J. Appl. Ichthyol. 24 (2008), 150–162
© 2008 The Authors
Journal compilation © 2008 Blackwell Verlag, Berlin
ISSN 0175–8659

Received: December 3, 2006
Accepted: June 14, 2007
doi: 10.1111/j.1439-0426.2007.01039.x

Estimating fish community diversity from environmental features in the Tagus estuary (Portugal): Multiple Linear Regression and Artificial Neural Network approaches

By J. C. Gutiérrez-Estrada[1], R. Vasconcelos[2] and M. J. Costa[2]

[1]*Departamento de Ciencias Agroforestales, Escuela Politécnica Superior, Campus Universitario de La Rábida, Universidad de Huelva, Palos de la Frontera, Huelva, Spain;* [2]*Instituto de Oceanografia / Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, Lisboa, Portugal*

Summary

Relationships between environmental variables and diversity (Shannon-Weaver index) of the fish communities in the Tagus estuary and adjacent coastal areas were analyzed. The focus was on the linearity or nonlinearity of these abiotic / biotic characteristics, with the aim to obtain an accurate short–medium term time-scale diversity prediction from habitat variables alone. Multiple Linear Regressions (MLR) were used for the linear approach and Artificial Neural Networks (ANNs) for the nonlinear approach. MLR results in the external validation phase indicated a lack of model accuracy (R^2 = 0.0710; %SEP = 47.5868; E = −0.0217; ARV = 1.0217; N = 43). Results of the best of the Artificial Neural Networks used in this study (12-15-15-1 architecture) in the external validation phase (ANN: R^2 = 0.9736; %SEP = 7.8499; E = 0.9722; ARV = 0.0278; N = 43) were more accurate than those obtained with MLR. This indicates a clear nonlinear relationship between variables. In the best ANN model, nitrate concentration, depth, dissolved oxygen and temperature were the most important predictors of fish diversity in the Tagus estuary. The sensibility analysis indicated that the remaining variables (silicate, nitrite, transparency, salinity, slope, phosphate, water particulate organic matter, and chlorophyll *a*) played lesser roles in the model.

Introduction

Estuaries are transition zones with an extreme spatial and temporal variability of physical, chemical and biological characteristics. In these transition systems, the diversity of fish communities is an important feature in the system dynamics because changes in diversity reflect changes in the ecosystem processes, such as productivity, energy pathways and material flow, disturbance regimes, abiotic stress and, biological interactions. Thiel et al. (1995), Marshall and Elliott (1998), Paperno and Brodie (2004), and Vega-Cendejas and Hernández-de-Santillana (2004) reported that distribution, abundance and composition of estuarine fish communities are determined by the physico-chemical characteristics of the water. This suggests that diversity is dependent on ecosystem characteristics such as salinity, temperature, water oxygen concentration, and consequently is affected by ecological processes (Brown et al., 2001; Kupschus and Tremain, 2001; Kupschus, 2004). Therefore, from an ecological point of view, the impact analysis of water physico-chemical properties on the distribution and abundance patterns of estuarine fish populations is very important (Guisan and Zimmermann, 2000; Lekve et al., 2003; Prista et al., 2003). Furthermore, these characteristics are not only being influenced by the dynamics of the estuarine system (Margalef, 1968; Vinagre et al., 2004; Baeta et al., 2005), but are also related to human disturbances of the habitat. Many human uses of estuaries (i.e. transportation, wastes, pollution) are potentially in conflict with the aquatic living resources. In addition, human activities in estuaries not only have an impact on fish communities, but in the same way also rebound in human communities associated with the exploitation of these resources (Costa and Elliott, 1991). Thus, fish community diversity is a basic ecological aspect, knowledge of which is necessary for the correct exploitation, regulation and management of fishing resources since it can provide a first approach to the health level of the estuarine system and allows for the identification of response patterns to possible environmental impacts.

The calculation of a diversity index (such as the Shannon-Weaver diversity index) is generally simple as it only requires information on the species number (richness) and the number of individuals for each species (abundance). However, the collection of explicit data of richness and abundance in large and complex environments, as in the Tagus estuary, requires a major sampling effort which is undoubtedly expensive and time-consuming. It is therefore very difficult to have a long and continuous time series diversity index allowing identification of temporal variations (i.e. trend changes to short and medium time of the average diversity) in response to possible environmental impacts. A possible alternative to mitigating this problem could be to develop a diversity index forecast, working with variables of habitat characteristics that are easy to measure and where measuring costs are low.

Few studies have approached diversity forecasting from variables of habitat characteristics (Angermeier and Winston, 1998; Lek-Ang et al., 1999; Attrill, 2002). Some authors proposed different kind of models for prediction of diversity and other population variables such as richness, abundance or biomass. Tilman (1982), Schoener (1983), MacArthur et al. (1996), Maes et al. (1998) and Prista et al. (2003) used a wide range of multivariate techniques such as principal component analysis, canonical correspondence analysis and logistic regressions. Power et al. (2000a,b, 2002), Power and Attrill (2002, 2003), and Maes et al. (2004) used linear multiple regressions to forecast some aspects of community structure of

U.S. Copyright Clearance Centre Code Statement: 0175–8659/2008/2402–0150$15.00/0

fish starting from environmental variables. In these cases, a direct linear relationship between independents and dependent variables was obtained.

Stenseth et al. (2002) reported that ecological systems may be heavily influenced by substantial stochastic processes, as well as by nonlinearities. In this state, the conceptual framework on which the formulation of the ecological model is developed may be based on pseudoequilibrium or non-equilibrium assumptions. In this way, Lek et al. (1996a,b), Baran et al. (1996), Guégan et al. (1998), Gozlan et al. (1999), Lek-Ang et al. (1999), Gutiérrez-Estrada et al. (2000) and Brosse et al. (2001) proposed a nonlinear approach to the ecological-biological variables forecasting, and obtained better results than with the classical linear models.

Significant progress in the fields of nonlinear pattern recognition and system control theory have recently been made possible through the advances in a branch of nonlinear system theoretic modelling called Artificial Neural Networks (ANNs; Lek et al., 1996b). In recent years, ANNs have attracted increasing attention from both academic researchers and industrial practitioners. The reason for their popularity is the powerful pattern recognition and flexible nonlinear modelling capacity (Qi and Zhang, 2001). ANNs have a great capacity to fit highly scattered data, and produce powerful models from few data, thus providing reliable predictions (Govindajaru, 2000).

In this paper some hypotheses relating to the diversity of fish communities in the Tagus estuary were investigated:

1. The relationship between environmental variables and diversity of fish communities is mainly nonlinear. For this purpose, the ANN approach is applied to forecast the diversity of fish communities from environmental variables in a short–medium time-scale. To contrast the nonlinear hypothesis the ANN approach is compared with traditional linear multiple regression analysis.
2. The diversity of fish communities in the Tagus estuary can be estimated solely from environmental variables.
3. The mechanisms for the variation of diversity can be identified.

Methods

Study area and sampling

The Tagus estuary is the largest estuary on the Portuguese coast and one of the largest in Europe. It extends for almost 80 × 4 km (approximately 320 km^2), of which about 40% is intertidal with a mesotidal regime (1.5 to 3.5 m). The middle of the estuarine area is very wide, whereas the estuarine mouth is formed by a narrow channel 12 km in length, with a width of 2 km and a depth of about 30 m. The average river flow in recent decades was around 300 ± 120 m^3 s^{-1} (average ± SD).

Sampling was carried out every 2 months between May 2001 and March 2002 at 30 sampling sites located throughout the Tagus estuary and its adjacent coastal area (Fig. 1). Fish were captured using an otter trawl with an 18-mm mesh size at the cod-end operated aboard commercial fishing vessels. A 15-min haul was made in each site during ebb tide. Depth, temperature, salinity and dissolved oxygen concentration were measured at 50 cm from the estuary bottom before and after each tow using a multiparameter probe. Before the tow, water transparency was measured with a Secchi disc and 10-L water samples from the bottom (50 cm) and surface were collected with a horizontal bottle for determining chlorophyll *a* and suspended solids. The average bottom slope in the drag trajectory was estimated from the initial and final depth of each sampling site.

Water samples were filtered through fiberglass filters (three replicate filters used for chlorophyll *a* determination; three replicate filters used for water particulate organic matter determination). Chlorophyll *a* extraction was performed in

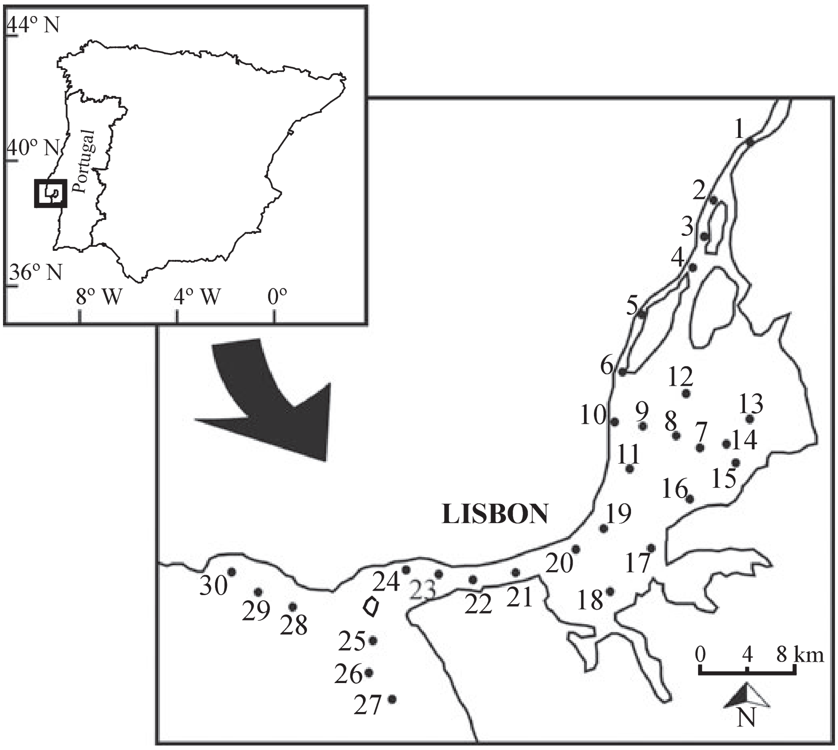

Fig. 1. Tagus estuary and adjacent coastal area with numbered locations of sampling sites (1–24 inside the estuary; 25–30 in adjacent coastal area)

90% aqueous acetone solution and the chlorophyll *a* quantified according to the Lorensen spectrophotometric method (Lorensen, 1967). Water particulate organic matter (WPOM, mg L^{-1}) was determined according Cleresci et al. (1989) by drying the filters in an oven (60°C, 24 h) followed by incineration in a muffle (500°C, 4 h). Around 200 ml of each water sample were filtered through cellulose acetate filters and frozen (−20°C) for nutrient determination. Water nitrate (NO_3), nitrite (NO_2), phosphate (PO_4) and silicate (SiO_4) concentrations were determined according to Grasshoff et al. (1999).

Diversity modeling

All captured fish were identified, counted and measured. From these data, fish diversity (*H*, dependent variable) was calculated for each day and sampling site according to the Shannon-Weaver formula:

$$H = -\sum_{z=1}^{s} p_z \log_2 p_z \tag{1}$$

where *p* is the relationship between the total number of fishes of the species *z* and total number of fishes in the sampling, and *s* is the number of species.

The set of independent variables comprised: depth (*D*, m), slope (*S*, ‰), temperature (*T*, °C), salinity (Sal), dissolved oxygen (Ox, mg L^{-1}), nitrite (NO_2, mg L^{-1}), nitrate (NO_3, mg L^{-1}), phosphate (PO_4, mg L^{-1}), silicate (SiO_4, mg L^{-1}), chlorophyll *a* (Ch *a*, μg L^{-1}), Secchi disc (Se, m) and water particulate organic matter (WPOM, mg L^{-1}). The data matrix was composed of 174 records (6 months × 30 sample sites = 180 records minus 6 anomalous cases = 174 records) and the 12 environmental variables. Before the calibration of the models, an exploratory analysis of the correlation matrix of all variables was performed to examine expected and unexpected significant relationships. Later, the models were obtained with Multiple Linear Regression (MLR) and Artificial Neural Networks (ANNs). In both cases, a random calibration set (75% of the records; 131 patterns) and an external validation set (25% of the records; 43 patterns) were selected.

Linear approach: Multiple Linear Regression (MLR)

The multiple regression procedure was utilized to estimate the regression coefficients $b_0, b_1,..., b_q$ of the linear equation:

$$e = b_0 + b_1x_1 + \ldots\ldots + b_qx_q \tag{2}$$

where the regression coefficients $b_0, b_1,..., b_q$ represent the independent contributions of each independent variable $x_1,...., x_q$ to the prediction of the dependent variable *e*. The global statistical significance of the relationship between *e* and the independent variables was analysed by means of analysis of variance (ANOVA, α level = 0.05) to ensure the validity of the model. A stepwise procedure of the STATISTICA 6.0 tool (Statsoft, Inc., 1984–2002) was used for the MLR calibration.

Nonlinear approach: Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are mathematical models inspired by the neural architecture of the human brain. The ANNs can recognize patterns and learn from their interactions with the environment. The most widely studied and used structures are multilayer feed forward networks (Rumelhart et al., 1986). A typical four-layer feed forward ANN is shown in Fig. 2.

The processing elements in each layer are called nodes or neurons. In Fig. 2, there are *g*, *n*, *m* and *s* nodes in the input, first hidden, second hidden and output layers, respectively [the notation of the neural network is (*g*, *n*, *m*, *s*)]. The parameters associated with each of these connections are called weights. All connections are 'feed forward'; that is, they allow information transfer only from an earlier layer to the next consecutive layers.

Each node *j* receives incoming signals from every node *i* in the previous layer. Associated with each incoming signal (x_i) is a weight (W_{ji}). The effective incoming signal (I_j) to node *j* is the weighted sum of all the incoming signals, according to:

$$I_j = \sum_{i=1}^{g} x_i W_{ji} \tag{3}$$

The effective incoming signal I_j is passed through an activation function (sometimes called a transfer function) to produce the outgoing signal (y_j) of the node *j*. In this study, the linear function ($y_j = I_j$) was used in the output layer and the sigmoid nonlinear function, in the hidden layers:

$$y_j = \mathrm{f}(I_j) = \frac{1}{1 + \exp(-I_j)} \tag{4}$$

in which I_j can vary in range (-∞, ∞), and y_j is bound between 0 and 1.

To determine the set of weights a corrective-repetitive process called learning or training of the ANN is performed. This training forms the interconnections among neurons, and is accomplished by using both known inputs and outputs

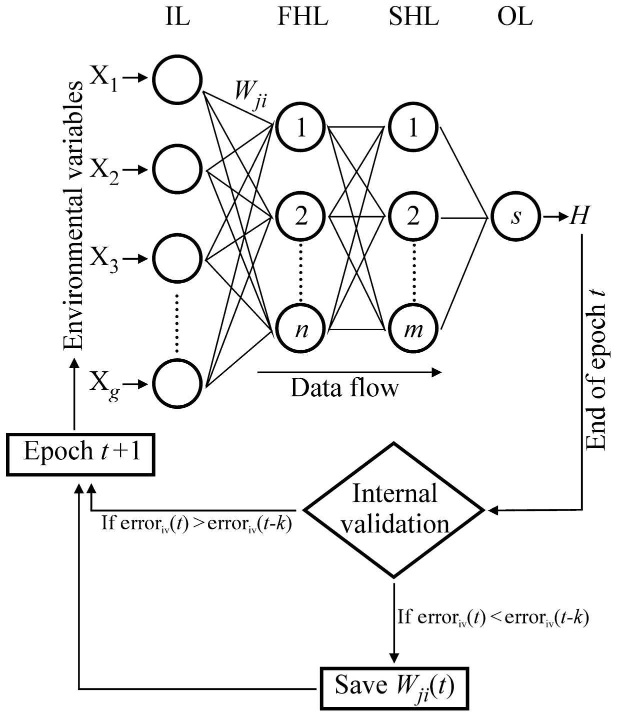

Fig. 2. Four-layer feed forward artificial neural network (ANN) with *g*, *n*, *m* and *s* nodes in the input (IL), first hidden (FHL), second hidden (SHL) and output (OL) layers, respectively. The learning process shows at the end of epoch *t* that an internal validation phase is carried out. If the internal validation error in epoch *t* is lower than the error of internal validation in epoch *t* − *k* (*k* = 1, 2,..., *t* − 1), then save the weights (W_{ji}). If this condition is not completed, then carry out a new epoch. In each epoch, weights are recalculated

(training sets or training patterns), and presenting these to the ANN in some ordered manner, adjusting the interconnection weights until the desired outputs are reached. The strength of these interconnections is adjusted using an error convergence technique so that a desired output will be produced for a given input. There are many training methods. In this work, a variation of back-propagation algorithm (Rumelhart et al., 1986), known as the Levenberg-Marquardt algorithm (Shepherd, 1997) was applied. This is a second-order nonlinear optimization algorithm with very fast convergence and is recommended by several authors (Tan and van Cauwenberghe, 1999; Anctil and Rat, 2005).

The Levenberg-Marquardt algorithm uses the following formula that is continuously updated:

$$\Delta W = -(Z^T Z + \lambda I)^{-1} Z^T \varepsilon \tag{5}$$

where ϵ is the vector of errors, Z is the matrix of the partial derivatives of these errors with respect to the weights W, and I is the identity matrix. The first term of the second member of the Levenberg-Marquardt formula represents the linear assumption and the second, the gradient-descent step. The control parameter λ governs the relative influence of these two approaches.

An epoch denotes the time period that encompasses all iterations performed after all patterns are displayed. The learning process was controlled by the method of internal validation (about 20% of calibration data to test the error at the end of each epoch; Tsoukalas and Uhrig, 1997; Gutiérrez-Estrada et al., 2004). The weights are updated at the end of each epoch (Fig. 2). The number of epochs with the smallest error of the internal validation indicates the weights to select.

Thirty neural networks were calibrated for each neural architecture (12-5-5-1; 12-5-10-1; 12-10-10-1; 12-10-15-1 and 12-15-15-1). Thus, 150 neural networks in total were calibrated and validated. Although an ANN with a single hidden layer is a structure capable of identifying complex nonlinear relationships between input and output data sets, here only ANNs with two hidden layers were used. This choice can be explained because *a priori* we do not know the type of relationship among the variables. In this way, it can happen that with a single hidden layer the number of necessary intermediate nodes to reach a certain error is so high that its application is unapproachable in practice (Müller and Reinhardt, 1990).

Average error term values of each neuronal topology on the validation phase were then compared by means of one-way ANOVA (α level = 0.05). A *post hoc* Tukey test was later performed to detect the optimum neuronal topology. The ANN models were implemented using STATISTICA 6.0 (Statsoft, Inc., 1984–2002).

External validation, measures of accuracy and sensibility analysis

Important aspects of any model are the capture of mapping data and its capacity to generalise from examples. Generalisation refers to the capacity of a model to provide a correct response to patterns that have not been employed in its calibration. Thus, a Multiple Linear Regression will have a minimum square errors sum for determined coefficient and an unknown pattern set. On the other hand, an ANN will learn until reaching the optimum point at which the generalisation error for neural network architecture and an unknown pattern set is minimum. In this study, the correct calibration of the models is controlled by an external validation, in which the estimation from a common data set not employed during the calibration of the different models is compared with observed values. This procedure allows the correct model calibration and avoids over-calibrating and over-parameterising (Tsoukalas and Uhrig, 1997).

Several measures of accuracy were calculated in the calibration and external validation phase. A measure of correlation between the observations and the predictions is the coefficient of correlation (R). The error between observation and prediction can be measured with the root mean square error (RMSE). On the other hand, measures of variances are the coefficient of determination (R^2), the percent standard error of prediction (%SEP; Ventura et al., 1995), the coefficient of efficiency (E; Kitanidis and Bras, 1980), and the average relative variance (ARV; Griñó, 1992). These estimators are not biased by the range of variations of its elements and were employed to see how far the model was able to explain the total variance of the data. The percent standard error of prediction is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \qquad \%\mathrm{SEP} = \frac{100}{\bar{y}}\mathrm{RMSE} \tag{6}$$

where y_i is the observed value; $\hat{y}_i$ is the estimated value; N is the total number of observations and $\bar{y}$ is the mean of the observed values. The main advantage of %SEP is that it has no dimension. This characteristic allows the comparison of the forecast from different models and different problems. The efficiency E coefficient and ARV were used to see how the model explains the total variance of the data and represents the proportion of variation of the observed data considered for modeling. The efficiency coefficient and ARV are defined as:

$$E = \frac{S_{obs} - CE}{S_{obs}} \qquad \mathrm{ARV} = \frac{CE}{S_{obs}} \tag{7}$$

$$S_{obs} = \sum_{i=1}^{N}(y_i - \bar{y})^2$$

$$CE = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where $\bar{y}$ is mean of the observed data. For a perfect fit, R^2 and E values should be close to 1 and %SEP and ARV values close to 0.

Other indexes used in the identification of the best linear and nonlinear models were the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (Qi and Zang, 2001):

$$\mathrm{AIC} = \log(\hat{\sigma}^2_{\mathrm{MVE}}) + \frac{2m}{N} \qquad \mathrm{BIC} = \log(\hat{\sigma}^2_{\mathrm{MVE}}) + \frac{m\log(N)}{N} \tag{8}$$

$$\hat{\sigma}^2_{\mathrm{MVE}} = \frac{CE}{N}$$

In this case, $\hat{\sigma}^2_{\mathrm{MVE}}$ is the maximum likelihood estimate of residual term variance and m is the number of parameters of the model (in ANN, the number of weights is considered the number of parameters). The first part $[\log(\sigma^2_{\mathrm{MVE}})]$ measures the goodness-of-fit of the model to the data; the second part $[2m/\mathrm{N}$ and $m\ \log(\mathrm{N})/\mathrm{N}]$ sets a penalty for model over-parameterisation. The optimal model is selected when AIC and BIC are minimized.

In this work, an alternative form of sensibility analysis based on the approach of the missing value problem was used. The sensibility analysis was carried out by replacing each variable by missing values and assessing the effect upon the output error (ratio = error of the model with a variable with missing values / error of the model with all variables). In this way, for a variable *x*, a ratio with a value equal to 1 or very close to 1 indicates that this variable has a very low weight in the general structure of the model (Hunter et al., 2000).

Results

Totals of 8211 fishes belonging to 31 families and 63 species were caught in the Tagus estuary and adjacent coastal area. Total biomass was 1404 kg. Concerning absolute abundance, the most important species were: *Callionymus lyra* (12.37%), *Pomatoschistus minutus* (11.63%), *Diplodus bellottii* (9.6%), *Liza ramada* (8.91%), *Halobatrachus didactylus* (7.28%), and *Argyrosomus regius* (6.9%). Higher contributions in biomass were *A. regius* (23.35%), *H. didactylus* (21.83%), *Barbus bocagei* (16.75%), *L. ramada* (8.13%), *Raja undulata* (7.12%), and *D. bellottii* (4.43%). In relation to the percentages of occurrences, *Argyrosomus regius* was fished in 93.3% of landings followed by *Pomatoschistus minutus* (90%), *Sardina pilchardus* (86.6%), *Halobatrachus didactylus* (73.3%), *Solea solea* (73.3%), *Solea senegalensis* (70%), *Diplodus bellottii* (66.6%), *Pomatoschistus microps* (66.3%) and *Diplodus vulgaris* (60%). The environmental variables, chlorophyll *a* concentration and transparency, showed the highest variation coefficients (145.1% and 137.6%, respectively). The average Shannon-Weaver index value was 1.14 ± 0.52 ($n = 173$), with a maximum value of 2.29 and minimum of 0 (Table 1).

The exploratory correlation analysis among variables showed that in some cases the correlation coefficients were significant ($P < 0.05$ and $P < 0.01$). However, in general the correlation values were not very high (Table 2). Among all variables, nitrate concentration showed a significant correlation with all other variables, except with chlorophyll *a*. The highest correlation value ($R = -0.9$; $P < 0.01$) was found between nitrate concentration and salinity. Absolute correlation values higher than 0.6 were found between the concentration of nitrate and phosphate (R NO_3-PO_4 = 0.64, $P < 0.01$), nitrate and silicate (R NO_3-SiO_4 = 0.64, $P < 0.01$), phosphate and salinity (R Sal-PO_4 = −0.69, $P < 0.01$), and silicate and salinity (R Sal-SiO_4 = −0.61, $P < 0.01$). Correlation values very close to 0.6 were found between phosphate and silicate concentrations ($R = 0.59$; $P < 0.01$), depth and transparency ($R = 0.59$; $P < 0.01$) and water temperature and dissolved oxygen ($R = -0.58$; $P < 0.01$; Table 2).

Correlation values between independent variables and the Shannon-Weaver index were significant only concerning salinity ($R = 0.26$; $P < 0.01$), nitrite ($R = -0.23$; $P < 0.01$), and nitrate ($R = -0.24$; $P < 0.01$) (Table 2).

Multiple regression analysis

Multiple regression analysis between the Shannon-Weaver index and the 12 environmental variables revealed five variables with a significant weight in the calibration phase (D, $P < 0.001$; T, $P = 0.0345$; Sal, $P = 0.0254$; Ox, $P < 0.01$; NO_2, $P = 0.0119$). The adjusted variance level was lower than 0.18. In the validation phase, the results were similar to those obtained in the calibration phase and the determination coefficient was lower than 0.07, with an insignificant relationship between the dependent (Shannon-Weaver index estimated) and independent variable (Shannon-Weaver index observed in the external validation set;

Table 1
Average, standard deviation (SD), maximum, minimum and Pearson coefficient of variation (SD / average; %) of independent variables (*D*, depth; *S*, slope; *T*, water temperature; Sal, Salinity; Ox, oxygen dissolved; NO_2, nitrite concentration; NO_3, nitrate concentration; PO_4, phosphate concentration; SiO_4, silicate concentration; Ch a, chlorophyll *a*; Se, Sechhi; WPOM, water particulate organic matter), and Shannon-Weaver index (*H*)

	D	*S*	*T*	Sal	Ox	NO_2	NO_3	PO_4	SiO_4	Ch *a*	Se	WPOM	*H*
Average	7.49	0.13	17.26	27.17	7.31	1.75	1.18	2.63	16.09	3.25	99.71	11.59	1.14
SD	4.60	0.12	2.45	8.41	1.40	1.21	1.35	1.92	10.80	4.72	137.20	10.46	0.52
Max	18.74	0.82	22.90	36.24	9.63	7.81	5.15	14.93	53.76	28.95	690.00	103.00	2.29
Min	0.48	0.00	13.49	0.34	3.53	0.16	0.00	0.29	1.29	0.00	0.00	0.00	0.00
CV	61.33	87.04	14.17	30.97	19.08	68.95	113.67	72.92	67.17	145.10	137.60	90.21	45.77

Table 2
Correlation matrix between Shannon-Weaver index (*H*) and independent variables

	D	*S*	*T*	Sal	Ox	NO_2	NO_3	PO_4	SiO_4	Ch *a*	Se	WPOM
S	0.19*											
T	−0.08	0.00										
Sal	0.46**	0.21**	−0.04									
Ox	−0.11	0.00	−0.58**	−0.14								
NO_2	−0.42**	−0.09	−0.13	−0.44**	0.11							
NO_3	−0.45**	−0.20**	−0.21**	−0.90**	0.31**	0.55**						
PO_4	−0.47**	−0.07	−0.12	−0.69**	0.18*	0.35**	0.64**					
SiO_4	−0.54**	−0.15*	−0.05	−0.61**	0.11	0.52**	0.64**	0.59**				
Ch a	−0.16*	−0.07	0.46**	−0.27**	0.01	−0.05	0.10	0.17*	0.09			
Se	0.59**	−0.04	−0.10	0.38**	0.13	−0.37**	−0.35**	−0.39**	−0.45**	−0.09		
WPOM	−0.31**	−0.10	−0.12	−0.16*	0.07	0.10	0.18*	0.33**	0.18*	0.07	−0.27**	
H	−0.02	0.10	−0.07	0.26**	−0.11	−0.23**	−0.24**	−0.09	−0.10	−0.09	0.12	0.08

Values are Pearson correlation coefficients.
$^*P < 0.05$; $^{**}P < 0.01$.

Table 3
Multiple regression model (calibration and external validation phase) for Shannon-Weaver index (H) and the environmental variables

Dependent variable	Independent variables	B	P-value
Calibration phase: $R = 0.5017$; $R^2 = 0.2517$; adjusted $R^2 = 0.1756$; $F(12,118) = 3.3077$; $P < 0.01$; $N = 131$			
H	Intercept	2.3439	0.0187*
	D	−0.0510	0.0002**
	S	0.4982	0.2349
	T	−0.0598	0.0346*
	Sal	0.0336	0.0254*
	Ox	−0.1165	0.0084**
	NO_2	−0.1110	0.0119*
	NO_3	0.1210	0.2061
	PO_4	0.0115	0.7705
	SiO_4	0.0006	0.9202
	Ch a	0.0054	0.6190
	Se	0.0007	0.0910
	WPOM	0.0028	0.5232
External validation phase: $R = 0.2665$; $R^2 = 0.0710$; adjusted $R^2 = 0.0483$; $F(1,41) = 3.1336$; $P = 0.0841$; $N = 43$; RMS = 0.5233; %SEP = 47.5868; $E = -0.0217$; ARV = 1.0217; AIC = −0.0044; BIC = −0.1067			
H observed	Intercept	0.4230	0.2845
	H estimated	0.5578	0.0841

*$P < 0.05$; **$P < 0.01$.

$F(1,44) = 3.1336$; $P = 0.0841$). Error terms showed the lack of accuracy between the estimated and observed Shannon-Weaver index. The percent standard error of prediction was 47.59% (Table 3, Fig. 3).

Multiple regression between the Shannon-Weaver index and the five significant variables described above exhibited similar results. In this case, the adjusted explained variance in the calibration phase was slightly higher than 0.18. The error magnitudes were similar to those obtained in the previous model. However, the Akaike information criteria and Bayesian information criteria were significantly lower as a consequence of the decrease in the number of parameters (AIC = −0.3245 and BIC = −0.3672, respectively).

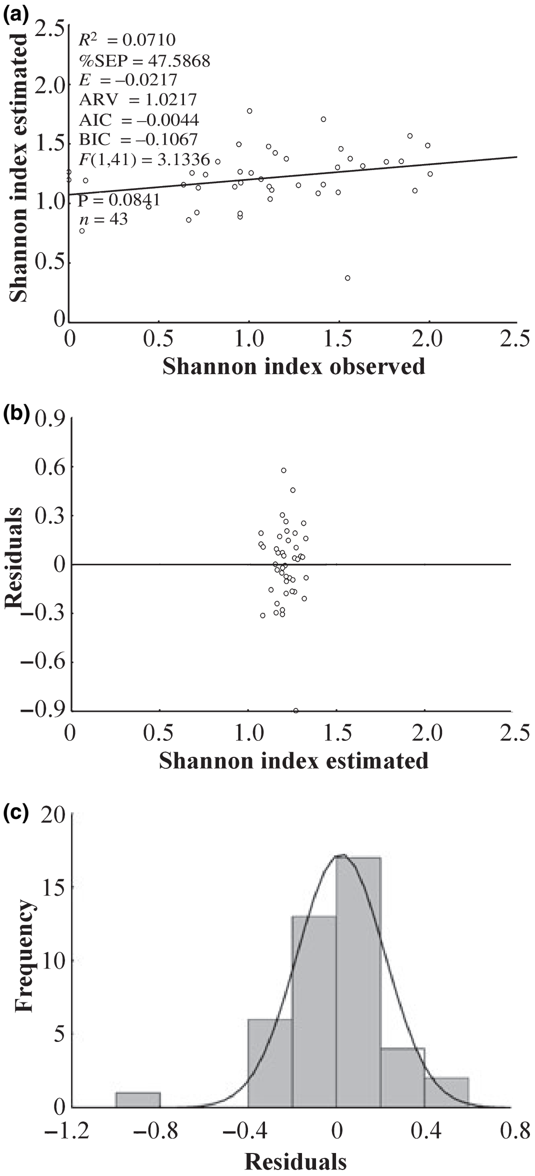

Fig. 3. External validation results of multiple regression model. (a) Scatterplot between observed and estimated Shannon Index. (b and c) Residual analyses

Artificial neural networks analysis

Analysis of variance (ANOVA) of the error terms (R^2, %SEP, E, ARV, AIC and BIC) showed that they significantly differed from each other ($P < 0.05$). A *post hoc* test (Tukey test) of determination coefficient showed no statistical differences between the average values of 12-10-15-1 and 12-15-15-1 architectures (average R^2 of 12-10-15-1 = 0.9294; average R^2 of 12-15-15-1 = 0.9536; $P = 0.1426$). No significant differences were found between average values of 12-10-10-1 and 12-10-15-1 architectures (average R^2 of 12-10-10-1 = 0.9115; average R^2 of 12-10-15-1 = 0.9294; $P = 0.4144$). A similar behaviour was observed in E coefficient and ARV. However, the Tukey test for %SEP only showed non-significant differences between the 12-10-10-1 and 12-10-15-1 architectures ($P = 0.1705$). In this case, statistical differences were observed between mean %SEP of 12-10-15-1 and mean %SEP of 12-15-15-1 ($P = 0.0113$). On the other hand, AIC and BIC showed significant differences in all cases (Fig. 4).

Since no statistical differences were observed among mean values of R^2, ARV and E coefficient for 12-10-15-1 and 12-15-15-1 architectures, and significant differences were found among mean values of %SEP for 12-10-15-1 and 12-15-15-1 architectures, the selection of the best neural network was carried out among the 30 models calibrated in the 12-15-15-1 architecture. The validation of the best model showed a highly significant relationship between predicted and observed Shannon-Weaver index values ($N = 43$; $R^2 = 0.9736$; $P < 0.001$). In this neural network, the %SEP, E and ARV were the best of all models (%SEP = 7.8499; $E = 0.9722$; ARV = 0.0278). The scatter plot of observed and predicted Shannon-Weaver index values and its regression line were very close to the line 1 : 1. The relationship between residuals and estimated values indicated no sign of residual dependence ($N = 43$; $R^2 = 0.0$; $P = 1$), which indicates that the neural network model fits the data well. Additionally, the frequency histogram of residuals showed values centred near zero ($N = 43$; mean = 0.0 ± 0.0852; Fig. 5).

Sensibility analysis of the 12 predictors indicated that the most important variable of the modelled variables was nitrate concentration (ratio = 5.1652) followed by depth (ratio = 4.1365), dissolved oxygen (ratio = 3.4120), temperature (ratio = 3.2514), silicate concentration

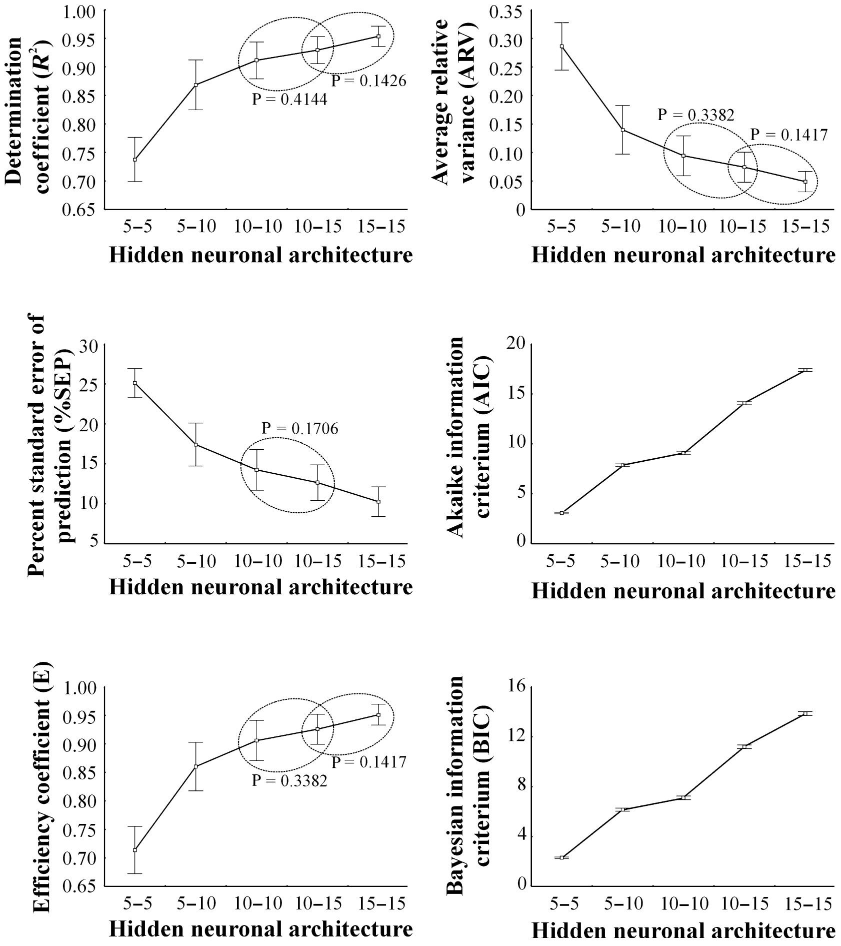

Fig. 4. ANOVA analysis (α level = 0.05) of error terms for artificial neural networks in the calibration phase. Dotted circles indicate no significant differences in the Tukey test

(ratio = 2.4380), nitrite concentration (ratio = 2.4281), transparency (ratio = 2.4105), and salinity (ratio = 2.3554). Parameters with ratios lower than 2 were slope (ratio = 1.8931), phosphate concentration (ratio = 1.4114), water organic particulate material (ratio = 1.1707), and lastly the chlorophyll *a* amount (ratio = 1.1564).

Contribution profiles of each independent variable to the Shannon-Weaver index differed greatly. The nitrate concentrate profile showed a linear decrease between 0 and 1.8 mg L^{-1}. Therefore, a Gaussian behaviour can be observed with a maximum value between 2.7 and 3.6 mg L^{-1}. A sigmoid behaviour was found in the silicate concentration profile. The average profile of depth showed a clear bimodal function with two maxima values. The first maximum value (the most important) was around 3.3 m and the second around 13.3 m. Between them, a transition zone of about 10 m can be observed. Unimodal functions such as log-normal, normal and quadratic functions were observed for dissolved oxygen, temperature and transparency, respectively. For dissolved oxygen the maximum diversity was reached at 6 mg L^{-1} while the maximum value in the temperature profile was around 17.8°C. A descending profile can be observed in the concentrations of nitrite, water organic particulate material and chlorophyll *a*. Smooth ascending profiles were found for the parameters slope and phosphate concentration, whereas salinity did not show a clear variation tendency (Figs 6–9).

Discussion

As the results show, nonlinear methods such as ANNs are more suitable than linear methods to describe the relationships between diversity of fish communities and environmental variables. In the same way, our results indicate that diversity of the fish community in the Tagus estuary can be estimated from physico-chemical variables alone. Similar conclusions may be extracted from the Lek et al. (1996a,b) studies. Lek-Ang et al. (1999) applied neural networks to relate the structure and diversity of an assemblage of hydrophilous *Collembola* to microhabitat characteristics. In this study, one of the six dependent variables forecasted was the Shannon-Weaver index; results showed the high forecasting capacity of neural network from environmental variables alone and the strong nonlinear relationships between diversity and habitat characteristics. Marshall and Elliott, 1998; Maes et al., 1998 and Cabral et al., 2001 also reported significant relationships between environmental variables and fish abundance in estuaries. Maes et al. (2004) indicated that these studies were based on short-term datasets and ignored the seasonal patterns in the data, which favored the correlation between environmental variables and biotic characteristics. Thus, these authors used a long-term database and incorporated in their models seasonal instrumental variables suggesting that the temporal distribution of fishes in estuaries is highly predictable from seasonal components of the model, largely independent of estuarine environmental conditions. Similar conclusions on the seasonal instrument variables were obtained by Power et al. (2000a,b, 2002) and Power and Attrill (2002, 2003). Our results showed that if the seasonal component is one of the most important control elements of estuarine fish abundance, then the artificial neural network models could have detected this seasonal component implicit in the environmental variables.

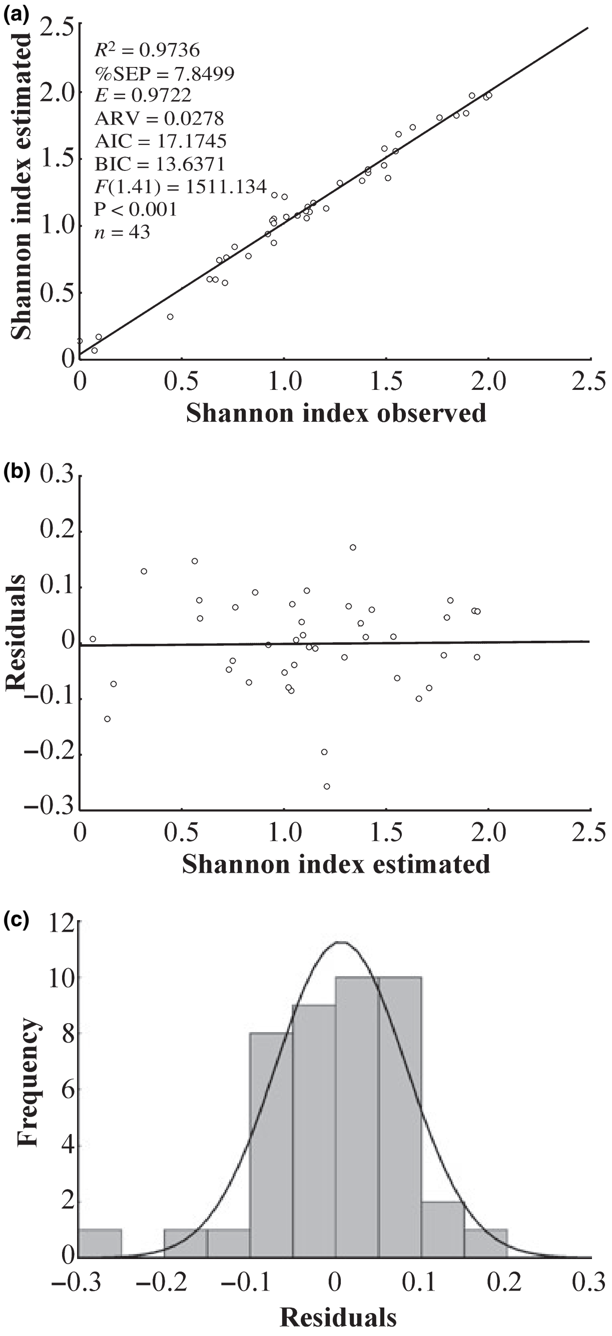

Fig. 5. External validation results of best neural network model. (a) Scatterplot between observed and estimated Shannon Index. (b and c) Residual analyses

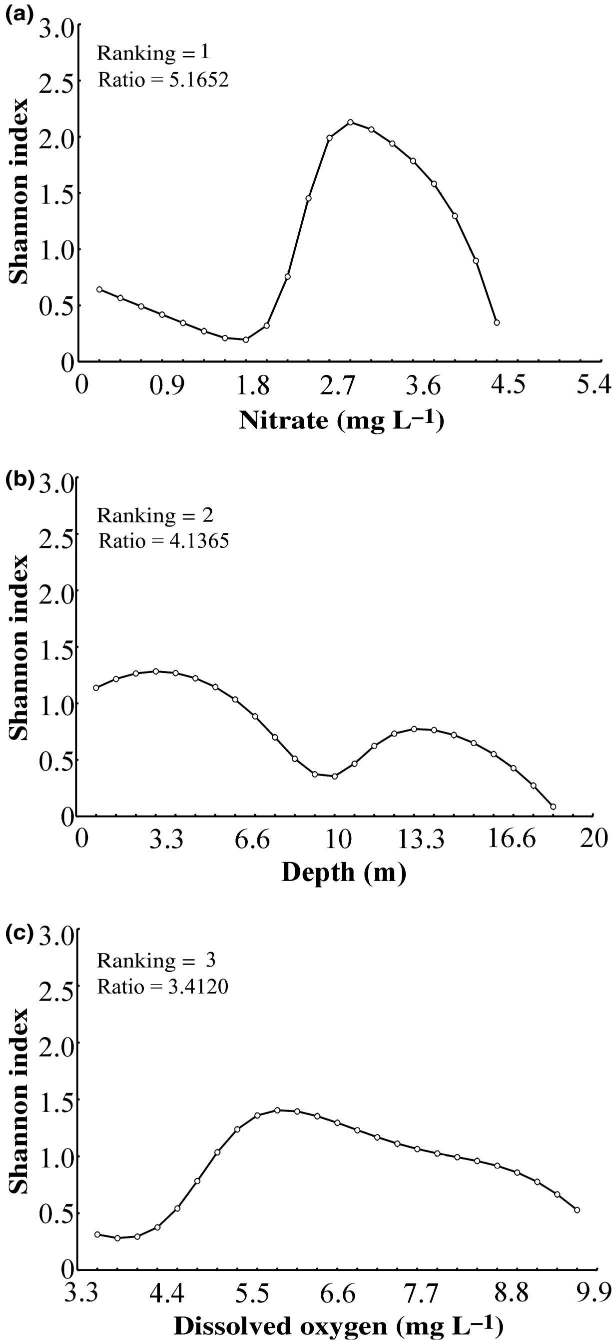

Fig. 6. Contribution profiles of (a) nitrate; (b) depth; (c) dissolved oxygen from the best neural network model

The contribution profiles of independent variables showed that fish diversity in the Tagus estuary was highly influenced by a variety of factors. Results of the sensibility analysis allowed identification of the relative importance of each independent variable and therefore of the most important mechanisms for variation of diversity. In this way, the sensibility analysis indicated that parameters related to water quality such as nitrates and nitrites had a high weight on the model (nitrate: ranking = 1, ratio = 5.1652; nitrite: ranking = 6, ratio = 2.4281). Nitrates influence in the Tagus estuary fish diversity seems difficult to explain, because maximum nitrate concentration was 7.35 mg L^{-1}. In addition, some studies report that the majority of fishes can tolerate nitrate concentrations higher than 500 mg L^{-1} (Hunnam et al., 1982). Ammonia, nitrites and nitrates are clearly related, since the bacterial decomposition of ammonia yields nitrites and nitrates. For this reason, it is possible that the profile and relative importance of nitrate is directly related to ammonia, nitrite and their decomposition cycles. Nitrite is considerably more toxic than ammonia, whereas nitrate is relatively nontoxic to fish. Because bottom waters may become anaerobic or nearly so during summer, nitrite accumulates in bottom sediments, reaching especially high levels at the bottom of stratified waters. Nitrite also accumulates in the water column in winter, when bacterial activity is reduced by low temperatures. Both conditions can also produce toxic levels of nitrite when the water is mixed during spring and autumn. Since the bacterial decomposition of ammonia (toxic) yields

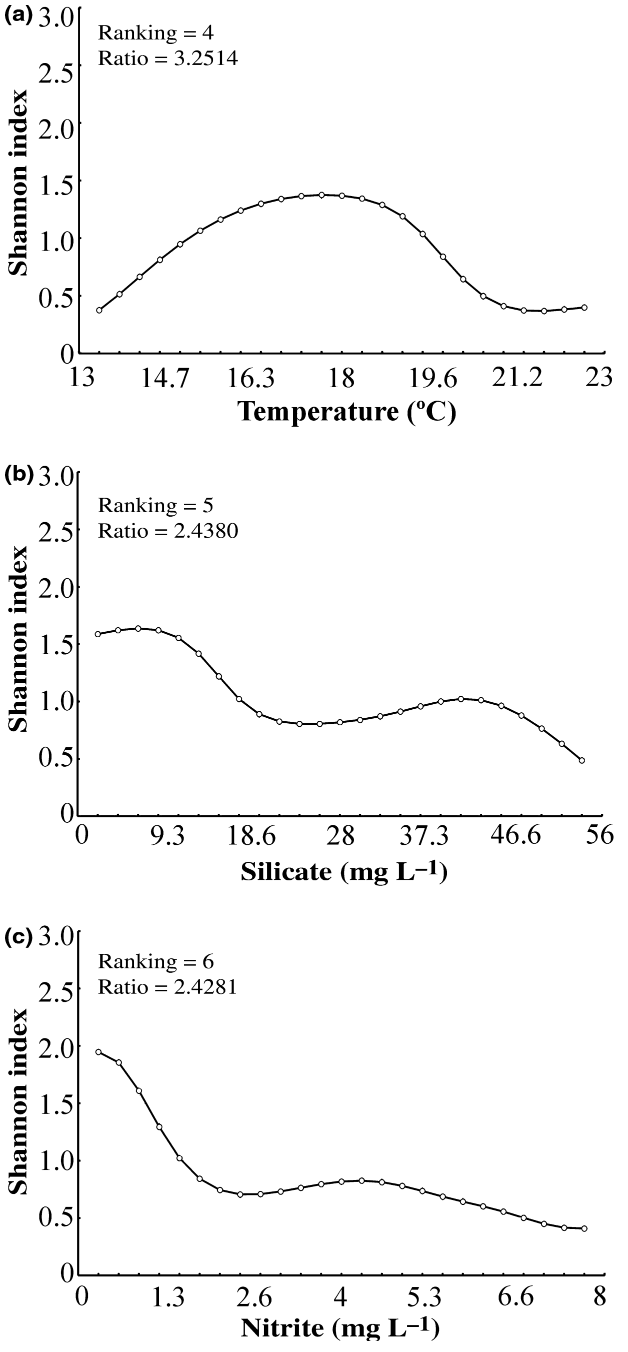

Fig. 7. Contribution profiles of (a) temperature; (b) silicate; (c) nitrite from the best neural network model

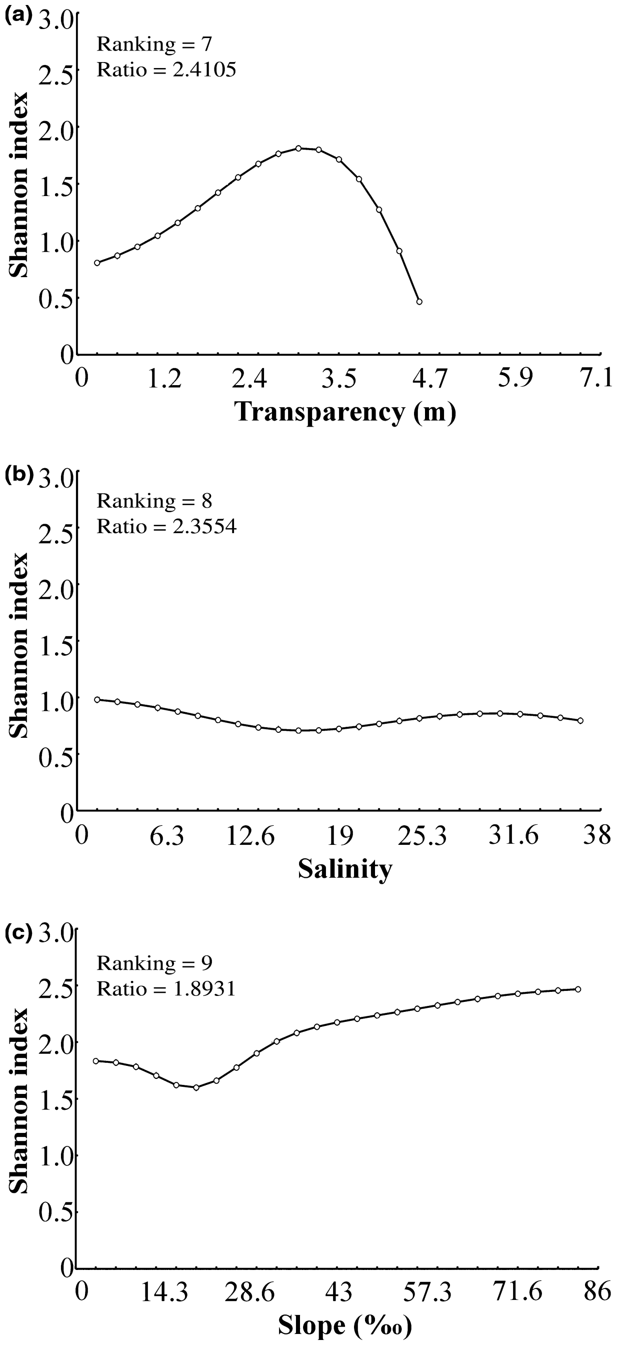

Fig. 8. Contribution profiles of (a) transparency; (b) salinity; (c) slope from the best neural network model

nitrite (very toxic), and bacterial decomposition of nitrite yields nitrate (relatively nontoxic), low levels of nitrate may indicate toxic levels of ammonia and nitrite and therefore, low fish diversity. On the other hand, nitrate concentration increments may indicate that ammonia is being transformed into nitrite and the latter into nitrate, which would allow an increment of diversity. Although ammonia concentration data were not available, contribution profiles of nitrate and nitrite concentrations could support this hypothesis.

Similar explanations can be given for other parameters such as silicate concentrations (ranking = 5; ratio = 2.4380). Here, the contribution profile is related with freshwater flow because the Tagus River crosses wide watershed areas containing siliceous materials (ITGE, 1994). The results show that maximum diversity is found for silicate concentration values around 9 mg L^{-1}, which implies low freshwater inflow in the estuary. Accordingly, other studies have indicated that the numbers of species and fish abundance are negatively correlated with freshwater inflow (Maes et al., 1998). Marshall and Elliott (1998) and Whitfield (1999) reported that freshwater discharge and high salinity fluctuations may act as barriers to estuarine penetration by stenohaline marine species, which could induce a decrease in diversity. This possibility was also noted by Costa and Elliott (1991) for the Tagus and Forth estuaries. On the other hand, other authors have observed positive relationships between fish abundance and freshwater inflow for low-medium flow rates (Flannery et al., 2002).

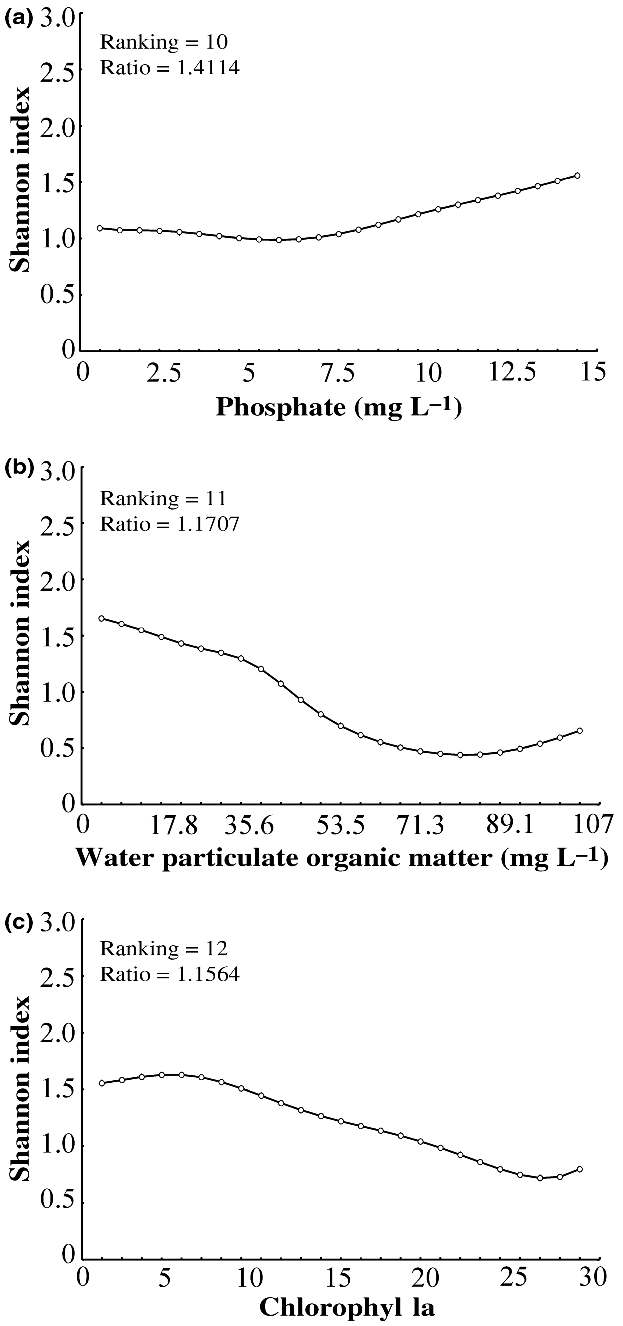

Fig. 9. Contribution profiles of (a) phosphate; (b) water organic particulate matter; (c) chlorophyll *a* from the best neural network model

The depth profile showed a negative effect on the diversity pattern. This profile had an average bimodal contribution which could be related to two aspects. Firstly, it could indicate a different catch capacity of the fishing gear used for demersal and benthic fish. Elliott and Hemingway (2002) reported that otter trawl gear is considered useful for community approaches to demersal fish assemblages, providing good quantitative data on both demersal and benthic fish species. Secondly, it could be related to the amount of available and profitable energy in the system (Oberdorff et al., 1995; Guégan et al., 1998). For example, in deep locations and in some year periods, water turbidity limits light penetration, which affects net primary productivity and food availability (also, the relationship between fish diversity and available energy could explain the transparency behaviour). Our model indicates that diversities higher than 1.5 are reached with chlorophyll *a* concentrations lower than 10 mg m^{-3}. From there on, the diversity decreases in a linear form to 30 mg m^{-3}. Together with the relationship between the Shannon-Weaver index and the phosphate concentration, it may indicate that a slight level of eutrophication can raise fish diversity. However, very high energy levels (chlorophyll *a* concentration >10 mg m^{-3} and water organic particulate material >20 mg L^{-1}) can be useful to a small number of species, which implies low diversities. Persson et al. (1991) obtained similar conclusions studying shifts in fish communities along the productivity gradient of temperate lake ecosystems.

With regard to the variable slope (ranking = 9; ratio = 1.8931), the smooth increment of diversity from 21% seems to be related to the habitat heterogeneity increase. Habitat heterogeneity affects communities because different species may specialize on different habitats. Thus, high heterogeneity favours the presence of more species (MacArthur, 1972). Additionally, the presence of different habitats may also change the habitat selection of a species and may create more opportunities for coexistence (Rosenzweig, 1991). Therefore, in the absence of other factors or for same energy level, local habitat heterogeneity may induce the increase in diversity.

The effects of salinity, temperature and dissolved oxygen gradients upon fish communities are well documented (Thiel et al., 1995; Maes et al., 1998, 2004; Marshall and Elliott, 1998; Lappalainen et al., 2000; Power et al., 2000a; Power and Attrill, 2002; Prista et al., 2003; Attrill and Power, 2004; Paperno and Brodie, 2004; Vega-Cendejas and Hernández-de-Santillana, 2004). Our results showed that dissolved oxygen (ranking = 3; ratio = 3.4120) and temperature (ranking = 4; ratio = 3.2514) were more important than salinity (ranking = 8; ratio = 2.3554) in the Tagus estuary. Maes et al. (2004) results showed of all environmental variables included in their models, that dissolved oxygen was a primary determinant of fish abundance in the Scheldt estuary. In the Tagus estuary, fish diversity exhibited an increasing trend with concentrations of dissolved oxygen in the range of 3.3–5.7 mg L^{-1}. Possibly the diversity increment in this dissolved oxygen range is related to the response of marine species to hypoxic events. Poxton and Allouse (1982) reported that marine fish become stressed at dissolved oxygen levels below 4.5 mg L^{-1}. The smooth diversity decrease when the dissolved oxygen level is higher than 4.5 mg L^{-1} may be a consequence of the negative response of some species to medium-high levels of this parameter.

As with dissolved oxygen, water temperature had a strong weight in the ANN model. Studies in other estuaries have highlighted the importance of temperature in fish species abundance. Attrill and Power (2004) reported that the abundance response of Thames fish species to water temperature was of quadratic form. In this case, the quadratic models were fitted to temperature-abundance data for eight species of fish, indicating distinctive preferences for mid-range temperatures, with the optima range between 11 and 16°C. In the Tagus estuary, the temperature profile shows that the diversity optima was obtained when the temperature fluctuated between 16 and 19°C, indicating acclimation of species.

Regarding salinity, our results showed a low variation range of diversity (0.73–1) vs the salinity gradient (0–38). This finding supports the hypothesis of minimal influence of these salinity levels on the estuarine fish structure or the high osmotic regulation capacity of the estuarine fishes. However, the sensibility analysis classified salinity as a parameter with a medium weight in the model. It may indicate that the salinity effect (more than that of other parameters) could only be explained in the framework of its interaction with other physico-chemical parameters such as temperature, dissolved oxygen, or freshwater flow. Some authors indicate that this parameter plays an important role in the determination of estuarine fish community structures (Henderson, 1988; Elliott et al., 1990; Power et al., 2000a; Attrill, 2002), although several study results are contradictory. For example, Henderson and Holmes (1991) reported no significant correlation between salinity and abundance of dab, sole and flounder in the Severn estuary. In the same estuary, Potter et al. (1986) indicated that salinity acted indirectly, moderating seasonal abundance of fish species with more obvious effects on abundance in wet and dry years. Maes et al. (2004) indicated that physico-chemical parameters, and salinity in particular, were not the major determinants of fish abundance in the Scheldt estuary. In opposition, Thorman (1986) and Thiel et al. (1995) observed a strong influence of salinity on species richness and total biomass in the Bothnian Sea and Elbe estuary, respectively. Additionally, Maes et al. (1998) and Marshall and Elliott (1998) reported that salinity had a great influence on species composition in the Zeeschelde and Humber estuaries, respectively.

Success of the obtained model suggests that further and more accurate studies focusing on relationships between environmental parameters and biotic characteristics of estuary fish communities should be developed. These future studies should give further attention to the role of salinity, chlorophyll *a* and other physical and chemical parameters, the interactions and synergies between them, plus the inclusion of new parameters not considered in this work such as ammonia concentration, photoperiod and life-history characteristics.

Conclusions

The present work focused on three principal objectives: determination of the type of relationship (linear or nonlinear) existing among abiotic (environmental variables) and biotic (diversity) factors of the fish community in the Tagus estuary; the possibility of carrying out an accurate prediction of the Shannon-Weaver diversity index starting solely from environmental variables; and the determination of mechanisms for diversity variation in the fish community. Multiple Linear Regressions and Artificial Neural Network models were applied. The good results obtained with ANNs support the hypothesis that relationships among fish diversity and environmental variables are highly nonlinear in the Tagus estuary. The low error levels obtained with data not used in the calibration phase (validation phase) allowed for a good, realistic short–medium time prediction of diversity in fish communities based on a set of environmental variables. The difficulty of this good, realistic prediction is not only dependent on the nature of inter-relationships between environmental and biological variables and the spatial complexity of environmental effects (as in this case) but also on the type of relationship among variables assumed and the modeling procedures used. Finally, the mechanisms for diversity variations in the Tagus estuary fish community have been explained, at least in part, because it was possible to analyse the existing interaction between the diversity index and an environmental variable; however, for some environmental variables it was not possible to interpret this interaction outside the framework of their interactions with other factors.

With the obtained results, it is necessary to frame the usefulness of the model (its forecast capacity) in the context of a short–medium time period. In this framework, the model can be used to estimate the diversity value when the real value cannot be obtained (for example if abundance and richness values are misleading or not possible to obtain), to estimate interpolated data between two consecutive samples and to simulate different scenarios. Furthermore, the model may be used to detect important changes in the ecosystem organization, since a lack of fit between observed and estimated data will indicate that a new pattern of diversity must be incorporated into the model and therefore the model should be recalibrated and revalidated. In this situation, comparison of the contribution profiles could give relevant information, since it should be possible to quantify the level of the change from the new profiles obtained.

Acknowledgements

This study was development within the scope of the ERIC Project, 'Effects of river flow changes on the fish communities of the Douro, Tagus and Guadiana estuaries and adjoining coastal areas. Ecological and socio-economic predictions' (PDCTM/C/MAR15263/1999), financially supported by Fundação para a Ciência e a Tecnologia. An additional research fellowship was granted to Rita Vasconcelos by Fundação para a Ciência e a Tecnologia.

References

Anctil, F.; Rat, A., 2005: Evaluation of neural network streamflow forecasting on 47 watersheds. J. Hydrol. Eng. **10,** 85–88.

Angermeier, P. L.; Winston, M. R., 1998: Local vs. regional influences on local diversity in stream fish communities of Virginia. Ecology **79,** 911–927.

Attrill, M. J., 2002: A testable linear model for diversity trends in estuaries. J. Anim. Ecol. **71,** 262–269.

Attrill, M. J.; Power, M., 2004: Partitioning of temperature resources amongst an estuarine fish assemblage. Estuar. Coast. Shelf Sci. **61,** 725–738.

Baeta, F.; Pinheiro, A.; Corte-Real, M.; Costa, J. L.; Raposo de Almeida, P.; Cabral, H.; Costa, M. J., 2005: Are the fisheries in the Tagus estuary sustainable? Fish. Res. **76,** 243–251.

Baran, P.; Lek, S.; Delacoste, M.; Belaud, A., 1996: Stochastic models that predict trout population densities or biomass on macrohabitat scale. Hydrobiologia **337,** 1–9.

Brosse, S.; Lek, S.; Townsend, C. R., 2001: Abundance, diversity, and structure of freshwater invertebrates and fish communities: an artificial neural network approach. N. Z. J. Mar. Freshw. Res. **35,** 135–145.

Brown, J. H.; Morgan-Ernest, S. K.; Parody, J. M.; Haskell, J. P., 2001: Regulation of diversity: maintenance of species richness in changing environments. Oecologia **126,** 321–332.

Cabral, H. N.; Costa, M. J.; Salgado, J. P., 2001: Does the Tagus estuary fish community reflect environmental changes? Clim. Change **18,** 119–126.

Cleresci, L. S.; Greenberg, A. E.; Trussell, R., 1989: Standard methods for the examination of water and wastewater. American Public Health Association, American Waterworks Association, Water Pollution Control Federation, Washington, DC, 2134 pp.

Costa, M. J.; Elliott, M., 1991: A fish usage and feeding in two industrialised estuaries – the Tagus, Portugal, and Forth, Scotland. In: Estuaries and coast: spatial and temporal intercomparisons. H. Elliot & J. P. Ducrotoy (Eds). Olsen and Olsen, Fredensborg, Denmark, pp. 289–297.

Elliott, M.; Hemingway, K., 2002: Fishes in estuaries. Blackwell Science, Oxford, 636 pp.

Elliott, M.; O'Reilly, M. G.; Taylor, C. J. L., 1990: The Forth estuary: a nursery and overwintering area for North Sea fishes. Hydrobiologia **195,** 89–103.

Flannery, M. S.; Peebles, E. B.; Montgomery, R. T., 2002: A percent-of-flow approach for managing reductions of freshwater inflows from unimpounded rivers to southwest Florida estuaries. Estuaries **25,** 1318–1332.

Govindajaru, R. S., 2000: Artificial neural networks in hydrology: I. Preliminary concepts. J. Hydrol. Eng. **5,** 115–123.

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. Freshw. Biol. **41,** 809–820.

Grasshoff, K.; Kremling, K.; Ehrhardt, M., 1999: Methods of seawater analysis. Wiley-VCH, Weinheim, 600 pp.

Griñó, R., 1992: Neural networks for univariate time series forecasting and their application to water demand prediction. Neural Netw World **2,** 437–450.

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

Guisan, A.; Zimmermann, N. E., 2000: Predictive habitat distribution models in ecology. Ecol. Modell. **135,** 147–186.

Gutiérrez-Estrada, J. C.; Pulido-Calvo, I.; Prenda, J., 2000: Gonadosomatic index estimates of an introduced pumpkinseed (*Lepomis gibbosus*) population in a Mediterranean stream, using computational neural networks. Aquat. Sci. **62,** 350–363.

Gutiérrez-Estrada, J. C.; De-Pedro-Sanz, E.; López-Luque, R.; Pulido-Calvo, I., 2004: Comparison between traditional methods and artificial neural networks for ammonia concentration forecasting in an eel (*Anguilla anguilla* L.) intensive rearing system. Aquacult. Eng. **31,** 183–203.

Henderson, P. A., 1988: The structure of estuarine fish communities. J. Fish Biol. **33,** 223–225.

Henderson, P. A.; Holmes, R. H. A., 1991: On the population dynamics of dab, sole and flounder within Bridgewater Bay in the lower Severn estuary, England. Neth. J. Sea Res. **27,** 337–344.

Hunnam, P.; Milne, A.; Stebbing, P., 1982: El acuario vivo. Agua dulce y salada. Libros del Carballo, Madrid, 240 pp.

Hunter, A.; Kennedy, L.; Henry, J.; Ferguson, I., 2000: Application of neural networks and sensitivity analysis to improved prediction of trauma survival. Comput. Methods Programs Biomed. **62,** 11–19.

ITGE, 1994: Mapa geológico de la Península Ibérica, Baleares y Canarias. Instituto Tecnológico Geominero de España e Instituto Geológico e Mineiro de Portugal, Madrid.

Kitanidis, P. K.; Bras, R. L., 1980: Real time forecasting with a conceptual hydrological model: 2. Applications and results. Water Resour. Res. **16,** 1034–1044.

Kupschus, S., 2004: A temperature-dependent reproductive model for spotted seatrout (*Cynoscion nebulosus*) explaining spatio-temporal variations in reproduction and young-of-the-year recruitment in Florida estuaries. J. Mar. Sci. **61,** 3–11.

Kupschus, S.; Tremain, D., 2001: Associations between fish assemblages and environmental factors in nearshore habitats of a subtropical estuary. J. Fish Biol. **58,** 1383–1403.

Lappalainen, A.; Shurukhin, A.; Gregory, A.; Jukka, R., 2000: Coastal fish communities along the northern coast of the Gulf of Finland, Baltic sea: responses to salinity and eutrophication. Int. Rev. Hydrobiol. **85,** 687–696.

Lek, S.; Belaud, A.; Baran, P.; Dimopoulos, I.; Delacoste, M., 1996a: Role of some environmental variables in trout abundant models using neural networks. Aquat. Living Resour. **9,** 23–29.

Lek, S.; Delacoste, M.; Baran, P.; Dimopoulos, I.; Lauga, J.; Aulanier, S., 1996b: Application of neural networks to modelling non-linear relationships in ecology. Ecol. Modell. **90,** 39–52.

Lek-Ang, S.; Deharveng, L.; Lek, S., 1999: Predictive models of collembolan diversity and abundance in a riparian habitat. Ecol. Modell. **120,** 247–260.

Lekve, K.; Stenseth, N. C.; Gjöæter, J., 2003: Regulation of species richness by advection and richness-dependent processes in a coastal fish community. Hydrobiologia **503,** 93–102.

MacArthur, R. H., 1972: Geographical ecology. Harper and Row, New York, 269 pp.

MacArthur, R. H.; Reicher, H.; Cody, M. L., 1996: On the relation between habitat selection and bird species diversity. Am. Nat. **100,** 319–332.

Maes, J.; Van Damme, P. A.; Taillieu, A.; Ollevier, F., 1998: Fish communities along an oxygen-poor salinity gradient (Zeeschelde estuary, Belgium). J. Fish Biol. **52,** 534–546.

Maes, J.; Van Damme, S.; Meire, P.; Ollevier, F., 2004: Statistical modeling of seasonal and environmental influences on the population dynamics of an estuarine fish community. Mar. Biol. **145,** 1033–1042.

Margalef, R., 1968: Perspectives in ecology theory. Univ. Chicago Press, Chicago, 111 pp.

Marshall, S.; Elliott, M., 1998: Environmental influences on the fish assemblage of the Humber estuary, U.K. Estuar. Coast. Shelf Sci. **46,** 175–184.

Müller, B.; Reinhardt, J., 1990: Artificial Neural Networks: electronic implementations. Springer-Verlag, Berlin.

Oberdorff, T.; Guégan, J. F.; Hugueny, B., 1995: Global scale patterns of fish species richness in rivers. Ecography **18,** 345–352.

Paperno, R.; Brodie, R. B., 2004: Effects of environmental variables upon the spatial and temporal structure of a fish community in a small, freshwater tributary of the Indian River Lagoon, Florida. Estuar. Coast. Shelf Sci. **61,** 229–241.

Persson, L.; Diehl, S.; Johansson, L.; Andersson, G.; Hamrin, S. F., 1991: Shifts in fish communities along productivity gradient of temperature lakes – patterns and importance of size-structured interactions. J. Fish Biol. **38,** 281–293.

Potter, I. C.; Claridge, P. N.; Warwick, R. M., 1986: Consistency of seasonal changes in an estuarine fish assemblage. Mar. Ecol.-Prog. Ser. **32,** 217–228.

Power, M.; Attrill, M. J., 2002: Factors affecting long-term trends in the estuarine abundance of pogge (*Agonus cataphractus*). Estuar. Coast. Shelf Sci. **54,** 941–949.

Power, M.; Attrill, M. J., 2003: Long-term trends in the estuarine abundance of Nilsson's pipefish (*Syngnathus rostellatus* Nilson). Estuar. Coast. Shelf Sci. **57,** 325–333.

Power, M.; Attrill, M. J.; Thomas, R. M., 2000a: Environmental factors and interactions affecting the temporal abundance of juvenile flatfish in the Thames estuary. J. Sea Res. **43,** 135–149.

Power, M.; Attrill, M. J.; Thomas, R. M., 2000b: Temporal abundance patterns and growth of juvenile herring and sprat from the Thames estuary 1977–1992. J. Fish Biol. **56,** 1408–1426.

Power, M.; Attrill, M. J.; Thomas, R. M., 2002: Environmental influences on the long-term fluctuations in the abundance of gadoid species during estuarine residence. J. Sea Res. **47,** 185–194.

Poxton, M. G.; Allouse, S. B., 1982: Water quality criteria for marine fisheries. Aquacult. Eng. **1,** 153–191.

Prista, N.; Vasconcelos, R. P.; Costa, M. J.; Cabral, H., 2003: The demersal fish assemblage of the coastal area adjacent to the Tagus estuary (Portugal): relationships with environmental conditions. Oceanol. Acta **26,** 525–536.

Qi, M.; Zhang, G. P., 2001: An investigation of model selection criteria for neural network time series forecasting. Eur. J. Oper. Res. **132,** 666–680.

Rosenzweig, M. L., 1991: Habitat selection and population interactions: the search for mechanism. Am. Nat. **137,** 5–28.

Rumelhart, D. E.; Hinton, G. E.; Williams, R. J., 1986: 'Learning' representations by back propagation errors. Nature **323,** 533–536.

Schoener, T. W., 1983: Field experiments on interspecific competition. Am. Nat. **122,** 240–285.

Shepherd, A. J., 1997: Second-order methods for neural networks. Springer, New York, 208 pp.

Stenseth, N. C.; Lekve, K.; Gjöæter, J., 2002: Modeling species richness controlled by community-intrinsic and community-extrinsic processes: coastal fish communities as an example. Popul. Ecol. **44,** 165–178.

Tan, Y.; van Cauwenberghe, A., 1999: Neural-network-based d-step-ahead predictors for nonlinear systems with time delay. Eng. Appl. Artif. Intell. **12,** 21–25.

Thiel, R.; Sepúlveda, A.; Kafemann, R.; Nellen, W., 1995: Environmental factors as forces structuring the fish community of the Elbe Estuary. J. Fish Biol. **46,** 47–69.

Thorman, S., 1986: Seasonal colonization and effects of salinity and temperature on species richness and abundance of fish of some brackish and estuarine shallow waters in Sweden. Holartic Ecol. **9,** 126–132.

Tilman, D., 1982: Resource competition and community structure. Monography population biology, **vol. 17**. Princeton University, Princeton, NJ, pp. 296.

Tsoukalas, L. H.; Uhrig, R. E., 1997: Fuzzy and neural approaches in engineering. Wiley Interscience, New York, 587 pp.

Vega-Cendejas, M. E.; Hernández-de-Santillana, M., 2004: Fish community structure and dynamics in a coastal hypersaline lagoon: Rio Lagartos, Yucatán, Mexico. Estuar. Coast. Shelf Sci. **60,** 285–299.

Ventura, S.; Silva, M.; Pérez-Bendito, D.; Hervás, C., 1995: Artificial neural networks for estimation of kinetic analytical parameters. Anal. Chem. **67,** 1521–1525.

Vinagre, C.; França, S.; Costa, M. J.; Cabral, H. N., 2004: Accumulation of heavy metals by flounder *Platichthys flesus* (Linnaeus 1758), in a heterogeneously contaminated nursery area. Mar. Pollut. Bull. **49,** 1109–1126.

Whitfield, K. A., 1999: Ichthyofaunal assemblages in estuaries: a South African case study. Rev. Fish Biol. Fish. **9,** 151–186.

Author's address: Juan C. Gutiérrez-Estrada, Departamento de Ciencias Agroforestales, Escuela Politécnica Superior, Campus Universitario de La Rábida, Universidad de Huelva, E-21819 Palos de la Frontera, Huelva, Spain.
E-mail: juanc@uhu.es