ELSEVIER

Ecological Modelling 120 (1999) 299–311

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecomodel

The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake

Sébastien Brosse [a,*], Jean-François Guegan [b], Jean-Nöel Tourenq [a], Sovan Lek [a]

[a] *CNRS, UMR 5576 CESAC, Université Paul Sabatier, 118 Route de Narbonne 31062 Toulouse cedex, France*
[b] *Centre de recherche IRD de Montpellier, CEPM/UMR CNRS-IRD 9926, BP 5045, 34032 Montpellier cedex 1, France*

Abstract

The present work describes a comparison of the ability of multiple linear regression (MLR) and artificial neural networks (ANN) to predict fish spatial occupancy and abundance in a mesotrophic reservoir. Models were run and tested with 306 observations obtained by the sampling point abundance method using electrofishing. For each of the 306 samples, the relationships between physical parameters and the abundance and spatial occupancy of various fish species were studied. For the 15 fish species occurring in the lake, six main fish populations were retained to perform comparisons between ANN and MLR models. Each of the six MLR and ANN models had eight independent environmental variables (i.e. depth, distance from the bank, slope of the bottom, flooded vegetation cover, percentage of boulders, percentage of pebbles, percentage of gravel and percentage of mud) and one dependent variable (fish density for the considered population). To determine the population assemblage, principal component analysis (PCA) was performed on the partial coefficients of the MLR and on the relative contribution of each independent variable of ANN models (determined using Garson's algorithm). The results stress that ANN are more suitable for predicting fish abundance at the population scale than MLR. In the same way, a higher level of ecological complexity, i.e. community scale, was reliably obtained by ANN whereas MLR presented serious shortcomings. These results show that ANN are an appropriate tool for predicting population assemblage in ecology. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Artificial neural networks; Multiple linear regression; Principal component analysis; Population assemblage; Fish ecology; Lake

1. Introduction

Interactions between organisms and their biotic and abiotic environmental characteristics strongly influence the habitat use, the spatial occupancy of species, the proportion of each species within the community and, thus, the community composition and structure (Schoener, 1989; Eklöv, 1997). Modelling and simulation are useful tools to roughly mimic the ecosystem structuration and functioning but their ability to model individual

* Correponding author. Fax: +33-5-61556096.
E-mail address: brosse@cict.fr (S. Brosse)

0304-3800/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0304-3800(99)00110-6

distribution, populations and ecosystems depends on the available modelling techniques and computing power (Giske et al., 1998). For example, Ricker (1975) used correlation analysis to assess the influence of the environment on recruitment using abundance data. Canonical correspondence analysis (ter Braak and Verdonschot, 1995) and multiple least-square regression (Binns and Eiserman, 1979) have frequently been used as qualitative methods to explore the relationships between biological assemblages of species and their habitat preferences. The MLR method is now a statistical tool which is used in routine in ecology, but it suffers from some drawbacks in that the relationships between variables in environmental sciences are often non-linear (James and McCulloch, 1990), while the method used is based on linear principles. Transformation of non-linear variables by logarithmic, power or exponential functions can appreciably improve the results, but have often failed to fit data (Lek et al., 1996b). The artificial neural network (ANN), with the error back-propagation procedure, is at the origin of an interesting approach comparable with regression analysis, but particularly efficient for non-linear data (Rumelhart et al., 1986). Up to now, ANN have been used in ecology for modelling phytoplankton production (Scardi, 1996), fish species richness prediction (Guegan et al., 1998), and prediction of density and biomass of various fish populations (Baran et al., 1996; Lek et al., 1996a,b; Mastrorillo et al., 1997). Nevertheless, ANN have scarcely been applied at the community scale, and the work of Tan and Smeins (1996) is probably the only study at this scale which used ANN performance to predict grassland community changes. Moreover, their work only predicted the density of each species taken one by one, and did not deal with the existence of interactions between species.

The aim of the present study is to model the spatial distribution and abundance of six fish populations according to measurable environmental characteristics. Here, we use two distinct modelling methods and we compare their respective capacities to fit observed patterns: (1) multiple linear regression (MLR); (2) artificial neural networks (ANN). Then we quantified the influence of the eight environmental variables on the spatial distribution and habitat use of each population, leading to an approach of the spatial assemblage of the six fish populations studied.

2. Materials and methods

2.1. Study site and sampling

Lake Pareloup is located in the southwest of France, near the city of Rodez. It covers a total surface area of 1350 ha for a volume of about 168 10^6 m^3. The maximum depth is 37 m and the average depth is 12.5 m. It is a warm monomictic lake, which therefore undergoes a summer thermal stratification, with a low oxygen content below the thermocline (located at about 10 m depth from early June to mid-September) preventing the fish from colonising deep water during this period. Fish sampling was performed weekly from late June to late August in a restricted littoral zone of the lake providing a wide range of topographical characteristics. Point abundance sampling by electrofishing (Nelva et al., 1979) modified for young fish (Copp, 1989) was employed to evaluate the microhabitat of the main fish populations. Each week, 30–40 sampling points were investigated in the same area of the lake. For each of the resulting 306 sampling points, nine habitat variables were taken into account: distance from the bank (DIS) in metres, depth (DEP) in metres, local slope of the bottom at each sampling point (SLO) expressed in four classes from zero (nil slope) to three (sheer slope), percentage of flooded vegetation cover (VEG) and percentages of five substrata: boulders (BOU), pebbles (PEB), gravel (GRA), sand (SAN) and mud (MUD). Fishes collected were preserved in 4% formaldehyde solution. In the laboratory, 0+ roach (*Rutilus rutilus*, L. 1758), 0+ perch (*Perca fluviatilis*, L. 1758), 0+ rudd (*Scardinius erythrophthalamus*, L. 1758), 0+ gudgeon (*Gobio gobio*, L. 1758), 0+ pike (*Esox lucius*, L. 1758) and adult perch were identified and numbered for each sampling point.

2.2. Modelling techniques

Modelling was carried out after $\log_{10}(x+1)$ transformation of the dependent variables. This transformation was applied to avoid an undue influence of outliers on the models (ter Braak and Looman, 1995). The Pearson correlation matrix showed a strong correlation between SAN and MUD ($r = -0.98$) and therefore, the variable SAN was removed from the data matrix in order to deal with colinearity. MLR and ANN models were set up using the same dataset (i.e. 306 samples × (eight environmental variables + six fish populations)) with the aim of comparing the two methods.

For MLR, models were set up using all the variables simultaneously. Calculations were done using SPSS software (Norusis, 1993). For 0+ pike, which is a top-predator fish with low density, we considered its absence (coded 0) and presence (coded 1). To process these categorical variables, a logistic regression was used to model 0+ pike distribution. For each of the six models, final values of the partial standardised regression coefficients of MLR were retained to define the influence of environmental factors at the population scale. Then, they were used to perform principal component analysis (PCA) in order to assess the spatial occupancy of fish populations within the entire fish assemblage.

For ANN modelling, a multilayer feed-forward neural network was used. The processing elements in the network, called neurons, are arranged in a layered structure. The first layer, called the input layer, connects with the input variables. In our case, it comprises eight input neurons corresponding to the eight environmental variables, respectively. The last layer, called the output layer, comprises a single neuron which corresponds to the dependent variable to be predicted (fish density for the population considered) (Fig. 1). The layer between input and output layers is called the hidden layer. We could have used a single neural network with six output neurons (one for each of the six fish populations), but we preferred to use six networks with the same architecture, each one predicting the abundance of one fish population, as to easily extract from the models the influence of the eight environmental variables on each fish population. The network configuration is approached empirically by testing various possibilities and selecting the solution that provides the best compromise between bias and variance (Geman et al., 1992; Kohavi, 1995). Training the network consists of using a training data set to adjust the connection weights in order to minimise the error between observed and predicted values. This training was performed according to the back-propagation algorithm (Rumelhart et al., 1986). The computational program was written in a Matlab® environment and computed with an Intel Pentium® processor.

The modelling was carried out in two steps: first, model training was performed using the whole data matrix. This step was used to estimate the performance of the ANN to learn data. Second, we used the 'leave-one-out' bootstrap cross-validation test (Efron, 1983; Efron and Tibshirani, 1995), where each sample is left out of the model formulation in turn and predicted once,

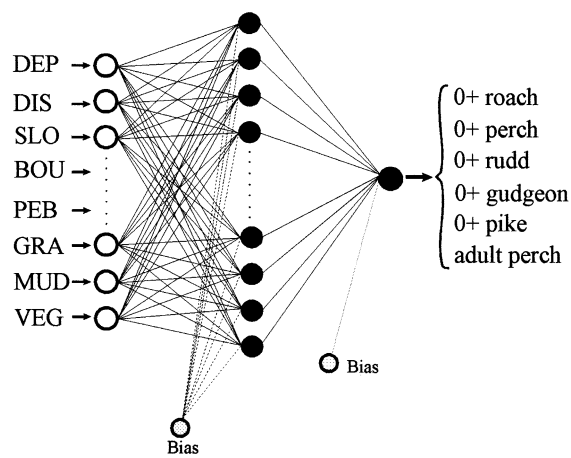

Fig. 1. Typical three-layered feed-forward artificial neural network. Eight input neurons corresponding to eight independent environmental variables (DEP = depth, SLO = slope, DIS = distance from the bank, BOU = boulders, PEB = pebbles, GRA = gravel, MUD = mud, VEG = flooded vegetation), ten hidden layer neurons and one output neuron for estimating one fish population density. Each of the six fish populations was predicted in turn. Connections between neurons are shown by solid lines: they are associated to synaptic weights that are adjusted during the training procedure. The bias neurons are also shown; their input value is one.

to validate the models. This procedure is appropriate when the amount of data is quite small and/or when each sample is likely to have 'unique information' (Efron and Tibshirani, 1995; Kohavi, 1995). This step allows the prediction capabilities of the network to be assessed.

One disadvantage of ANN is their lack of explanatory power. Classical analyses, like MLR, can identify the contribution each independent variable (i.e. input) has on the dependent variable (i.e. output) and can also give some measures of confidence about the estimated coefficients. On the other hand, currently, there is no theoretical or practical way of accurately interpreting the weights attributed in ANN. For example, weights cannot be interpreted as regression coefficients. Therefore, ANN are generally better suited for forecasting or prediction than for explanatory analysis. Some authors have proposed methods for interpreting neural network connection weights to illustrate the importance of explanatory variables in the ANN (Garson, 1991; Dimopoulos et al., 1995; Goh, 1995; Lek et al., 1996a,b). These studies have demonstrated the potential of ANN approaches to explain non-linear interactions between variables in complex systems, and have proposed a procedure for partitioning the connection weights to determine the relative importance of the various input variables. In the present work, Garson's algorithm (Garson, 1991), modified by Goh (1995), was used to determine the influence of the environmental variables. Ten models were set up for each of the six fish populations studied. Then the influence of environmental variables was defined for the ten models and used to assess the spatial distribution of the six populations within the entire community using PCA. In this case, each model was considered as a statistical unit. Thus, PCA was performed on a data matrix containing 60 units (ten units per population × six populations) and the eight environmental variables. Finally, to separate fish population spatial occupancy within the community, cluster analysis was performed on the PCA results using the coordinates of the 60 units on the first two PCA axes.

3. Results and discussion

3.1. Performance of the models

3.1.1. Multiple linear regression models

Examination of Fig. 2 shows some pitfalls which may exist when developing MLR models. Two of the six models were not significant to fit the relationships between fish density and the eight environmental variables: 0+ pike ($r = 0.15$, $P = 0.54$) and adult perch ($r = 0.19$, $P = 0.22$). In both these models, the predicted values showed only nil or close-to-nil values (except one point for adult perch) (see Fig. 2). Overall, we obtained 94% of correct performance estimated using a performance index (PI), based on the proportion of responses within plus or minus 10% of the actual value, but samples with fish were never well-predicted. For the four significant models, correlation coefficients were quite low. Only two models gave a correlation coefficient higher than 0.5 (0.59 for 0+ rudd and 0.70 for 0+ gudgeon), furthermore, the best of these two coefficients was biased as this high value was due to only one non-nil sample well-predicted. Moreover, for the six models, most of the high values of fish abundance were always underestimated and some low predicted values were aberrant, i.e. negative fish densities. The points were not well-distributed along the line of perfect prediction (coordinates 1:1). The residuals tended to increase with estimated values, and their distribution was far from normal. To determine the optimal predictive capacity of traditional methods, we used a non-parametric regression technique: generalized additive models (GAM) (Hastie and Tibshirani, 1990), using the locally-weighted smoother of Cleveland (1979) currently called 'lowess', were set up for the six populations. With this method, the six models were significant ($P < 0.01$) and we obtained a clear improvement of the correlation coefficients: $r = 0.54$ for 0+ roach, $r = 0.38$ for 0+ perch, $r = 0.74$ for 0+ rudd, $r = 0.74$ for 0+ gudgeon, $r = 0.27$ for 0+ pike and $r = 0.37$ for adult perch. These improvements of the quality of the model's predictions testifies to the non-linear behaviour of the relationships between dependent (i.e. fish populations) and independent (i.e. envi-

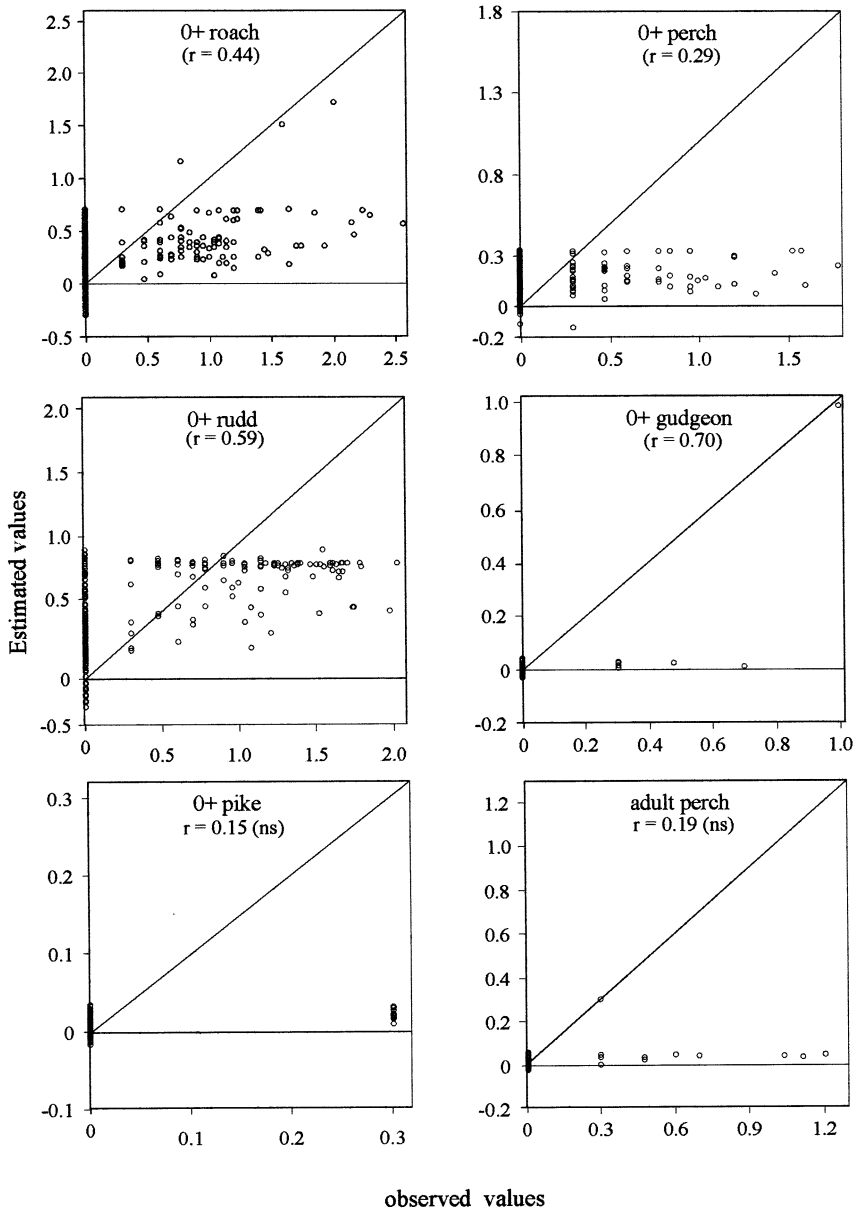

Fig. 2. Recognition performance of the MLR models for the six fish populations. Scatter plots of predicted values vs. observed values. The solid line indicates the perfect fit line of prediction (coordinates 1:1).

ronmental variables) variables. In addition, it justifies the use of ANN, which are known to be able to deal with non-linear relationships between dependent and independent variables when compared with classical MLR methods.

3.1.2. Artificial neural network models

The ANN structure used was a three-layered ($8 \rightarrow 10 \rightarrow 1$) feed-forward network with bias (Fig. 1). There were eight input neurons to code the eight different independent variables. The hidden

layer had ten neurons, determined as the optimal configuration giving the lowest error in the training and testing sets of data with minimal computing time (Geman et al., 1992; Lek et al., 1996b,c). The output neuron computed the value of the dependent variable (fish density). We thus had a total of 101 parameters: (eight input neurons × ten hidden neurons) + (ten hidden neurons × one output neuron) + 11 bias parameters.

The ANN with back-propagation gave much higher correlation coefficients between observed and predicted values (Fig. 3) than MLR. Fig. 3 shows that both low and high values of fish densities were well-predicted even for scarce pop-

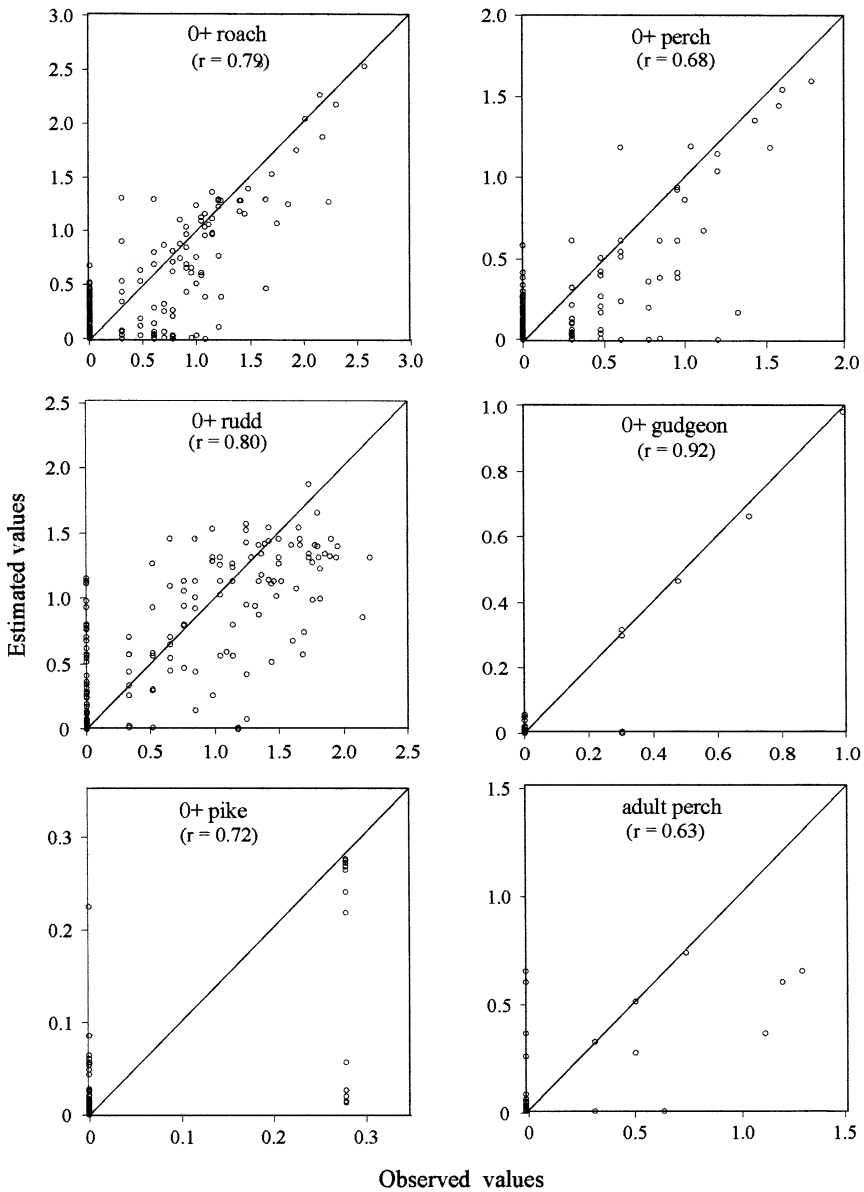

Fig. 3. Recognition performance of the ANN models for the six fish populations. Scatter plots of predicted values vs. observed values in the training procedure. The solid line indicates the perfect fit line of prediction (coordinates 1:1).

Table 1
Performance index (PI) and sum of squared errors (SSE) in ANN training and testing and in MLR training for the six populations[a]

	ANN				MLR	
	Training		Testing		Training	
	PI	SSE	PI	SSE	PI	SSE
0+ Roach	66	7.32	63	11.98	50	70.74
0+ Perch	72	6.23	69	14.02	65	29.46
0+ Rudd	69	9.26	61	13.46	46	65.27
0+ Gudgeon	97	0.66	96	2.65	98	1.04
0+ Pike	90	2.26	91	1.80	90	5.50
Adult Perch	94	7.74	91	8.25	95	1.18

[a] PI is the percentage of well-predicted values with an error rate lower than 10%.

ulations such as for 0+ gudgeon, for 0+ pike and for adult perch. For these three fish populations, non-nil values were rarely predicted as nil values by the network (only two samples for 0+ gudgeon and for adult perch) and a large proportion of the high values were well or perfectly predicted. For 0+ roach, 0+ rudd and 0+ gudgeon, points were well-distributed along the diagonal of best fit. 0+ perch, adult perch and 0+ pike abundances were underestimated, but the results remained clearly better than those obtained using MLR. Moreover the distribution of residuals was close to normal with a mean value of 0.007 (S.D. = ± 0.152) for 0+ roach, 0.017 (S.D. = ± 0.128) for 0+ perch, 0.006 (S.D. = ± 0.183) for 0+ rudd, -0.004 (S.D. = ± 0.056) for 0+ gudgeon, -0.001 (S.D. = ± 0.172) for 0+ pike and 0.001 (S.D. = ± 0.010) for adult perch.

A cross-validation testing procedure (i.e. leave-one-out bootstrap) was performed to validate the ANN models. Models could have been evaluated using the determination coefficients (r^2) or correlation coefficients (r), but because of the scarcity of high values of fish densities (especially for 0+ gudgeon, 0+ pike and adult perch), we preferred to use performance index (PI) and sum of squared errors (SSE) to assess model prediction performance. The PI was based on the proportion of responses within plus or minus 10% of the actual value.

The PIs obtained after the testing procedure were very close to those obtained after training for each of the six species (Table 1). SSE of the test were low and close to those obtained during the training procedure. MLR gave high PIs due to the abundance of nil values; however the SSE values were clearly higher than for ANN, except for 0+ gudgeon and 0+ perch due to the scarcity of non-nil values. Thus, compared with MLR, ANN gave better results both in training and testing procedures.

3.2. Importance of the environmental variables in population abundance

In MLR, the influence of each variable can be roughly assessed by checking the final values of the partial standardized regression coefficients. Each coefficient of a linear model is the partial derivative of the response of the model with respect to the variable of that coefficient. The standardized coefficients of MLR therefore generally give a way to compare the relative influence of each independent variable on the dependent variable, when all other independent parameters have been kept constant in the models. Table 2 shows the MLR standardised partial coefficients of the eight variables for each population. Few among these coefficients were significant (5 for 0+ roach, 3 for 0+ perch and 0+ rudd, 2 for 0+ gudgeon, 1 for 0+ pike and adult perch). Moreover, three of the eight variables are usually considered as essential for 0+ fish microhabitat choice: distance from the bank (DIS), depth

(DEP), and flooded vegetation (VEG), but MLR considered only distance from the bank (DIS) as significant (except for adult perch). MLR shows that 0+ fish abundance was significantly correlated to low values of distance from the bank (DIS) (i.e. negative coefficients), this is in accordance with ecological studies (Haberlehner, 1988; Copp, 1992). Nevertheless, according to MLR models, 0+ roach and 0+ gudgeon abundance increase with depth (DEP), which seems illogical, as deep littoral areas are usually avoided by 0+ fish. Finally, the flooded vegetation (VEG) was never considered as a significant variable, whereas it is logically one of the most important variables for 0+ fish (Persson and Eklöv, 1995; Eklöv, 1997).

For ANN, the results of Garson's algorithm stress the importance of environmental variables in the model (Fig. 4). Standard errors calculated for each variable after ten training procedures were very low, showing the stability of the network models. The contribution of each environmental variable to the model for the six populations was in accordance with previous ecological studies (Holland and Huston, 1984; Haberlehner, 1988; Copp, 1992; Mastrorillo et al., 1996): 0+ roach, 0+ perch, 0+ rudd and 0+ pike are closely linked to the flooded vegetation (VEG) and the distance from the bank (DIS) whereas 0+ gudgeon is indifferent to the flooded vegetation (VEG) but strongly influenced by the distance from the bank (DIS). Finally, adult perch habitat is known to be largely governed by the depth (DEP) and the distance from the bank (DIS) (Persson, 1983; Persson and Eklöv, 1995). Moreover, fish microhabitat is defined by several variables showing that microhabitat results from a complex combination of different habitat characteristics (only 0+ gudgeon show a quite simple diagram, with only one important variable, the distance from the bank (DIS), which contributes more than 50%). The main processes that determine fish habitat and distribution can be approximated by linear functions only to a limited extent. Even when simple (e.g. logarithmic) transformations of variables to linearize their distribution are used, the results remain unsatisfactory. The use of complex transformations of the variables (e.g. GAM) improves the results, but they remain lower than those obtained by ANN. On the other hand, ANN with only one hidden layer can model non-linear systems in ecology without complex transformations of the data (Goh, 1995; Lek et al., 1996b; Scardi, 1996). The microhabitat of the six fish populations studied here was reliably fitted by ANN to the measured environmental characteristics of the points sampled in the lake. The ANN models clearly show the influence of each variable on the microhabitat of each population whereas MLR gives aberrant values which are irrelevant from an ecological point of view. Thus, MLR models are unable to represent ecological reality due to non-linear relationships such as those which probably exist between the densities

Table 2
MLR partial standardised coefficients for the six fish populations studied[a]

	0+ Roach	0+ Perch	0+ Rudd	0+ Gudgeon	0+ Pike	Adult perch
DEP	0.117*	−0.050	−0.011	0.689**	−0.016	0.003
DIS	−0.418**	−0.246**	−0.167**	−0.213**	−0.126*	0.008
SLO	0.054	0.180**	−0.294**	0.033	−0.033	0.075
BOU	−0.033	−0.019	−0.073	−0.013	−0.014	0.123*
PEB	−0.112*	−0.061	−0.031	−0.019	−0.018	−0.021
GRA	0.184**	−0.047	−0.049	−0.014	−0.013	−0.013
MUD	0.263**	0.144*	−0.256**	−0.073	0.094	0.067
VEG	−0.098	−0.087	0.044	−0.066	0.025	−0.037

[a] Environmental variables were lettered as follows: DEP = depth, SLO = slope, DIS = distance from the bank, BOU = boulders, PEB = pebbles, GRA = gravel, MUD = mud, VEG = flooded vegetation.

* Significant coefficient ($P<0.05$).

** Highly significant coefficient ($P<0.01$).

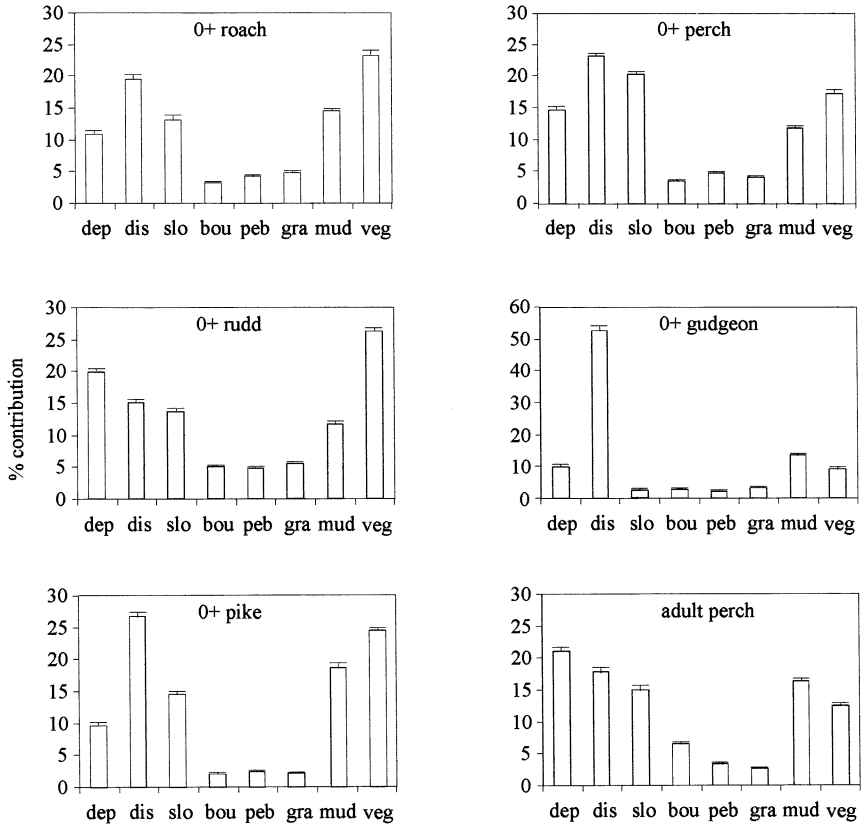

Fig. 4. Percentage contribution of each of the eight independent variables to the prediction of fish densities, obtained by Garson's algorithm (variables lettered as in Fig. 1). Bars indicate the mean value of the results of the ten models for each fish population, horizontal lines represent standard errors of the mean.

of the six fish populations considered and some environmental variables.

3.3. Population assemblage studies

To visualise the spatial distribution of the six fish populations studied within their environment (i.e. population assemblage), on the basis of the information provided by the models, PCA techniques were used.

On its first and second axes which accounted for 49.2 and 29.5% of the total information, respectively (Fig. 5a), the PCA performed on the partial coefficients of the MLR revealed a significant correlation ($P < 0.01$) between distance from the bank (DIS), pebbles (PEB), flooded vegetation (VEG) and 0+ pike and 0+ roach; gravel (GRA), mud (MUD) and 0+ perch and 0+ roach; slope of the bottom (SLO), distance from the bank (DIS), boulders (BOU) and adult perch. We can notice, on the first axis, an opposition between (0+ pike, 0+ rudd) and (0+ roach, 0+ perch). The second axis shows an opposition between adult perch, and 0+ rudd (Fig. 5b). These results based on MLR models conflict with general agreement on habitat use by both 0+ roach and 0+ perch individuals since, during the larval and juvenile periods, they are generally located close to shelters such as flooded vegetation (Haberlehner, 1988; Persson and Eklöv, 1995).

Concerning ANN, the PCA performed on the contribution factors (Goh's algorithm results) allowed the microhabitat of the six fish populations

to be taken into account simultaneously to better define their spatial occupancy and thus to approach the population assemblage. On its first and second axes, which accounted for 43.1 and 20.6% of the total information, respectively (Fig. 6a), the PCA revealed a significant correlation ($p < 0.01$) between flooded vegetation (VEG) and 0+ roach, 0+ rudd and 0+ pike; between depth (DEP) and adult perch; between distance from the bank (DIS) and 0+ gudgeon. We can see, on the first axis an opposition between 0+ gudgeon individuals and the other fish species individuals except for 0+ pike. The second axis shows an opposition between adult perch and the group 0+ roach, 0+ rudd, 0+ pike and 0+ perch (Fig. 6b). The representation of the ten statistical units for each population reveals the range of microhabitat variation for each fish population. Moreover, the cluster analysis distinguishes several groups and enables an approach to be made to the spatial range of microhabitat characteristics for each population (Fig. 6c). The separation of some fish populations such as 0+ gudgeon or for top-predators (i.e. 0+ pike and adult perch) has already been observed in natural environments, and the spatial occurrence of 0+ roach, 0+ rudd and 0+ perch, as illustrated by the cluster analysis, is well-known by ichthyologists. The fish assemblage visualised in the PCA was in accordance with various ecological studies concerning the microhabitat of these species (Persson, 1983; Haberlehner, 1988; Copp, 1992; Hosn and Dowing, 1994; Persson and Eklöv, 1995; Mastrorillo et al., 1996). As a consequence, the fish assemblage was reliably predicted using ANN. This predicted spatial occupancy can be easily visualised on a PCA plane. Thus, ANN are more suitable than MLR to reproduce the operation of real complex multispecies systems (i.e. population assemblage) on the basis of the ecological variables introduced in the model.

4. Conclusion

The back-propagation of ANN constitutes a more efficient tool than MLR to predict fish abundance and spatial occupancy from the environmental characteristics of the littoral area of a lake. The selection of input variables introduced into the modelling procedures, their ecological significance and the constitution of testing sets of

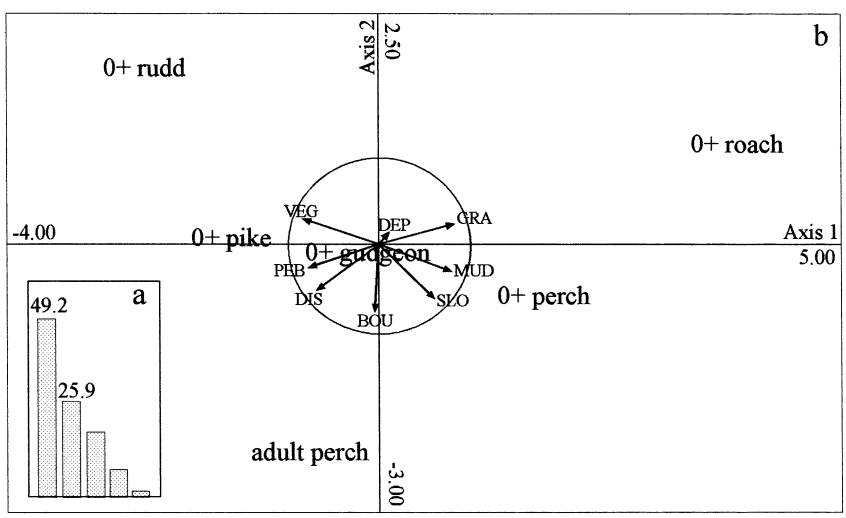

Fig. 5. Principal component analysis (PCA) performed on MLR results using the standardised partial regression coefficients for the six fish populations. (a) Histogram of eigenvalues; (b) distribution of the six samples (i.e. populations) and the eight environmental variables (DEP = depth, SLO = slope, DIS = distance from the bank, BOU = boulders, PEB = pebbles, GRA = gravel, MUD = mud, VEG = flooded vegetation) on the F1 × F2 plane.

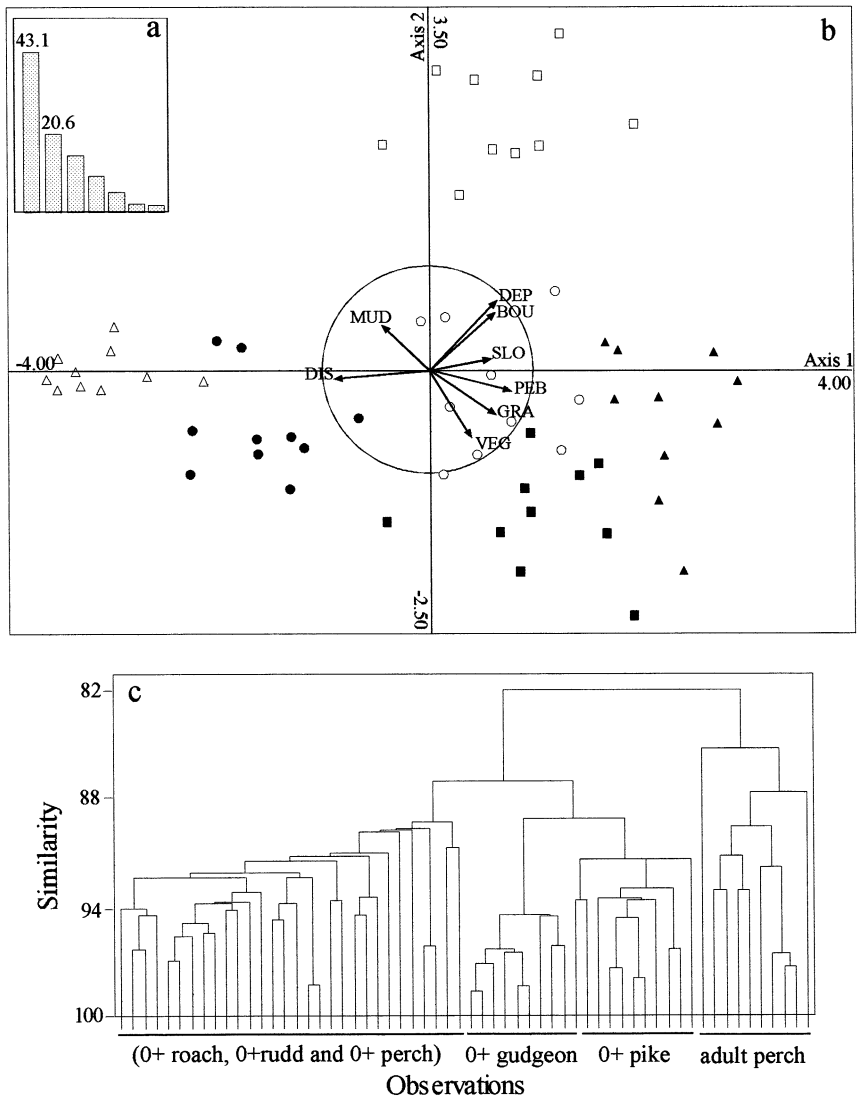

Fig. 6. Principal Component Analysis (PCA) performed on ANN results using Garson's algorithm for the six fish populations. For each population, the statistical units (samples) were the results of the ten ANN models. (a) Histogram of eigenvalues; (b) distribution of the 60 samples and the eight environmental variables ((DEP = depth, SLO = slope, DIS = distance from the bank, BOU = boulders, PEB = pebbles, GRA = gravel, MUD = mud, VEG = flooded vegetation) on the F1 × F2 plane. (□) adult perch; (■) 0+ roach; (o) 0+ perch; (●) 0+ pike; (△) 0+ gudgeon; (▲) 0+ rudd; (c) cluster analysis of the first two coordinates of PCA showing a separation between adult perch, 0+ pike and 0+ gudgeon, the three other populations are dispersed across the similarity gradient.

data to assess the performance of the model are important elements for this type of approach (Faush et al., 1988). The ANN modelling approach used here is a fast and flexible way to incorporate multiple input parameters into a single model. In addition to the predictive value of the model, the combination of ANN and multivariate analysis simultaneously visualise the results provided by several ANN models with the same data matrix at the input. It is this ability to

deal with multiple information sources that provides the power of this approach, resulting in a significant improvement in ANN modelling over conventional techniques. These results on the use of ANN for population assemblage analyses are promising and open new fields for their applications to ecology.

Acknowledgements

The authors are grateful to S. Beker for correcting the English version. The authors thank William Silvert and an anonymous reviewer for helpful discussions on the subject leading to improve the manuscript. This research was supported in part by a doctoral grant (S. Brosse) provided by the French electricity agency (E.D.F.).

References

Baran, P., Lek, S., Delacoste, M., Belaud, A., 1996. Stochastic models that predict trouts population densities or biomass on microhabitat scale. Hydrobiologia 337, 1–9.

Binns, N.A., Eiserman, J.P., 1979. Quantification of fluvial trout habitat in Wyoming. Trans. Am. Fish. Soc. 198, 215–228.

Cleveland, W.S., 1979. Robust locally-weighted regression and scatterplot smoothing. J. Am. Stat. Assoc. 74, 829–836.

Copp, G.H., 1989. Electrofishing for fish larvae and juveniles: equipment modifications for increased efficiency with short fishes. Aquacult. Fish. Manage. 20, 453–462.

Copp, G.H., 1992. Comparative microhabitat use of cyprinid larvae and juveniles in a lotic floodplain channel. Environ. Biol. Fishes 33, 181–193.

Dimopoulos, Y., Bourret, P., Lek, S., 1995. Use of some sensitivity critera for choosing networks with good generalization ability. Neural Process. Lett. 2, 1–4.

Efron, B., 1983. Estimating the error rate of a prediction prediction rule: some improvements on cross-validation. J. Am. Stat. Assoc. 78, 316–331.

Efron, B., Tibshirani, R., 1995. Cross-validation and the Bootstrap: estimating the error rate of a prediction rule. Tech. Rep. 176. Department of statistics, Stanford Univ., 27 pp. ftp://utstat.toronto.edu/pub/tibs/cvboot.ps.

Eklöv, P., 1997. Effects of habitat complexity and prey abundance on the spatial and temporal distributions of perch (*Perca fluviatilis*) and pike (*Esox lucius*). Can. J. Fish. Aquat. Sci. 54, 1520–1531.

Faush, K.D., Hawkes, C.L., Parsons, M.G., 1988. Models that predict the standing crop of stream fish from habitat variables: 1950–85. Gen. Tech. Rep. PNW-GTR-213. U.S. Department of agriculture, Forest service, Pacific north reaserch station, Portland, OR, 52 pp.

Garson, G.D., 1991. Interpreting neural network connection weights. Artif. Intel. Expert. 6, 47–51.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilema. Neural Comput. 4, 1–58.

Giske, J., Huse, G., Fiksen, O., 1998. Modelling spatial dynamics of fish. Rev. Fish. Biol. Fish. 8, 57–91.

Goh, A.T.C., 1995. Back-propagation neural networks for modeling complex systems. Artif. Intel. Eng. 9, 143–151.

Guegan, J.F., Lek, S., Oberdorff, T., 1998. Energy availability and habitat heterogeneity predict global riverine fish diversity. Nature 391, 382–384.

Haberlehner, E., 1988. Comparative analysis of feeding and schooling behaviour of the Cyprinidae *Alburnus alburnus* (L., 1758), *Rutilus rutilus* (L., 1758), and *Scardinius erythrophthalamus* (L. 1758) in a backwater of the Danube near Vienna. Int. Rev. Hydrobiol. 73, 537–546.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized additive models. Chapman and Hall, London, p. 333.

Holland, L.E., Huston, M.L., 1984. Relationship of Young-of-the-Year northern pike to aquatic vegetation types in backwaters of the upper Mississippi river. North Am. J. Fish. Manage. 4, 514–522.

Hosn, W.A., Dowing, J.A., 1994. Influence of cover on the spatial distribution of littoral-zone fishes. Can. J. Fish. Aquat. Sci. 51, 1832–1838.

James, F.C., McCulloch, C.E., 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box? Ann. Rev. Ecol. Syst. 21, 129–166.

Kohavi, R., 1995. A study of the cross-validation and bootstrap for accuracy estimation and model selection. In: Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI), Montreal, pp. 1137–1143

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 Res. 9, 23–29.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagner, S., 1996b. Application of neural networks to modeling non-linear relationships in ecology. Ecol. Model. 90, 39–52.

Lek, S., Dimopoulos, I., Fabre, A., 1996c. Predicting phosphorus concentration and phosphorus load from watershed characteristics using back-propagation neural networks. Acta Oecol. 17, 43–53.

Mastrorillo, S., Dauba, F., Belaud, A., 1996. Utilisation des microhabitats par le vairon, le goujon et la loche franche dans trois rivières du sud-ouest de la France. Ann. Limnol. 32, 185–195.

Mastrorillo, S., Lek, S., Dauba, F., Belaud, A., 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. Freshwater Biol. 38, 237–246.

Nelva, A., Persat, H., Chessel, D., 1979. Une nouvelle méthode d'étude des peuplements ichtyologiques dans les grands cours d'eau par échantillonnage ponctuel d'abondance. C. R. Acad. Sci. Paris Serie III 289, 1295–1298.

Norusis, M.J., 1993. SPSS for Windows. Base system user's guide release 6.0, SPSS Inc., 828 pp.

Persson, L., 1983. Food consumption and competition between age classes in a perch *Perca fluviatilis* population in in a shallow eutrophic lake. Oikos 40, 197–207.

Persson, L., Eklöv, P., 1995. Prey refuges affecting interactions between piscivorous perch and juvenile perch and roach. Ecology 76, 70–81.

Ricker, W.E., 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 191, 1–382.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating error. Nature 323, 533–536.

Scardi, M., 1996. Artificial neural networks as empirical models for estimating phytoplankton production. Mar. Ecol. Prog. Ser. 139, 289–299.

Schoener, T.W., 1989. Food webs from the small to the large. Ecology 70, 1559–1589.

Tan, S.S., Smeins, F.E., 1996. Predicting grassland community changes with an artificial neural network model. Ecol. Model. 84, 91–97.

ter Braak, C.J.F., Looman, C.W.N., 1995. Regression. In: Jongman, R.G.H., ter Braak, C.J.F., Van Tongeren, O.F.R. (Eds.), Data analysis in community and landscape ecology. Cambrige University Press, pp. 29–77.

ter Braak, C.J.F., Verdonschot, F.M., 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. Aquat. Sci. 57, 254–289.