

Abundance, diversity, and structure of freshwater invertebrates and fish communities: an artificial neural network approach

SEBASTIEN BROSSE^{1,2}

SOVAN LEK¹

COLIN R. TOWNSEND²†

¹CNRS, UMR 5576 CESAC

Université Paul Sabatier

118 Route de Narbonne

31062 Toulouse cedex, France

²Department of Zoology

University of Otago

P. O. Box 56

Dunedin, New Zealand

email: colin.townsend@stonebow.otago.ac.nz

that ANN is an appropriate tool for both prediction and explanation of ecological relationships at various spatial scales (microhabitat and macrohabitat), and for a range of aquatic ecosystems (lakes and rivers), organisms (invertebrates and fish), and ecological descriptors (abundance, Shannon diversity index, and community composition).

Keywords macroinvertebrates; fish; community composition; abundance; diversity; river; lake; spatial scales; modelling; artificial neural networks; back-propagation

Abstract Artificial neural networks (ANN) are models inspired by the structure and processes of biological cognition and learning. To illustrate the ecological applications of ANN, we present analyses of two complementary examples. ANN is first used to predict the diversity of macroinvertebrates, at the macrohabitat scale, in tributaries of a large river in New Zealand and, second, to predict the distribution and abundance of several fish species at the microhabitat scale in a French lake. The predictive abilities of the models were high, with correlation coefficients between observed and estimated values from 0.61 and 0.92. Moreover, the environmental variables found to be associated with invertebrate diversity and fish abundance were in accord with results of previous studies. The combination of ANN with a multivariate analysis of fish community composition provided both for accurate prediction of fish assemblages and effective visualisation of their relationships with environmental variables. On the basis of these studies in different locations (New Zealand streams, French lake), involving various population and community attributes, we conclude

INTRODUCTION

Artificial neural networks (ANN), inspired by the structure and processes of biological cognition and learning, were developed initially to model biological functions. They learn from experience and can rapidly solve difficult computational problems. In the past decade, research into ANN has shown explosive growth. ANN models have often been applied in chemistry (Kvasnicka 1990) and in physics, for example in studies of speech and image recognition (Rahim et al. 1993; Dekruger & Hunt 1994; Chu & Bose 1998). Most applications of ANN in biology have concerned medicine and molecular biology (Albiol et al. 1995; Faraggi & Simon 1995; Lo et al. 1995) but with a few examples in the ecological and environmental sciences beginning in the 1990s. For instance, Colasanti (1991) perceived similarities between ANN and ecosystem structure and functioning and recommended the utilisation of this tool in ecological modelling. In a review of computer-aided research in biodiversity, Edwards & Morse (1995) also emphasised the potential importance of ANN. Other examples can be found in different fields of ecology, such as modelling of the greenhouse effect (Seginer et al. 1994), predicting phytoplankton production (Scardi 1996; Recknagel et al. 1997), and predicting various parameters in fish ecology (Baran et al. 1996; Lek et al. 1996a,b; Guégan et al. 1998; Brosse et al.

† Author for correspondence.

M00010

Received 13 March 2000; accepted 27 July 2000

1999b,c). Most of these studies demonstrated that ANN performed better than classical linear and non-linear modelling methods such as multiple linear regression or generalised additive models. However, very few studies have dealt with the application of ANN to the ecology of stream invertebrates or of multi-species assemblages (Brosse et al. 1999b; Schleiter et al. 1999).

In this paper we assess the ability of ANN to predict ecological parameters relating to freshwater invertebrates and fish. We provide a general description of the ANN approach and then test the capacity of ANN models to predict the diversity of macroinvertebrates in tributaries of the Taieri River in New Zealand and the distribution and abundance of several fish species at the microhabitat scale in Lake Pareloup in France.

MATERIALS AND METHODS

Study sites and sampling

The analysis of macroinvertebrate diversity was performed on 97 samples taken during summer 1990 from sites dispersed throughout the Taieri River basin. This river, which is 318 rectilinear km in length and ranges in elevation from sea level to 1150 m, lies between latitudes 44°55'S and 46°05'S in the south-eastern quarter of the South Island of New Zealand. Its drainage area is 5650 km², the fifth largest in New Zealand. For each of the 97 sampling sites, 10 environmental descriptors that operate at various spatial scales, ranging from the river catchment to the bedform scale, were selected to model the spatial distribution of invertebrate diversity. These were elevation (m above sea level), drainage density (total stream length per unit area of basin in the catchment area of the sample site), stream order at the sample site, percentage of the site's catchment area that is barren (i.e., cleared of vegetation and subject to major disturbances because composed of roads or urban areas), percentage of stream bank adjacent to each sample site that is composed of exposed bedrock, percentage of the riparian zone adjacent to the sampling site composed of exotic pasture grasses, percentage composed of native tussock grasses, mean water depth (m) at baseflow at each sampling site, mean channel width (m), and median particle size of the streambed. Channel width was measured at six cross-sections at each sampling site, water depth measured at three points across each cross-section, and median particle size estimated using the method of Wolman (1954),

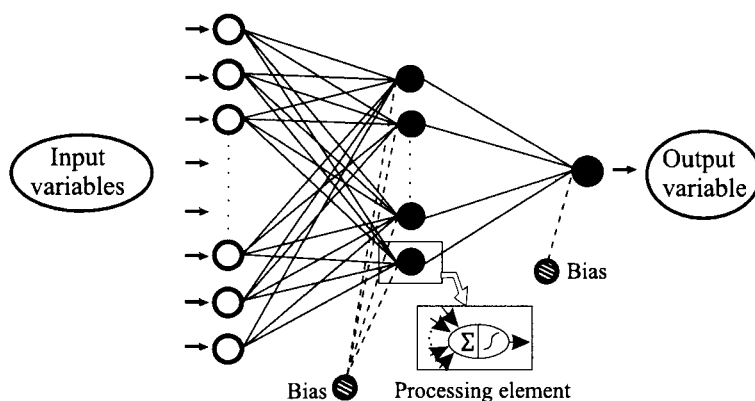
in which 100 randomly chosen particles are measured at each site. This last variable can be considered to be representative of current velocity which is closely related to the median particle size of the streambed. Benthic macroinvertebrates were collected using a standard Surber sampler (mesh size 250 µm, surface sampled 0.06 m²), with two samples per site. The samples were fixed in 5% formaldehyde and in the laboratory macroinvertebrates were sorted and identified to species level or to the lowest taxonomic level possible on the basis of keys in Winterbourn & Gregson (1989). The database includes 85 taxa. The Shannon diversity index (H) was the biological variable chosen to be predicted by the ANN. Although the usefulness of this index has been questioned (Green 1979; Norris 1995), it still constitutes one of the most commonly used diversity measurements that combines two independent pieces of biological information, species richness and species abundance (Legendre & Legendre 1998; Townsend et al 2000). For each sample site, diversity was calculated as:

$$H = - \sum P_i \ln P_i$$

where P_i is the proportion of individuals in the community belonging to the i th taxon.

The assessment of fish distribution and abundance was performed on data gathered in Lake Pareloup (maximum depth 37 m, average depth 12.5 m, surface area 1350 ha, volume 168 million m³) in the south-west of France (44°12' N, 2°46' E). It is a warm monomictic lake, undergoing summer thermal stratification. Low oxygen concentrations below the thermocline (located at c. 10 m depth from early June to mid September) prevent the fish from colonising deep water during summer. Fish were collected weekly from late June to late August 1997 using point abundance sampling by electrofishing adapted for young fish (Nelva et al. 1979; Copp 1989). This method provides quantitative and comparable fish samples without the need for standardisation (Copp 1989). Sampling was performed in the littoral zone of the lake, which exhibits a wide range of local topographical characteristics. Each week, 30–40 sampling points were investigated, giving a total of 306 sampling points. Nine habitat variables were also assessed at each point: distance from the bank (m); depth (m); local slope of the bottom at each sampling point (four classes from zero (nil slope) to three (sheer slope)); percentage of inundated terrestrial vegetation (aquatic vegetation is very scarce in the lake and was not found in the sampling area) visually estimated as the percentage of bottom area covered; and substratum particle size,

Fig. 1 Typical 3-layered feed-forward artificial neural network (ANN) with one input layer corresponding to the input (i.e., independent) variables (open circles), one hidden layer and one output layer to estimate the output (i.e., dependent) variable (closed circles). Solid lines show connections between neurons. Bias neurons are also shown (hatched circles); their input value is 1. Number of hidden neurons was set to obtain optimal results.



determined using the Cailleux (1954) methodology, and expressed in percentage of bottom area composed of five types of substratum—boulders, pebbles, gravel, sand, and mud. The variables were measured in a 1 m² bottom area corresponding to each sample. A Pearson correlation matrix showed a strong correlation between sand and mud ($r = -0.98$); thus, the variable sand was removed from the data matrix in order to deal with colinearity and the models were set up using the eight remaining environmental variables. Fish were preserved in 4% formaldehyde solution. Underyearling (0+) roach (*Rutilus rutilus*, L.), 0+ perch (*Perca fluviatilis*, L.), 0+ rudd (*Scardinius erythrophthalmus*, L.), 0+ gudgeon (*Gobio gobio*, L.), 0+ pike (*Esox lucius*, L.), and adult perch, which together represent more than 90% of the fish in Lake Pareloup (Brosse et al. 1999a), were identified and numbers recorded for each sampling point. Throughout this paper we refer to analysis of six fish species, although two of the fish groups are different age classes of the same species.

Modelling procedure

The ANN architecture is a layered feed-forward network, in which the non-linear elements (neurons) are arranged in successive layers, with a one-way flow of information (i.e., weights) from input layer to output layer, through a hidden layer (Fig. 1). In ANN, the computational or processing elements are called neurons. Like a natural neuron, they have many inputs but only a single output, which can stimulate other neurons in the network. Neurons from one layer are connected to all neurons in the adjacent layer(s), but no lateral connections within a layer nor feed-back connections are possible (for more detail see Lek & Guégan 2000). The number

of input and output units depends on the representations of the input and the output objects, respectively. A “bias” neuron was added to each computational layer (i.e., hidden and output layer); these two neurons (Fig. 1) had a constant input value of one and were used to lower biases in the modelling procedure (Rumelhart et al. 1986). 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 by one of the authors (SL) in a Matlab[®] version 5.0 environment and computed with an Intel Pentium III[®] processor. Model reliability was assessed using the correlation coefficient (r) between the observed values (i.e., actual values) and the predicted values. We also used a performance index (PI) defined as the proportion of responses within $\pm 10\%$ of the actual value to estimate the percentage of samples well predicted by the models.

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 in learning data. Second, we used the “leave-one-out” bootstrap cross-validation test (Efron 1983), where each sample is left out of the model formulation in turn and predicted once, to validate the models. This procedure is appropriate when the amount of data is limited and/or when each sample is likely to have “unique information” (Efron 1983; Kohavi 1995); it has been found to be efficient for ANN modelling of small data sets (Guégan et al. 1998; Brosse et al. 1999b). This second step allows the prediction capabilities of the network to be assessed.

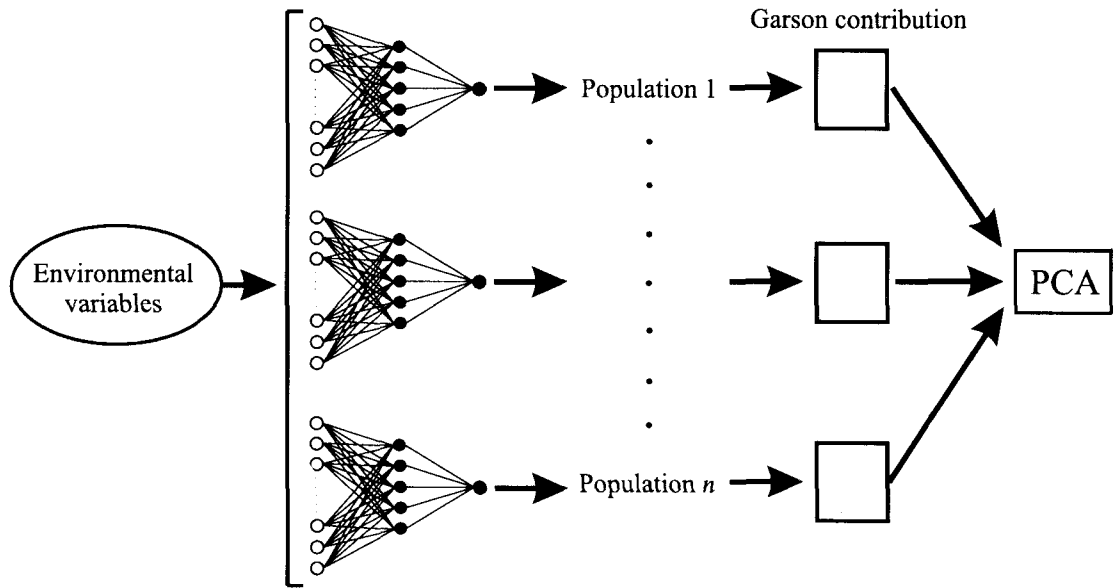


Fig. 2 Schematic representation of the modelling procedure used to assess the spatial occupancy of the six fish populations. The six artificial neural network (ANN) models (symbols as in Fig. 1) were performed using the same data matrix as the input (i.e., environmental variables). Each ANN model predicts abundance of one fish population. After the predictive modelling procedure, the Garson algorithm gave the percentage contribution of each input variable to the selected output. Finally, the Garson matrix obtained was used to perform a Principal Component Analysis (PCA) to provide a visual representation of fish spatial occupancy.

A disadvantage of ANN is a lack of explanatory power. Some analyses, such as multiple regression, can identify the contribution that each individual input makes to the output and can also give some measures of confidence about the estimated coefficients. However, there is currently no theoretical or practical way of accurately interpreting the weights attributed in ANN (Lek et al. 1996b). In ecology, it is desirable to understand the impact of the explanatory variables and some authors have proposed methods allowing the determination of the impact of the input variables in an ANN analysis (Garson 1991; Goh 1995; Lek et al. 1996a,b). In the present study, Garson's algorithm was used to quantify the percentage contribution of each variable in the models. This procedure is based on the partitioning of the connection weights to determine the relative importance of the input variables to the response of the model (output variable). The method essentially involves partitioning the hidden-output connection weights of each hidden neuron into components associated with each input neuron. The result, expressed as a percentage, gives the relative importance or distribution of all the output weights

attributable to a given input variable (see Garson 1991 and Goh 1995 for more detail).

Finally, the total fish population assemblage was investigated. Modelling was carried out after $\log_{10}(x+1)$ transformation of the dependent variables (six fish species abundances), to reduce the influence of outliers (ter Braak & Looman 1995). Ten models were run for each species to check the stability of the predictions. The influence of each of the eight environmental variables in the 60 resulting models (10 models per species \times 6 species) was quantified by means of the Garson algorithm. Then, we used this matrix to perform a normalised Principal Component Analysis (PCA) (Legendre & Legendre 1998), where the results of each model were considered as a statistical unit (Fig. 2). PCA was therefore performed on a data matrix containing 60 rows corresponding to the units (10 units per species \times 6 species) and eight columns, each accounting for one environmental variable. This analysis allowed the microhabitat of the six fish populations to be taken into account simultaneously to define their spatial occupancy and thus to describe the fish community assemblage.

RESULTS

Invertebrate diversity in tributaries of the Taieri River basin

The ANN consisted of a 3-layered (10→4→1) feed-forward network with bias. There were 10 input neurons to code the 10 independent variables. The hidden layer had four neurons, determined as the optimal configuration giving the lowest error in both training and testing procedures. The output neuron computed the value of the dependent variable (H). In the training procedure, results were highly significant for all the models with correlation coefficients and PI values close to perfection ($r = 0.94$ between observed and estimated diversity values, $P < 0.001$, $PI = 84\%$). Moreover, the majority of the points in the scatter plots are well aligned along the diagonal of best prediction (Fig. 3A). In the testing procedure (leave-one-out), prediction was less efficient, but still highly significant ($P < 0.001$). Although some samples were not accurately predicted by the models (Fig. 3B), aberrant values never appeared (i.e., negative values of the predicted variable) and both r and PI values remained high ($r = 0.71$ between observed and estimated diversity values, and $PI = 68\%$).

The results of Garson's algorithm applied to 10 models were stable as shown by low standard errors (Fig. 4). The contribution of the 10 considered variables ranged between 7 and 13%. A given variable was assumed to be important if its contribution was greater than 10% (the mean value of a theoretical homogeneous distribution of all the variables; i.e., 100% of contribution/10 variables = 10%). The results emphasise the important relative contribution of four variables in the model (each more than 10%), namely elevation, the percentage of the catchment that is barren, the median particle size of the riverbed, and lack of human influence in the riparian zone (% of tussock).

Fish microhabitat and species assemblages in Lake Pareloup

In this case the ANN consisted of a 3-layered (8→10→1) feed-forward network with bias. Eight input neurons coded the eight independent variables, the hidden layer had 10 neurons and one output neuron computed the value of the dependent variable (abundance of a fish species). We could have used a single neural network with six output neurons (one for each of the six fish species), but we preferred to use six networks with the same architecture, each one predicting the abundance of a single species. This

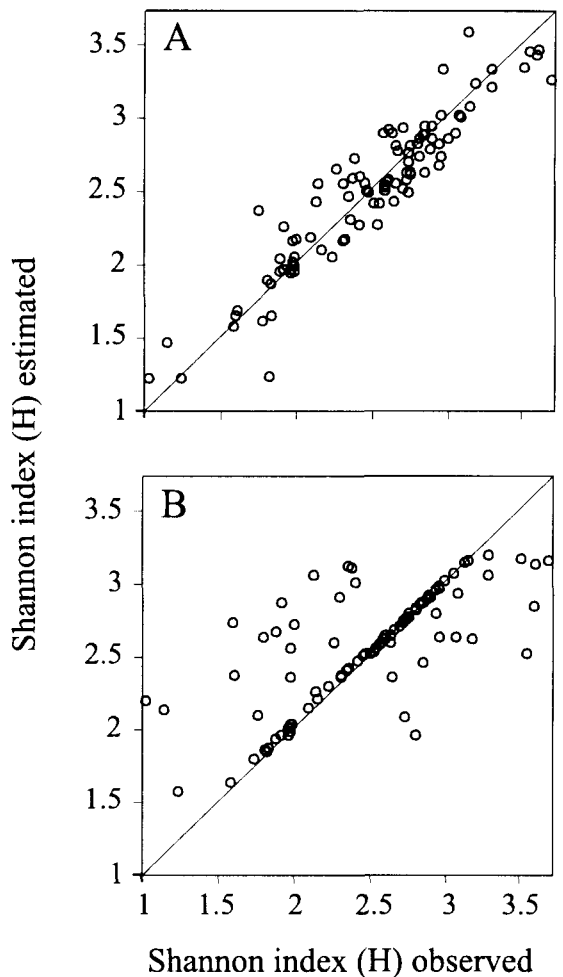


Fig. 3 Artificial neural network (ANN) model predictions of macroinvertebrate diversity (Shannon index, H) for the Taieri River study, New Zealand. Scatter plots of predicted values versus observed values are shown for the: **A**, training; and **B**, testing procedures. Solid line indicates the perfect line of fit (1:1 ratio). Statistics for the regressions are given in the text.

allowed us to easily extract from the models the influence of the eight environmental variables on each fish species and facilitated the visualisation of spatial distribution of the six species together. The ANN models yielded high correlation coefficients ($P < 0.001$) and PIs between observed and predicted values in both training and testing procedures (Table 1). In the training procedure, values for r and PI for the six species ranged from 0.63 to 0.92 and 66 to 97%, respectively. In the testing procedure, r and PI

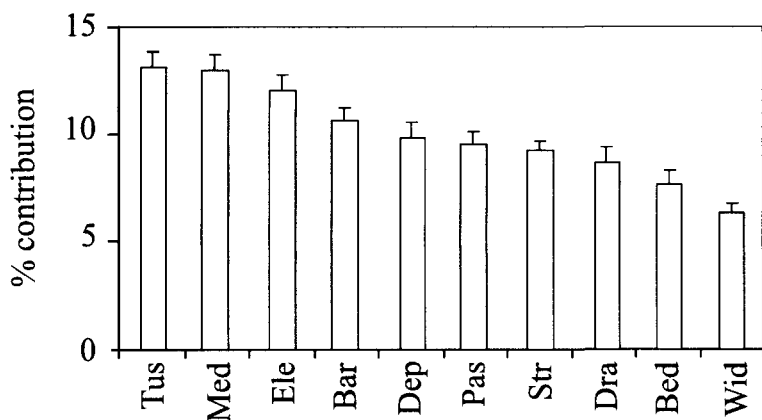


Fig. 4 Percentage contribution of each of the 10 independent variables to the prediction of macroinvertebrates diversity (Shannon index, H), obtained by Garson's algorithm. Bars indicate the mean value of the results of the 10 models and horizontal lines represent standard error. (Tus and Pas = land use (tussock and pastoral respectively); Med = median particle size in the streambed; Ele = elevation; Bar = percentage of basin barren (e.g., roads, urban areas); Dep = water depth; Str = stream order; Dra = drainage density (total stream length per unit area of basin); Bed = percentage of bedrock; and Wid = channel width.)

Table 1 Correlation coefficient (r) between observed and estimated values and performance index (PI) in artificial neural network (ANN) training and testing for the six fish populations (roach, *Rutilus rutilus*; perch, *Perca fluviatilis*; rudd, *Scardinius erythrophthalmus*; gudgeon, *Gobio gobio*; pike, *Esox lucius*). PI is the percentage of well-predicted values with an error rate lower than 10%.

	Training		Testing	
	r	PI	r	PI
0+ roach	0.79	66	0.74	63
0+ perch	0.68	72	0.61	69
0+ rudd	0.80	69	0.79	61
0+ gudgeon	0.92	97	0.84	96
0+ pike	0.72	90	0.62	91
Adult perch	0.63	94	0.61	91

ranged from 0.61 to 0.84 and 61 to 96%, respectively. Different sets of environmental variables in the 60 training models (10 models per species, 6 species), assessed using the Garson algorithm, evidently played important roles in the distribution of the various species (Table 2). Standard errors, calculated for each variable from 10 training procedures, were very low, indicating high stability of the ANN models. Moreover, for most species, fish microhabitat was defined by several variables, showing that microhabitat use is a result of a complex combination of habitat characteristics. Only 0+ gudgeon had a simple habitat relationship, with a single variable, distance from the bank, contributing in a major way (c. 50%; Table 2).

Table 2 Mean values of the percentage contribution (\pm standard error) of each of the eight independent variables to the prediction of the six fish population densities (roach, *Rutilus rutilus*; perch, *Perca fluviatilis*; rudd, *Scardinius erythrophthalmus*; gudgeon, *Gobio gobio*; pike, *Esox lucius*), obtained by Garson's algorithm applied to the results of the 10 models for each fish population. Mean contributions superior than 20% are bolded.

	0+ roach	0+ perch	0+ rudd	0+ gudgeon	0+ pike	Adult perch
Depth	11.77 \pm 0.67	15.55 \pm 0.51	19.25 \pm 0.48	10.30 \pm 0.62	9.78 \pm 0.50	21.17 \pm 0.60
Distance	19.36 \pm 0.75	22.52 \pm 0.68	14.47 \pm 0.35	52.29 \pm 0.93	26.50 \pm 0.87	17.94 \pm 0.99
Slope	14.03 \pm 0.89	19.50 \pm 0.71	12.96 \pm 0.41	2.72 \pm 0.26	14.20 \pm 0.61	15.53 \pm 0.81
Boulders	3.24 \pm 0.16	2.18 \pm 0.13	5.29 \pm 0.16	2.22 \pm 0.17	1.65 \pm 0.08	6.58 \pm 0.34
Pebbles	5.13 \pm 0.25	4.15 \pm 0.16	5.09 \pm 0.20	2.42 \pm 0.09	2.02 \pm 0.10	3.59 \pm 0.15
Gravel	5.31 \pm 0.34	3.37 \pm 0.14	5.16 \pm 0.26	2.72 \pm 0.07	1.83 \pm 0.10	2.75 \pm 0.15
Mud	16.35 \pm 0.45	14.20 \pm 0.44	13.01 \pm 0.50	16.99 \pm 0.46	19.48 \pm 0.87	19.93 \pm 0.56
Vegetation	24.80 \pm 0.81	17.91 \pm 0.67	24.78 \pm 0.65	10.35 \pm 0.46	24.55 \pm 0.58	12.51 \pm 0.37

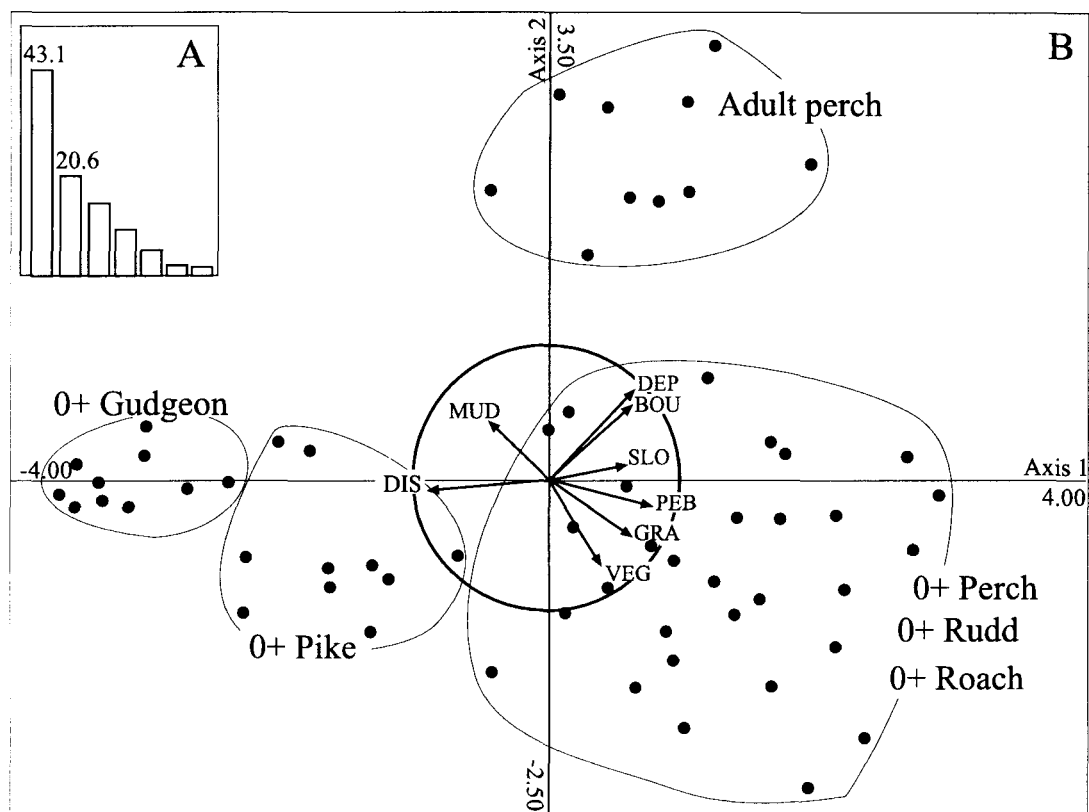


Fig. 5 Principal Component Analysis (PCA) performed on artificial neural network (ANN) results using Garson's algorithm for the six fish populations. For each population, the statistical units (i.e., fish populations) were the results of the 10 ANN models. **A**, Histogram of eigenvalues; **B**, distribution of the six fish populations (roach, *Rutilus rutilus*; perch, *Perca fluviatilis*; rudd, *Scardinius erythrophthalmus*; gudgeon, *Gobio gobio*; pike, *Esox lucius*) 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 \times F2 plane.

Finally, to visualise the simultaneous spatial distributions of the six fish species (i.e., community composition) a normalised PCA was performed on the relative contributions of each independent variable in the ANN models (determined using Garson's algorithm) (Fig. 2). The first two axes accounted for 43.1 and 20.6% of total variation, respectively (Fig. 5A). The PCA revealed significant correlations ($P < 0.01$): (1) between flooded vegetation (VEG) and 0+ roach, 0+ rudd, 0+ perch, and 0+ pike; (2) between depth (DEP) and adult perch; and (3) between distance from the bank (DIS) and 0+ gudgeon. Along the first axis, 0+ gudgeon and 0+ pike were separated from the other species. Along the second axis, adult perch were separated from the overlapping assemblage of 0+ roach, 0+ rudd, and 0+ perch (Fig. 5B).

DISCUSSION

The Taieri River basin is extensive and heterogeneous and the environmental variables in our analyses were derived from geographic and catchment scales (elevation and catchment characteristics) as well as local features (bed particle size, stream width, and depth). In this large-scale study of streams, the results of our analysis of macroinvertebrate diversity were very satisfactory, with predictions from the ANN being close to perfection for most of the samples during the training procedure. The ability to make successful predictions for new samples, as tested by the cross-validation procedure, was lower than during the training phase, but more than 65% of samples were still perfectly predicted (i.e., within $\pm 10\%$ of actual). Nevertheless, some median values of invertebrate diversity were over- or underestimated

by the ANN (Fig. 3B); these were from habitats identified by the model as able to sustain a higher or lower diversity than that observed. Such discrepancies are partly because of unmeasured environmental variables; in other words, the 10 environmental variables were not able to account for the entire variability of diversity in the sampling sites. For example, the high instability of some streambeds is known to influence macroinvertebrate diversity (Townsend et al. 1997b; Matthaei et al. 1999), but this variable could not be taken into account because such information was only available for a limited number of the Taieri sites. ANN prediction capabilities are also limited by the scope of information contained in the training data set. Thus, the scarcity of samples with high Shannon index values induced an underestimation of the most diverse sites. Nevertheless, most of the points in the scatter plot are well aligned along the diagonal of best prediction of coordinates (1:1 ratio). Moreover, all observed samples with medium or high diversity values were predicted as able to sustain moderate or high invertebrates diversity, even if some values were over- or underestimated (Fig. 3).

The patterns of invertebrate diversity indicated by the sensitivity analysis of the environmental variables (Fig. 4) are in accord with other results in the literature. For example, sites in the Taieri associated with a high percentage of native tussock can be considered least disturbed by human activities and correspondingly high diversities are to be expected (Ormerod et al. 1993; Allan 1995; Townsend et al. 1997a). In the same way, the percentage of barren areas (e.g., roads, urban areas) is likely to relate to a decrease in diversity as a result of point-source pollution (Hynes 1960; Williams & Feltmate 1992). Median particle size was also a major variable in the models. Streams with a higher median particle size also possess greater heterogeneity in particle sizes and streambed habitat heterogeneity (C. R. Townsend unpubl. data), factors that can be expected to be associated with higher faunal diversity (Williams & Mundie 1978; Williams 1980; Townsend 1989; Townsend et al. 1997b; Vinson & Hawkins 1998). Finally, elevation proved to be a highly influential macro-scale factor, as it has in other studies (Ward 1986; Allan 1995; Jacobsen et al. 1997). Within individual basins, insect diversity has been reported to decrease, change irregularly or increase with elevation, as a result of a complex of concomitant changes in temperature regime and other influential factors (see review by Vinson & Hawkins 1998). It appears that the ANN model

approach can not only predict stream invertebrate diversity, but also identify the ecological importance of the environmental variables introduced in the model. On this basis, an interesting future step in the application of ANN to ecology will be to identify the influence of anthropogenic disturbance parameters on biotic communities, with the expectation of developing bio-assessment methodologies using a "reference condition" approach (Reynoldson et al. 1997). Thus, ANN models could be developed using information measured in undisturbed reference sites (environmental parameters and invertebrate diversity), and deviations between reference and test sites may be interpreted with respect to potential anthropogenic impacts.

In contrast to the large scale of the Taieri River study, analysis of fish populations in the French lake involved microhabitat variables. Nevertheless, fish abundances were, once again, reliably fitted by ANN to the measured environmental characteristics of the points sampled in the lake (Table 1). From an ecological point of view, the contribution of each environmental variable to the models for the six species adds weight to patterns described in the literature. Thus, 0+ roach, rudd, and perch were strongly influenced by distance from the bank, depth, and the presence of vegetation cover, being associated with shallow littoral areas that provide both shelter from predators and a rich foraging habitat (Copp 1992; Christensen & Persson 1993; Hosn & Downing 1994; Persson & Eklöv 1995; Brosse & Lek 2000). The influence on 0+ pike of distance from the bank and vegetation cover accords with its feeding behaviour, pike usually staying hidden under vegetation waiting for prey (Holland & Hudson 1984; Turner & Mackay 1985). Moreover, according to Eklöv (1997), occupation by 0+ pike of shallow areas is likely to afford protection against predation by larger pike. The importance for 0+ gudgeon of distance from the bank and its indifference to vegetation cover parallels results from lowland and piedmont rivers (Mastrorillo et al. 1996). Finally, adult perch distribution was strongly influenced by depth, percentage of mud, and distance from the bank (boulder cannot be considered to be an important variable as it contributes only c. 7% of the total information in the adult perch models (Table 2)); this species usually colonises open water located outside vegetation cover where it preys on invertebrates (e.g., chironomid larvae) that colonise muddy areas (Persson 1983; Persson & Eklöv 1995). In the same way, the fish spatial assemblage visualised on the PCA plane using the ANN results

(Fig. 5) adds weight to conclusions from various ecological studies concerning the microhabitat of these species (Persson 1983; Copp 1992; Hosn & Downing 1994; Persson & Eklöv 1995; Mastrorillo et al. 1996). For example, the separation of 0+ gudgeon and top-predators (i.e., 0+ pike and adult perch) from the other fish has been observed in natural environments (Persson 1983; Persson & Eklöv 1995; Mastrorillo et al. 1996). Similarly, the syntopy of 0+ roach, 0+ rudd, and 0+ perch has been described elsewhere (e.g., Diehl & Eklöv 1995; Persson & Eklöv 1995). We conclude that fish community composition was reliably predicted using ANN and this predicted spatial occupancy could be accurately visualised on a PCA plane. Thus, ANN is able to reproduce the operation of real, complex multi-species systems on the basis of the ecological variables introduced to the models.

The back-propagation ANN approach is evidently an efficient tool to predict abundance, diversity, and community composition in both large and small spatial scale studies. It is this ability to deal with multiple information sources that provides the power of the approach, resulting in a significant improvement in ANN modelling over conventional predictive techniques (see Lek et al. 1996a,b; Mastrorillo et al. 1997; Brosse et al. 1999b,c; Brosse & Lek 2000 for comparisons of techniques). However, some drawbacks need to be borne in mind, such as the need for a large data set with numerous observations to build the models. Further work is also needed to provide more explanatory power concerning the relationships between independent and dependent variables in the models. The work of Dimopoulos et al. (1999), which proposes a method based on partial derivatives of the model response to illustrate the sensitivity of the variables, may provide a way forward. Nevertheless, the successful application of ANN at various spatial scales, and for a range of aquatic ecosystems (lakes and rivers), organisms (invertebrates and fish), and ecological descriptors (abundance, Shannon diversity index, and community composition) demonstrated in this study opens new fields for the application of ANN in aquatic ecology.

ACKNOWLEDGMENTS

We are grateful to Chris Arbuckle for help with access to the Taierni invertebrate and GIS data bases. We thank two anonymous reviewers for their valuable and constructive comments which

improved the manuscript. This research was supported in part by a post-doctoral grant to S. Brosse from the Ecological Research Group of the University of Otago.

REFERENCES

- Albiol, J.; Campmajo, C.; Casas, C.; Poch, M. 1995: Biomass estimation in plant cell cultures: a neural network approach. *Biotechnology Progress* 11: 88–92.
- Allan, J. D. 1995: Stream ecology: structure and function of running waters. London, Chapman & Hall. 338 p.
- Baran, P.; Lek, S.; Delacoste, M.; Belaud, A. 1996: Stochastic models that predict trout population densities or biomass on microhabitat scale. *Hydrobiologia* 337: 1–9.
- 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.; Guégan, 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. 2000: Modelling roach (*Rutilus rutilus* L.) microhabitat using linear and non-linear techniques. *Freshwater Biology* 44: 441–452.
- 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 antipredator behaviours: effect of prey choice in different habitats. *Behavioural Ecology and Sociobiology* 32: 1–9.
- Chu, W. C.; Bose, N. K. 1998: Speech signal prediction using feedforward neural-network. *Electronics Letters* 34: 999–1001.
- Colasanti, R. L. 1991: Discussions of the possible use of neural network algorithms in ecological modelling. *Binary* 3: 13–15.
- 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. 1992: Comparative microhabitat use of cyprinid larvae and juveniles in a lotic floodplain channel. *Environmental Biology of Fishes* 33: 181–193.
- Dekruger, D.; Hunt, B. R. 1994: Image-processing and neural networks for recognition of cartographic area features. *Pattern Recognition* 27: 461–483.
- Diehl, S.; Eklöv, P. 1995: Effects of piscivore-mediated habitat use on resources, diet, and growth of perch. *Ecology* 76: 1712–1726.
- Dimopoulos, I.; Chronopoulos, J.; Chronopoulou-Sereli, A.; Lek S. 1999: Neural networks models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). *Ecological Modelling* 120: 157–165.
- Edwards, M.; Morse, D. R. 1995: The potential for computer-aided identification in biodiversity research. *Trends in Ecology and Evolution* 10: 153–158.
- Efron, B. 1983: Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association* 78: 316–330.
- 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.
- Faraggi, D.; Simon, R. 1995: A neural network model for survival data. *Statistics in Medicine* 14: 73–82.
- Garson, G. D. 1991: Interpreting neural network connection weights. *Artificial Intelligence Expert* 6: 47–51.
- Goh, A. T. C. 1995: Back-propagation neural networks for modelling complex systems. *Artificial Intelligence Engineering* 9: 143–151.
- Green, R. H. 1979: Sampling design and statistical methods for environmental biologists. Cambridge, Cambridge University Press. 258 p.
- Guégan, J. F.; Lek, S.; Oberdorff, T. 1998: Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature* 391: 382–384.
- 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 American Journal of Fisheries Management* 4: 514–522.
- 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.
- Hynes, H. B. 1960: The biology of polluted waters. Liverpool University Press. 202 p.
- Jacobsen, D.; Schultz, R.; Encalada, A. 1997: Structure and diversity of stream invertebrate assemblages: the influence of temperature with altitude and latitude. *Freshwater Biology* 38: 247–261.
- Kohavi, R. 1995: A study of the cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Montreal, Canada. Pp. 1137–1143.
- Kvasnicka, T. 1990: An application of neural networks in chemistry. *Chemical Papers* 44: 775–792.
- Legendre, P.; Legendre, L. 1998: Numerical ecology. Elsevier. 853 p.
- 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 modeling nonlinear relationships in ecology. *Ecological Modelling* 90: 39–52.
- Lek, S.; Guégan, J. F. 2000: Artificial neuronal networks, applications to ecology and evolution. Springer-Verlag. 244 p.
- Lo, J. Y.; Baker, J. A.; Kornguth, P. J.; Floyd, C. E. 1995: Application of artificial neural networks to interpretation of mammograms on the basis of the radiologists impression and optimized image features. *Radiology* 197: 242–242.
- 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. *Annales de Limnologie* 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 Biology* 38: 237–246.
- Matthaei, C. D.; Peacock, K. A.; Townsend, C. R. 1999: Patchy surface stone movement during disturbance in a New Zealand stream and its potential significance for the fauna. *Limnology and Oceanography* 44: 1091–1102.
- Nelva, A.; Persat, H.; Chessel, D. 1979: Une nouvelle méthode d'étude des peuplements ichthyologiques dans les grands cours d'eau par échantillonnage ponctuel d'abondance. *Comptes Rendus de l'Académie des Sciences, Paris, Serie III* 289: 1295–1298.
- Norris, R. H. 1995: Biological monitoring: the dilemma of data analysis. *Journal of the North American Benthological Society* 14: 440–450.

- Ormerod, S. J.; Rundle, S. D.; Lloyd, E. C.; Douglas, A. A. 1993: The influence of riparian management on the habitat structure and macroinvertebrate communities of upland streams draining plantation forests. *Journal of Applied Ecology* 30: 13–24.
- Persson, L. 1983: Food consumption and competition between age classes in a perch *Perca fluviatilis* population 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.
- Rahim, M. G.; Goodyear, C. C.; Kleijn, W. B.; Schroeter, J.; Sondhi, M. M. 1993: On the use of neural networks in articulatory speech synthesis. *Journal of the Acoustical Society of America* 93: 1109–1121.
- Recknagel, F.; French, M.; Harkonen, P.; Yabunaka, K. I. 1997: Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling* 96: 11–28.
- Reynoldson, T. B.; Norris, R. H.; Resh, V. H.; Rosenberg, D. M. 1997: The reference condition: a comparison of multimetric and multivariate approaches to assess water quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society* 16: 833–852.
- Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. 1986: Learning representations by back-propagating errors. *Nature* 323: 533–536.
- Scardi, M. 1996: Artificial neural networks as empirical models for estimating phytoplankton production. *Marine Ecology Progress Series* 139: 289–299.
- 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. *Ecological Modelling* 120: 271–286.
- Seginer, I.; Boulard, T.; Bailey, B. J. 1994: Neural network models of the greenhouse climate. *Journal of Agricultural Engineering Research* 59: 203–216.
- 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. ed. *Data analysis in community and landscape ecology*. Cambridge University Press. Pp. 29–77.
- Townsend, C. R. 1989: The patch dynamics concept of stream community ecology. *Journal of the North American Benthological Society* 8: 36–50.
- Townsend, C. R.; Arbuttle, C. J.; Crowl, T. A.; Scarsbrook, M. R. 1997a: The relationship between land use and physicochemistry, food resources and macroinvertebrates communities in tributaries of the Taieri River, New Zealand: a hierarchically scaled approach. *Freshwater Biology* 37: 177–191.
- Townsend, C. R.; Harper, J. L.; Begon, M. 2000: *Essentials of ecology*. Blackwell Science. 552 p.
- Townsend, C. R.; Scarsbrook, M. R.; Doledec, S. 1997b: The intermediate disturbance hypothesis, refugia and biodiversity in streams. *Limnology and Oceanography* 42: 938–949.
- Turner, L. J.; Mackay, W. C. 1985: Use of visual census for estimating population size in northern pike (*Esox lucius*). *Canadian Journal of Fisheries and Aquatic Sciences* 42: 1835–1840.
- Vinson, M. R.; Hawkins, C. P. 1998: Biodiversity of stream insects: variations at local, basin and regional scales. *Annual Review of Entomology* 43: 271–293.
- Ward, J. V. 1986: Altitudinal zonation in a Rocky mountain stream. *Archiv fur Hydrobiologie* 74: 133–199.
- Williams, D. D. 1980: Some relationships between stream benthos and substrate heterogeneity. *Limnology and Oceanography* 25: 166–172.
- Williams, D. D.; Feltmate, B. W. 1992: *Aquatic insects*. CAB International. 358 p.
- Williams, D. D.; Mundie, J. H. 1978: Substrate size selection by stream invertebrates, and the influence of sand. *Limnology and Oceanography* 23: 1030–1033.
- Winterbourn, M. J.; Greston, K. L. D. 1989: Guide to the aquatic insects of New Zealand. *Bulletin of the Entomological Society of New Zealand* 9. 97 p.
- Wolman, M. G. 1954: A method of sampling coarse river-bed material. *Transactions of the American Geophysical Union* 35: 951–956.