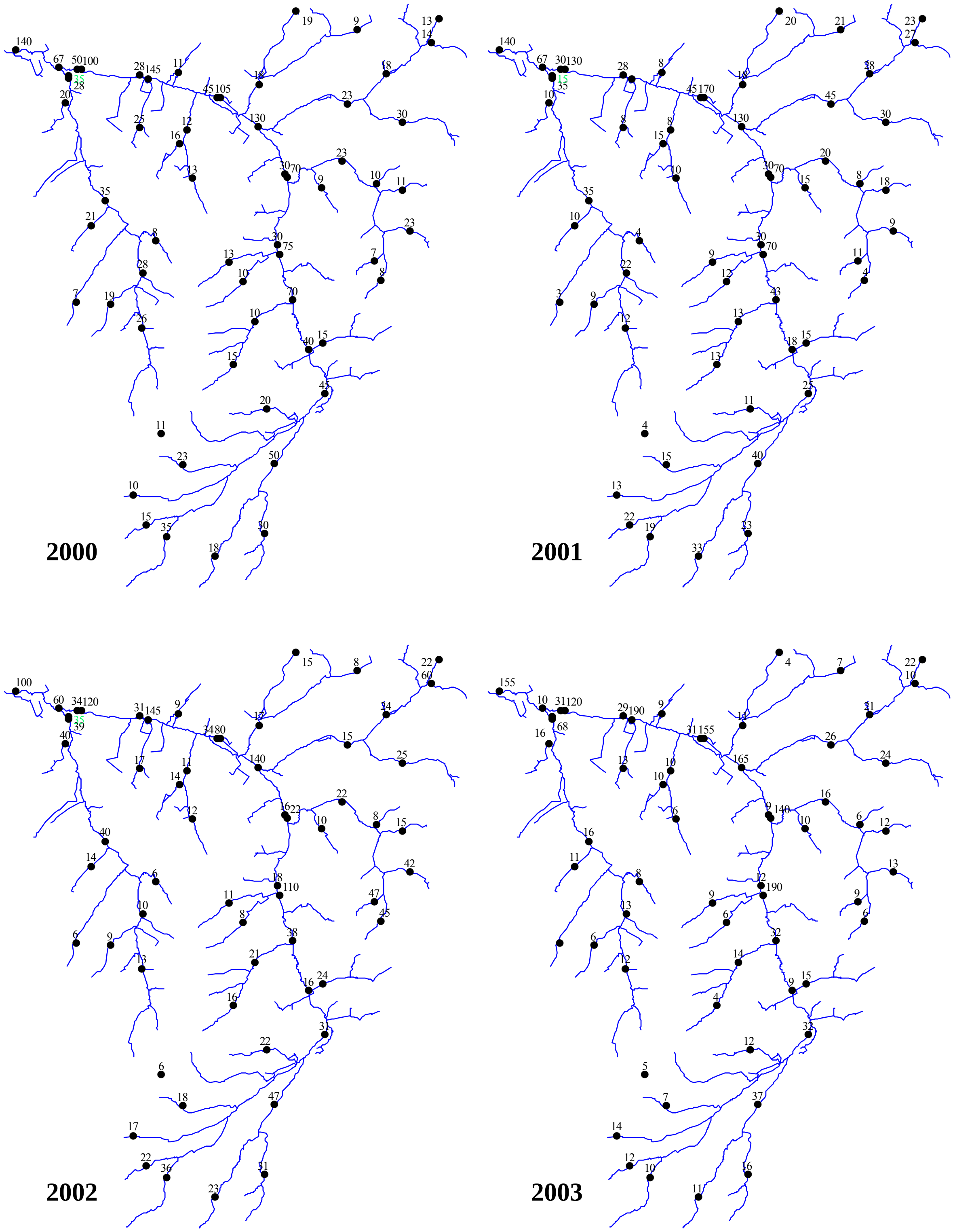


2002

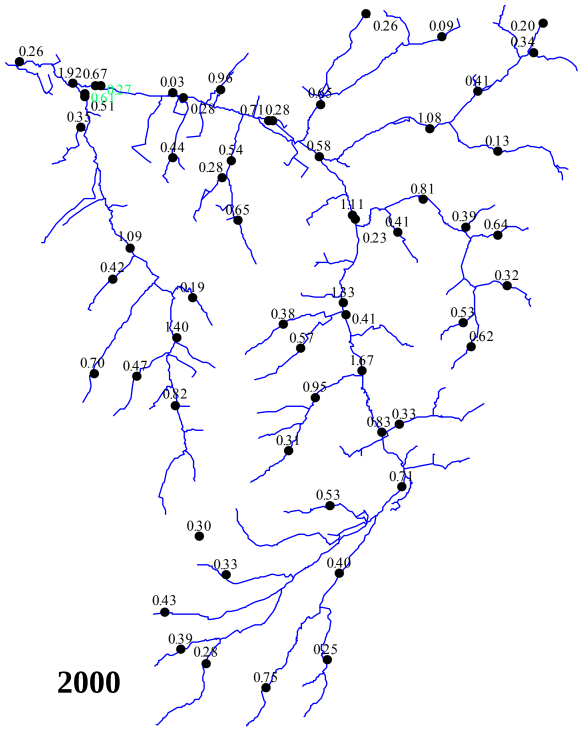


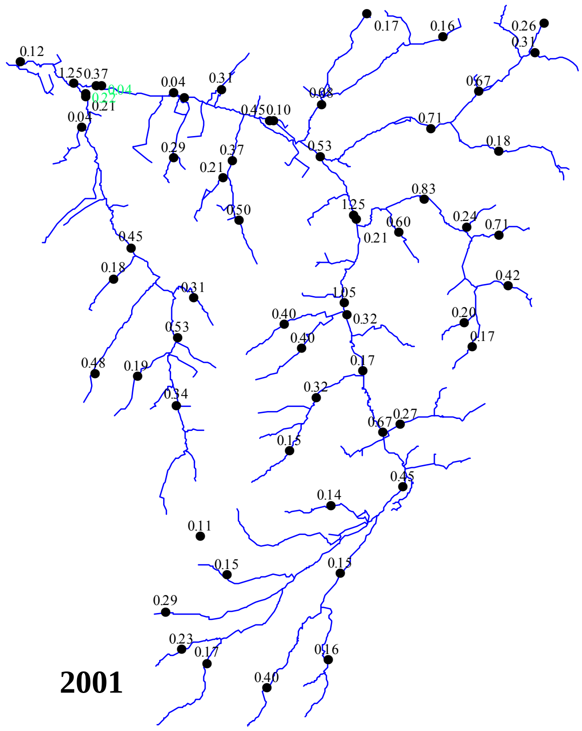
2003

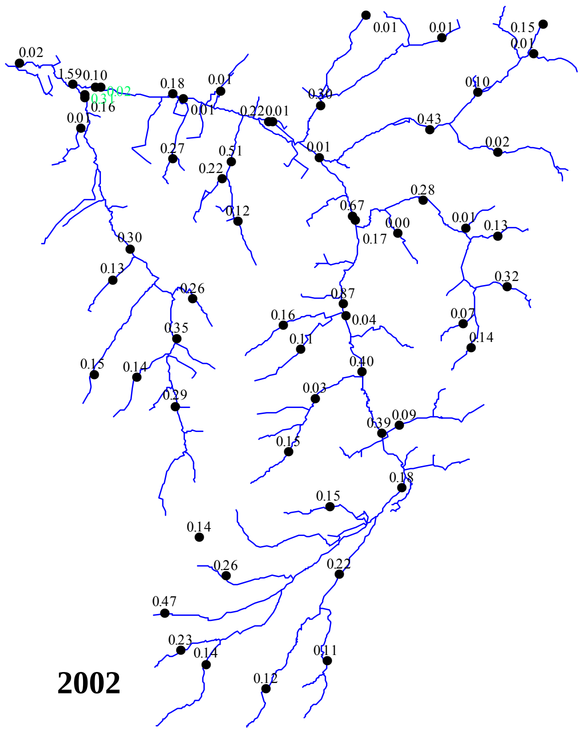
Depth (the Zwalm river basin):

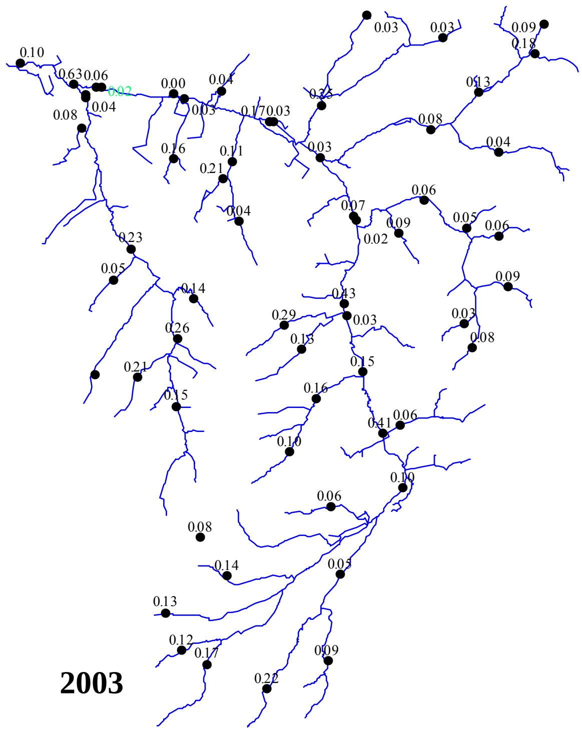


Flow velocity (the Zwalm river basin):

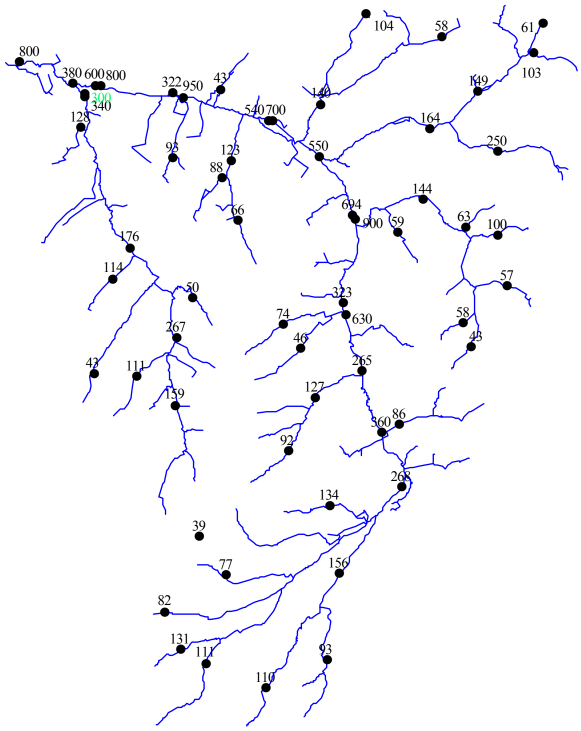




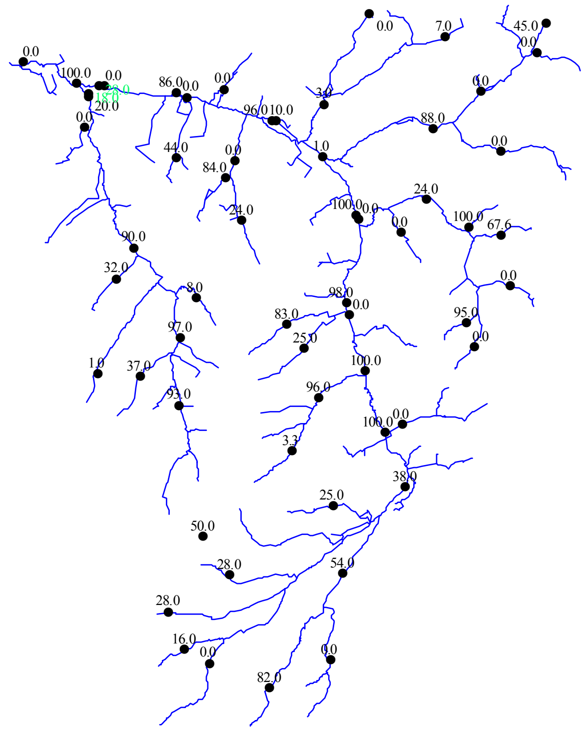




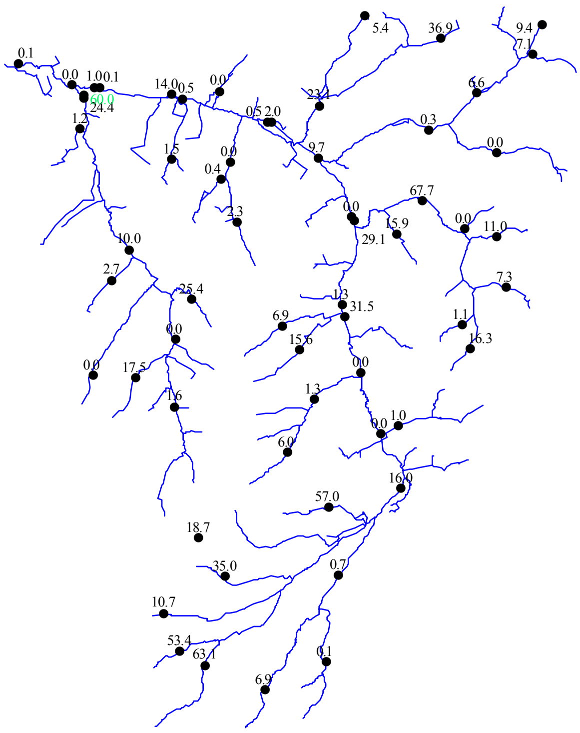
Width (the Zwalm river basin):



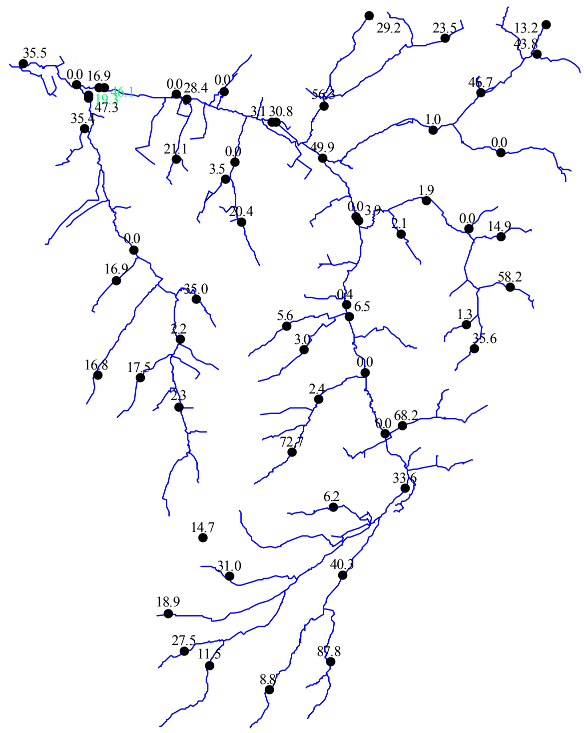
Fraction pebbles (the Zwalm river basin):



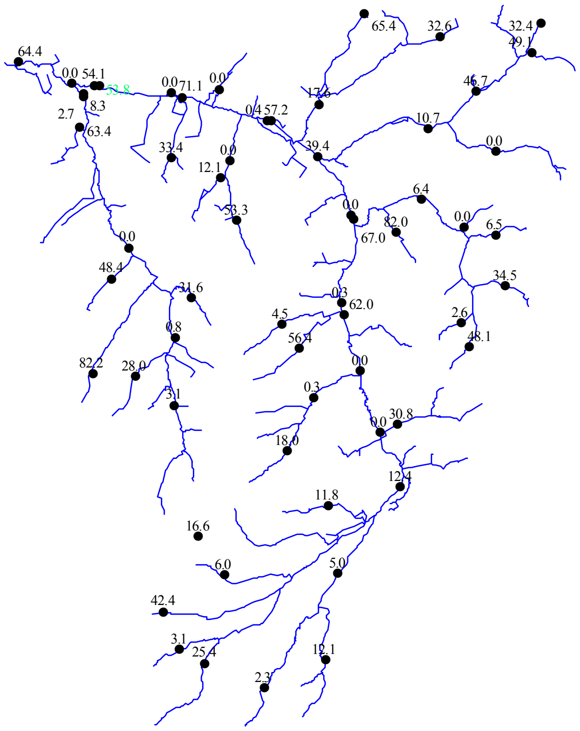
Fraction gravel (the Zwalm river basin):



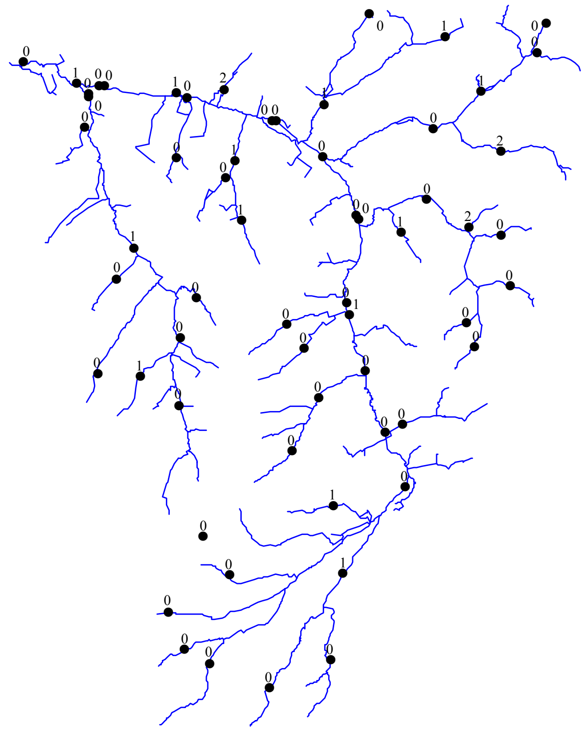
Fraction sand (the Zwalm river basin):



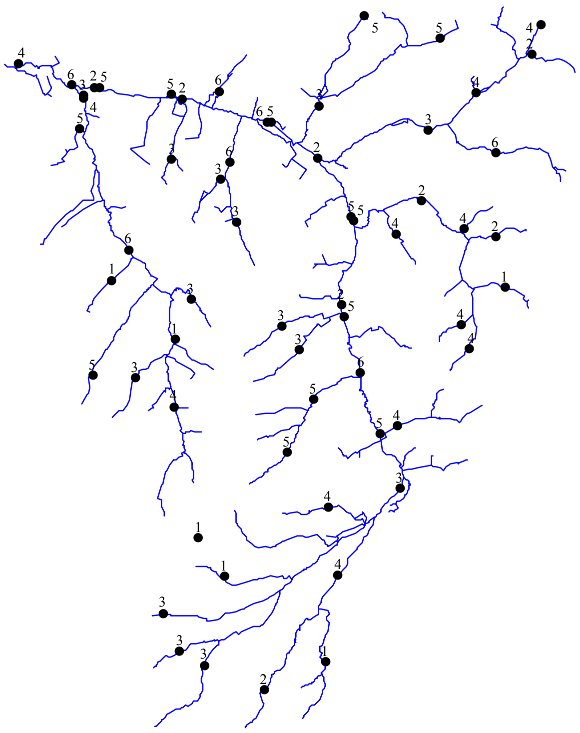
Fraction loam/clay (the Zwalm river basin):



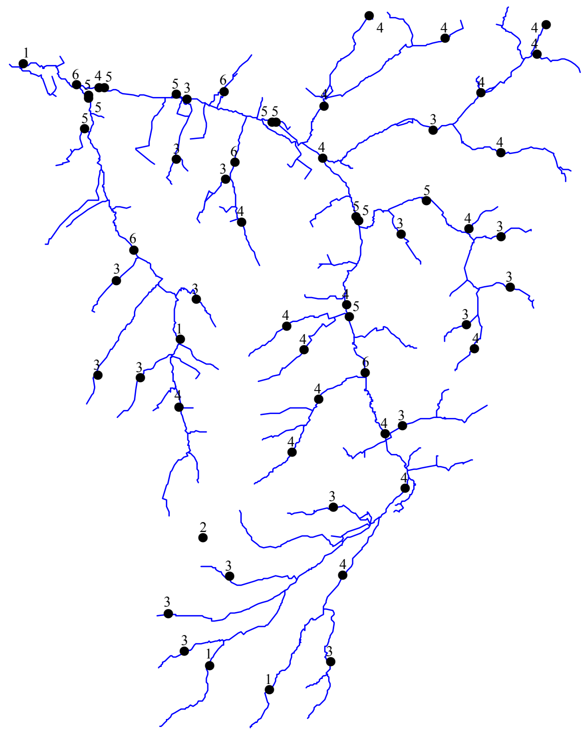
Embankment (the Zwalm river basin):



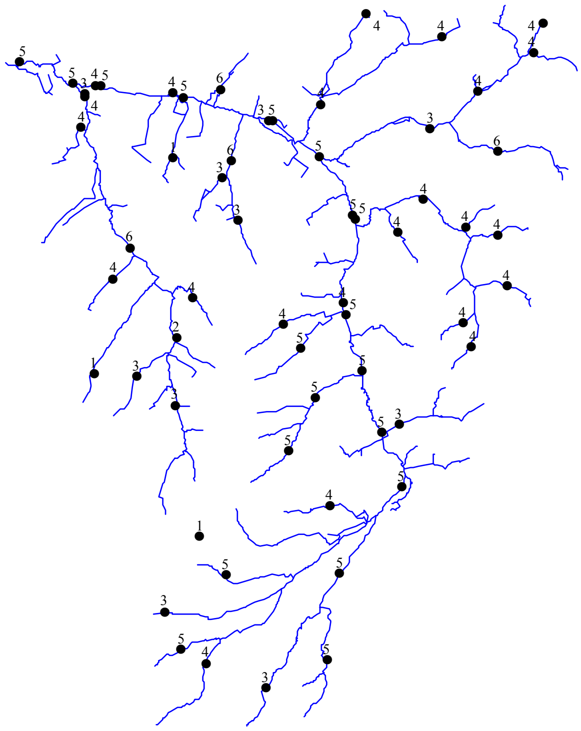
Meandering (the Zwalm river basin):



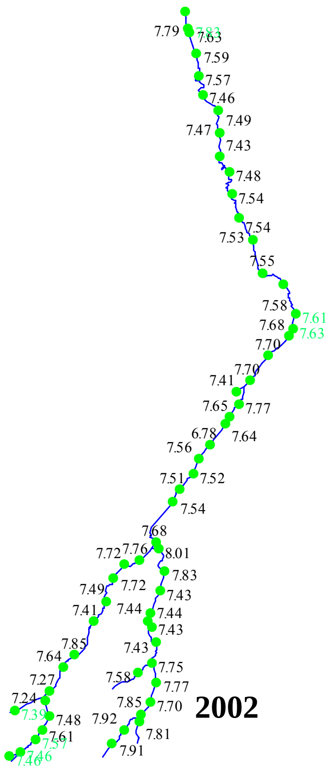
Hollow banks (the Zwalm river basin):

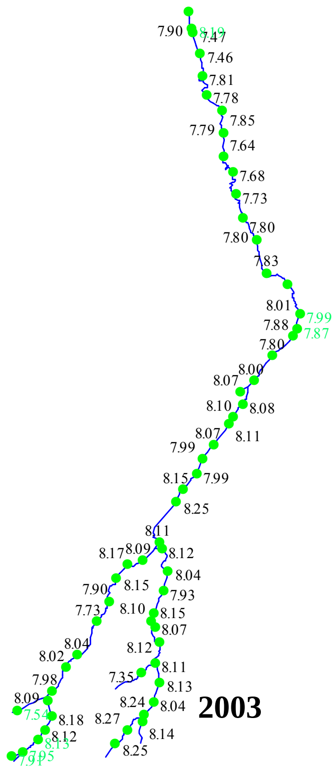


Pools-riffles (the Zwalm river basin):

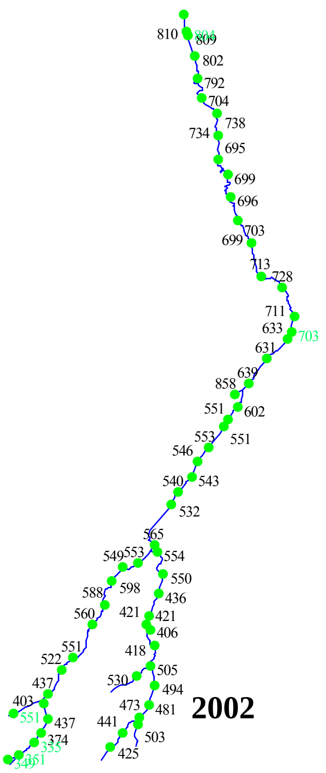


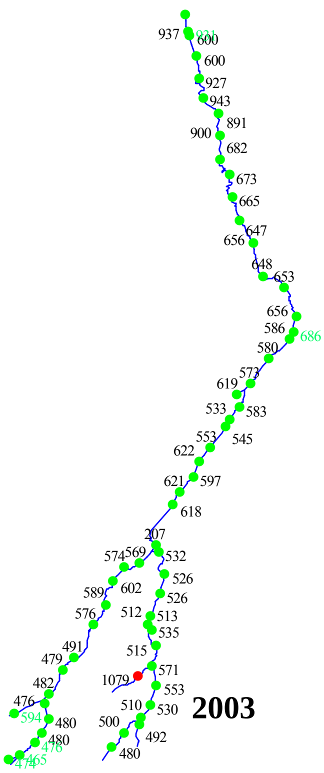
pH ('short distance' monitoring network):



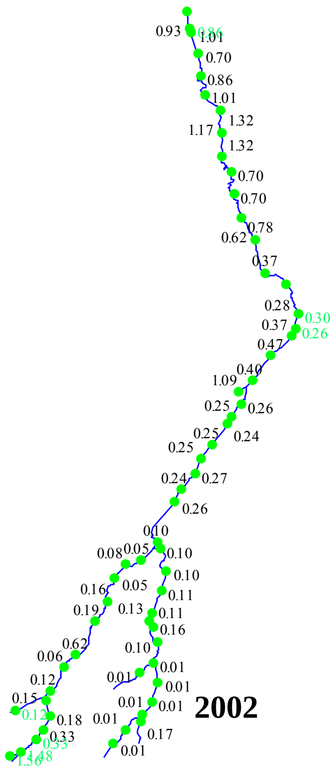


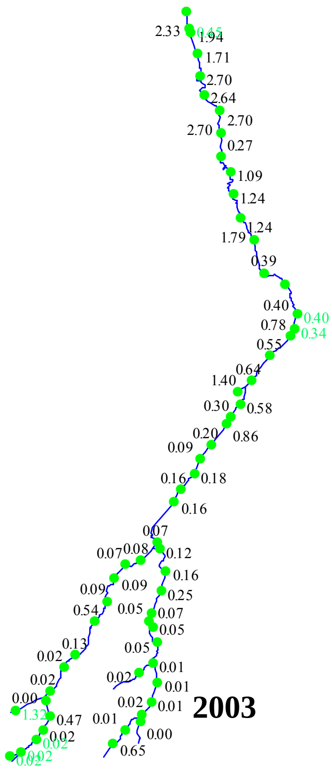
Conductivity ('short distance' monitoring network):



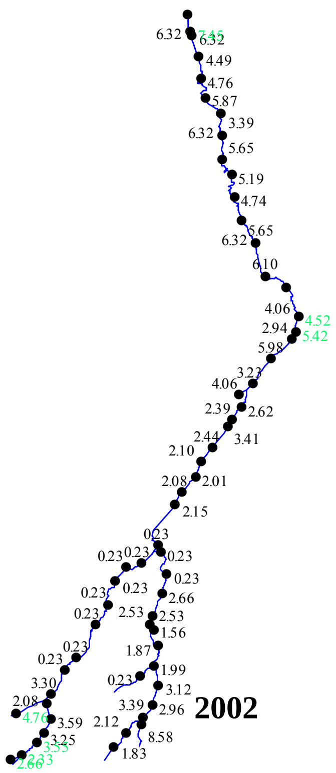


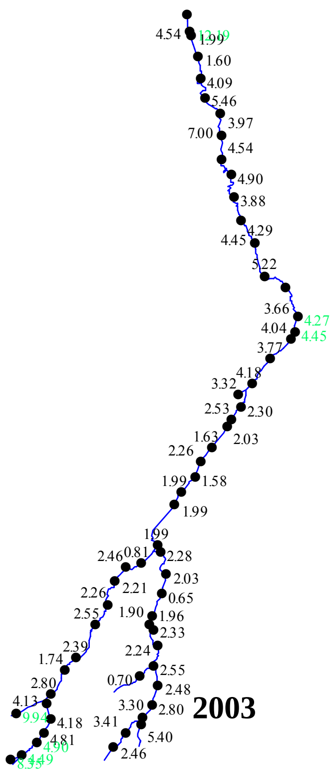
Ammonium ('short distance' monitoring network):



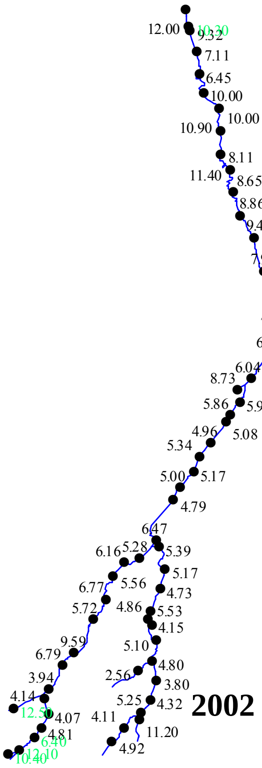


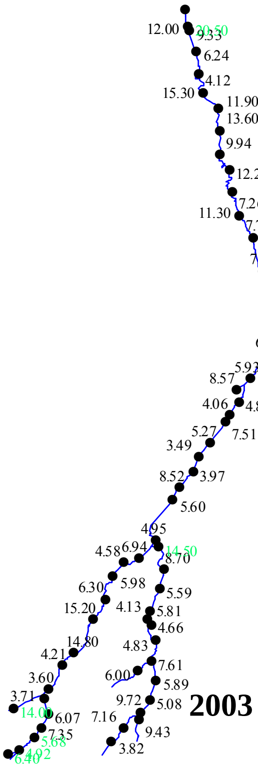
Nitrate ('short distance' monitoring network):



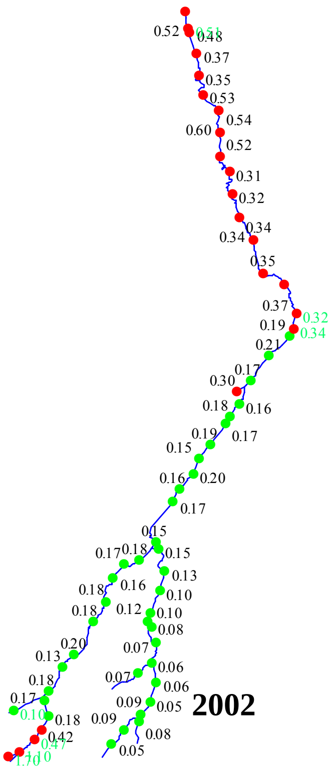


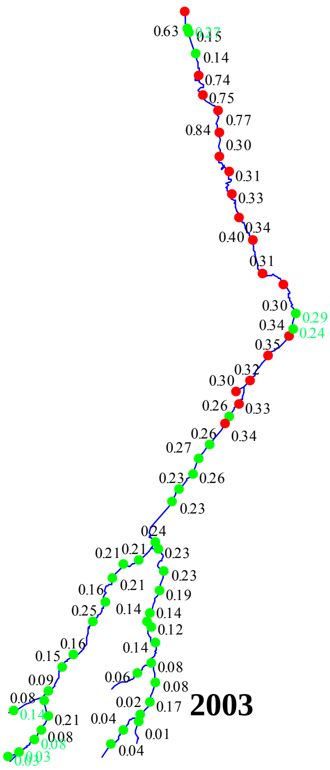
Total nitrogen ('short distance' monitoring network):



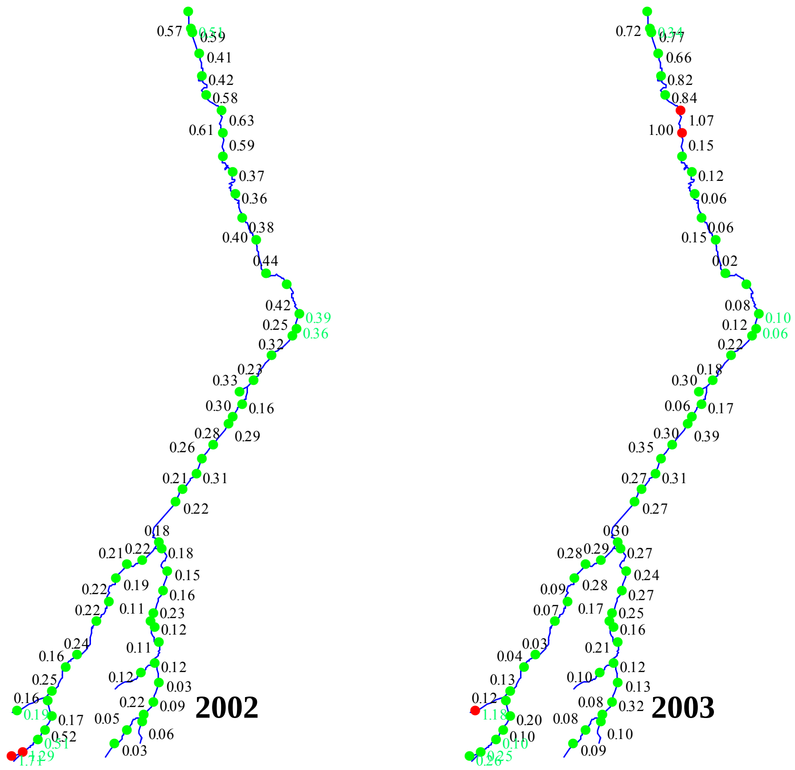


Ortho phosphate ('short distance' monitoring network):

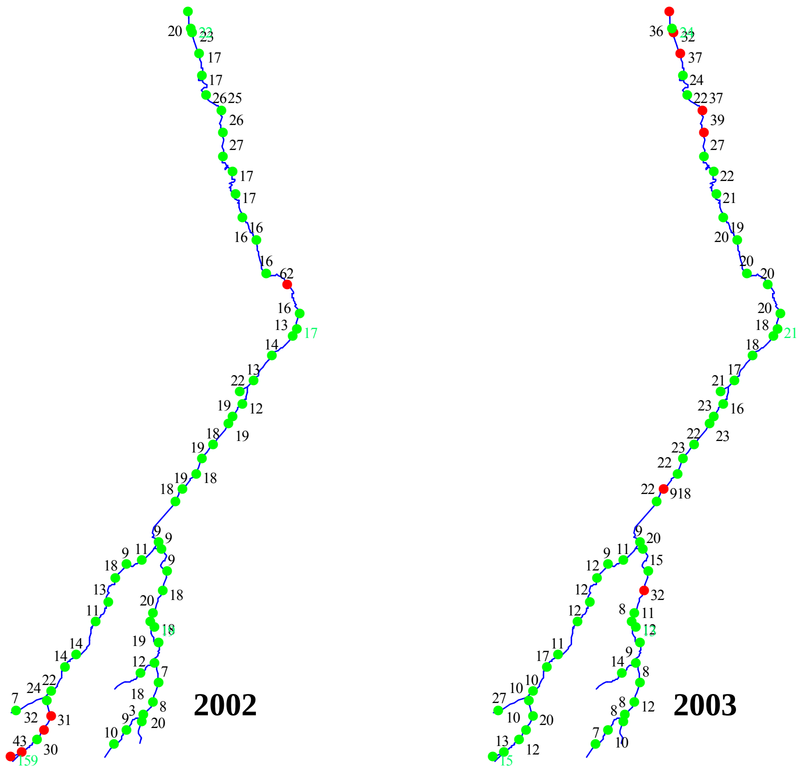




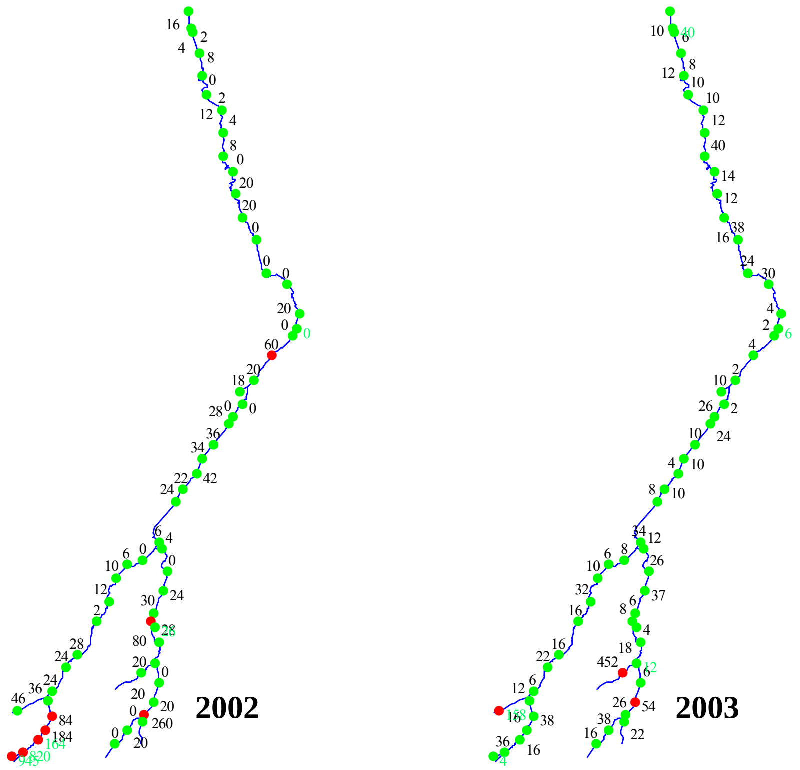
Total phosphorus (‘short distance’ monitoring network):



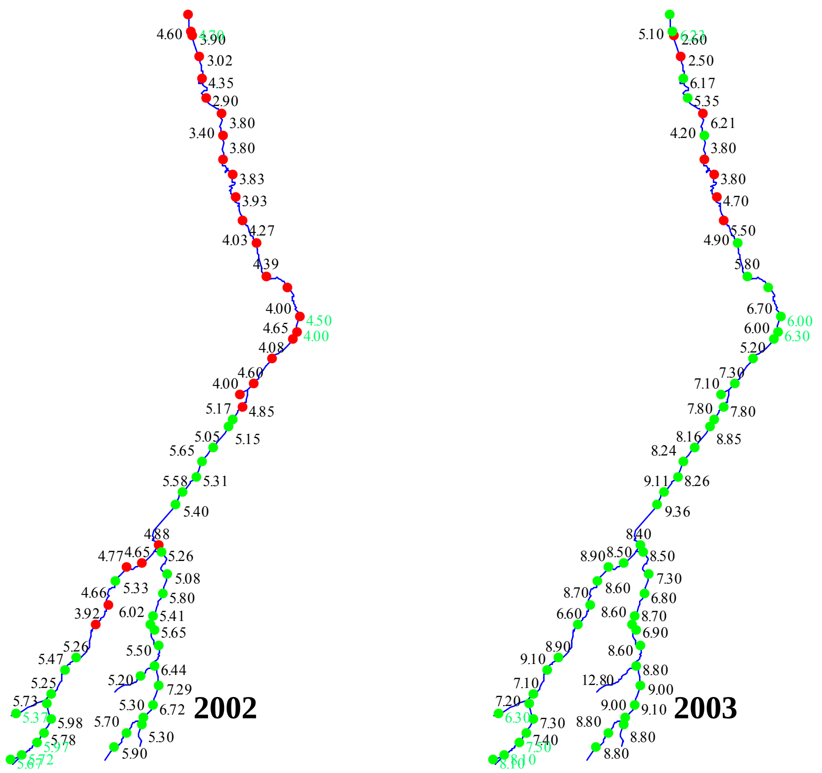
Chemical oxygen demand (‘short distance’ monitoring network):



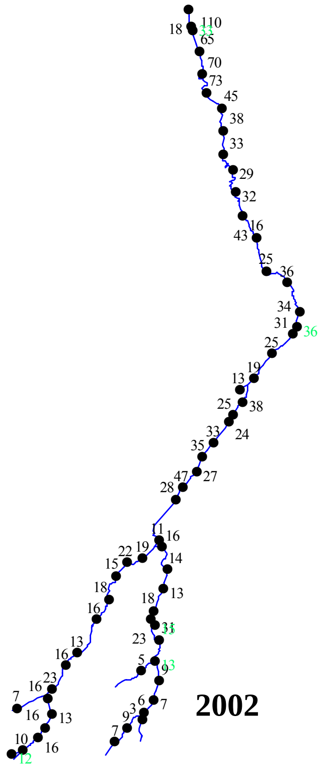
Suspended solids ('short distance' monitoring network):



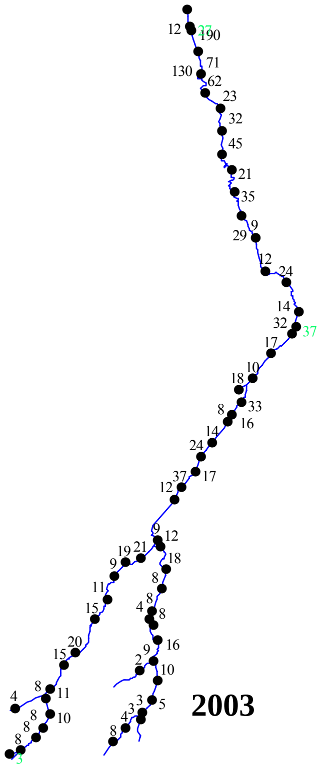
Dissolved oxygen ('short distance' monitoring network):



Depth ('short distance' monitoring network):

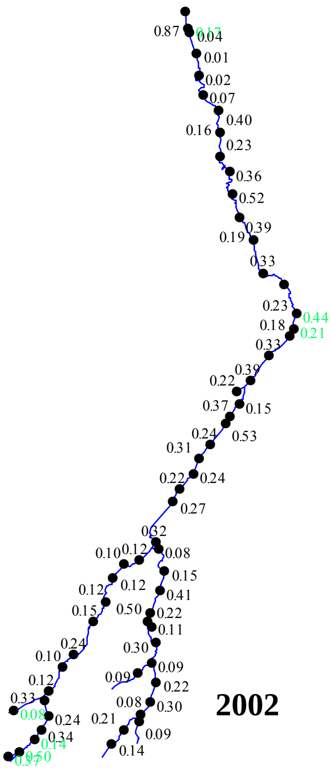


2002

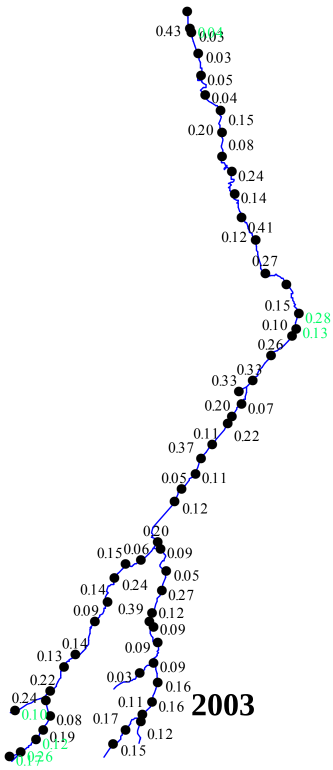


2003

Flow velocity ('short distance' monitoring network):

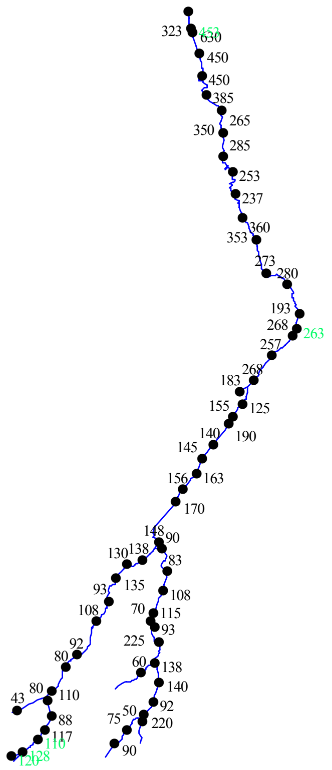


2002

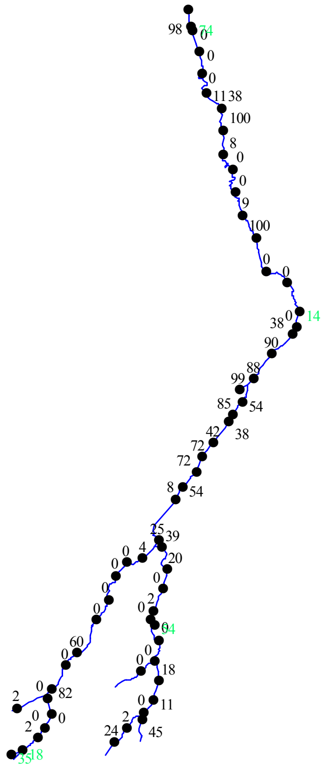


2003

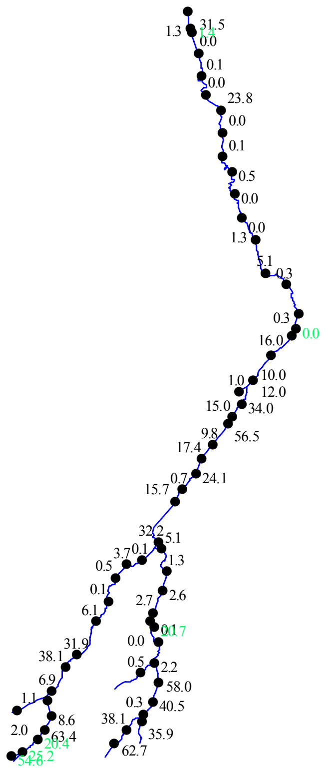
Width ('short distance' monitoring network):



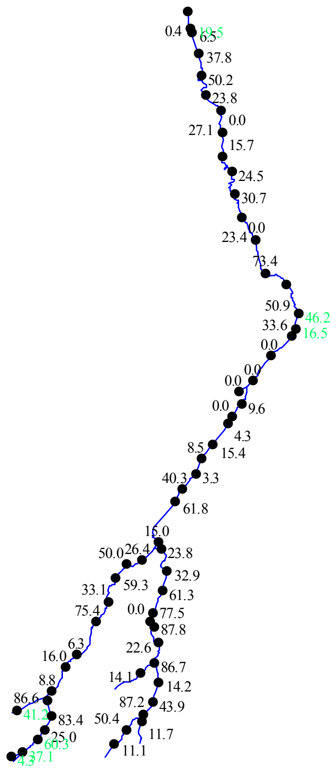
Fraction pebbles ('short distance' monitoring network):



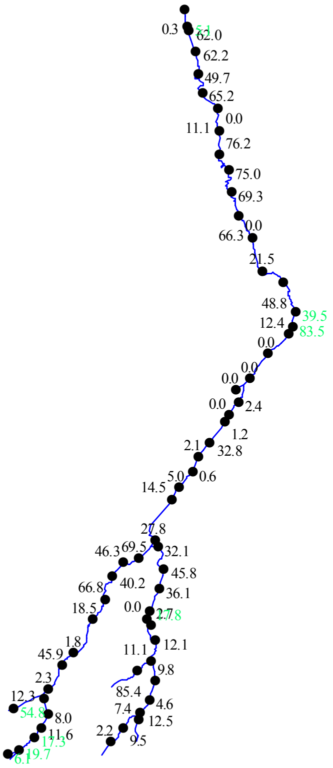
Fraction gravel ('short distance' monitoring network):



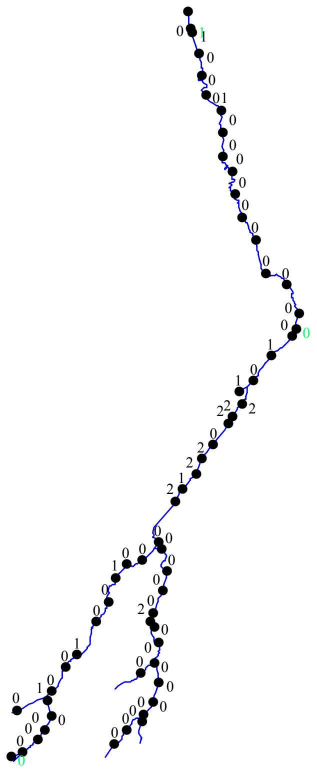
Fraction sand ('short distance' monitoring network):



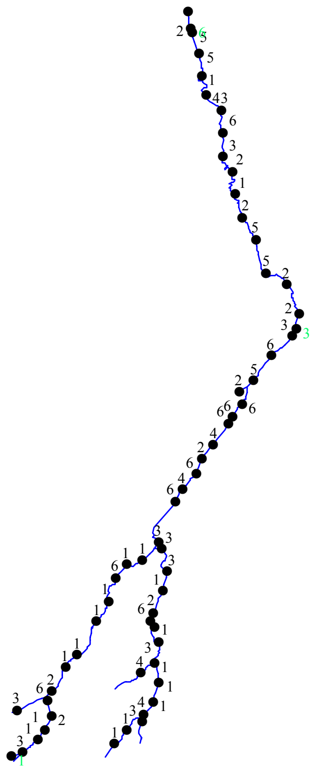
Fraction loam/clay ('short distance' monitoring network):



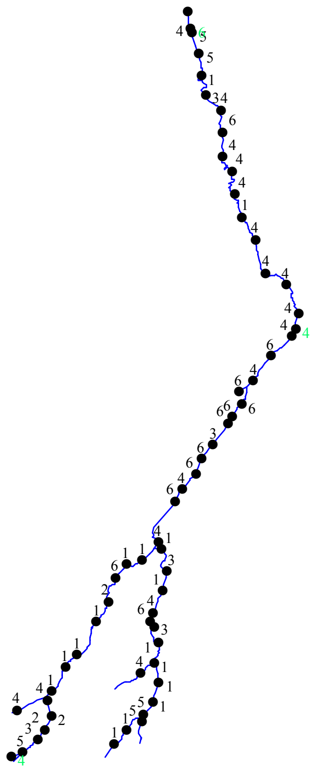
Embankment ('short distance' monitoring network):



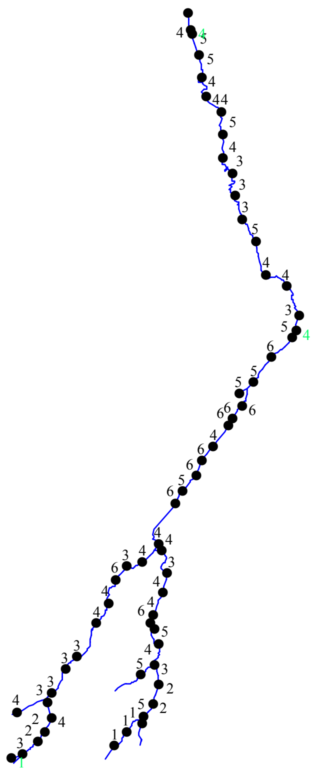
Meandering ('short distance' monitoring network):



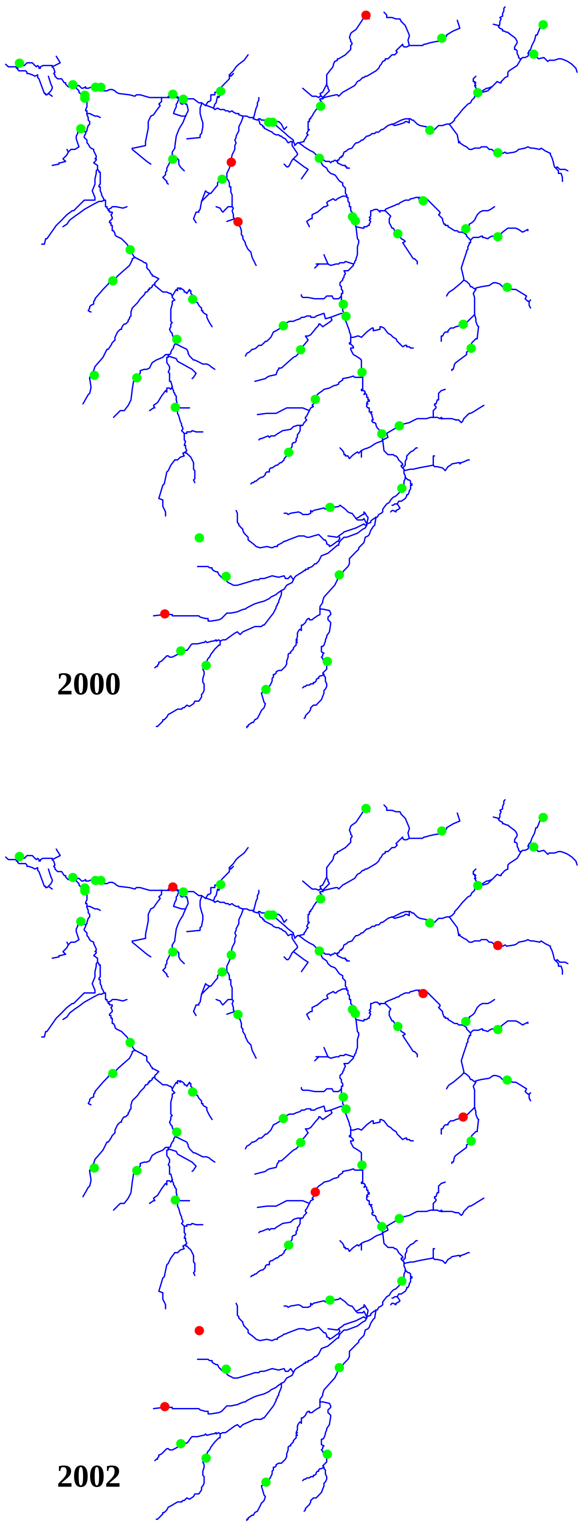
Hollow banks ('short distance' monitoring network):



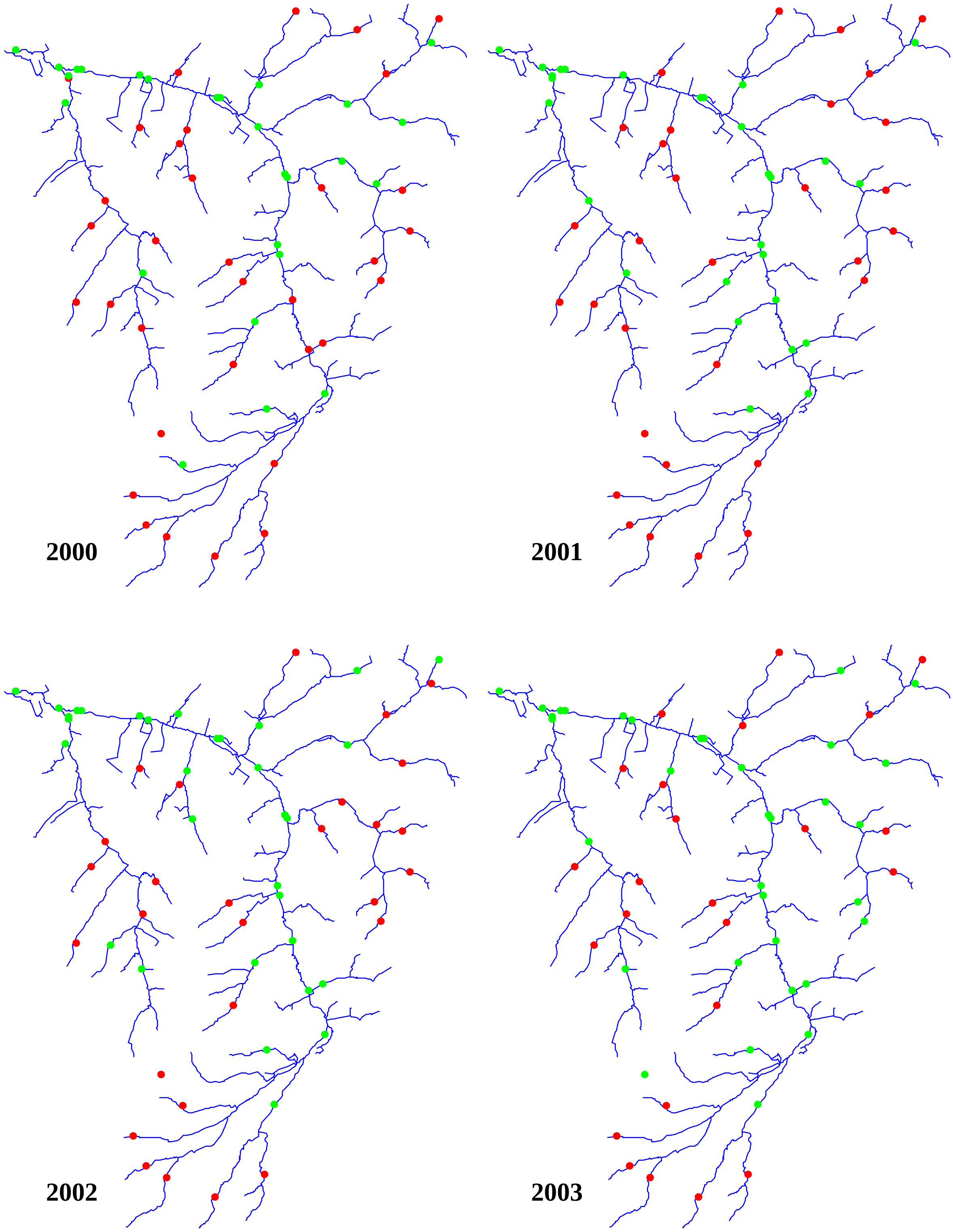
Pools-riffles ('short distance' monitoring network):



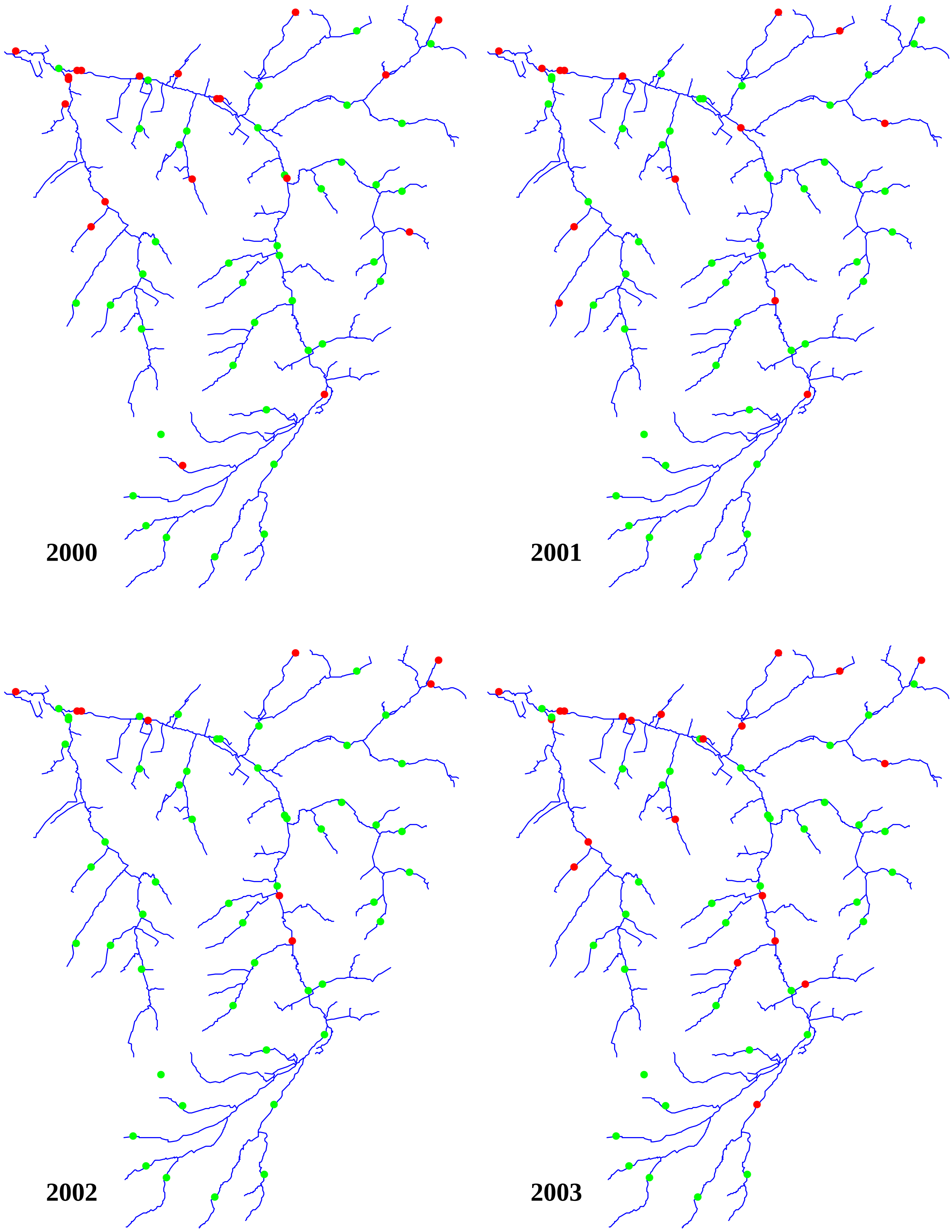
Tubificidae (the Zwalm river basin):



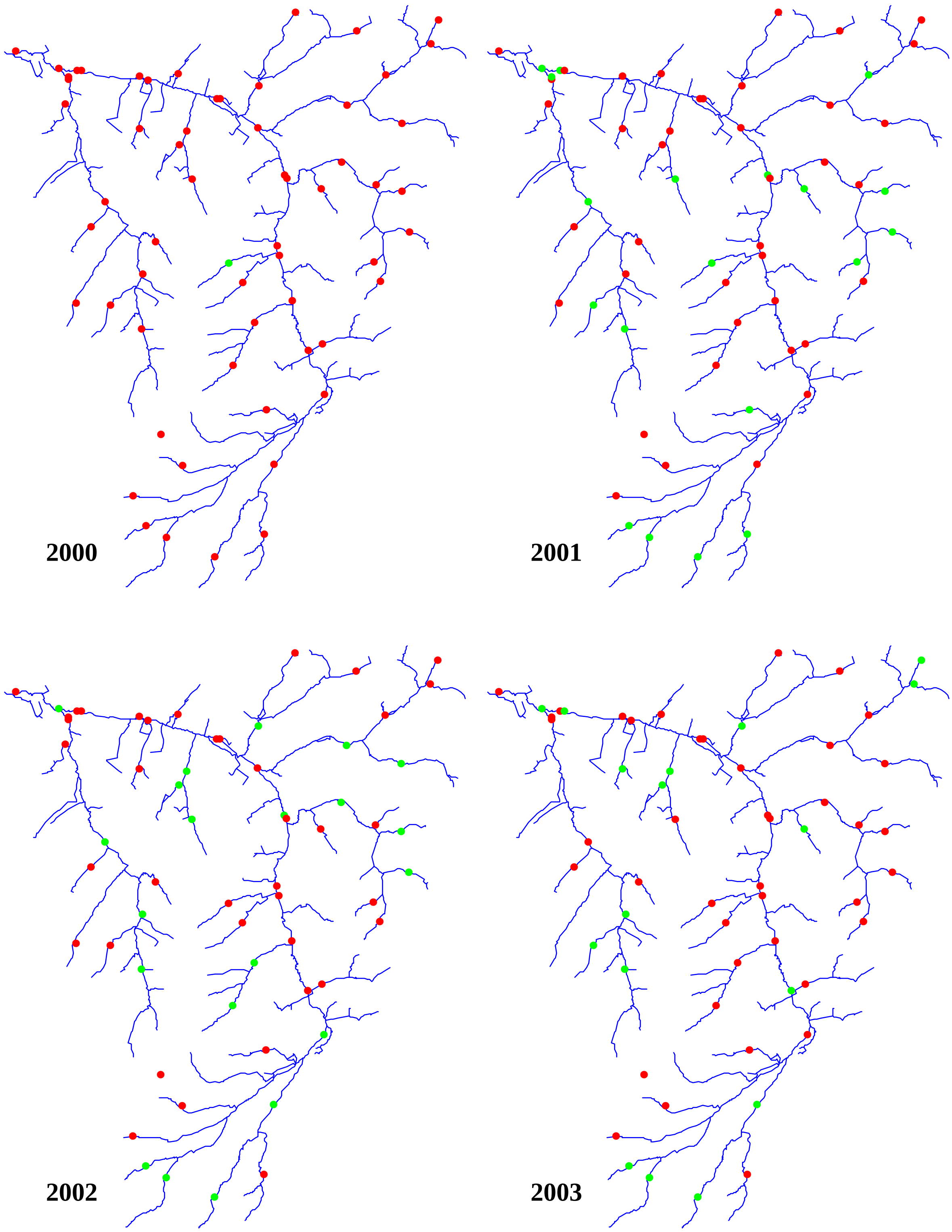
Asellidae (the Zwalm river basin):



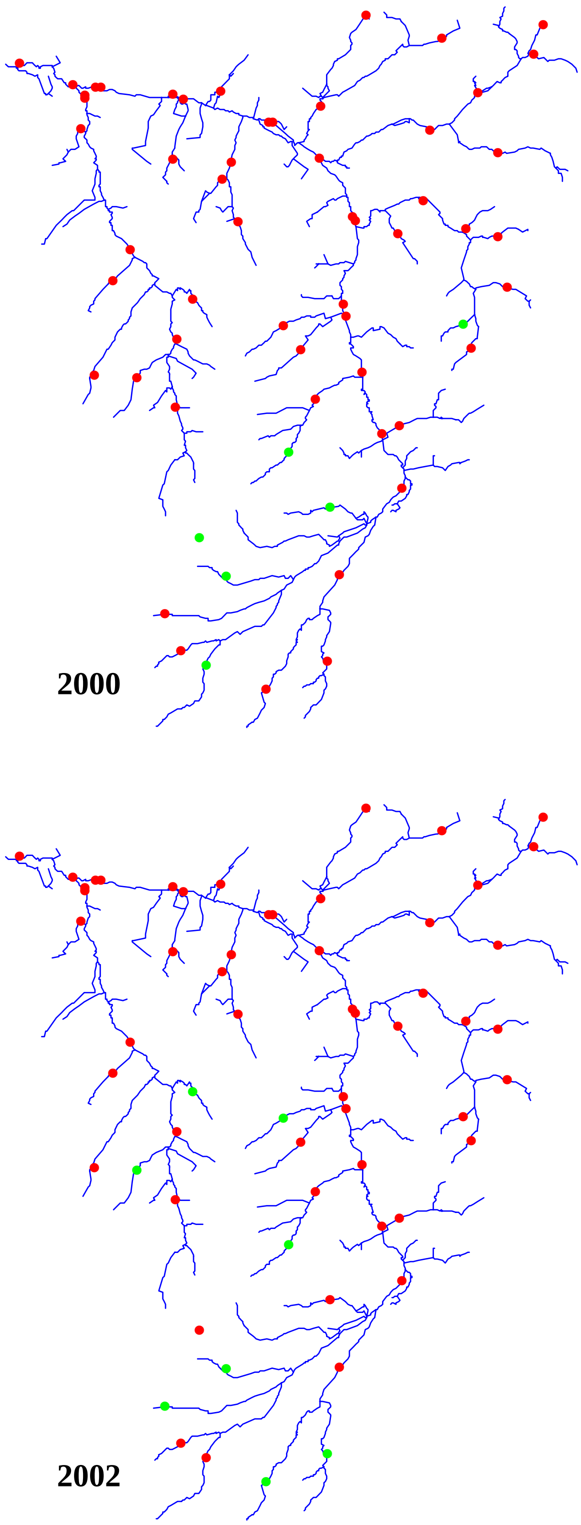
Gammaridae (the Zwalm river basin):



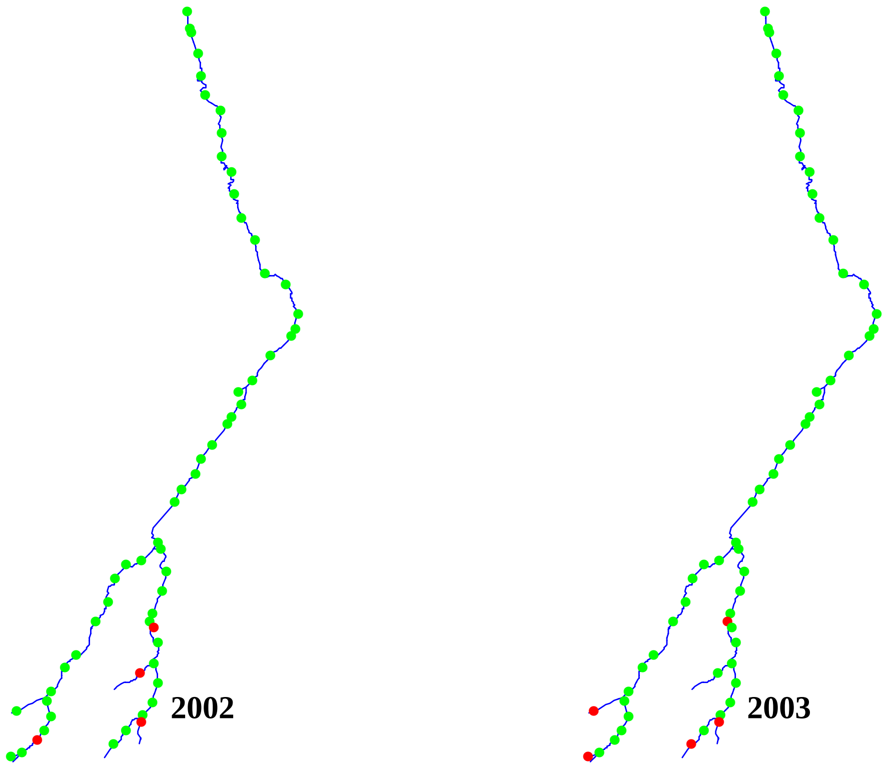
***Baetis* (the Zwalm river basin):**



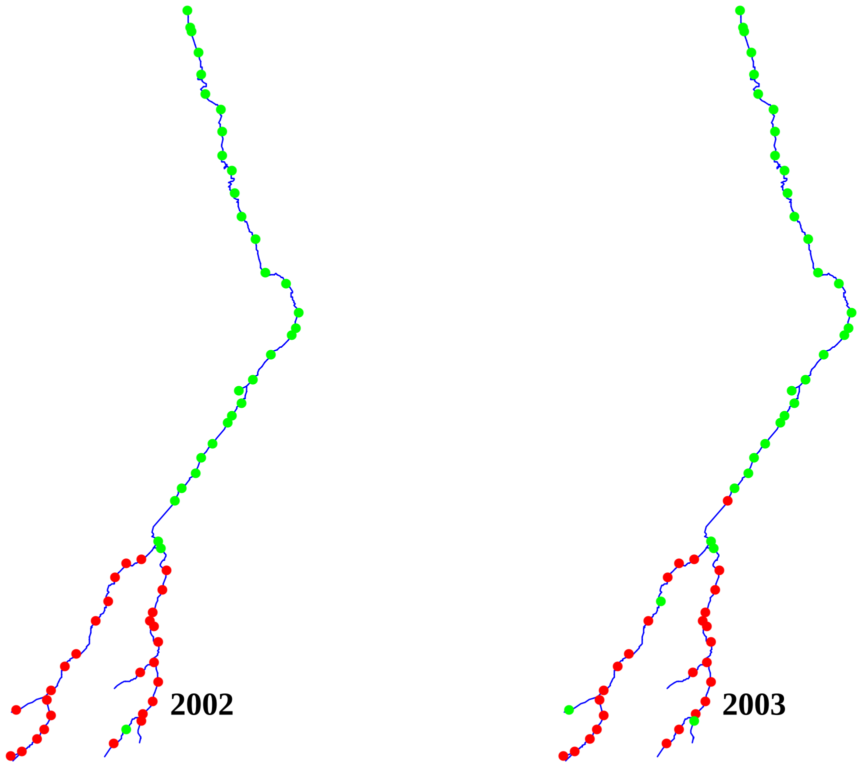
Limnephilidae (the Zwalm river basin):



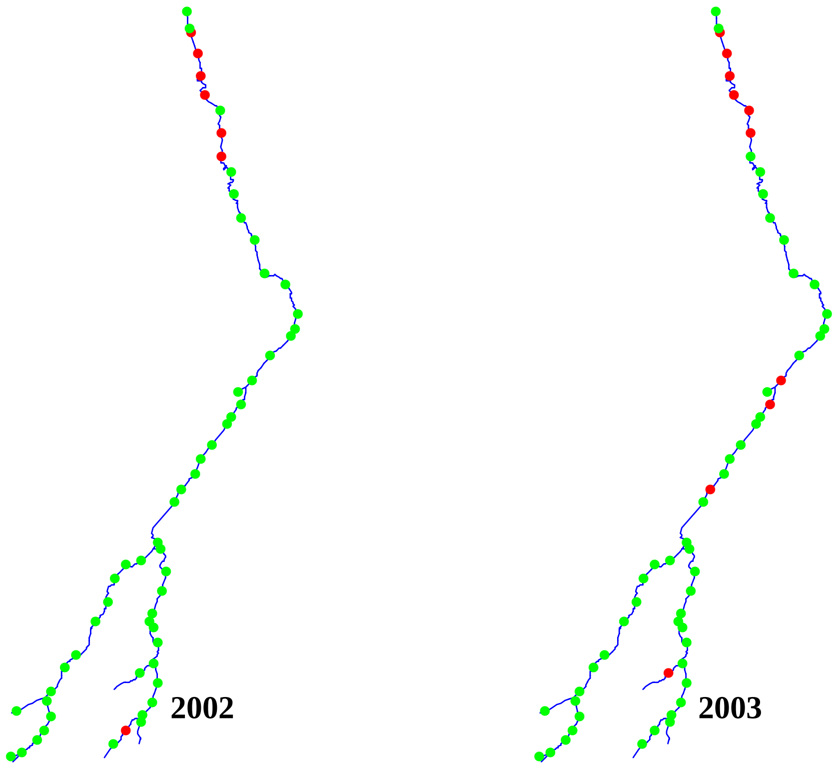
Tubificidae ('short distance' monitoring network):



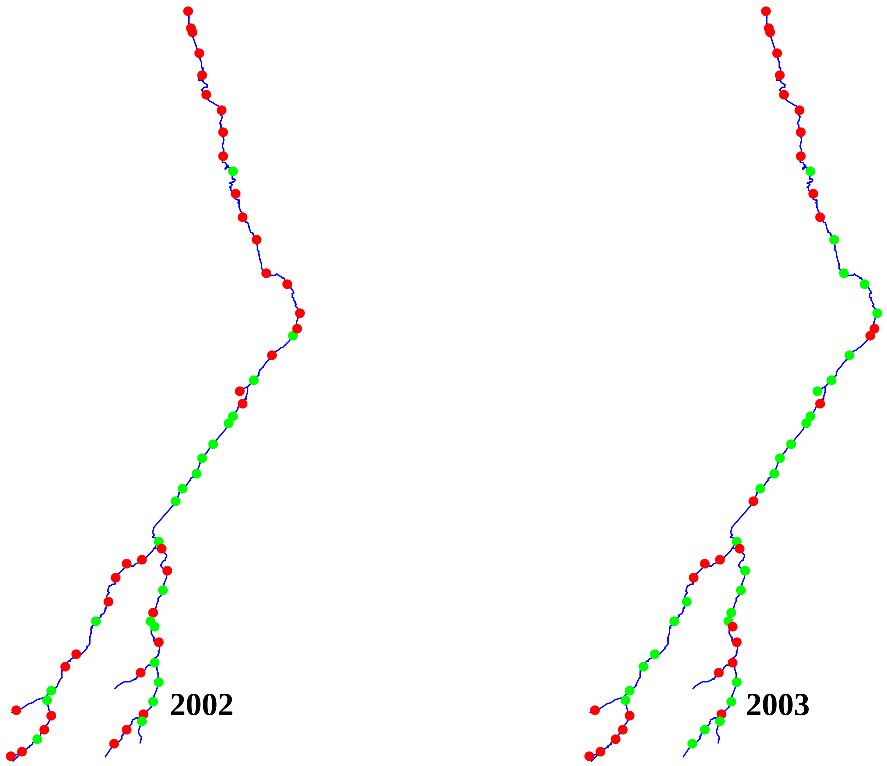
Asellidae ('short distance' monitoring network):



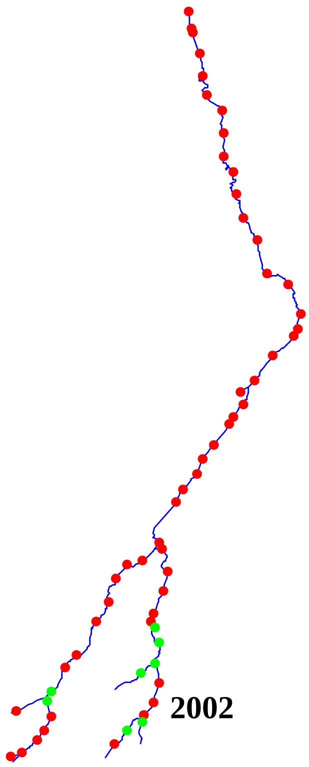
Gammaridae ('short distance' monitoring network):

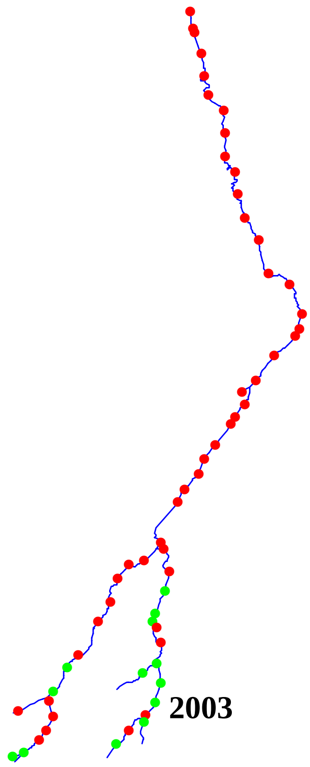


Baetis **('short distance' monitoring network)**:



Limnephilidae (‘short distance’ monitoring network):





ISBN 90-5989-068-X