

Modelling roach (*Rutilus rutilus*) microhabitat using linear and nonlinear techniques

SEBASTIEN BROSSE AND SOVAN LEK

CNRS, UMR 5576 CESAC-Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, France

SUMMARY

1. Multiple linear regression (MLR), generalised additive models (GAM) and artificial neural networks (ANN), were used to define young of the year (0+) roach (*Rutilus rutilus*) microhabitat and to predict its abundance.
2. 0+ Roach and nine environmental variables were sampled using point abundance sampling by electrofishing in the littoral area of Lake Pareloup (France) during summer 1997. Eight of these variables were used to set up the models after $\log_{10}(x + 1)$ transformation of the dependent variable (0+ roach density). Model training and testing were performed on independent subsets of the whole data matrix containing 306 records.
3. The predictive quality of the models was estimated using the determination coefficient between observed and estimated values of roach densities. The best models were provided by ANN, with a correlation coefficient (r) of 0.83 in the training procedure and 0.62 in the testing procedure. GAM and MLR gave lower prediction in the training set ($r = 0.53$ for GAM and $r = 0.32$ for MLR) and in the testing set ($r = 0.48$ for GAM and $r = 0.43$ for MLR). In the same way, samples without fish were reliably predicted by ANN whereas GAM and MLR predicted absence unreliably.
4. ANN sensitivity analysis of the eight environmental variables in the models revealed that 0+ roach distribution was mainly influenced by five variables: depth, distance from the bank, local slope of the bottom and percentage of mud and flooded vegetation cover. The nonlinear influence of these variables on 0+ roach distribution was clearly shown using nonparametric lowess smoothing procedures.
5. Non-linear modelling methods, such as GAM and ANN, were able to define 0+ fish microhabitat precisely and to provide insight into 0+ roach distribution and abundance in the littoral zone of a large reservoir. The results showed that in lakes, 0+ roach microhabitat is influenced by a complex combination of several environmental variables acting mainly in a nonlinear way.

Keywords: artificial neural networks, fish, generalised additive models, lake, microhabitat

Introduction

The spatial distribution and abundance of organisms in ecosystems is of crucial importance for understanding ecosystem functioning (Rosenzweig, 1991; Hayes, Ferreri & Taylor, 1996). Interactions between animals and their biotic and abiotic environment influence habitat use and the specific composition of

communities (Schoener, 1974; Begon, Harper & Townsend, 1996; Eklöv, 1997). Concerning fish, habitat and resource partitioning are regarded as key factors (Werner *et al.*, 1977). For several years, increasing interest has been taken in the study of habitat and spatial distribution of lake and river fish (Copp, 1990; Fausch, 1992; Rossier, 1995; Fischer & Eckmann, 1997). Within lakes, the littoral zone is an important area for fish (Bohl, 1980), where abundance is often much greater than elsewhere (Werner *et al.*, 1977; Brosse *et al.*, 1999a). Young of the year (0+) fish especially use the littoral zone during the summer period as these

Correspondence: S. Brosse, CNRS, UMR 5576 CESAC-Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, France. E-mail: brosse@cict.fr

areas provide food and shelter (Savino & Stein, 1989). In lakes, most studies of fish habitat have been performed using a large spatial scale (Rossier, 1995; Imbrock, Appenzeller & Eckmann, 1996; Brosse *et al.*, 1999a) whereas microhabitat studies are still limited to streams and rivers (Copp, 1990, 1992; Fausch, 1992; Baran *et al.*, 1996; Mastrorillo *et al.*, 1997; Gozlan *et al.*, 1999). The lack of knowledge concerning fish microhabitat in lakes is probably due to the complexity of lake fish microhabitat features which are probably regulated by a complex interaction of environmental variables acting mainly in a nonlinear way (Brosse *et al.*, 1999b).

Multiple linear regression (MLR) has frequently been used as a quantitative method to explore the relationships between species and their habitat (Binns & Eiserman, 1979; Fausch, 1992) but this statistical tool suffers from drawbacks when relationships between variables are nonlinear (James & McCulloch, 1990). Transformation of nonlinear variables by logarithmic, power or exponential functions can appreciably improve the normality of data, but the model results often do not reflect ecological reality (Lek *et al.*, 1996b). Some statistical methods such as generalised additive models (GAM) and artificial neural network (ANN), provide an interesting approach comparable to regression analysis, but are particularly efficient for predicting nonlinear data and explaining complex relationships between the variables (Rumelhart, Hinton & Williams, 1986; Hastie & Tibshirani, 1990). Recently, many applications of GAM and ANN in aquatic ecology have been published, e.g. for modelling stream hydrobiological and ecological responses to climate change (Poff, Tokar & Johnson, 1996), prediction of global fish species richness (Guégan, Lek & Oberdorff, 1998), prediction of density of brown trout redds (Lek *et al.*, 1996b), prediction of density and biomass of fish in streams (Baran *et al.*, 1996; Lek *et al.*, 1996a; Mastrorillo *et al.*, 1997; Gozlan *et al.*, 1999) and lakes (Brosse *et al.*, 1999b; Brosse, Lek & Dauba, 1999c).

In this study, we used MLR, GAM and ANN to study 0+ roach (*Rutilus rutilus* L.) microhabitat use in the littoral zone of a large reservoir (lake Pareloup, France). Roach was chosen as it represents one of the most common fish species in European lowland rivers and lakes. Although roach diet and growth have been frequently studied (Hammer, 1985; Cryer, Peirson & Townsend, 1986; Townsend *et al.*, 1986; Jamet *et al.*, 1990; Angelibert *et al.*, 1999), microhabitat studies have

been limited to running waters (Copp, 1990, 1992). These studies used only linear statistical methods whereas microhabitat can be defined by a complex combination of several factors acting in a nonlinear way (Brosse *et al.*, 1999b). Our aim was first to model the relationships between environmental variables and 0+ roach distribution and to predict its abundance from environmental characteristics. To reach this goal, we used three distinct modelling methods and we compare their performance: (i) a linear method (i.e. MLR), (ii) a traditional nonlinear method (i.e. GAM) and (iii) a more advanced modelling technique (i.e. ANN). Then, we quantified the influence of the environmental variables on 0+ roach microhabitat use, thus defining its microhabitat preferences in the littoral zone of the lake, and enables a discussion of the ecological significance of 0+ roach distribution.

Methods

Study site and sampling

The study was undertaken during summer 1997 in Lake Pareloup, a reservoir in the south-west of France near Rodez (44°12' N, 2°46' E; surface area of 1250 ha at a volume of about 168×10^6 m³; maximum depth 37 m and mean depth 12.5 m). Lake Pareloup is a warm monomictic lake, which stratifies thermally at 10 m depth and develops a low oxygen content in the hypolimnion; the thermocline is about 10 m deep from early June to mid-September. This prevents fish from colonising deep water in summer.

Fish sampling was performed weekly from the late larval period (June) to the juvenile period (August) in the littoral zone of the lake. Sampling point abundance by electrofishing (Nelva, Persat & Chessel, 1979) modified for young fish (Copp, 1989) was employed to evaluate the microhabitat of 0+ roach. Electrofishing was performed using a backpack electroshocker, with a small 10-cm ring anode, to provide reproducible and quantifiable samples. Such equipment can be used in a large range of situations and is efficient for the entire range of 0+ fish sizes (Copp, 1989). For each sampling point, the anode was swiftly immersed about 50 cm into the water (less at shallower points) and stunned fish were collected with a fine meshed (1 mm) dipnet. Each week, 30–40 sampling points, each separated by 5–10 m as to avoid biases due to fish escaping from one sample being

taken in the next, were randomly chosen and investigated in the same area of the lake. This area was a 250-m long littoral zone selected for its topographical heterogeneity. For each of the 306 resulting 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, i.e. from 0 (nil slope) to 3 (sheer slope), percentage of flooded vegetation cover (VEG) visually estimated as the percentage of bottom area covered and substratum particle size determined using the Cailleux (1954) methodology and expressed in percentage of bottom area covered by each of the five types of substratum: boulders (BOU), pebbles (PEB), gravel (GRA), sand (SAN) and mud (MUD). Fish collected were preserved in 4% formaldehyde solution. In the laboratory, 0+ roach were identified and counted for each sampling point.

Data preparation

Modelling was carried out after $\log_{10}(x + 1)$ transformation of the dependent variable, this transformation was applied to avoid any undue influence of outliers on the models (ter Braak & Looman, 1995). The Pearson correlation matrix showed a strong correlation between SAN and MUD ($r = -0.98$) thus, the variable SAN was removed from the data matrix. The correlation between the eight remaining variables was low and only seven out of the 28 coefficients were significant ($P < 0.05$). Among these significant coefficients, all the determination coefficients were under 0.30 (e.g. VEG–DEP = 0.28, DIS–DEP = 0.02).

Then, the whole data matrix (i.e. 306 records \times eight environmental variables) was divided into two submatrices. First, all the records with no zero values for roach density (i.e. 93 records) were isolated from the samples without roach (i.e. 213). Nil values are often removed entirely from analyses, as they make the data noisy, greatly affect the statistical analysis and can thus induce bias in the determination of the abundance and the spatial distribution of the species by the models (Pennington, 1996). Nevertheless nil values actually occur in reality and, according to ter Braak & Looman (1995), should not be discarded entirely. Thus, a quarter of these records (i.e. 53 records without roach) were selected randomly and added to the first submatrix, leading to a final SM1 matrix

containing 146 records. In the same way, the removal of three-quarters of the records without roach avoided an excessive influence of nil values in the model. The 160 nil records removed constitute the second submatrix (SM0), which was used to test model prediction of the absence of roach.

Model quality was judged by the hold-out, cross-validation procedure (Efron, 1983; Efron & Tibshirani, 1995; Kohavi, 1995) to determine recognition performance (training set) and prediction performance (testing sets). Modelling was carried out in four steps: (i) the models were trained after isolation, by random selection, of a training set from SM1 (SM1train: three quarters of the records from SM1, i.e. 110). This step allowed the performance of the three modelling techniques to be estimated. The correlation coefficient between observed and predicted values was used to quantify the ability of the models to produce the right answer through the training procedure (recognition performance). (ii) The models obtained during the training procedure were tested with the first test set (SM1test) made up of the remaining one quarter of the records from SM1 (i.e. 36 records). This step allowed the prediction capabilities of the models to be assessed. (iii) Sub-matrix SM0 (i.e. 160 records without roach) was used as a second test set. The models obtained with SM1train were tested with SM0. This step allowed us to estimate the capacity of the models to estimate samples where no fish were collected (i.e. to predict absence). (iv) Finally, after the previous calibration steps, the three models (i.e. MLR, GAM and ANN) were trained using the whole SM1 matrix (i.e. 146 records) to determine the influence of the eight environmental variables in 0+ roach microhabitat use.

Modelling techniques

For MLR, models were set up using all the variables simultaneously. Calculations were done using Splus® software. Final values of the partial standardised regression coefficients were retained to define the influence of environmental factors on microhabitat use by 0+ roach.

To improve the model's performance, we also used GAM (Hastie & Tibshirani, 1990) carried out with Splus® software. The GAM are a generalisation of multiple linear regression and of generalised linear models. They are nonparametric regression methods

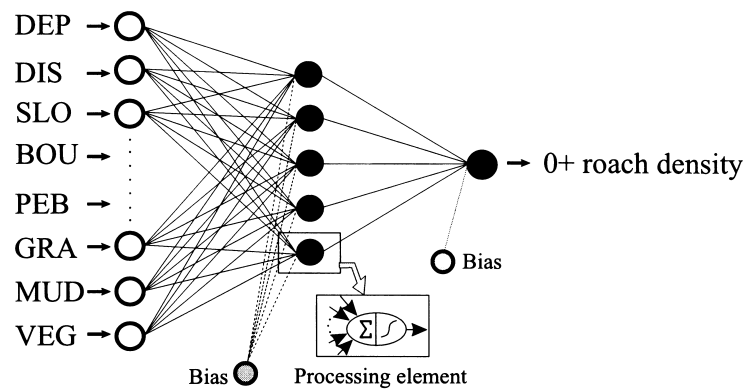


Fig. 1 A typical, three-layered feed-forward artificial neural network with eight input neurons corresponding to the eight independent environmental variables (DEP = depth, DIS = distance from the bank, SLO = slope, BOU = boulders, PEB = pebbles, GRA = gravel, MUD = mud, VEG = flooded vegetation), five hidden layer neurons and one output neuron for estimating 0+ roach densities. Connections between neurons are shown by solid lines. The bias neurons are also shown, their input value is 1. Five hidden neurons were used to obtain optimal results between bias and variance (see text for details).

which model the dependent variable as an additive sum of unspecified functions of covariates. Least squares and maximum likelihood methods used in multiple linear regression and generalised linear models are replaced by quasi-likelihood methods which rely on a locally weighted scatterplot smoother usually called 'lowess' (Cleveland, 1979). In the lowess procedure, each sample is smoothed using a defined proportion of the nearest-neighbours to the target point. Optimal fitting is obtained iteratively to minimise the residuals between observed and estimated values. The f -value of each fitted distribution indicates the proportion of samples perfectly fitted by the lowess smoother. f varies between 0 and 1 according to the sensitivity of the analysis and is determined empirically by testing various possibilities and selecting the one which provides the best generalisation ability aiming to visualise general data tendencies (Trexler & Travis, 1993). One of the major advantages of this method is that it automatically shows the dependence of the response on each of the predictors.

For ANN modelling, the processing elements in the network, called neurons, are arranged in a layered structure. Each neuron is connected to all neurons of adjacent layers. Neurons receive and send signals through these connections, they are transmitted in only one direction: from input layer to output layer through the hidden layers. The connections are given a weight which modulates the intensity of the signal they transmit. The network configuration is

approached empirically by testing various possibilities and selecting the one that provides the best compromise between bias and variance, i.e. best prediction in both training and testing sets (Geman, Bienenstock & Drousat, 1992; Kohavi, 1995). The model architecture used here was a three-layered feed-forward network with bias (Fig. 1). The first layer, called the input layer, connects with the input variables. In our case, it comprised eight neurons corresponding to the eight environmental variables. The last layer, called the output layer, connects to the output variable. It comprises a single neuron corresponding to the value of the dependent variable to be predicted (roach density). The layer between the input and output layers is called the hidden layer. A five hidden neurons configuration was determined as optimal (networks with two hidden layers were not significantly better). We thus had a total of 51 parameters ((eight input neurons \times five hidden neurons) + (five hidden neurons \times one output neuron) + six bias parameters).

One disadvantage of ANN is the lack of explanation power. Classical analyses, like MLR, can identify the contribution each individual input makes on the output and can also give some measure of confidence about the estimated coefficient. On the other hand, there is currently no theoretical or practical way to accurately interpret the weights attributed in ANN. For example, weights cannot be interpreted as regression coefficients and are difficult to use to compute causal impacts or elasticities. But in ecology,

it is necessary to know the impact of the explanatory variables. Some authors have proposed methods allowing the determination of the impact of the input variables (Garson, 1991; Dimopoulos, Bourret & Lek, 1995; Goh, 1995; Lek *et al.*, 1996a,b). In this work, Garson's algorithm modified by Goh was used to determine the relative importance of environmen-

tal variables on roach microhabitat. The computational program was realised in a Matlab® environment and computed with an Intel Pentium® processor. Finally, when variables relevant to 0+ roach microhabitat were identified by Garson-Goh's algorithm, their influence was visualised using lowess smoothing functions.

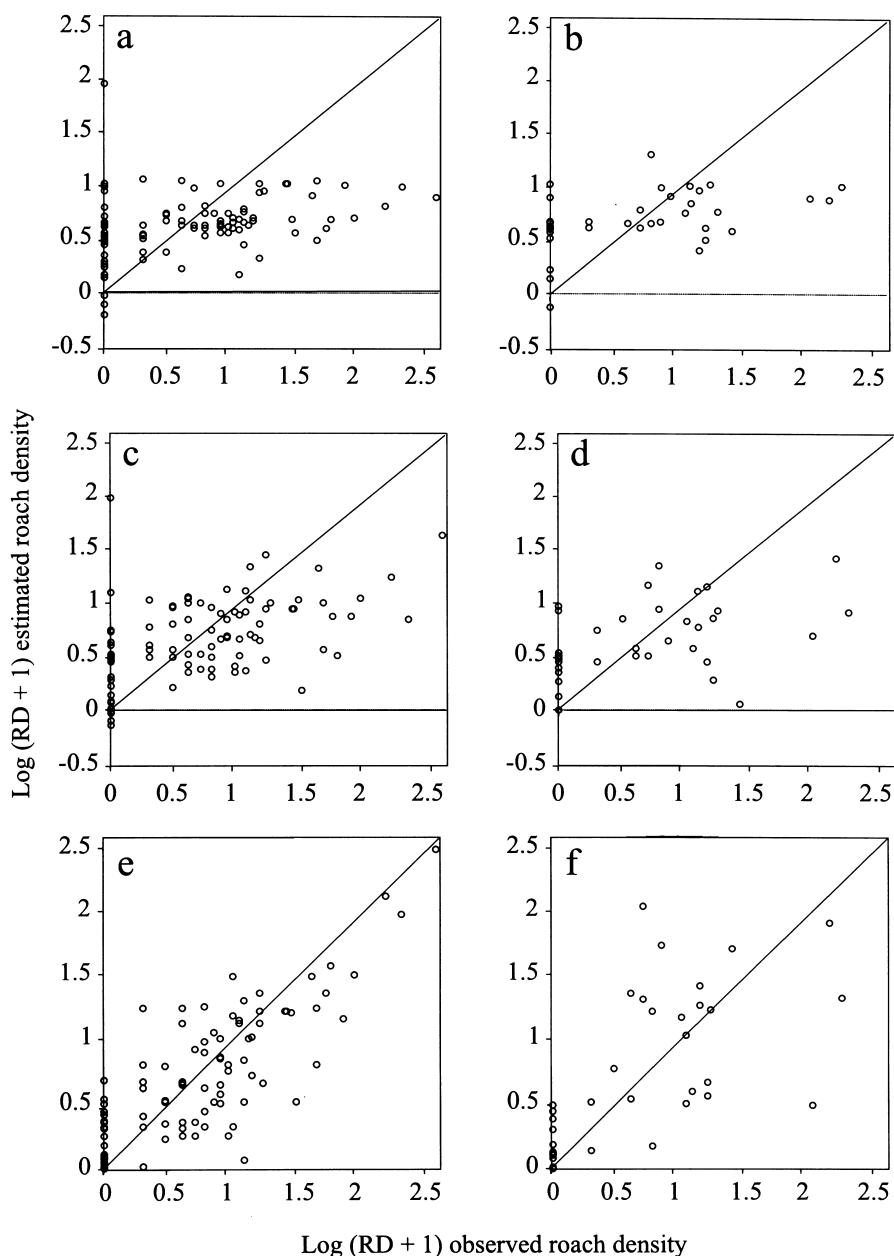


Fig. 2 Calibration of the models. MLR (a,b) GAM (c,d) and ANN (e,f) model predictions of 0+ roach densities (RD, i.e. number of roach per sample). Scatter plots of predicted values vs. observed values, in the training set (SM1train) for MLR (a), GAM (c) and ANN (e) training procedures and in the testing set (SM1test) for MLR (b), GAM (d) and ANN (f) testing procedures. The solid line indicates the perfect fit line (co-ordinates 1 : 1). See text for details.

Results

Calibration of the models

In the MLR analysis, the correlation coefficients between observed and predicted values were the following: in the training set (SM1train), $r = 0.32$ (Fig. 2a); and in the testing set (SM1test), $r = 0.43$ (Fig. 2b). Figure 2 shows some pitfalls that may exist when developing MLR models. Even though the correlation coefficients were highly significant ($P < 0.001$), the predicted values showed, both in training and testing, that the median and high values

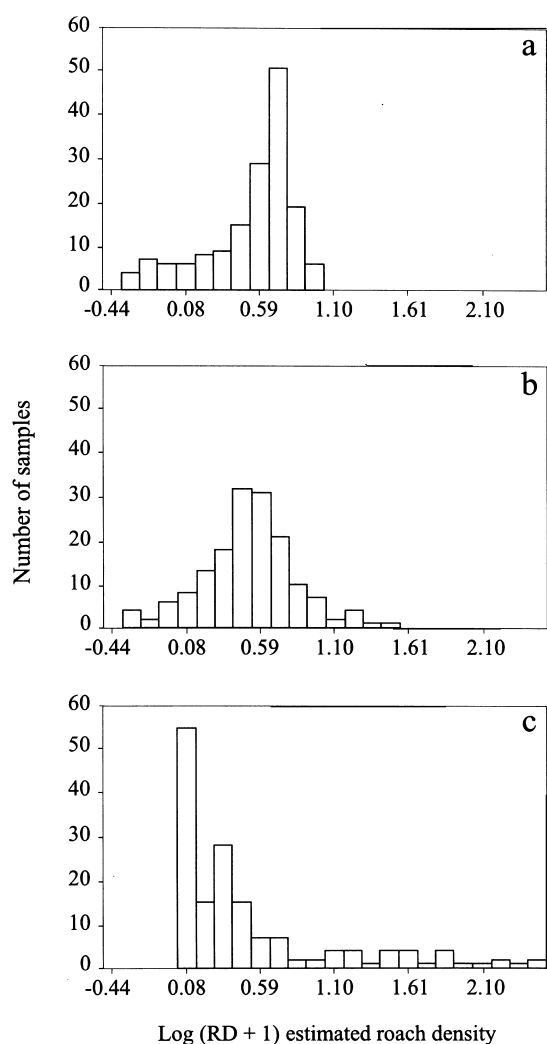


Fig. 3 Second tests on the second testing data set (SM0), i.e. samples without fish. Bars indicate the number of samples for each class of predicted values of 0+ roach densities (RD). (a) MLR, (b) GAM, (c) ANN.

of fish abundance were always underestimated and most of the nil values were either overestimated or given aberrant values, i.e. negative fish density. The points were not well distributed along the line of perfect prediction (co-ordinates 1 : 1). Nevertheless, the residuals were independent of estimated values ($r = -0.15$, $P = 0.13$ for SM1train and $r = 0.01$, $P = 0.96$ for SM1test). The second test revealed that most of the samples where no 0+ roach were collected were predicted as samples containing a medium fish density (Fig. 3a), thus the performance of the MLR model was poor.

The results of the GAM showed a slight improvement with respect to MLR models, both in training and testing sets ($r = 0.53$, $P < 0.001$ in the training set (SM1train) and $r = 0.48$, $P < 0.001$ in the testing set (SM1test)). Even though residuals were independent of estimated values ($r = -0.01$, $P = 0.32$ for SM1train and $r = -0.07$, $P = 0.67$ for SM1test), like for MLR, high values of fish densities were systematically underestimated, and a large proportion of nil values were either overestimated or given aberrant values both in training (Fig. 2c) and testing sets (Fig. 2d). In the same way, the second test revealed that most of the samples where no 0+ roach was collected were predicted to contain a medium fish density (Fig. 3b), thus even though the performance of the GAM model was slightly better than that of MLR, it was still not optimal.

The ANN prediction ability is visualised in Fig. 2(e) which shows the scatter plot between values of 0+ roach density observed and predicted by the ANN models after a training procedure of 1000 iterations. The correlation coefficient (r) was 0.83 ($P < 0.001$) in the training set and 0.62 ($P < 0.001$) in the testing set. In the training set the majority of the points in the scatter plot were well aligned along the diagonal of best prediction of co-ordinates [1 : 1] (Fig. 2e) and the residuals were independent of estimated values ($r = -0.01$, $P = 0.93$). In the testing set, prediction was less efficient, but clearly better than for GAM and MLR. Even if most samples were over- or underestimated, aberrant values never appeared (i.e. negative values of 0+ roach density) and the assumption of independence of residuals was verified ($r = -0.10$, $P = 0.56$). The underestimation of some medium or high values of 0+ roach density was due to the scarcity of samples containing medium or high fish densities. The network prediction capabilities were limited by the information contained in the training

data set, thus the limited number of points with fish did not allow the network to deal with all the factors influencing fish distribution. Nevertheless all the samples with medium and high values of 0+ roach density were always predicted as suitable habitat, even if fish density was sometimes over- or underestimated (Fig. 2f). Finally, the second test showed that most of the samples without 0+ roach were perfectly predicted (Fig. 3c), showing that the ANN was able to recognise and predict samples without fish on the basis of the eight environmental variables introduced in the model. Regression methods applied to the SM0 data matrix predicted high (for MLR) or medium (for GAM) values of 0+ roach densities (Fig. 3a,b), whereas ANN predicted nil or close to nil fish densities for most of the samples. Thus, ANN provided an accurate prediction of the samples where no 0+ roach were found.

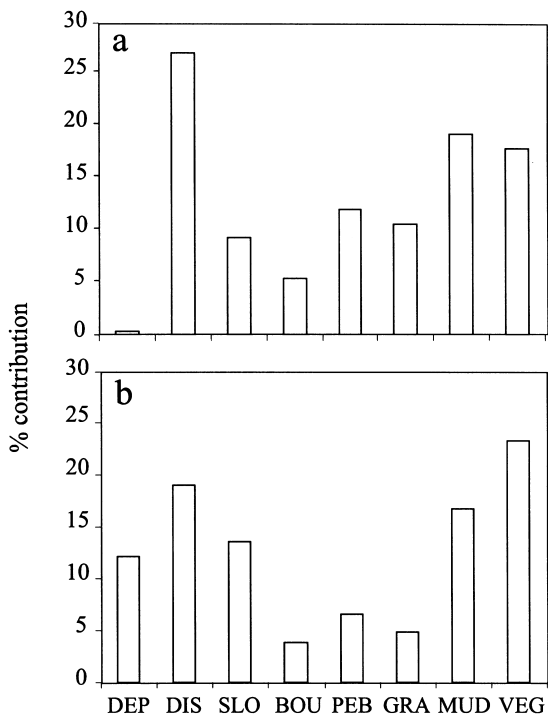


Fig. 4 Percentage contribution of each of the eight independent variables (DEP = depth, DIS = distance from the bank, SLO = slope, BOU = boulders, PEB = pebbles, GRA = gravel, MUD = mud, VEG = flooded vegetation) to the prediction of 0+ roach densities determined using the results of the training procedure on the whole SM1 data matrix. (a) Percentage contribution calculated using the absolute values of the standardised partial coefficients of the MLR; (b) percentage contribution obtained by Garson-Goh's algorithm applied on the ANN results.

Influence of environmental variables

Concerning MLR, the correlation coefficient obtained after a training phase on the whole SM1 data matrix was $r = 0.51$ ($P < 0.001$). The influence of each variable can be roughly assessed by checking the final standardised values of the regression coefficients. Among the eight coefficients, three were not significant ($P > 0.05$): BOU, SLO and DEP. Figure 4(a) shows the percentage contribution of each of the independent variables (i.e. the relative importance of each environmental variable on roach microhabitat), on the basis of the absolute values of the standardised partial coefficients of MLR. It revealed that the model was strongly influenced by five variables: mainly DIS (27%), MUD (19%), VEG (18%) and to a lesser extent PEB and GRA which each contributed about 11%. Thus, according to the results of MLR, 0+ roach microhabitat was best described by the distance from the bank (DIS), the percentage of mud (MUD) and the flooded vegetation cover (VEG).

Concerning ANN, the correlation coefficient was $r = 0.85$ ($P < 0.001$). The results of Garson-Goh's algorithm stress the weight of five environmental variables in the model (Fig. 4b) with contributions of roughly 24% for VEG, 20% for DIS and between 11 and 17% for DEP, SLO and MUD. The three remaining variables each contributed less than 10%.

Both analyses indicate that the same three variables (DIS, MUD and VEG) were most strongly associated with roach density, however, considering the predictive efficiency of the models, ANN provided more realistic information. Aiming to visualise how each of the five most influential variables identified by ANN acts, we used the information provided by the GAM model, which reached a correlation coefficient of $r = 0.63$ ($P < 0.001$). Even if the GAM result was slightly lower than that of ANN, it is slightly better than the coefficient obtained by MLR. Moreover, in GAM models, we can directly plot the influence each predictor has on the dependent variable by representing the lowest smoothing function on the plot of 0+ roach density vs. the considered variable. Then, the data fitted using a lowest smoothing function clearly showed the influence of each of the five important environmental variables, identified using ANN, on 0+ roach abundance according to the range of variation of each environmental variable considered.

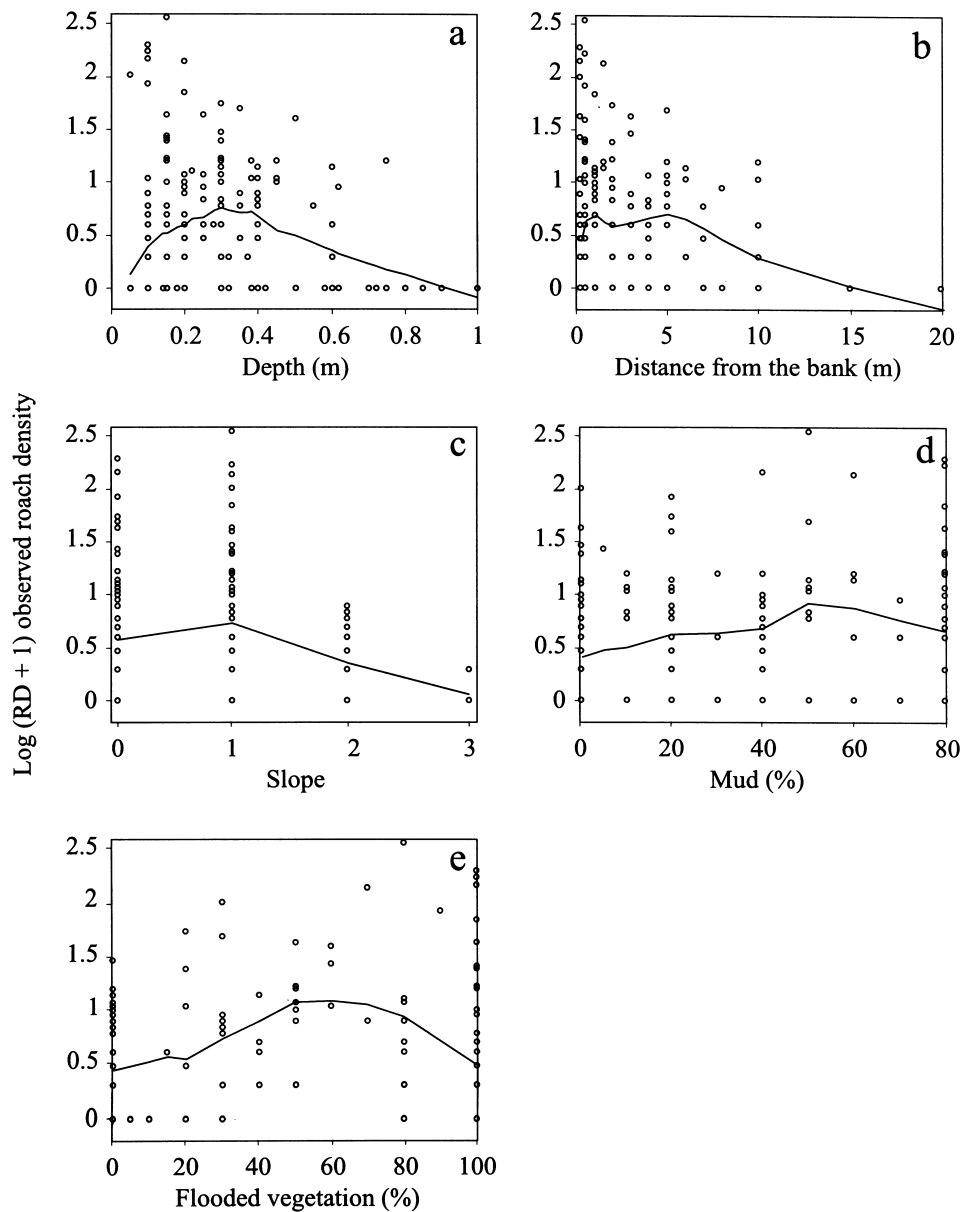


Fig. 5 Contribution profiles (responses) of each of the five most important independent variables for the prediction of 0+ roach densities (RD) using the information provided by the GAM model trained on the whole SM1 data matrix. Observed values of RD are plotted vs. each independent variable. On the same plots, the lowess smoothing function is represented by a solid line. (a) Depth, $f = 0.40$, (b) distance from the bank, $f = 0.40$, (c) slope, $f = 0.30$, (d) mud, $f = 0.35$, (e) flooded vegetation, $f = 0.45$. The f -value indicates the proportion of samples perfectly fitted by the lowess smoother; f varies between 0 and 1 according to the sensitivity of the analysis and is determined empirically by testing various possibilities and selecting the one which provides the best generalisation ability to visualise general data tendencies.

Age 0+ roach preferred median depth and low and median distance from the bank (Fig. 5a,b). In the same way, nil to gentle slopes were preferred to steep slopes (Fig. 5c). Moreover, 0+ roach preferentially inhabited areas with median and high cover of

flooded vegetation and avoided areas with no vegetation and, to a lesser extent, areas poorly and very highly vegetated (Fig. 5e). Finally, 0+ roach exhibited a preference for fine organic substrata (MUD) (Fig. 5d). As a consequence, 0+ roach were

mainly located in shallow, vegetated, gently sloping areas located close to the bank and, preferably, with a muddy substratum.

Discussion

The main associations measured between roach density and microhabitat features in lake Pareloup can be approximated by linear functions only to a limited extent (Fig. 2). Unlike ANN, regression models, even those which use nonlinear transformations (i.e. GAM), are not able to faithfully reproduce the relationships occurring in real systems (Fig. 2c,d). In the present study, the ANN was able to predict reliably, during the training procedure on the SM1train matrix, and also to predict new data, during the testing procedure on the SM1test matrix (Fig. 2e,f). Nevertheless, some low values of fish density were overestimated by the network (Fig. 2e,f); these samples could have come from habitats suitable for 0+ roach but where no fish were actually collected. This could have had two origins: first, a bias in the sampling methodology, i.e. fish escapes or underestimates, induced by the electrofishing technique (Bain, Finn & Booke, 1985; Dewey, 1992); second, there may not be enough fish to fill all suitable habitats. This second hypothesis is supported by Angelibert *et al.* (1999) who reported that the fish community in lake Pareloup is not saturated. Finally, the predictive accuracy of the models can be checked by studying the second test on the data matrix, where no 0+ roach were found (SM0) (Fig. 3). This second test validates the ability of ANN to predict accurately 0+ roach density, whereas MLR, and to a lesser extent GAM, revealed serious shortcomings. However, the samples predicted as being densely populated by the ANN model, but where no fish were captured, could represent areas where the environment is suitable for 0+ roach but not inhabited by fish, supporting our second hypothesis concerning roach density in the lake. These results allowed an estimate to be made of the potential fish density in the samples where no 0+ roach were actually found.

Considering the influence of each environmental variable on predictions of 0+ roach density using the training procedure on the whole SM1 data matrix, the study of the partial standardised coefficients provided by MLR gave misleading information for two variables (Fig. 3a): flooded vegetation (VEG) was con-

sidered as the third most influential variable, but the partial standardised coefficient is negative (-0.28), which seems illogical as macrophytes are widely assumed to be important as fish nursery areas (Conrow, Zale & Gregory, 1990; Hosn & Downing, 1994), especially for roach which is usually strongly associated with aquatic vegetation (Garner, 1995; Rossier, Castella & Lachavanne, 1996). In the same way, depth was never revealed as a significant variable in MLR, whereas it usually constitutes an important feature in 0+ roach habitat choice considering the requirement for shelters against predation (Brabrand & Faafeng, 1994; Garner, 1995). Concerning ANN results, the influence of environmental variables on fish distribution, assessed using sensitivity analysis of the variables supplied by the ANN model, was in accordance with ecological factors reported in previous studies, highlighting the importance of flooded vegetation, distance from the bank, depth, bottom slope and fine substrata on 0+ roach microhabitat use. This result demonstrates the suitability of the ANN approach to describe nonlinear relationships in ecological systems. The assumptions provided by Garson-Goh's algorithm are supported by the lowest analysis of the variables introduced in the model, showing that the 0+ roach microhabitat is clearly influenced by several environmental variables acting in a nonlinear way. This reveals an inability to define fish microhabitat precisely using linear methods such as MLR.

A more precise study of 0+ roach microhabitat using the information provided by the lowest curves (Fig. 5) showed the avoidance of deep water (more than 0.5 m), areas more than 10 m distant from the bank and steeply sloping areas, which are usually colonised by predators, such as adult perch (*Perca fluviatilis* L.), pike (*Esox lucius* L.) and, occasionally, pike-perch (*Stizostedion lucioperca* L.) (Brabrand & Faafeng, 1994; Eklöv, 1997). In the same way, areas without flooded vegetation were not preferred by 0+ roach as open areas are dangerous due to the enhanced efficiency of predators in open water (Brabrand & Faafeng, 1994; Eklöv, 1997). Although 0+ fish are generally located in vegetated areas, which afford protection from predators and provide a rich foraging habitat (Rozas & Odum, 1988; Christensen & Persson, 1993; Persson & Eklöv, 1995; Eklöv, 1997), the use of flooded vegetation by 0+ roach is quite complex. Median vegetation densities were preferred

since, though vegetation provides shelter against predators, it also reduces feeding efficiency of 0+ roach (Johanson, 1987). Generally, 0+ roach feed opportunistically on phytoplankton, macrophytes, zooplankton, insect larvae and sediment (Prejs, 1978; Persson, 1983, 1986; Cryer *et al.*, 1986; Townsend *et al.*, 1986; Jamet *et al.*, 1990; Michel & Oberdorff, 1995) and occupy gently sloping areas with muddy bottoms, because organic deposits in these areas provide food such as chironomid larvae. Also, 0+ roach prefer shallow water, but this preference decreases in the shallower zones (less than 0.2 m) as these areas are also occupied by rudd (*Scardinius erythrophthalmus* L.) larvae. A lower roach density could be induced by intraspecific competition and/or by different feeding requirements between 0+ roach and 0+ rudd (Rheinberger, Hofer & Weiser, 1987). Finally, the use of shallow littoral areas could be related to the fish preferring high temperature which increases foraging ability (Persson, 1986) and digestion efficiency (Persson, 1982).

This study showed that 0+ roach are mainly located in the shallow, gently sloping littoral areas in moderately vegetated areas with muddy bottoms. Moreover, these habitat features favour the whole range of sizes of 0+ roach and therefore satisfy possible habitat changes occurring during the development of roach larvae and juveniles (Copp, 1992). This study therefore constitutes an overview of the main habitat characteristics of 0+ roach in lake Pareloup. However, previous studies of 0+ roach habitat use in lakes described only a larger spatial scale and the only experiments on roach habitat use were performed in an artificial environment (i.e. in enclosures with limited space and low fish density) and focussed on predator-prey relationships and interspecific competition (Eklöv & Persson, 1995; Persson & Eklöv, 1995). In the natural environment, roach habitat use was more complex than in these experiments although in agreement with Christensen & Persson (1993), roach tended to colonise preferentially areas where structural heterogeneity is maximal. Such habitat preferences suggest a trade-off between the search for food and the requirement for shelter from predation. Such trade-offs between costs and benefits are well known in ecology (Persson & Eklöv, 1995) but differ between organisms and environments. This may explain why habitat use by 0+ roach differs in lakes and streams, where 0+ roach habitat is

strongly influenced by current velocity (Moyle & Baltz, 1985; Copp, 1992). The complexity of habitat features in lakes could explain why classical linear methods are sometimes successful in predicting fish microhabitat in rivers, whereas in lakes more sophisticated statistical methods, such as GAM or ANN, are needed. These latter methods constitute alternative approaches in ecology, particularly in cases with nonlinearly related variables. This shows that overall trends in fish distribution can accurately be assessed using a few pertinent environmental variables and can provide insight into the ecological meaning of fish microhabitat.

Acknowledgments

We thank the two anonymous reviewers for their constructive comments on an earlier version of this paper. This research was supported in part by a doctoral grant (S. Brosse) provided by the French electricity agency (E.D.F.).

References

- Angelibert S., Brosse S., Dauba F. & Lek S. (1999) Changes in roach (*Rutilus rutilus* L.) population structure induced on draining a large reservoir. *Comptes Rendus de l'Académie des Sciences Paris Serie III*, **322**, 331–338.
- Bain M.B., Finn J.T. & Booke H.E. (1985) A quantitative method for sampling riverine microhabitats by electrofishing. *North American Journal of Fisheries Management*, **5**, 489–493.
- 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.
- Begon M., Harper J.L. & Townsend C.R. (1996) *Ecology: Individuals, Population, and Communities*, 3rd edn. Blackwell Science, Oxford.
- Binns N.A. & Eiserman J.P. (1979) Quantification of fluvial trout habitat in Wyoming. *Transactions of the American Fisheries Society*, **108**, 215–228.
- Bohl E. (1980) Diel pattern of pelagic distribution and feeding in planktivorous fish. *Oecologia*, **44**, 368–375.
- ter Braak C.J.F. & Looman C.W.N. (1995) Regression. *Data Analysis in Community and Landscape Ecology* (Eds R.G.H. Jongman, C.J.F. ter Braak & O.F.R. Van Tongeren), pp. 29–77. Cambridge University Press, Cambridge.
- Brabrand A. & Faafeng B. (1994) Habitat shift in roach (*Rutilus rutilus*) induced by the introduction of pike-

- perch (*Stizostedion lucioperca*). *Verhandlungen der Internationale Vereinigung Fur Theoretische und Angewandte Limnologie*, **25**, 2123.
- Brosse S., Dauba F., Oberdorff T. & Lek S. (1999a) Influence of some topographical variables on the spatial distribution of lake fish during summer stratification. *Archiv für Hydrobiologie*, **145**, 359–371.
- Brosse S., Guegan J.F., Tourenq J.N. & Lek S. (1999b) The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. *Ecological Modelling*, **120**, 299–312.
- Brosse S., Lek S. & Dauba F. (1999c) Predicting fish distribution in a mesotrophic lake by hydroacoustic survey and artificial neural networks. *Limnology and Oceanography*, **44**, 1293–1303.
- Cailleux A. (1954) Limites dimensionnelles des noms des fractions granulométriques. *Bulletin de la Société Géologique Française*, **4**, 643–646.
- Christensen B. & Persson L. (1993) Species-specific anti-predator behaviours: effect of prey choice in different habitats. *Behavioral Ecology and Sociobiology*, **32**, 1–9.
- Cleveland W.S. (1979) Robust locally-weighted regression and scatterplot smoothing. *Journal of the American Statistical Association*, **74**, 829–836.
- Conrow R., Zale A.V. & Gregory R.W. (1990) Distribution and abundance of early stages in a florida lake dominated by aquatic macrophytes. *Transactions of the American Fisheries Society*, **119**, 521–528.
- Copp G.H. (1989) Electrofishing for fish larvae and juveniles: equipment modifications for increased efficiency with short fishes. *Aquaculture and Fisheries Management*, **20**, 453–462.
- Copp G.H. (1990) Shifts in the microhabitat of larval and juvenile roach, *Rutilus rutilus* L. in a floodplain channel. *Journal of Fish Biology*, **36**, 683–692.
- Copp G.H. (1992) An empirical model for predicting microhabitat of 0+ juvenile fishes in a lowland river catchment. *Oecologia*, **91**, 338–345.
- Cryer M., Peirson G. & Townsend C.R. (1986) Reciprocal interactions between roach, *Rutilus rutilus*, and zooplankton in a small lake. Prey dynamics and fish growth and recruitment. *Limnology and Oceanography*, **31**, 1022–1038.
- Dewey M.R. (1992) Effectiveness of a drop net, a pop net and an electrofishing frame for collecting quantitative samples of juveniles fishes in vegetation. *North American Journal of Fisheries Management*, **12**, 808–813.
- Dimopoulos Y., Bourret P. & Lek S. (1995) Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters*, **2**, 1–4.
- 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.
- Efron B. & Tibshirani R. (1995) *Cross-validation and the Bootstrap: estimating the error rate of a prediction rule*. Technical Report 176, Department of Statistics, Stanford University, 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*). *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 1520–1531.
- Eklöv P. & Persson L. (1995) Species-specific antipredator capacities and prey refuges: interactions between piscivorous perch (*Perca fluviatilis*) and juvenile perch and roach (*Rutilus rutilus*). *Behavioral Ecology and Sociobiology*, **37**, 169–178.
- Fausch K.D. (1992) Experimental analysis of microhabitat selection by juvenile steelhead (*Onchorhynchus mykiss*) and coho salmon (*O. kisutch*) in a British Columbia stream. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**, 1198–1207.
- Fischer P. & Eckmann R. (1997) Spatial distribution of littoral fish species in a large European lake, Lake Constance, Germany. *Archiv für Hydrobiologie*, **140**, 91–116.
- Garner P. (1995) Suitability indices for juvenile 0+ roach (*Rutilus rutilus* (L.) using point abundance sampling data. *Regulated Rivers: Research and Management*, **10**, 99–104.
- Garson G.D. (1991) Interpreting neural network connection weights. *Artificial Intelligence Expert*, **6**, 47–51.
- Geman S., Bienenstock E. & Doursat R. (1992) Neural networks and the bias/variance dilemma. *Neural Computation*, **4**, 1–58.
- Goh A.T.C. (1995) Back-propagation neural networks for modeling complex systems. *Artificial Intelligence Engineer*, **9**, 143–151.
- Gozlan R.E., Mastrorillo S., Copp G.H. & Lek S. (1999) Predicting the structure and diversity of young of the year fish assemblages in large rivers. *Freshwater Biology*, **41**, 811–822.
- Guégan J.F., Lek S. & Oberdorff T. (1998) Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature*, **391**, 382–384.
- Hammer C. (1985) Feeding behaviour of roach (*Rutilus rutilus*) and the fry of perch (*Perca fluviatilis*) in Lake Lankau. *Archiv für Hydrobiologie*, **103**, 61–74.
- Hastie T.J. & Tibshirani R.J. (1990) *Generalized Additive Models*. Chapman & Hall, London.
- Hayes D.B., Ferreri C.P. & Taylor W.W. (1996) Linking fish habitat to their population dynamics. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 383–390.
- Hosn W.A. & Downing J.A. (1994) Influence of cover on the spatial distribution of littoral-zone fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1832–1838.

- Imbrock F., Appenzeller A. & Eckmann R. (1996) Diel and seasonal distribution of perch in Lake Constance: a hydroacoustic study and in situ observations. *Journal of Fish Biology*, **49**, 1–13.
- James F.C. & McCulloch C.E. (1990) Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review of Ecology and Systematics*, **21**, 129–166.
- Jamet J.L., Gres P., Lair N. & Lasserre G. (1990) Diel feeding cycle of roach (*Rutilus rutilus* L.) in eutrophic Lake Aydat (Massif Central, France). *Archiv für Hydrobiologie*, **118**, 371–382.
- Johanson L. (1987) Experimental evidence for interactive habitat segregation between roach (*Rutilus rutilus*) and rudd (*Scardinius erythrophthalmus*) in a shallow eutrophic lake. *Oecologia*, **73**, 21–27.
- Kohavi R. (1995) *A study of the cross-validation and bootstrap for accuracy estimation and model selection*. Proceedings of the International Joint Conference 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. *Aquatic Living Resources*, **9**, 23–29.
- Lek S., Delacoste M., Baran P., Dimopoulos I., Lauga J. & Aulagner S. (1996b) Application of neural networks to modelling non-linear relationships in ecology. *Ecological Modelling*, **90**, 39–52.
- 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 Biology*, **38**, 237–246.
- Michel P. & Oberdorff T. (1995) Feeding habits of fourteen European freshwater fish species. *Cybium*, **19**, 5–46.
- Moyle P.B. & Baltz D.M. (1985) Microhabitat use by an assemblage of California stream fishes: developing criteria for in-stream flow determinations. *Transactions of the American Fisheries Society*, **114**, 695–704.
- 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. *Comptes Rendus de l'Académie des Sciences Paris Serie III*, **289**, 1295–1298.
- Pennington M. (1996) Estimating the mean and variance from highly skewed marine data. *Fishery Bulletin*, **94**, 498–505.
- Persson L. (1982) Rate of food evacuation in roach (*Rutilus rutilus*) in relation to temperature, and the application of evacuation rate estimates for studies on the rate of food consumption. *Freshwater Biology*, **12**, 203–210.
- Persson L. (1983) Food consumption and the significance of detritus and algae in the intraspecific competition in roach (*Rutilus rutilus*) in a shallow eutrophic lake. *Oikos*, **41**, 118–125.
- Persson L. (1986) Temperature induced shift in foraging ability in two fish species, roach (*Rutilus rutilus*) and perch (*Perca fluviatilis*): implications for coexistence between poikilotherms. *Journal of Animal Ecology*, **55**, 829–839.
- Persson L. & Eklöv P. (1995) Prey refuges affecting interactions between piscivorous perch and juvenile perch and roach. *Ecology*, **76**, 70–81.
- Poff L., Tokar S. & Johnson P. (1996) Stream hydrological and ecological responses to climate change assessed with an artificial neural network. *Limnology and Oceanography*, **41**, 857–863.
- Prejs A. (1978) Lake macrophytes as the food of roach (*Rutilus rutilus* L.) and rudd (*Scardinius erythrophthalmus* L.). Daily intake of macrophyte food in relation to body size. *Ekologia Polska*, **26**, 537–553.
- Rheinberger V., Hofer R. & Weiser W. (1987) Growth and habitat separation in eight cohorts of three species of cyprinids in a subalpine lake. *Environmental Biology of Fishes*, **18**, 209–217.
- Rosenzweig M.L. (1991) Habitat selection and populations interactions: the search for mechanisms. *American Naturalist*, **137**, 5–28.
- Rossier O. (1995) Spatial and temporal separation of littoral zone fishes of Lake Geneva (Switzerland–France). *Hydrobiologia*, **300/301**, 321–327.
- Rossier O., Castella E. & Lachavanne J.-B. (1996) Influence of submerged aquatic vegetation on size class distribution of perch (*Perca fluviatilis*) and roach (*Rutilus rutilus*) in the littoral zone of Lake Geneva (Switzerland). *Aquatic Sciences*, **58**, 1–14.
- Rozas L.P. & Odum W.E. (1988) Occupation of submerged aquatic vegetation by fishes: testing the roles of food and refuge. *Oecologia*, **77**, 101–106.
- Rumelhart D.E., Hinton G.E. & Williams R.J. (1986) Learning representations by back-propagating error. *Nature*, **323**, 533–536.
- Savino J.F. & Stein R.A. (1989) Behavioural interactions between fish predators and their prey; effects of plant density. *Animal Behaviour*, **37**, 311–321.
- Schoener T. (1974) Resource partitioning in ecological communities. *Science*, **185**, 27–39.
- Townsend C.R., Winfield I.J., Pierson G. & Cryer M. (1986) The response of young roach *Rutilus rutilus* to seasonal changes in abundance of microcrustacean prey: a field demonstration of switching. *Oikos*, **46**, 372–378.
- Trexler J.C. & Travis J. (1993) Nontraditional regression analyses. *Ecology*, **74**, 1629–1637.
- Werner E.E., Hall D.J., Laughlin D.R., Wagner D.J., Wilmann L.A. & Funk F.C. (1977) Habitat partitioning in a freshwater fish community. *Journal of the Fisheries Research Board of Canada*, **34**, 360–370.

(Manuscript accepted 13 December 1999)