

# Development of artificial neural network models predicting macroinvertebrate taxa in the river Axios (Northern Greece)

E. Dakou<sup>1</sup>, P.L.M. Goethals<sup>2\*</sup>, T. D'heygere<sup>2</sup>, A.P. Dedecker<sup>2</sup>, W. Gabriels<sup>2</sup>, N. De Pauw<sup>2</sup>, M. Lazaridou-Dimitriadou<sup>1</sup>

<sup>1</sup>Laboratory of Zoology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece

<sup>2</sup>Department of Applied Ecology and Environmental Biology, Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, J. Plateaustraat 22, B-9000 Ghent, Belgium

Artificial Neural Network models (ANNs) were used to predict habitat suitability for 12 macroinvertebrate taxa, using environmental input variables. This modelling technique was applied to a dataset of 102 measurement series collected in 31 sampling sites in the Greek river Axios. The database consisted of seven physical-chemical and seven structural variables, as well as abundances of 90 macroinvertebrate taxa. A seasonal variable was included to allow the description of potential temporal changes in the macroinvertebrate communities. The induced models performed well for predicting habitat suitability of the macroinvertebrate taxa. Sensitivity analyses revealed that dissolved oxygen concentration and the substrate composition always played a crucial role in predicting habitat suitability of the macroinvertebrates. Although ANNs are often referred to as black box prediction techniques, it was demonstrated that ANNs combined with sensitivity analyses can provide insight in the relationship between river conditions and the occurrence of macroinvertebrates, and thus deliver new ecological knowledge. Consequently, these models can be useful in decision-making for river restoration and conservation management.

Keywords : Ecological modelling, environmental quality, habitat suitability, sensitivity analysis, river assessment

## Introduction

In Greece, the ecological deterioration of freshwater is becoming more and more pronounced during the last decades (Dimitriou et al. 2006) due to untreated urban and industrial wastewater discharges and diffuse pollution originating from agricultural activities. In addition, several structural and morphological disturbances are present (e.g. irrigation dams) (Argiropoulos 1991). Consequently, the need for protection and restoration of the ecological quality of freshwater systems becomes urgent. Knowledge on the relationship between environmental factors and freshwater organisms is vital for conservation management and restoration of rivers (Onaindia et al. 2005). Because restoration efforts entail drastic social and economical consequences, the decisions should be taken with enough forethought. Ecosystem models could act as promising tools to support decision-making in river restoration management (Goethals et al. 2006). In particular, models that can predict the

habitat requirements of organisms are of considerable importance to ensure that the planned actions have the desired effects on the aquatic ecosystems (Goethals & De Pauw 2001). In this context, Artificial Neural Networks (ANNs) have been recognized by several authors as valuable tools for predicting macroinvertebrates (e.g. Walley & Fontana 1998, Schleiter et al. 1999, Wagner et al. 2000, Hoang et al. 2001, Park et al. 2001, Arab et al. 2004, Dedecker et al. 2004, 2005a,b, 2006). On the other hand, ANN models have been labelled as 'black boxes' (Olden & Jackson 2002) because they do not provide direct insight in the habitat preferences of the taxa. However, several authors have explored this problem and proposed algorithms to clarify the role of input variables in ANN models (Olden & Jackson 2002, Gevrey et al. 2003, Dedecker et al. 2005a, 2006).

While Dakou et al. (2007) applied rule induction techniques to predict the habitat suitability of several macroinvertebrate taxa, the aim of this paper is to demonstrate the potential and limitations of Artificial Neural Network models to predict the habitat suitability of several benthic macroinvertebrate taxa in the Axios river in Northern

\* Corresponding author : E-mail : peter.goethals@UGent.be

Greece. To this end, the ANN models have been compared to ordinary probabilistic guesses. Additionally, sensitivity analyses have been carried out to examine the importance of the different input variables for the predictions of the macroinvertebrate taxa.

## Materials and Methods

### Study area

The river Axios originates in the Sar mountains of the Former Yugoslavian Republic of Macedonia (FYROM) (Fig. 1). It discharges into the Thermaikos Gulf in northern Greece. Only the last 80 km of the 320 km long river are within the Greek territory. At 49 km from the border with FYROM an irrigation dam (Fragma Ellis) has been constructed which remains closed from May to September. Due to this fact, discharge drops to  $1 \text{ m}^3 \text{ s}^{-1}$  during the dry season (Argiropoulos 1991). Agriculture is the main source of pollution in the watershed. Besides this diffuse pollution, also urban and industrial wastes are being discharged in several places along the river (Fig. 1).

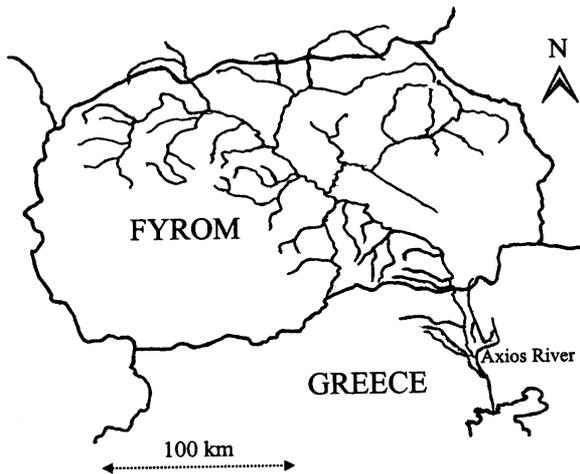


Fig. 1. Location of the Axios river in FYROM and Greece.

### Measurements and database set-up

The database consisted of measurements of 102 samples originating from the Axios River, collected in 31 sites located in the Greek part between 1997 and 2001. In some sites, the measurements were conducted several times a year (from 2 up to 12 measurements) to be able to detect seasonal changes in the macroinvertebrate communities, while other sites were sampled only once between 1997 and 2001. Nineteen variables were mea-

sured during the monitoring campaigns. Further information on the selection of the sampling sites and the sampling methodologies is given by Langrick et al. (1998), Kampa et al. (2000) and Chatzinikolaou (2001, 2002). Of those nineteen variables, fourteen were used as inputs to predict the habitat suitability of the macroinvertebrate taxa. The applied input variables included information on the physical-chemical as well as structural characteristics of the river. Also a variable was introduced to account for seasonal variation in the dataset (Table 1).

Table 1. Input variables used for the prediction of the habitat suitability of twelve macroinvertebrate taxa in the Axios River in Northern Greece

Variables	Units
pH	-
Dissolved Oxygen (DO)	$\text{mg l}^{-1}$
Biological Oxygen Demand ( $\text{BOD}_5$ )	$\text{mg l}^{-1}$
Temperature	$^{\circ}\text{C}$
Conductivity	$\mu\text{S cm}^{-1}$
Total Dissolved Solids (TDS)	$\text{mg l}^{-1}$
Total Suspended Solids (TSS)	$\text{mg l}^{-1}$
Flow velocity	$\text{m s}^{-1}$
Granulometric classification of substrate : boulders, cobbles, pebbles, gravel, sand, silt (%B, %C, %P, %G, %SA, %SI)	%
Season	4 classes (winter, spring, summer, autumn)

Macroinvertebrate samples were taken by means of the three-minute-kick-sweep method (Armitage et al. 1983), using a standard pond net (surface  $575 \text{ cm}^2$ , mesh size  $900 \mu\text{m}$ , depth  $27.5 \text{ cm}$ ). In total, 90 taxa were encountered, twelve of which were selected for modelling their habitat suitability. This selection was based on their frequency of occurrence and on their sensitivity to pollution. The pollution sensitivity grades of the 12 selected taxa are presented in Table 2. These grades were set according to the Greek Biotic Index (100 = high sensitivity, 10 = high tolerance to pollution) (Artemiadou & Lazaridou 2005). In the models, the absence or presence of macroinvertebrate taxa was represented by 0 or 1, respectively.

Table 2. Selected macroinvertebrate taxa used for the modelling approach and their respective pollution sensitivity grades according to the Greek Biotic Index (Artemiadou & Lazaridou 2005)

Taxon	Pollution sensitivity grades
Heptageniidae (Insecta, Ephemeroptera)	80
Limnephilidae (Insecta, Trichoptera)	80
Gomphidae (Insecta, Odonata)	60
Caenidae (Insecta, Ephemeroptera)	50
Gammaridae (Crustacea, Amphipoda)	50
Hydropsychidae (Insecta, Trichoptera)	50
Baetidae (Insecta, Ephemeroptera)	40
Piscicolidae (Hirudinea, Rhynchobdellida)	40
Asellidae (Crustacea, Isopoda)	30
Chironomidae (Insecta, Diptera)	30
Oligochaeta	30
Lymnaeidae (Mollusca, Gastropoda)	20

### Probabilistic guesses

The probabilistic guesses of the presence or absence of the macroinvertebrates were calculated based on the following formula (Dedecker et al. 2002) : Probabilistic guess = [(frequency of presence of the taxon in the validation set) x (probability of predicting at random the presence, based on the training set)] + [(frequency of absence of the taxon in the validation set) x (probability of predicting at random the absence, based on the training set)]

### Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are computational algorithms inspired on the physiology of the human brain. They were initially developed to model biological functions, which are often characterised by a non-linear nature (Lek & Guegan 1999). In this study, a multi-layer feed-forward network trained with backpropagation (BPN) (Rumelhart 1986) was used. This algorithm is based on supervised learning, as the model is constructed

using examples of data with known outputs. The examples used to train the model are assumed to contain all the information needed to establish the relations between input and output data (Lek & Guegan 1999). Before the BPN-training, the connection weights are set to small random values (Lek et al. 2000). Subsequently, the training set of examples is presented to the network. For each input vector, the output vector is calculated by the ANN, 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 (the "target"). Using this error term, the weights and biases are updated in order to decrease the error, so future outputs are more likely to be correct. This procedure is repeated until the errors become small enough (minimal) or a predefined maximum number of iterations is reached. After the training, the ANN can be validated using independent data, representative for the model application. In ANN models several parameters have to be considered, while adjusting any of the parameters usually influences the predictive capability of the network. Therefore, the optimisation and development of different neural networks is necessary to obtain the best model configuration for a given problem (Dedecker et al. 2004). In the present study, the ANNs were built using the machine learning software package WEKA (Witten & Frank 2000). The neural network consisted of one input layer, including 15 neurons (one for each input variable), one hidden layer with 10 neurons and one output layer consisting of 2 neurons (one for each class - presence or absence). The learning rate and momentum were set at 0.2 and the learning time at 300 iterations.

### Sensitivity Analysis

To perform a sensitivity analysis, a special variant of the BPN type, the so-called senso-net (Dapper 1998), was used. Senso-nets include an additional weight for each input neuron representing the relevance (sensitivity) of the corresponding input parameter for the neural network model. These weights are adapted during the training process of the network (Schleiter et al. 1999). By using the senso-net, the variables of major importance for the predictions can be selected because the higher the weight of an input variable is, the higher its importance on the absence or presence of a particular macroinvertebrate taxon.

In the present study the senso-net consisted of one input layer of 15 neurons, two hidden layers with 15 and 10 neurons, respectively, and one output layer with 2 neurons (Fig. 2). The learning rate and momentum were set at 0.2 and the learning time at 300 iterations. For each taxon, 10 senso-net models were built, each using a dif-

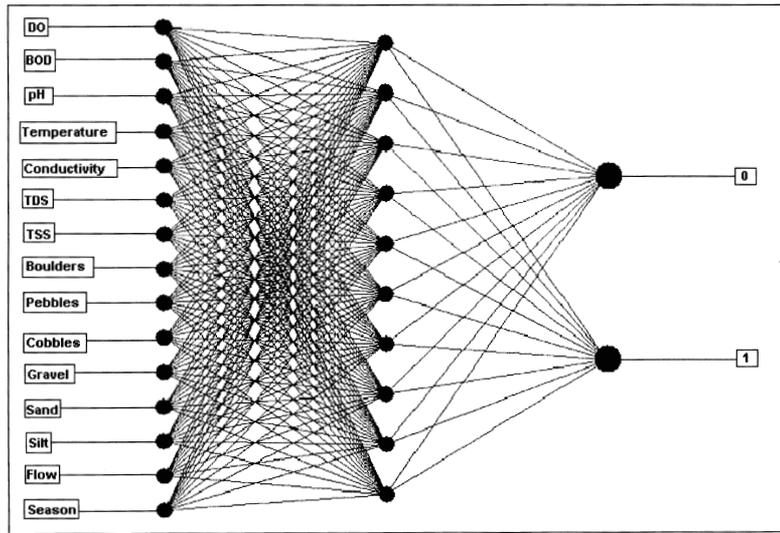


Fig. 2. The senso-net used for the sensitivity analysis.

ferent random number generator for setting the initial weights of the connections between nodes, and also for shuffling the training data. For comparing the relative importance of the input variables in the models, the average weights of these 10 senso-net models were used.

After input variable selection for the prediction of each taxon based on the senso-net weights, an experimental approach was used to unravel information on the impact of the most important input variables on the predictions of the macroinvertebrates. This approach included the induction of new ANN models of reduced complexity, allowing a better understanding of the results. These ANN models included only the five most important input variables for each taxon and consisted of one hidden layer with three neurons, while the learning rate and momentum were set at 0.2 and the learning time at 300 iterations. To determine the response of these models to each individual input variable, a range of variation of a single independent input variable was applied to the model while the other input variables were kept constant. In this manner it was possible to detect the relationship between the input variables and the presence or absence of each taxon (Laë et al. 1999, Dedecker et al. 2004).

**Model training and validation**

The predictive models were assessed based on two performance measures. This required the derivation of confusion matrices from the modelling results that identified

true positive (TP), false positive (FP), false negative (FN) and true negative (TN) cases obtained from each model (Fielding & Bell 1997) (Table 3). In this way, the presence/absence patterns were tabulated against those predicted.

Table 3. The derivation of the confusion matrix used as a basis of performance measures in presence-absence models with true positive (TP), false positive (FP), false negative (FN) and true negative values (TN)

		Actual	
		+	-
Predicted	+	TP	FP
	-	FN	TN

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

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

Another performance measure that was calculated was the Cohen's kappa statistic (Cohen 1960). It is a 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. It is calculated as :

$$Kappa = \frac{[(TP + TN) - (((TP + FN)(TP + FP) + (FP + TN)(FN + TN))/n)]}{[n - (((TP + FN)(TP + FP) + (FP + TN)(FN + TN))/n)]}$$

with  $n$  the total number of cases :  $n = TP+TN+FP+FN$

The training and validation of the ANN models was based on a stratified 10-fold-cross-validation (Kohavi 1995). To allow a reliable error estimate of the models, ten stratified 10-fold-cross validation experiments were conducted, from which the average predictive performance was calculated. These ten-times repeated 10-fold-cross-validation experiments allowed determining the 95 % confidence limit of the average predictive performance, assuming that the data were normally distributed. For the training and validation of the senso-net models the full dataset was used, since there was a special interest in the connection weights between the input layer and the first hidden layer and not their general applicability.

### Results

#### Model development and validation

Models based on the backpropagation algorithm were induced for the prediction of the habitat suitability of twelve macroinvertebrate taxa. The predictive results are presented in Figs 3 and 4. The predictive performance was estimated by the CCI (%) of the repeated ten 10-fold-cross validations with a 95 % confidence interval of the average (Fig. 3). The predictive performance was also estimated by Cohen’s kappa statistic (Fig. 4). The CCI ranged from 66.5 % to 95.9 % for Gammaridae and Chironomidae, respectively. ANN models outperformed ordinary probabilistic guesses in all the cases. However, this result was more obvious in organisms that were moderately present. The highest difference (27.1 %) between the performance of ANNs and probabilistic guesses was found for Gomphidae which had a frequency of occurrence of 52.9 %. For very frequent or rare taxa, the performance of ANN models did not differ significantly from ordinary probabilistic guesses. The predictive performance of the induced models was still high according to Cohen’s kappa statistic (Fig. 4). A value of Cohen’s kappa statistic above 0.3 is generally considered to indicate a fair model performance, while lower values indicate a low model performance. Thus, it can be concluded that most of the models performed reasonably well and that the predictions were not only based on chance. However, in the case of very rare organisms (Limnephilidae, Heptageniidae, Lymnaeidae and Piscicolidae) the predictions were based on pure chance, since Cohen’s kappa was always lower than 0.3. In these cases the organisms were almost always predicted as absent.

#### Sensitivity analysis

In order to investigate the importance of the input variables on the predictions of the macroinvertebrate taxa, a

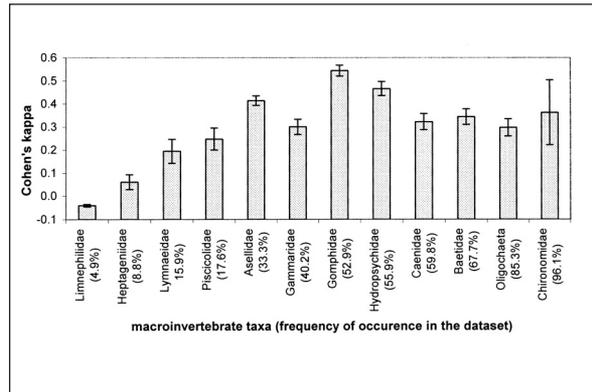


Fig. 4. Predictive performance of habitat suitability models of twelve taxa based on ANNs, when estimated by Cohen’s kappa statistic.

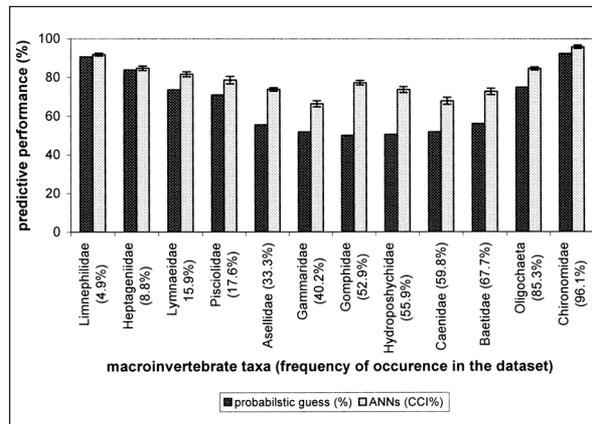


Fig. 3. Comparison of the predictive performance of habitat suitability models of twelve taxa based on ANNs, when estimated by the CCI (%) and an ordinary probabilistic guess.

sensio-net was used. For this purpose, three taxa were selected (Asellidae, Gammaridae and Gomphidae). The habitat suitability models for these taxa had a high or relatively high predictive performance when estimated by the CCI %, and proved to be reliable models since their Cohen’s kappa statistic was high. The input variables that were selected to predict Asellidae, Gammaridae and Gomphidae, and the weights of their neurons, indicating their importance, are presented in Table 4. This table suggests that the concentration of dissolved oxygen was the most important input variable for the prediction of the three taxa. Other dominant input variables are the concentration of total suspended solids, season and substrate composition.

To unravel information on the impact of each of the dominant input variables on the predictions of the

Table 4. Absolute values of the weights of the input variables in a senso-net for the prediction of three taxa in the Axios river (Northern Greece), ranked in descending order

Asellidae		Gammaridae		Gomphidae	
Input Variables	Weight	Input Variables	Weight	Input Variables	Weight
DO	5.98	DO	6.20	DO	6.65
TSS	4.46	TSS	3.21	Silt	6.39
Season	4.01	Season	2.93	Temperature	5.23
Pebbles	3.55	pH	2.64	Cobbles	3.55
Cobbles	3.27	Silt	2.25	TSS	3.35
pH	2.56	Sand	2.06	Season	3.32
Silt	2.23	BOD <sub>5</sub>	2.03	Flow velocity	3.18
BOD <sub>5</sub>	1.36	Pebbles	1.17	Pebbles	2.65
Sand	1.20	Cobbles	0.94	Conductivity	2.39
Temperature	0.96	Conductivity	0.72	TDS	1.81
Flow velocity	0.60	TDS	0.68	Gravel	1.78
Conductivity	0.57	Gravel	0.64	pH	1.68
TDS	0.49	Boulders	0.39	BOD <sub>5</sub>	0.42
Gravel	0.37	Flow velocity	0.18	Sand	0.30
Boulders	0.29	Temperature	0.12	Boulders	0.29

macroinvertebrates, a sensitivity analysis was used. This approach included the construction of ANN models by using only the five most important input variables for the prediction of each taxon. Each of these independent variables varied over their whole range of minimum and maximum measurements, while the rest of the variables were kept constant. The values of the variables that were kept constant were actual measurements at single sampling sites where the predicted taxon was encountered, to avoid that one of these variables would exclude the presence of the taxon. For each model, several measurements of different sampling sites were introduced. In this way, reliable models that were characterised by high Cohen's kappa statistic values (e.g. Gomphidae) allowed to get a very clear insight in the relationship between the input variables and the predictions of the macroinvertebrates. Fig. 5 suggests that the probability of presence of Gomphidae was very low when dissolved oxygen concentration was low. The modelling approach revealed that a critical amount of 2.5 mg DO l<sup>-1</sup> was required in order to obtain a probability of presence for Gomphidae above 0.5. However, this critical value also depended on the constant values of the other input variables, and altering these values may influence the results.

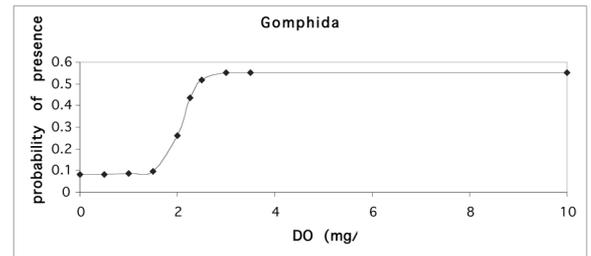


Fig. 5. The impact of dissolved oxygen (mg l<sup>-1</sup>) on the probability of presence of Gomphidae in the Axios river (Northern Greece) (water temperature = 10 °C).

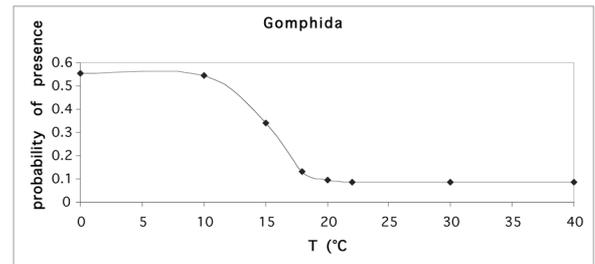


Fig. 6. Impact of water temperature on the probability of presence of Gomphidae in the Axios river (Northern Greece).

Gomphidae preferred low water temperatures (Fig. 6). In this case the critical value of water temperature was 10 °C. This means that the probability of presence of Gomphidae became very low when the water temperature was above 10 °C. It should be noted that for this model dissolved oxygen concentration was set at a constant value of 2.5 mg l<sup>-1</sup>. Thus, it can be concluded that Gomphidae can tolerate relatively low concentrations of dissolved oxygen, when water temperature is low.

If water temperatures increased, the requirements of this taxon for the concentration of dissolved oxygen in the water became higher. When the water temperature was 25 °C (Fig. 7), the critical value of dissolved oxygen concentration which allowed the presence of Gomphidae became 4.5 mg l<sup>-1</sup>.

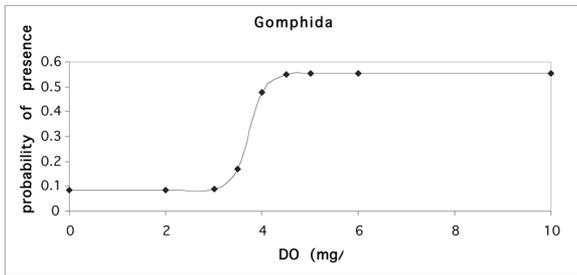


Fig. 7. Impact of dissolved oxygen (mg l<sup>-1</sup>) on the probability of presence of Gomphidae in the Axios river (Northern Greece) (water temperature = 25 °C).

It can be concluded that Gomphidae tended to be present when dissolved oxygen concentration was high and when the water temperature was relatively low. However, it was not possible to define the critical values of these variables that permit the presence of this family in general, as its presence is strongly related to other habitat characteristics (input variables) which were not available in the dataset.

The sensitivity analysis for Gammaridae demonstrated that water temperature had a very small impact on the predictions of this taxon, since the weight of this input variable was the lowest. When a model for the prediction of Gammaridae based on five input variables was constructed, silt, which was the fifth most important input variable in the sensitivity analysis, was replaced by temperature. In this model however, water temperature had a relatively high weight and appeared to have an impact on the prediction of Gammaridae, which was related to the concentration of dissolved oxygen. In Fig. 8 the impact of the concentration of dissolved oxygen on the proba-

bility of presence of Gammaridae in relation to water temperature is presented. In this graph, one can observe that the impact of the concentration of dissolved oxygen is similar as with Gomphidae. Thus, to be present, Gammaridae required a minimum of 5 mg l<sup>-1</sup> of dissolved oxygen when the water temperature was 10 °C and a minimum of 9 mg l<sup>-1</sup> of dissolved oxygen when the water temperature exceeded 25 °C.

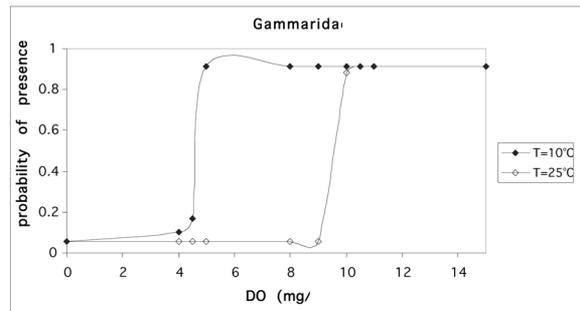


Fig. 8. Impact of dissolved oxygen (mg l<sup>-1</sup>) on the probability of presence of Gammaridae, in relation to water temperature in the Axios river (Northern Greece).

The same experimental approach was used to investigate the substrate preferences of Gomphidae, with respect to the percentage of silt and cobbles. Fig. 9 and Fig. 10 showed that Gomphidae can tolerate a maximum amount of silt in the substrate of 45%, when there are no cobbles in the substrate. However, the amount of silt in the substrate that can be tolerated increased and reached the value of 60% when the percentage of cobbles in the substrate was above 25%.

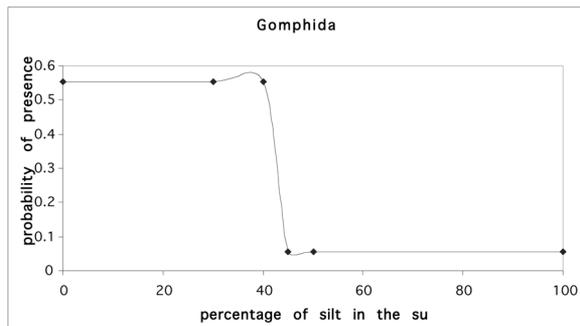


Fig. 9. Impact of the percentage of silt in the substrate on the probability of presence of Gomphidae in the Axios river (Northern Greece) (cobbles = 0%).

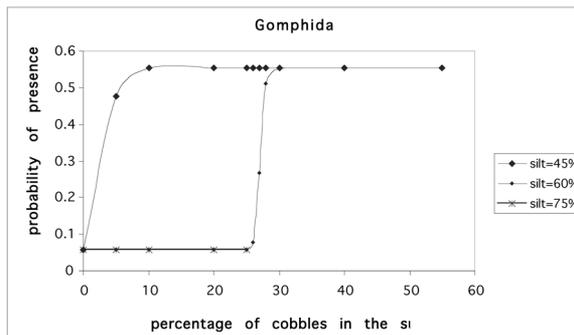


Fig. 10. Impact of the percentage of cobbles in the substrate on the probability of presence of Gomphidae in the Axios river (Northern Greece), in relation to the percentage of silt in the substrate.

## Discussion

### Model development and validation

The performance of the models was higher when predicting very frequent or rare organisms (according to CCI%), compared to the predictions of moderately present organisms. Goethals et al. (2002), Gabriels et al. (2006) and Manel et al. (2001) derived similar conclusions when predicting macroinvertebrate taxa using a limited set of environmental variables. On the other hand, their respective Cohen's kappa statistic had low values indicating that these predictions were based on chance. Consequently, the predictions of very frequent or rare organisms did not differ significantly from ordinary probabilistic guesses. It is thus likely that in these situations, the models fail to recognise patterns in the data. As a consequence, the predictions are based on chance. Thus, the relationship between the frequency of occurrence of the predicted organisms in the dataset and the predictive performance of the models seems to be clear.

### Sensitivity analysis

The sensitivity analysis revealed that the main input variables for the prediction of the habitat suitability were the same for the different macroinvertebrates. The dissolved oxygen concentration played the most important role in nearly all models. Other input variables of major importance were the concentration of total suspended solids and substrate composition. According to Hynes (1970), the concentration of dissolved oxygen is rarely a crucial factor in the ecology of macroinvertebrates in unpolluted rivers. However, a lack of oxygen in polluted rivers can be of major importance. On the other hand,

substrate composition is a very important factor influencing the distribution of benthic macroinvertebrates. Thus, the identification of these key variables in the models indicated a fairly reliable performance. Generally, there was also a tendency to recognise structural and other input variables not related to pollution as important. The ability of the models to distinguish patterns in the dataset relating presence or absence of the macroinvertebrates with environmental variables implies that these variables have a strong ecological impact on the macroinvertebrates. However, it is possible that other input variables also play an important role on the presence or absence of the predicted organisms, but a low variability of the values, or missing values and noise in the dataset did not allow the models to recognise any patterns that demonstrate their importance.

The experimental approach that was used to investigate the impact of each of the major input variables on the predictions of the macroinvertebrates revealed the existence of tendencies for presence or absence among the macroinvertebrates related to the values of the input variables. It was demonstrated that for some input variables the tendencies of the taxa were similar. For some others, they were not. For example, all the selected taxa showed a clear preference for high dissolved oxygen concentrations. This however varied in relation to the constant variables in the models, as well as within models for the prediction of the habitat suitability for different taxa. This technique proved to be a valuable tool for extracting information on the habitat preferences or the ecological requirements of the predicted taxa. Additionally, it increased the transparency of the developed models and improved the understanding of the recognised patterns in the models.

The modelling approach for Gomphidae demonstrated that this taxon had a tendency to be present when dissolved oxygen was relatively high and water temperature was low. Moreover, both input variables interacted in a way that within sites with low concentration of dissolved oxygen, a low water temperature was required for Gomphidae to be present, and vice versa. The fact that Gomphidae are very sensitive to pollution, can explain the preference of this taxon for high dissolved oxygen concentrations. Furthermore, according to Hynes (1970) the solubility of oxygen is inversely correlated with temperature. Thus, in the case of high water temperatures, even higher concentrations of dissolved oxygen in the water are required for the presence of Gomphidae. With the same technique, the substrate preferences of Gomphidae were investigated. Gomphidae were present when the percentage of silt did not exceed 45% if there were no cobbles in the substrate. If the percentage of cob-

bles reached 25%, higher amounts of silt, up to 60%, could be tolerated. According to McCafferty (1981), Gomphidae usually burrow in the sand, silt or sometimes gravel of streams where they lie in wait for prey while partially buried. Thus, the percentage of silt in the substrate appears to be highly important. However, according to the habitat suitability model of this species, a variation of the substrate particles is preferred.

The sensitivity analysis on Gammaridae demonstrated that water temperature did not have a strong impact on the predictions for this taxon. However, in the model in which water temperature was added to the four major input variables, the impact of water temperature on the predictions of Gammaridae was relatively high. This implies that the input variables that were the most important were also preferred for the pattern recognition by the model, but at the same time other input variables were also able to offer valuable information. It was observed that dissolved oxygen concentration in relation to water temperature had an impact on the presence or absence of Gammaridae, in such a way that Gammaridae tended to be present at high dissolved oxygen concentrations or low temperatures. The critical values of these input variables showed that Gammaridae were more sensitive to low values of dissolved oxygen, as they required higher values than Gomphidae. However, in addition to the fact that measurements of different sampling sites were used, the critical values which define the presence of the predicted taxa also depended on the constant values of the other input variables in the models. This implies that these critical values do not represent the general ecological requirements of the predicted taxa, but only their requirements in the examined river system. Therefore, the tendencies of the presence of the macroinvertebrate taxa in relation to environmental variables can be used for water management purposes, as the potential effect of alterations in the river on the benthic macroinvertebrates can be predicted. Similar studies have demonstrated that ANNs are powerful modelling tools with a high potential in ecological modelling (Chon et al. 1996, Mastrotrillo et al. 1997, Goethals et al. 2002).

## Conclusions

Our study based on the Axios river in Greece showed that ANN models performed better than ordinary probabilistic guesses, and that ANNs are relevant tools for modelling habitat requirements of macroinvertebrates. Additionally, the sensitivity analysis demonstrated that although ANN models function as a black box technique, they have the potential to offer insight into the relationships between input variables and the predicted output.

Due to the multivariate interaction between variables, it was not possible to determine fixed critical values that define the presence or absence of macroinvertebrates, but key variables and trends can be derived, thus delivering relevant information to support decision-making in river restoration and conservation management.

## References

- Arab A., Lek S., Lounaci A., & Park Y.S. 2004. – Spatial and temporal patterns of benthic invertebrate communities in an intermittent river (North Africa). *Ann. Limnol. – Int. J. Lim.*, 40, 317-327.
- Argiropoulos D. 1991. Axios River Basin, Water quality management. Report for the Ministry of Housing, Physical Planning and Environment of Greece. Delft hydraulics, Athens, Greece. (in Greek).
- Armitage P.D., Moss D., Wright J.F. & 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. *Wat. Res.*, 17, 333-347.
- Artemiadou V. & Lazaridou M. 2005. Evaluation Score and Interpretation Index for the ecological quality of running waters in Central and Northern Hellas. *Environ. Monit. Assess.*, 110, 1-40.
- Chatzinikolaou G. 2001. Monitoring the ecological quality of the waters and the habitat structure of the Axios River. M.Sc. thesis, Aristotle University of Thessaloniki, Thessaloniki, Greece. (in Greek).
- Chatzinikolaou G. 2002. Monitoring the quality of the waters and the quality of habitats of the Axios River in Greece and FYROM. Final report of the Project DAC: Transboundary co-operation and actions for the protection and management of the waters of the Axios River, by order of the Ministry of Housing, Physical Planning and Environment of Greece. Laboratory of Zoology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece. (in Greek).
- Chon T.S., Park Y.S., Moon K.H. & Cha E.Y. 1996. Patterning communities by using an artificial neural network. *Ecol. Model.*, 90, 69-78.
- Cohen J. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 20, 37-46.
- Dakou E., D'heygere T., Dedecker A.P., Goethals P.L.M., Lazaridou-Dimitriadou M. & De Pauw N. 2007. Decision tree models for prediction of macroinvertebrate taxa in the river Axios (Northern Greece). *Aquat. Ecol.* (in press).
- Dapper T. 1998. Dimension reducing study on neural networks applied in aquatic ecology. PhD thesis, University of Kassel, Aachen, Germany.
- Dedecker A.P., Goethals P.L.M. & De Pauw N. 2002. Comparison of artificial neural network (ANN) model development methods for prediction of macroinvertebrate communities in the Zwalm river basin in Flanders, Belgium. In Proceedings of the 2<sup>nd</sup> Symposium on European Freshwater Systems. *TheScientificWorldJOURNAL*, 2, 96-104.
- Dedecker A.P., Goethals P.L.M., Gabriels W. & De Pauw N. 2004. Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). *Ecol. Model.*, 174, 161-173.
- Dedecker A.P., Goethals P.L.M., D'heygere T., Gevrey M., Lek S. & De Pauw N. 2006. Habitat preference study of *Asellus* (Crustacea, Isopoda) by applying input variable contribution

- methods to Artificial Neural Network models. *Environ. Model. Assess.* (in press).
- Dedecker A.P., Goethals P.L.M., D'heygere T., Gevrey, M., Lek S. & De Pauw N. 2005a. Application of Artificial Neural Network models to analyse the relationships between *Gammarus pulex* L. (Crustacea, Amphipoda) and river characteristics. *Environ. Monit. Assess.*, 111, 223-241.
- Dedecker A.P., Goethals P.L.M. & De Pauw N. 2005b. Sensitivity and robustness of a stream model based on artificial neural networks for the simulation of different management scenarios. Pages 133-146 in *Modelling community structure in freshwater ecosystems*. Lek S., Scardi M., Verdonshot P.F.M., Descy J.-P. & Park Y.-S. (eds). Springer-Verlag, Heidelberg, Germany.
- Dimitriou E., Karaouzas I., Skoulidakis N. & Zacharias I. 2006. – Assessing the environmental status of Mediterranean temporary ponds in Greece *Ann. Limnol. – Int. J. Lim.*, 42, 33-41.
- Fielding A.H. & Bell J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, 24, 38-49.
- Gabriels W., Goethals P.L.M., Dedecker A.P., Lek S. & De Pauw N. 2006. Analysis of macrobenthic communities in Flanders, Belgium, using a stepwise input variable selection procedure with artificial neural networks. *Aquat. Ecol.* (submitted).
- Gevrey M., Dimopoulos I. & Lek S. 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Model.*, 160, 249-264.
- Goethals P. & De Pauw N. 2001. Development of a concept for integrated river assessment in Flanders, Belgium. *J. Limnol.*, 60, 7-16.
- Goethals P., Dedecker A., Gabriels W. & De Pauw N. 2002. Development and application of predictive river ecosystem models based on classification trees and artificial neural networks. Pages 91-107 in *Ecological Informatics: Understanding ecology by biologically-inspired computation*. Recknagel F. (ed) Springer-Verlag, Berlin.
- Goethals P.L.M., Dedecker A.P., Bouma J.J., François D., Verstraete A. & De Pauw N. 2006. The Water Ecology decision support system (WAECO-DSS) for integrated cost-benefit analyses in river restoration management: case study of the Zwalm river basin (Belgium). *J. Environ. Manage.* (submitted).
- Hoang H., Recknagel F., Marshall J. & Choy S. 2001. Predictive modelling of macroinvertebrate assemblages for stream habitat assessments in Queensland (Australia). *Ecol. Model.*, 195, 195-206.
- Hynes H.B.N. 1970. *The ecology of running waters*. Liverpool University Press, Liverpool, 555 p.
- Kampa E., Artemiadou V. & Lazaridou-Dimitriadou M. 2000. Ecological quality of the River Axios (N. Greece) during spring and summer, 1997. *Belg. J. Zool.*, 130, 23-29.
- Kohavi R. 1995. A study of cross-validation and bootstrap for estimation and model selection. Pages 1137-1143 in *Proceedings of the 14<sup>th</sup> International Joint Conference on Artificial Intelligence*. Mellish C.S. (ed). Morgan Kaufmann Publishers, Montreal.
- Langrick J.M., Artemiadou V., Yfantis G., Lazaridou-Dimitriadou M. & White K.N. 1998. An integrated water quality assessment of the river Axios, Northern Greece. Pages 135-143 in *Proceedings of the International Conference "Protection and Restoration of the Environment IV"*. Halkidiki, Vol 1.
- Laë R., Lek S. & Moreau J. 1999. Predicting fish yield of Afrikan lakes using neural networks *Ecol. Model.*, 120, 325-335.
- Lek S., Belaud A., Dimopoulos I., Lauga J. & Moreau J. 1995. Improved estimation, using neural networks, of the food consumption of fish populations. *Mar. Freshwat. Res.*, 46, 1229-1236.
- Lek S., Belaud A., Baran P., Dimopoulos I. & Delacoste M. 1996a. Role of some environmental variables in trout abundance models using neural networks. *Aquat. Living Resour.*, 9, 23-29.
- Lek S., Delacoste M., Baran P., Dimopoulos I., Lauga J. & Aulagnier S. 1996b. Application of neural networks to modelling nonlinear relationships in ecology. *Ecol. Model.*, 90, 39-52.
- Lek, S. & Guegan, J.F. 1999. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol. Model.*, 120, 65-73.
- Lek S., Giraudel J.L. & Guegan J.F. 2000. Neural Networks: Algorithms and Architectures for Ecologists and Evolutionary Ecologists. Pages 3-27 in *Artificial Neural Networks: Application to Ecology and Evolution*. Lek S. & Guegan J.F. (eds). Springer -Verlag, Berlin.
- Manel S., Williams H.C. & Ormerod S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.*, 38, 921-931.
- Mastrorillo S., Lek S., Bauba F. & Belaud A. 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwat. Biol.*, 38, 237-246.
- McCafferty W.P. 1981. *Aquatic Entomology: The fisherman's and ecologist's illustrated guide to insects and their relatives*. Jones and Bartlett publishers, Boston, 448 p.
- Olden J.D. & Jackson D.A. 2002. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Model.*, 154, 135-150.
- Onaindia M., Amezaga I., Garbisu C. & Garcia-Bikuna B. 2005. – Aquatic macrophytes as biological indicators of environmental conditions of rivers in north-eastern Spain. *Ann. Limnol. – Int. J. Lim.*, 41, 175-182
- Park Y.S., Kwak I.S., Chon T.S., Kim J.K. & 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. *Ecol. Model.*, 146, 143-157.
- Rumelhart D.E., Hinton G.E. & Williams R.J. 1986. Learning representations by back-propagation errors. *Nature*, 323, 533-536.
- Schleiter I.M., Borchardt D., Wagner R., Dapper T., Schmidt K.D., Schmidt H.H. & Werner H. 1999. Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks. *Ecol. Model.*, 120, 271-286.
- Wagner R., Dapper T. & 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.
- Walley W.J. & Fontama V.N. 1998. Neural network predictors of average score per taxon and number of families at unpolluted river sites in Great Britain. *Wat. Res.*, 32, 613-622.
- Witten I.H. & Frank E. 2000. *Data mining: practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann Publishers, San Francisco, 369 p.