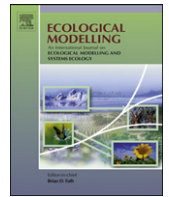




Contents lists available at ScienceDirect

## Ecological Modelling

journal homepage: [www.elsevier.com/locate/ecolmodel](http://www.elsevier.com/locate/ecolmodel)

## A heuristic approach to predicting water beetle diversity in temporary and fluctuating waters

Juan C. Gutiérrez-Estrada<sup>a,\*</sup>, David T. Bilton<sup>b</sup>

<sup>a</sup> Dpto. Ciencias Agroforestales, Universidad de Huelva, EPS, Campus de La Rábida, 21819 Palos de la Frontera, Huelva, Spain

<sup>b</sup> Marine Biology and Ecology Research Centre, School of Biological Sciences, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK

### ARTICLE INFO

#### Article history:

Received 8 December 2009

Received in revised form 1 March 2010

Accepted 7 March 2010

Available online 2 April 2010

#### Keywords:

Artificial neural network (ANN)  
Generalised additive model (GAM)  
Multiple linear regression (MLR)  
Physicochemistry  
Species diversity  
Macroinvertebrates

### ABSTRACT

An understanding of the causal mechanisms and processes that shape macroinvertebrate communities at a local scale has important implications for the management and conservation of freshwater biodiversity. Here we compare the performance of linear and non-linear statistics to explore diversity–environment relationships using data from 76 temporary and fluctuating ponds in two regions of southern England. We focus on aquatic beetle assemblages, which have been shown to be excellent surrogates of wider freshwater macroinvertebrate diversity. Ponds in the region contained a rich coleopteran fauna, totaling 68 species, which provided an excellent model system with which to compare the performance of two non-linear procedures (artificial neural networks—ANNs and generalised additive models—GAMs) and one more traditional linear approach (Multiple linear regression—MLR) to modelling diversity–environment relationships. Of all approaches employed, the best fit was obtained using an ANN model with only four input variables (conductivity, turbidity, magnesium concentration and depth). This model accounted for 82% of the observed variability in Shannon diversity index across ponds. In contrast, the best GAM and MLR models only explained 50% and 14% of this variation, respectively. Contribution profile analysis of conductivity, turbidity, magnesium concentration and depth, obtained from the best fit ANN through a hierarchical cluster analysis, allowed the identification of direct and proxy effects in relation to the environmental variables measured in this study. In each case, distinct clusters of ponds were identified in contribution profile analysis, suggesting that ponds across the two regions fall into a number of discrete groups, whose beetle faunas respond in subtly yet significantly different ways to key environmental variables. Aquatic coleopteran diversity in ponds in the two regions appears to be driven at a local scale by changes in relatively few physicochemical gradients, which are related to diversity in a clearly non-linear manner.

© 2010 Elsevier B.V. All rights reserved.

### 1. Introduction

In recent years, ponds have been recognised as one of the most important inland aquatic ecosystems from a biodiversity perspective (Collison et al., 1995), with temporary and fluctuating waters supporting a number of rare and specialist plant and animal taxa, particularly insects (Nicolet et al., 2004; Williams, 2005; Bilton et al., 2009). In these ecosystems it is recognised that differential sensitivity to local conditions (e.g. water chemistry, pond size—Wellborn et al., 1996, Bilton et al., 2009) generate patterns of species turnover along environmental gradients in the field. Such patterns may be generated by differential dispersal, or selection on dispersers once they arrive at individual localities (Cottenie, 2005). Additionally understanding the processes that determine commu-

nity diversity can help us predict the response of these ecosystems to global environmental change.

Several factors have been identified as key influences on the diversity and composition of macroinvertebrate communities in temporary and fluctuating ponds. These factors may be grouped together, depending on the spatial scale at which they operate, as regional and local, or biogeographical and habitat factors (Friday, 1987). The relative importance of factors operating at these different scales in determining community composition is still poorly understood, although many authors such as Wellborn et al. (1996), Moss (1998), Lahr et al. (1999), Rundle et al. (2002), Kiflawi et al. (2003) and Bilton et al. (2009) have reported that much of the variation in pond macroinvertebrate community composition is dependent on local physicochemistry. Physical factors such as pond size, water temperature or hydroperiod have been significantly correlated with species richness (Bilton et al., 2001; Ripley and Simovich, 2009), and different chemical factors including pH, conductivity or heavy metal concentrations have been shown to have

\* Corresponding author. Tel.: +34 959217528; fax: +34 959217528.  
E-mail address: [juanc@uhu.es](mailto:juanc@uhu.es) (J.C. Gutiérrez-Estrada).

a strong influence on the richness and diversity of macroinvertebrates communities in individual ponds (Bechara, 1996; Williams, 1996; Miserendino and Archangelsky, 2006).

The bulk of the literature analyzing the relationships between environmental variables and ecological properties of aquatic macroinvertebrate communities employ classical statistical methodologies. These include a wide variety of linear approaches including linear regression (LR) (Lahr et al., 1999; Marchant et al., 2006), multiple linear regression (MLR) (Friday, 1987), principal component analysis (PCA) (Rundle et al., 2002; Jeffries, 2005; Kasangaki et al., 2008; Haidekker and Hering, 2008; Blocksom and Johnson, 2009) and multidimensional scaling (MDS) (Cowell et al., 2004). Also, more advanced multivariate techniques like detrended correspondence analysis (DCA) (Valladares et al., 2002) and canonical correspondence analysis (CCA) (Fernández-Aláez et al., 2002; Irmiler et al., 2002; Rundle et al., 2002; Miserendino and Archangelsky, 2006; Bilton et al., 2006; Rohasliney and Jackson, 2008) have been used in the past. Despite the fact that these methods may reveal significant relationships between community composition and environmental parameters, they possess a number of features that limit their explanatory, estimation or forecasting power (James and McCulloch, 1990). The degree of variance explained by these approaches is often relatively low, although whether this reflects the limited importance of local physicochemistry, or the inability of linear approaches to describe species–environment interactions remains unclear. As Stenseth et al. (2002) have pointed out, in most studies these limitations may result from the stochastic and non-linear nature of ecological systems, meaning that a non-equilibrium systems framework is required to tackle the (typically multivariate) problem of understanding species–environment relationships.

At present, few studies have explored the relationship between species diversity and environmental variation from a stochastic and non-linear point of view. Kiflawi et al. (2003) employed non-linear regression techniques to explore the impact of local and regional processes on macroinvertebrate species richness in Israel. Lek-Ang et al. (1999, 2007), Cérèghino et al. (2001), Park et al. (2003), Goethals et al. (2007) and Lencioni et al. (2007) have studied the diversity, richness, abundance and distribution of several aquatic macroinvertebrate groups using heuristic techniques including artificial neural networks (ANNs). Also, Amiaud et al. (2005) designed an expert system based in fuzzy logic to predict the presence probability of grassland plant species. In all these cases, the authors reported the high fit capacity of these modelling techniques.

In this paper we evaluate the performance of multilayer feed-forward ANN to analyse the relationships between water beetle diversity and environmental factors in temporary ponds in two areas of southern England. Amongst the macroinvertebrate taxa that inhabit temporary and fluctuating ponds, aquatic Coleoptera are particularly well suited both to assess habitat quality and explore ecological patterns, since this group is relatively diverse, ecologically well understood and occurs across a wide spectrum of pond types (Bilton et al., 2006; Sánchez-Fernández et al., 2006); characteristics which have enabled them to be employed as indicators of overall community diversity and ecosystem status. The performance of ANN is compared with more classic non-linear and linear fit techniques, namely generalised additive models (GAMs) and multiple linear regressions (MLRs), allowing us to evaluate the relative ability of these techniques to model diversity–environment interactions. Our approach reveals the highly non-linear nature of the relationship between species diversity and environmental parameters, and demonstrates that the beetle communities of different groups of ponds respond in subtly, yet significantly, different ways to environmental gradients, an insight which has implications for habitat management and our

**Table 1**

Mean, standard deviation (SD) and Pearson coefficient of variation (%CV) of physical, chemical characteristics and coleopteran Shannon index ( $H'$ ) of the 76 ponds sampled. Cu = copper concentration; Zn = zinc concentration; Fe = iron concentration; Co = cobalt concentration; Cr = chromium concentration; Ni = nickel concentration; Al = aluminium concentration; Mg = magnesium concentration; Ca = calcium concentration; TON = total organic nitrogen; SRP = soluble reactive phosphorous.

	Mean	SD	%CV
Area (m <sup>2</sup> )	755.060	2269.190	300.531
Depth (cm)	19.475	10.679	54.836
Turbidity (NTU)	16.882	20.820	123.325
pH	6.172	0.883	14.306
Conductivity (μS cm <sup>-1</sup> )	474.420	290.558	61.245
Cu (mg L <sup>-1</sup> )	0.004	0.003	92.177
Zn (mg L <sup>-1</sup> )	0.050	0.132	264.827
Fe (mg L <sup>-1</sup> )	0.430	1.423	331.175
Co (mg L <sup>-1</sup> )	0.156	0.369	236.001
Cr (mg L <sup>-1</sup> )	0.091	0.188	207.073
Ni (mg L <sup>-1</sup> )	0.057	0.054	95.565
Al (mg L <sup>-1</sup> )	0.084	0.115	136.731
Mg (mg L <sup>-1</sup> )	10.609	12.012	113.223
Ca (mg L <sup>-1</sup> )	3.470	3.712	106.974
TON (mg L <sup>-1</sup> )	0.391	1.112	284.223
SRP (mg L <sup>-1</sup> )	0.011	0.015	142.353
Coleoptera $H'$	1.999	0.896	44.821

understanding of community responses to ongoing environmental change.

## 2. Methods

### 2.1. Study area

The ponds analysed in this study were located in two regions of southern England, the New Forest (Hampshire) and the Lizard Peninsula (Cornwall). Both regions contain a high density of temporary and fluctuating ponds, described in detail by Bilton et al. (2001, 2006, 2009) and Rundle et al. (2002). Ponds sampled varied substantially in their biological, physical and chemical characteristics (Table 1). Lizard ponds were above ultra-basic serpentine geology in heathland/unimproved grassland, whereas New Forest ponds were in heathland/grassland above eroded sedimentary beds with superficial deposits of sand and gravel.

### 2.2. Coleopteran sampling

Beetles were sampled during February and March 2000 in a total of 76 temporary ponds, a time when the spatial extent and the presence of ponds were at their maximum (Bilton et al., 2006). Ponds were sampled using a hand net (1 mm mesh, dimensions 20 cm × 25 cm), taking semi-quantitative 1 m sweeps amongst aquatic vegetation. Each 1 m sweep involved approximately 10 s of back and forth netting over the same area of habitat (Rundle et al., 2002; Foggo et al., 2003; Bilton et al., 2006). Two or three such samples were taken from the largest sites according to their area. Sweeps were pooled and samples preserved in 95% ethanol in the field.

In the laboratory beetles were extracted from washed samples, determined to species level, and counted. Shannon's index ( $H'$ ) was calculated for Coleoptera from each pond following Brower et al. (1998). This diversity measure was selected because it reflects both species richness and the relative abundance of species within assemblages.  $H'$  normally varies between 1.5 and 3.5, with values higher than 3 being seen as representing diverse communities whilst those below 2 are relatively uniform (Cowell et al., 2004).

A range of environmental variables was also recorded from each pond. Before Coleoptera were sampled pH, temperature compensated conductivity and turbidity readings were taken on-site using

a Solomat 520C probe (Zellweger Analytics, Poole, U.K.). Water depth in the area sampled was estimated using a 1 m rule (mean of five measurements). Two water samples from each pond were also collected, in acid washed polypropylene bottles, for analysis of metal cation and nutrient concentrations. Metal cation concentrations were analysed in the laboratory by atomic absorption spectroscopy. Cations measured were calcium, magnesium, aluminium, nickel, chromium, cobalt, iron, zinc and copper, which were chosen to represent the main ions present in the two regions. Water samples were also analysed for total organic nitrate (TON) and soluble reactive phosphorus (SRP) using a Dionex autoanalyser (Camberley, U.K.). Estimates of Lizard pond areas were derived by using a differential Global Positioning System (GPS; Trimble) to map the margin of each pond. In the New Forest area was estimated by either pacing pond perimeter or measurement from 1:10,000 Ordnance Survey maps.

### 2.3. Modelling techniques

The determination of beetle diversity patterns was performed using artificial neural networks (ANNs). ANNs are mathematical models inspired by the neural architecture of the biological nervous system. The most widely studied and used structures are multilayer feed-forward networks or multilayer perceptrons (Rumelhart et al., 1986). These models 'learn' in an iterative way in which the data are introduced to the neural network a number of times until a pre-determined error level (calculated as the sum of the squared errors) is reached (the iteration where all the data are introduced to the ANN is termed the epoch). These supervised ANNs allow the analysis of complex data sets and the determination of non-linear relationships between dependent and independent variables. A detailed description of the performance of multilayer perceptron ANNs can be found in Lek and Guegan (1999), Gutiérrez-Estrada et al. (2000), Dedecker et al. (2005), Goethals et al. (2007), Pulido-Calvo and Portela (2007) and Gutiérrez-Estrada et al. (2008). There are many ANN calibration or learning methods. In this work, the Levenberg–Marquardt algorithm (Shepherd, 1997) was applied. This is a second-order non-linear optimization algorithm that guarantees local convergence and which is recommended by several authors (Tan and Van Cauwenbergh, 1999; Ancia and Rat, 2005).

In order to test the coherency of the results obtained with ANNs, the behaviour of ANNs was compared with generalised additive models (GAMs) (Hastie and Tibshirani, 1990) and multiple linear regressions (MLRs). In the same way as ANN, GAMs and MLRs were applied to describe Coleoptera–habitat relationships. In the case of GAMs, normal and gamma distribution models with identity, inverse and logit link-functions and cubic spline smooth function with three degrees of freedom were fitted.

### 2.4. General procedure

The general procedure employed for the calibration of neural network models is outlined in Fig. 1. Before the calibration of any ANN model, a basic correlation and cross-correlation analysis between variables was carried out. Later on the data set was randomly divided in two subsets: the first one (calibration subset [CS]+select subset [SS], CSS) was composed of 66 ponds and the second one (test subset, TS) was composed of the 10 remaining ponds (Tsoukalas and Uhrig, 1997). In the first subset (CSS), 10 ponds (also randomly selected) composed the select subset (SS) and these were used to avoid overtraining or over-calibration of the ANN. The best method to ensure that overtraining does not occur is to monitor periodically (at the end of each epoch) the sum square error for both the CS subset and the SS subset (internal validation). It is normal that the sum square error for the CS subset decreases continuously with training. However, this may be forcing the neu-

ral network to fit the noise in the CS subset. To avoid this problem, training is stopped at the end of each epoch and the sum square error of the SS subset is calculated. When the sum square error of the SS subset begins to increase, training must be stopped and the weights of the epoch which provided a minimum error for the SS subset should be tested with the TS subset. This last phase is also known as the generalisation phase, or external validation. Iyer and Rhinehart (1999) recommend repeating this process at least 30 times for each model, and this recommendation was followed here (Fig. 1, step 1).

Once each ANN was calibrated and tested a neural ensemble process was carried out. This process combines the 30 ANNs at the level of the output neuron. Ensemble, together with internal validation are the most important means of combating over-learning and improving the generalisation capacity of the ANN (Watts and Worner, 2008).

In a second phase, each ensemble ANN was subjected to a selectivity analysis in order to select the most weighted input variables (Fig. 1, step 2). The sensitivity analysis was carried out by replacing each variable by missing values and assessing the effect of this on the output error. Following this, the new error calculated was compared with the original error to obtain a ratio value (ratio = error of the model with a variable with missing values/error of the model with all variables). In this way, for a given variable  $x$ , a ratio with a value equal to or very close to 1 indicates that this variable has a very low weight in the general structure of the model and therefore it will not be selected (Hunter et al., 2000). Once the independent variables were selected, step 1 was repeated. This way, a new ensemble neural network was obtained with the input variables selected.

The procedure described above was carried out for each neuronal configuration tested. In this work, ANNs with one and two hidden layers were tested. In both cases, a number of neurons between 5 and 20 were tested (Gutiérrez-Estrada et al., 2008).

Six accuracy measures were calculated in the calibration, selection and validation phases of each ANN examined: the correlation coefficient ( $R$ ), the determination coefficient ( $R^2$ ), the square root of the mean square error (RMSE), the mean absolute error (MAE), the standard error of prediction (SEP), and the efficiency coefficient ( $E_2$ ) (Nash and Sutcliffe, 1970; Kitanidis and Bras, 1980; Ventura et al., 1995; Legates and McCabe, 1999).

### 2.5. Coleopteran diversity patterns

For ANN models, the contribution profiles, or the impacts, of each environmental variable in relation to beetle diversity were obtained following and extending the method proposed by Laë et al. (1999). For each environmental variable (input variables) and each ANN, the model response for each pond  $i$  was extracted by applying arbitrary values contained in the variation range of the input variable  $k$  selected (25 values equally spaced between the minimum and maximum of  $k$ ) whilst the remaining environmental variables took their real value. Therefore, one contribution profile was obtained for each pond  $i$  and each environmental variable  $k$ .

Subsequently, in order to determine whether groups of ponds within our dataset differed in the way their beetle diversity responds to environmental gradients, the contribution profiles obtained from each ANN were subjected to a hierarchical cluster analysis using Ward's linkage method. This method was selected since it uses an analysis of variance approach to evaluate the distances between clusters. Thus, this method attempts to minimize the sum of squares (SS) of any two (hypothetical) clusters that can be formed at each step. The final number of clusters was extracted by analysing the graph of the linkage distances across consecutive steps of the linking process (i.e. a graph of the amalgamation sched-

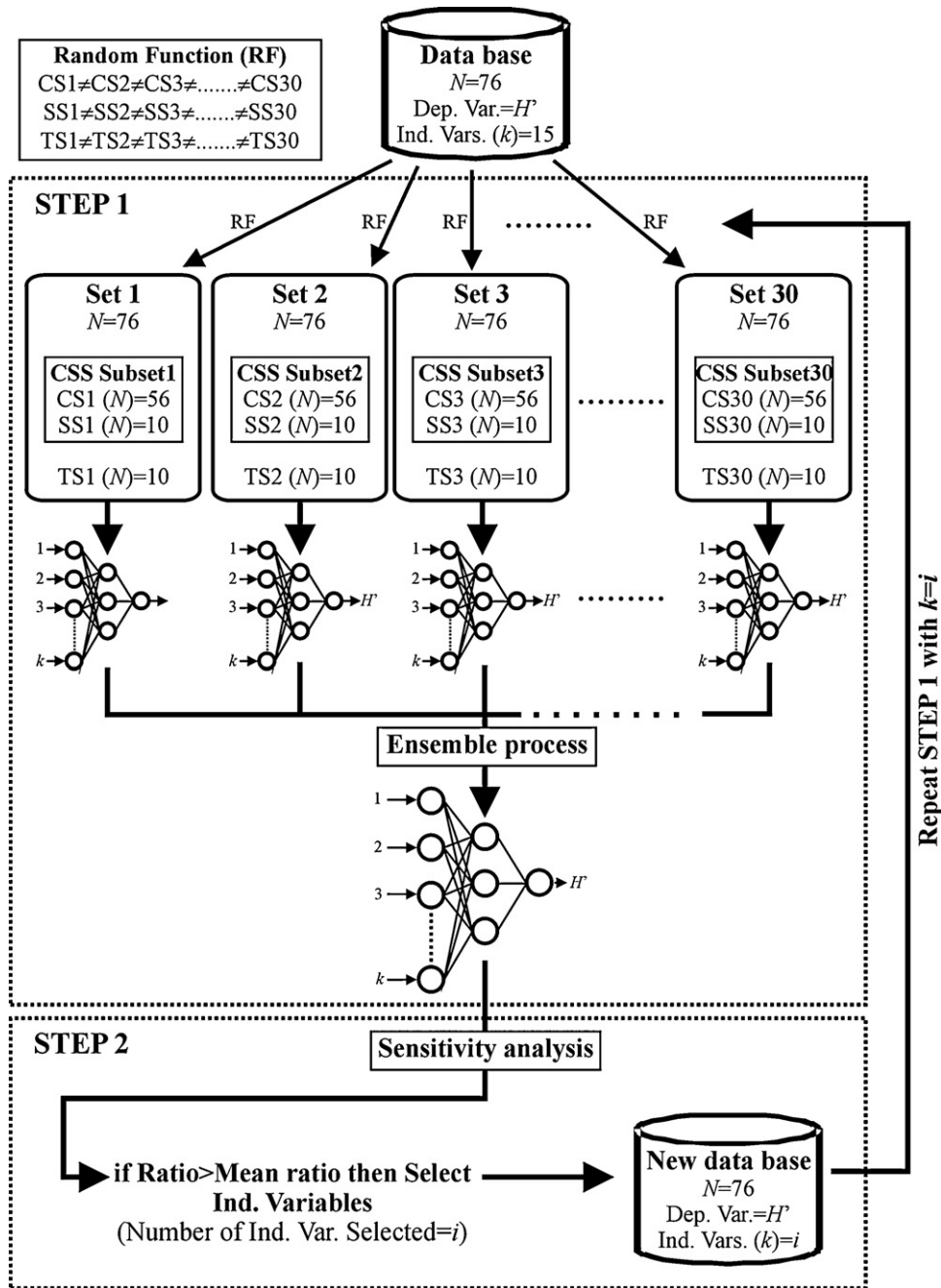


Fig. 1. Schematic representation of the general procedure followed for the calibration of neural network models. CS = calibration set; SS = select set; TS = test set; CSS = calibration + select subset; N = number of patterns.

ule) (Hartigan, 1975). Finally, the centroid of each identified cluster was obtained and analysed.

In the case of GAM models, the response of a specific predictor variable was obtained by analysing the behaviour of the smoothed predictor variable against the residuals after removing the effect of all other predictor variables (using a partial residuals plot) (Hastie and Tibshirani, 1990). In the case of MLR the beta coefficients were analysed (Hair et al., 1998)

Finally, the possible hidden or proxy effects of variables rejected by the ANN models were examined for selected physicochemical variables after sensitivity analysis by means of ANOVA and *post hoc* Tukey HDS tests. All analyses were conducted in Statistica version 7.0 (Statsoft, Inc. 1984–2004).

### 3. Results

Across the two regions a total of 68 Coleoptera species were recorded, of which 24 were of conservation importance (see Bilton et al., 2009). In relation to Shannon's diversity index, the maximum value ( $H' = 3.968$ ) was observed on the Lizard, whilst high values were also found in the New Forest ( $H'$  max = 3.471). In both regions, some ponds lacked Coleoptera, or contained a single species resulting in diversity values of zero. The overall mean diversity was  $1.999 \pm 0.896$  SD with a coefficient of variation of 44.8% (Table 1). Amongst the environmental variables, only five (depth, pH, conductivity, copper concentration and nickel concentration) had coefficients of variation <100%, reflecting the high

**Table 2**

Cross-correlation matrix between environmental variables and Shannon index ( $H'$ ). Turb. = turbidity; Cond. = conductivity; Cu = copper concentration; Zn = zinc concentration; Fe = iron concentration; Co = cobalt concentration; Cr = chromium concentration; Ni = nickel concentration; Al = aluminium concentration; Mg = magnesium concentration; Ca = calcium concentration; TON = total organic nitrogen; SRP = soluble reactive phosphorous.

	Area	Depth	Turb.	pH	Cond.	Cu	Zn	Fe	Co	Cr	Ni	Al	Mg	Ca	TON	SRP
Coleoptera $H'$	0.14	0.16	-0.02	0.14	-0.21	-0.11	-0.02	0.11	-0.25*	-0.24*	-0.20	0.05	0.06	-0.08	-0.06	-0.01

\*  $p < 0.05$ .

**Table 3**

Cross-correlation matrix between all environmental factors. Turb. = turbidity; Cond. = conductivity; Cu = copper concentration; Zn = zinc concentration; Fe = iron concentration; Co = cobalt concentration; Cr = chromium concentration; Ni = nickel concentration; Al = aluminium concentration; Mg = magnesium concentration; Ca = calcium concentration; TON = total organic nitrogen; SRP = soluble reactive phosphorous.

	Area	Depth	Turb.	pH	Cond.	Cu	Zn	Fe	Co	Cr	Ni	Al	Mg	Ca	TON
Depth	0.20														
Turb.	0.19	-0.14													
pH	0.06	0.34*	-0.34*												
Cond.	-0.24*	0.11	-0.25*	0.03											
Cu	-0.05	-0.14	0.22	-0.20	-0.14										
Zn	-0.07	-0.11	0.06	-0.02	0.16	0.14									
Fe	0.02	0.17	-0.04	0.05	0.11	0.29*	0.26*								
Co	0.03	-0.17	0.46*	-0.16	-0.31*	0.13	0.00	-0.08							
Cr	0.02	-0.18	0.45*	-0.16	-0.33*	0.14	0.00	-0.08	0.99*						
Ni	-0.04	-0.27*	0.27*	-0.12	-0.16	0.07	0.01	-0.09	0.83*	0.85*					
Al	0.07	0.11	-0.02	-0.16	0.39*	-0.09	0.02	0.28*	-0.24*	-0.26*	-0.20				
Mg	-0.08	0.14	-0.17	0.03	0.77*	-0.21	0.12	0.07	-0.30*	-0.32*	-0.16	0.38*			
Ca	-0.10	0.04	-0.14	-0.05	0.78*	-0.12	0.13	0.09	-0.31*	-0.33*	-0.15	0.44*	0.74*		
TON	-0.06	-0.11	0.03	-0.04	0.34*	-0.08	0.48*	0.00	-0.09	-0.10	-0.05	0.15	0.35*	0.36*	
SRP	-0.04	-0.11	-0.06	0.13	-0.10	0.19	0.01	-0.04	0.24*	0.25*	0.35*	-0.12	-0.06	-0.11	-0.07

\*  $p < 0.05$ .

degree of physicochemical habitat heterogeneity analysed in this study.

The exploratory linear correlation analysis of diversity and environmental variables revealed some significant correlations, although in general these were not very high (Table 2). In the case of chemical variables, cobalt and chromium concentrations showed significant linear correlations with diversity ( $R_{H'-Co} = -0.25$ ,  $p < 0.05$ ;  $R_{H'-Cr} = -0.24$ ,  $p < 0.05$ ).

A cross-correlation analysis between environmental variables showed that some were significantly correlated. In some cases as with conductivity, cobalt, chromium and magnesium concentrations the correlation with other variables were higher than 0.6 (Table 3). On the other hand, some variables such as area or copper concentration only were significantly correlated with a single other variable (conductivity and iron concentration, respectively).

### 3.1. ANN, GAM and MLR analysis

In the first phase the results of the ensemble ANNs which considered all environmental variables as inputs and the Shannon diversity index as an output, showed that any ANNs with a complexity level lower or higher than a hidden architecture with two layers and 15 neurons in each layer (16 inputs-15 neurons in the first hidden layer-15 neurons in the second hidden layer-1 output) provided significantly worse results. The validation of this ensemble model (model 16-15-15-1) showed a significant relationship between the estimated and observed Shannon index values ( $R = 0.844$ ;  $R^2 = 0.712$ ;  $MAE = 0.396$ ;  $RMSE = 0.452$ ;  $SEP = 20.517\%$ ;  $E_2 = 0.555$ ). However, the sensitivity analysis of this model showed that only four variables (conductivity, depth, turbidity and magnesium concentration) had significant ratios and therefore, the remaining variables were removed (Table 4). In this way, following the procedure shown in Fig. 2, a new model set was calibrated and validated considering conductivity, depth, turbidity and magnesium concentration as inputs.

In this second phase, the best results were obtained with a simpler hidden architecture (two layers and 10 neurons in each layer; 4

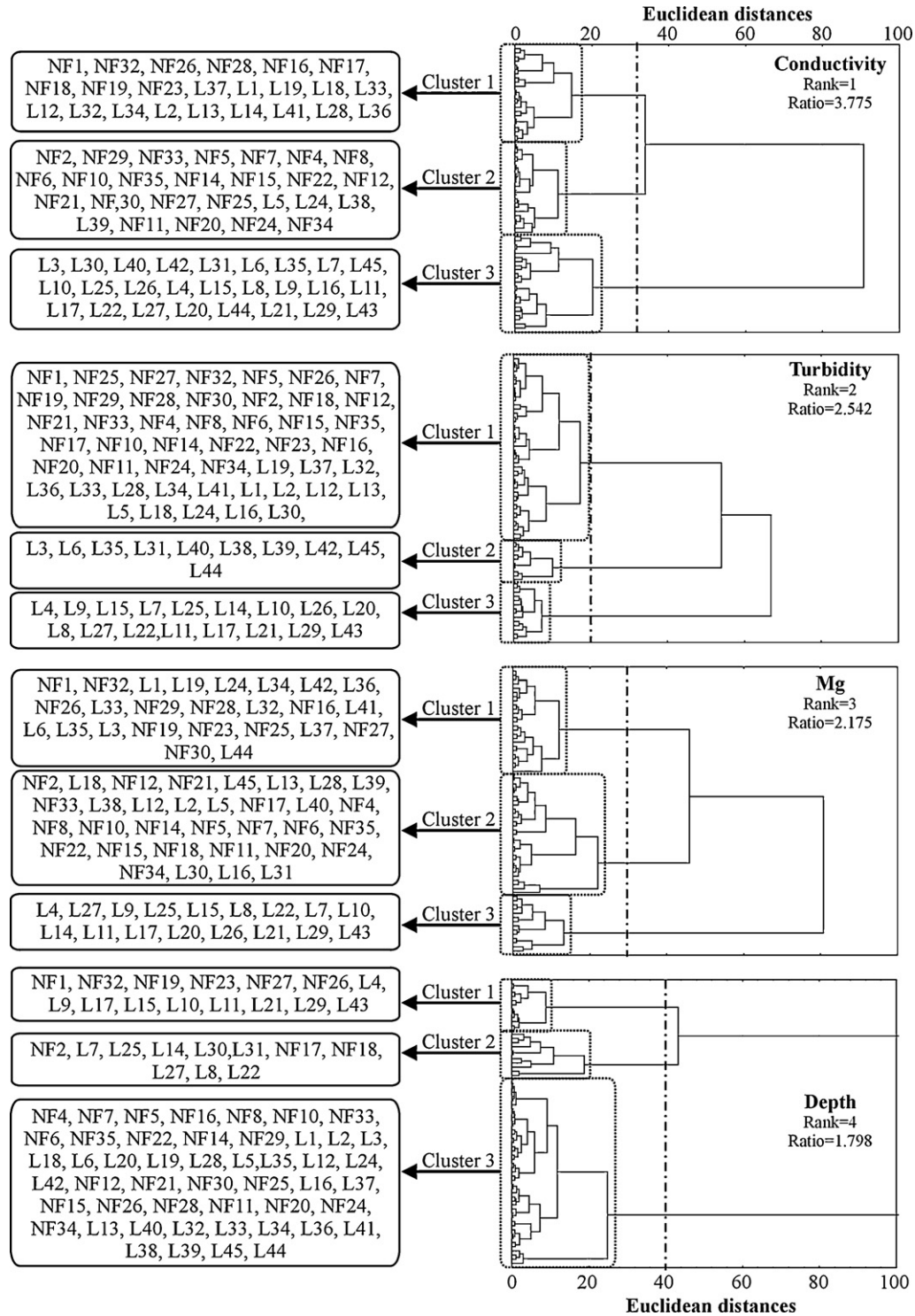
inputs-10-10-1 output). In this case (model 4-10-10-1) all the error terms, except the determination coefficient, were improved significantly in relation to the best previous model (16-15-15-1). The explained variance in the validation phase was slightly higher than 82% ( $R^2 = 0.821$ ) and the standard error was improved 7% points ( $SEP = 13.615\%$ ). The RMSE and MAE decreased to 0.223 and 0.307, respectively, and the efficiency coefficient increased to 0.789. The sensitivity analysis of this model indicated that to remove any variable would decrease the model fit capacity significantly. On the other hand, although the elimination of 12 variables in the first phase did not change the dependence relationships between conductivity, depth, turbidity and magnesium concentration, part of model fit capacity was moved toward conductivity (conductivity:

**Table 4**

First sensitivity analysis of the best ensemble ANN considering all environmental variables. Cu = copper concentration; Zn = zinc concentration; Fe = iron concentration; Co = cobalt concentration; Cr = chromium concentration; Ni = nickel concentration; Mg = magnesium concentration; Al = aluminium concentration; Ca = calcium concentration; TON = total organic nitrogen; SRP = soluble reactive phosphorous.

	Ratio	Ranking
Conductivity	1.813*	1
Turbidity	1.554*	2
Mg	1.490*	3
Depth	1.362*	4
Cr	1.195	5
Ni	1.178	6
Ca	1.143	7
Fe	1.110	8
Co	1.108	9
Cu	1.068	10
Area	1.065	11
Al	1.062	12
pH	1.047	13
TON	1.036	14
SRP	1.034	15
Zn	1.004	16
Average	1.204	

\* Selected variables.



**Fig. 2.** Hierarchical cluster analysis of ANN contribution profiles on conductivity, turbidity, magnesium concentration and depth. The dotted line indicates the number of clusters selected from the plot of amalgamation schedule analysis. In boxes, each pond is arranged in relation to its position in the dendrogram (NF = New Forest ponds; L = Lizard ponds).

rank = 1, ratio = 3.775; turbidity: rank = 2, ratio = 2.542; Mg concentration: rank = 3, ratio = 2.175; depth: rank = 4; ratio = 1.798).

The results of the fits obtained with the generalised additive models are shown in Table 5. The highest level of explained variance ( $R^2 = 0.486$ ) was provided by a GAM which considered a gamma distribution of the diversity and an identity function as link-function, but globally the results obtained with all

the GAM models were very similar. In the case of the best fit model, the relatively low explained variance implied that any non-linear  $p$ -value associated to independent variables was significant.

MLR models provided the worst results of all the approaches explored here ( $R^2$  adjusted = 0.137). The values of the beta coefficients in MLRs were also not coherent with the results obtained

**Table 5**

Generalised additive model results. In the lower panel the degrees of freedom (df), GAM coefficient and non-linear *p*-value of the best fit model are shown. Mg = magnesium concentration.

Model family	Link-function	Final deviance	Residual df	R <sup>2</sup>
Normal	Identity	31.093	52.993	0.422
Normal	Inverse	32.039	53.196	0.404
Normal	Log	31.380	52.937	0.417
Gamma	Identity	3.151	46.469	0.486*
Gamma	Inverse	3.219	47.011	0.475
Gamma	Log	3.193	46.997	0.479
Variable	df	GAM coefficient	Non-linear <i>p</i> -value	
Intercept	1.000	2.352		
Conductivity	3.145	−0.001	0.766	
Turbidity	3.154	0.005	0.105	
Mg	3.092	0.023	0.934	
Depth	3.138	0.026	0.793	

\* GAM gamma-identity.

with ANN and GAM models. For example, contrary to the findings of ANNs and GAMs, the relationships between diversity and both conductivity and turbidity were negative, whilst relationships with depth and magnesium concentration were positive. In addition, only the relationships between diversity and conductivity and magnesium concentration were statistically significant (Table 6).

### 3.2. Cluster analysis and contribution profiles

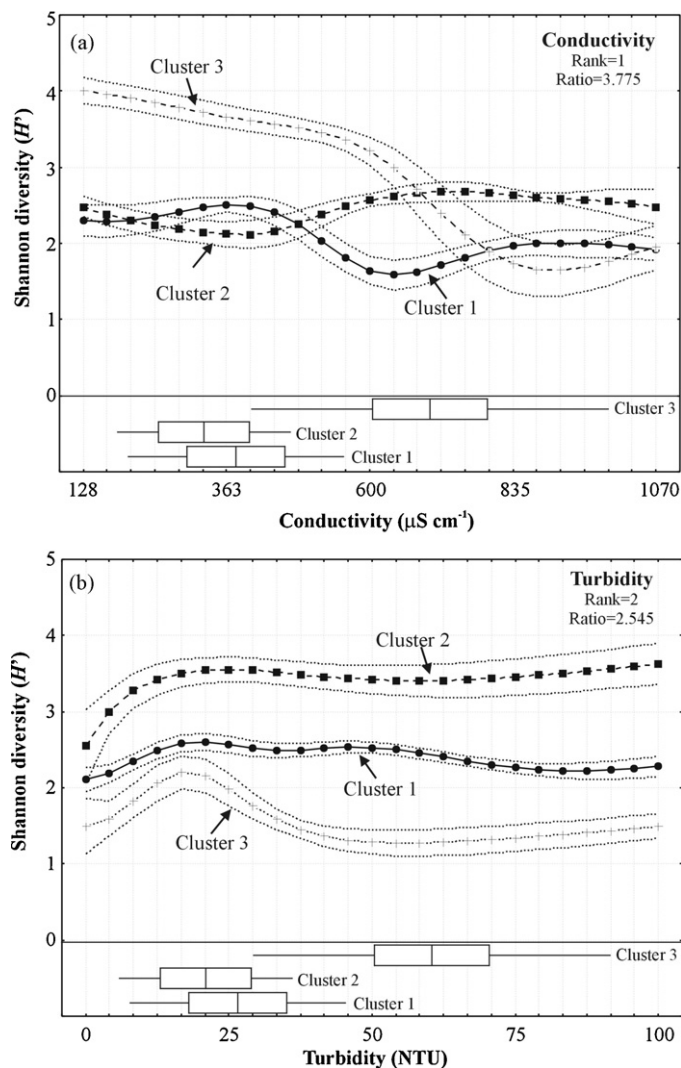
Through hierarchical cluster analysis with the Ward linkage method, the responses of the ANN model (4-10-10-1) for each variable versus the Shannon diversity index were classified according to their behaviour in each pond (Fig. 2). The analysis of the amalgamation schedule graph of each cluster analysis indicated that three clusters should be selected for each of the four variables. For conductivity, only one cluster (conductivity: cluster 3) contained ponds from a single region (the Lizard), whilst cluster 2 was almost entirely composed of New Forest ponds (only 5 ponds were from the Lizard region) and cluster 1 was made up of an equal mix of Lizard and New Forest ponds. In the case of turbidity, one cluster (cluster 1) grouped all ponds from the New Forest with some ponds from the Lizard, whilst cluster 2 and cluster 3 only contained Lizard ponds. A similar behaviour was found for depth, although for this variable, cluster 1 and cluster 2 contained a more equal configuration of ponds from both localities. In relation to magnesium concentration, a very similar grouping was observed to the one seen for conductivity.

The contribution profiles for the two main variables and each cluster versus Shannon diversity are shown in Fig. 3a and b. The profiles associated with clusters 1 and 2 for conductivity showed a low level of response of diversity to the conductivity gradient. For ponds in these two clusters, diversity oscillated between 1.5 and 2.6. In contrast, the cluster 3 profile showed a clear decrease in diversity with increasing conductivity, this trend becoming much stronger between 550 and 600  $\mu\text{S cm}^{-1}$ . For the variable ranked second in our models (turbidity), three very similar contribution

**Table 6**

Multiple linear regression results.  $N=76$ ;  $F(4,71)=3.980$ ;  $p<0.05$ ;  $R=0.428$ ;  $R^2=0.183$ ;  $R^2$  adjusted = 0.137.

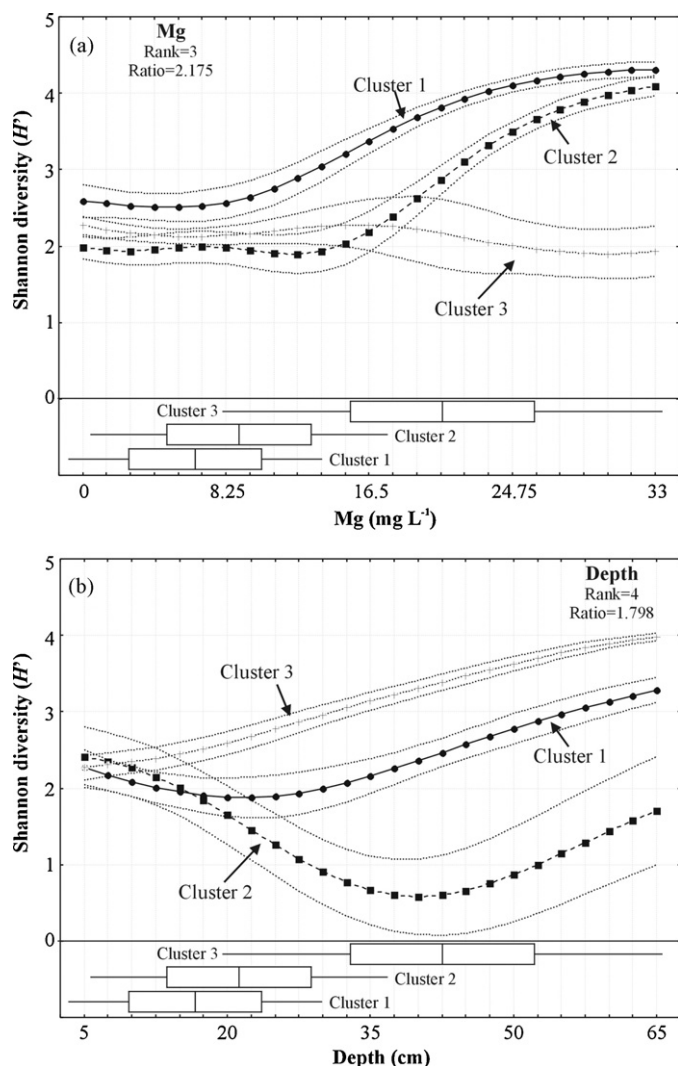
Variable	Beta coefficient	Coefficient	<i>p</i> -Value
Intercept		2.313	<0.01
Conductivity	−0.622	−0.002	<0.01
Turbidity	−0.067	−0.003	0.55
Magnesium concentration	0.500	0.038	<0.01
Depth	0.153	0.012	0.16



**Fig. 3.** Cluster centroids (continuous lines + markers)  $\pm$  95% confidence limits (dotted lines) of (a) conductivity and (b) turbidity. For each cluster and variable, the average values (vertical lines), 95% confidence limits (boxes) and standard deviations (bars) are shown below the centroids.

profiles were obtained, which differed in the level of response. The highest diversity values were observed in cluster 2 where asymptotic values ( $H' = 3.5$ – $3.6$ ) were reached at between 15 and 20 NTU. A similar relationship was found for cluster 1 although the maximum response level here was centred on  $H'$  values of 2.5. On the other hand, cluster 3 switched from an increase to a decrease in diversity at around 15 NTU, reaching minimum values of  $H'$  around 50–60 NTU. In the case of magnesium concentration (Fig. 4a), one cluster (cluster 3) did not show significant changes in  $H'$  across the range of concentrations of this variable, and clusters 1 and 2 both showed a positive sigmoid response to increases in magnesium, with a trend change between 10 and 15  $\text{mg L}^{-1}$ . Two different trends were also found for depth (Fig. 4b). Here clusters 3 and 1 showed an almost linear relationship between depth and Shannon diversity whilst ponds contained in cluster 2 displayed a parabolic relationship, with a minimum diversity value around 40 cm.

The conductivity response plot obtained from the GAM model (Fig. 5a) showed that this variable had a positive effect on diversity at values between the minimum recorded and 500–550  $\mu\text{S cm}^{-1}$ , whilst that the model indicated negative effects above these values. Turbidity had a negative effect on beetle diversity below 10–15 NTU



**Fig. 4.** Cluster centroids (continuous lines + markers)  $\pm$  95% confidence limits (dotted lines) of (a) magnesium concentration and (b) depth. For each cluster and variable, the average values (vertical lines), 95% confidence limits (boxes) and standard deviations (bars) are shown below centroids.

and an asymptotic response around 20–25 NTU, although above 85 NTU again the model indicated a negative effect (Fig. 5b). On the other hand, the response curve of magnesium concentration showed negative effects on diversity from 0 to 12–15  $\text{mg L}^{-1}$  and positive effects above these values (Fig. 5c). For this variable, the GAM model provided the more linear response (GAM  $df=3.092$ , Table 5) of all variables examined. In the case of depth, the response curve showed a strong positive effect above 20 cm and a quasi negative linear response below this value (Fig. 5d).

The hidden effect of the non-selected variables on the response of diversity to conductivity, turbidity, magnesium concentration and depth were analysed for three restrictions established through visual analysis of the ANN contribution profiles and the partial residuals of the GAM models (see below Table 7). In this way, ANOVA and Tukey HSD tests were carried out for each variable in the ponds grouped in each cluster. Only the significant effects are shown in Table 7. The results indicate that a large proportion of the variability in pond area was accounted for by conductivity, turbidity and magnesium concentration, whilst the effects of iron, cobalt, chromium, nickel, aluminium and calcium concentration seemed to be largely accounted for by conductivity. On

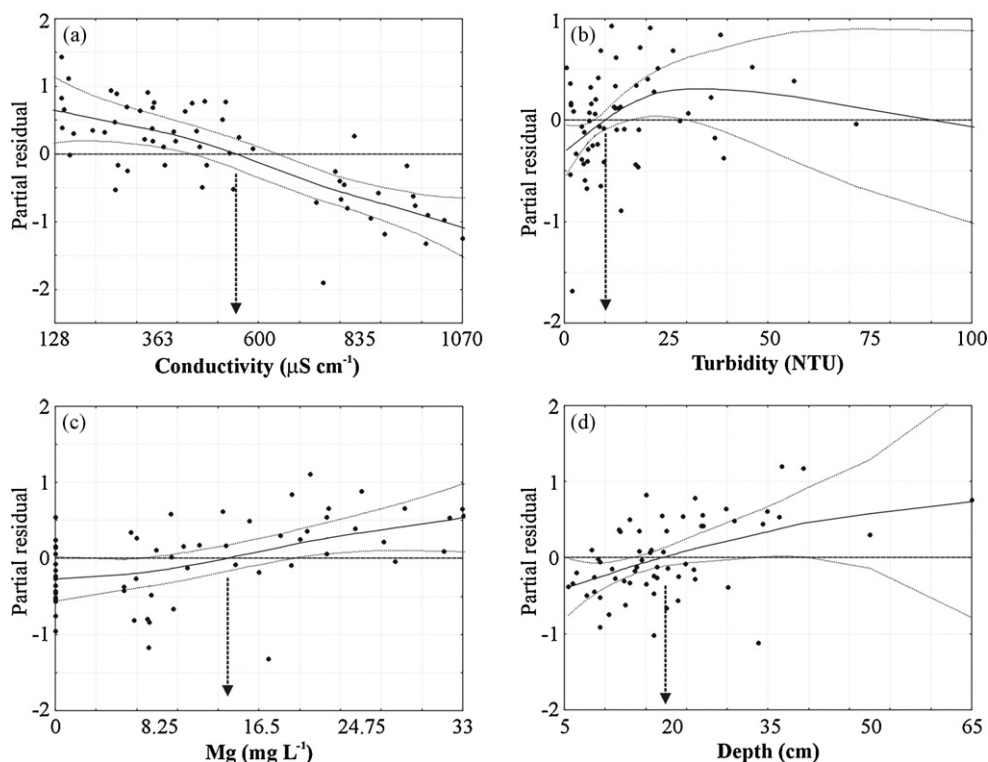
the other hand, turbidity and magnesium concentration accounted for variation in total organic nitrogen (TON). Non-significant hidden effects were found for the variable ranked in fourth position (depth).

#### 4. Discussion

In this paper, the effects of environmental parameters on coleopteran diversity in a series of temporary and fluctuating ponds were explored using multilayer perceptron neural networks, generalised additive models and multiple linear regressions. Results indicate that water beetle diversity is driven by both physical and chemical gradients and that these gradients clearly influence diversity in a non-linear fashion. The ANN models obtained here clearly indicate that a high proportion of the variation in species diversity can be explained by models containing relatively few (in our case four) environmental parameters. Our study also highlights the fact that water beetle communities across the sampled ponds fall into a number of discrete groups, which respond in subtly yet significantly different ways to environmental gradients. The ANN approach applied here resulted in models explaining >80% of the variance in diversity observed across ponds. This relatively high fit is in line with previous applications of ANNs in other systems, such as Lek et al. (1996a,b) and Lek-Ang et al. (1999, 2007). 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 case, the results showed that ANNs explained on average 79% of the variation in collembolan diversity variability versus 28% explained by linear approaches, indicating a clear non-linear relationship between diversity and environmental factors.

The results obtained with ANNs were clearly better than those obtained with GAM or MLR models, demonstrating the advantage of this approach in understanding and predicting community–environment relationships in these systems. However, in spite of the fact that the explained variance and remaining error terms were worse in the GAM models, both these and ANNs forecasted changes in coleopteran diversity in a broadly similar fashion. Few studies have explicitly examined the relative performance of ANNs and GAMs in an ecological framework, but the few which have (Thomaes et al., 2008; Schlink et al., 2003) report similar findings. Schlink et al. (2003) demonstrated that both ANN and GAM models perform better than other types of model such as linear and multiple linear regressions, transfer function models, dynamic regressions implemented with a Kalman filter, wavelet approaches and deterministic models.

Cluster analyses of Lizard and New Forest ponds based on contribution profiles of conductivity, turbidity, magnesium concentration and depth, showed a congruent ordination related to the weight of each variable in the model. Sampled ponds could be divided into a series of well defined groups, whose beetle diversity responded in different ways to changes in these variables. For conductivity, diversity across one group of ponds (cluster 3), entirely from the Lizard, responded in quite a different manner to that observed elsewhere, showing a clear decline in diversity with increasing conductivity. This group was largely composed of small ephemeral pools associated with old trackways on the Lizard, which are characterised by relatively high conductivities due to their mineral substrate being derived from serpentine bedrock (see Bilton et al., 2009). The influence of conductivity on macroinvertebrate community composition and diversity is well documented (Kapoor, 1978; Lemly, 1982; Williams et al., 1997; Williams and Williams, 1998; Blasius and Merritt, 2002; Biggs et al., 2005; De Jonge et al., 2008), and at low values, the vast majority of studies document a positive relationship between conductivity and species



**Fig. 5.** GAM model partial residual analysis. Continuous lines are spline fits and dotted lines are 95% confidence bands of (a) conductivity (b) turbidity (c) magnesium concentration and (d) depth. Dotted arrows indicate trend change.

richness. Very high conductivities are associated with a decline in the diversity of freshwater macroinvertebrates, this operating largely through the osmotic challenges of high ion concentrations (Macan, 1974; Blasius and Merritt, 2002). Such a decline in diversity with increasing conductivity can be clearly observed for the

ponds contained in cluster 3 and was also globally predicted by GAM and MLR models. For ANNs and GAMs, the models indicated that a trend change occurred at around 550  $\mu\text{S cm}^{-1}$ . In contrast, the ANN showed that the ponds grouped in clusters 1 and 2 were not very sensitive to the conductivity gradient recorded, even below

**Table 7**

ANOVA and Tukey HSD tests of the influence of non selected factors on conductivity, turbidity and magnesium concentration. The restrictions were established from the visual analysis of ANN contribution profiles and the partial residuals of GAM models. In Tukey HSD tests the average value for each cluster is shown. Different letters indicate statistically significant differences ( $\alpha$  level = 0.05). Fe = iron concentration; Co = cobalt concentration; Cr = chromium concentration; Ni = nickel concentration; Al = aluminium concentration; Ca = calcium concentration; TON = total organic nitrogen.

Variable	Restrictions <sup>a</sup>	ANOVA	Post hoc Tukey HSD test		
			Cluster 1	Cluster 2	Cluster 3
Area	Conductivity < 550 $\mu\text{S cm}^{-1}$	$F(2,50) = 3.475; p < 0.05$	Area = 1130 m <sup>2</sup> (a)	Area = 492 m <sup>2</sup> (a,b)	Area = 3325 m <sup>2</sup> (b)
	All range of turbidity	$F(2,73) = 3.204; p < 0.05$	Area = 766 m <sup>2</sup> (a,b)	Area = 2291 m <sup>2</sup> (b)	Area = 143 m <sup>2</sup> (a)
	Turbidity < 15 NTU	$F(2,49) = 5.346; p < 0.01$	Area = 800 m <sup>2</sup> (a)	Area = 4917 m <sup>2</sup> (b)	Area = 435 m <sup>2</sup> (a)
	All range of Mg	$F(2,73) = 4.483; p < 0.05$	Area = 1855 m <sup>2</sup> (a)	Area = 384 m <sup>2</sup> (b)	Area = 144 m <sup>2</sup> (b)
	Mg > 15 mg L <sup>-1</sup>	$F(2,21) = 7.629; p < 0.01$	Area = 1509 m <sup>2</sup> (a)	Area = 42 m <sup>2</sup> (b)	Area = 88 m <sup>2</sup> (b)
Fe	Conductivity < 550 $\mu\text{S cm}^{-1}$	$F(2,50) = 4.154; p < 0.05$	Fe = 0.214 mg L <sup>-1</sup> (a,b)	Fe = 0.112 mg L <sup>-1</sup> (a)	Fe = 0.325 mg L <sup>-1</sup> (b)
Co	All range of conductivity	$F(2,73) = 4.673; p < 0.05$	Co = 0.144 mg L <sup>-1</sup> (a,b)	Co = 0.451 mg L <sup>-1</sup> (a)	Co = 0.007 mg L <sup>-1</sup> (b)
	Conductivity > 550 $\mu\text{S cm}^{-1}$	$F(2,20) = 3.613; p < 0.05$	Co = 0.030 mg L <sup>-1</sup> (a)	Co = 0.000 mg L <sup>-1</sup> (a,b)	Co = 0.008 mg L <sup>-1</sup> (b)
Cr	All range of conductivity	$F(2,73) = 5.490; p < 0.01$	Cr = 0.082 mg L <sup>-1</sup> (a,b)	Cr = 0.251 mg L <sup>-1</sup> (a)	Cr = 0.006 mg L <sup>-1</sup> (b)
Ni	All range of conductivity	$F(2,73) = 6.012; p < 0.01$	Ni = 0.033 mg L <sup>-1</sup> (a)	Ni = 0.101 mg L <sup>-1</sup> (b)	Ni = 0.049 mg L <sup>-1</sup> (a)
	Conductivity < 550 $\mu\text{S cm}^{-1}$	$F(2,50) = 4.956; p < 0.05$	Ni = 0.033 mg L <sup>-1</sup> (a)	Ni = 0.103 mg L <sup>-1</sup> (b)	Ni = 0.025 mg L <sup>-1</sup> (a,b)
Al	All range of conductivity	$F(2,73) = 9.823; p < 0.01$	Al = 0.063 mg L <sup>-1</sup> (a)	Al = 0.029 mg L <sup>-1</sup> (a)	Al = 0.150 mg L <sup>-1</sup> (b)
	Conductivity < 550 $\mu\text{S cm}^{-1}$	$F(2,50) = 4.693; p < 0.01$	Al = 0.066 mg L <sup>-1</sup> (a,b)	Al = 0.027 mg L <sup>-1</sup> (a)	Al = 0.130 mg L <sup>-1</sup> (b)
Ca	All range of conductivity	$F(2,73) = 24.915; p < 0.01$	Ca = 2.552 mg L <sup>-1</sup> (a)	Ca = 0.886 mg L <sup>-1</sup> (a)	Ca = 6.295 mg L <sup>-1</sup> (b)
	Conductivity < 550 $\mu\text{S cm}^{-1}$	$F(2,50) = 5.463; p < 0.01$	Ca = 2.411 mg L <sup>-1</sup> (a)	Ca = 0.662 mg L <sup>-1</sup> (b)	Ca = 3.046 mg L <sup>-1</sup> (a)
TON	All range of turbidity	$F(2,73) = 3.160; p < 0.05$	TON = 0.205 mg L <sup>-1</sup> (a)	TON = 0.195 mg L <sup>-1</sup> (a,b)	TON = 0.891 mg L <sup>-1</sup> (b)
	Turbidity > 15 NTU	$F(2,22) = 5.126; p < 0.01$	TON = 0.303 mg L <sup>-1</sup> (a)	TON = 0.188 mg L <sup>-1</sup> (a)	TON = 2.887 mg L <sup>-1</sup> (b)
	All range of Mg	$F(2,73) = 3.620; p < 0.05$	TON = 0.252 mg L <sup>-1</sup> (a,b)	TON = 0.161 mg L <sup>-1</sup> (a)	TON = 0.939 mg L <sup>-1</sup> (b)

<sup>a</sup> Restrictions in the conductivity analysis: (a) all conductivity range (128–1070  $\mu\text{S cm}^{-1}$ ); (b) <550  $\mu\text{S cm}^{-1}$ ; (c) >550  $\mu\text{S cm}^{-1}$ . Restrictions in the turbidity analysis: (a) all turbidity range (0.1–100 NTU); (b) <15 NTU; (c) >15 NTU. Restrictions in the magnesium concentration analysis: (a) all Mg concentration range (0–33 mg L<sup>-1</sup>); (b) <15 mg L<sup>-1</sup>; (c) >15 mg L<sup>-1</sup>. Restriction in the depth analysis: (a) all range of depth (5.6–65 cm); (b) <20 cm; (c) >20 cm.

550  $\mu\text{S cm}^{-1}$ , which may result from the effect of other variables in these sites.

In this context, it is possible to understand the statistical functional relationships found for the selected variables in relation to beetle diversity, through the proxy effects of additional variables. In the case of conductivity, the differences observed between cluster 3 and clusters 1 and 2 may be a consequence of the combination of high conductivities and high concentrations of metals such as aluminium and iron (cluster 3), and low conductivities and low concentrations of other metals such as cobalt, chromium and nickel (clusters 1 and 2), which is consistent with findings of for other animal taxa (Driscoll et al., 1980; Bervoets et al., 2005).

In the same way, different proxy effects can be suggested for the remaining selected variables. For example, it is difficult to provide a direct functional explanation for the increase in diversity predicted by the ANN from 15  $\text{mg L}^{-1}$  magnesium in the ponds contained in clusters 1 and 2, particularly bearing in mind the fact that magnesium concentration was significantly correlated with conductivity. However, when other variables were analysed within the framework of the observed clusters based on magnesium, the response of diversity to this ion is more clear. It would appear that at least two factors, total organic nitrogen (TON) and pond area could account for the apparent relationship between diversity and magnesium concentration. TON was significant higher for the ponds grouped in cluster 3, where the ANN predicted low diversities with higher magnesium concentrations. This result indicates, that, as has been observed in fish communities (Persson et al., 1991; Gutiérrez-Estrada et al., 2008), moderate eutrophication (around 0.200  $\text{mg N L}^{-1}$ ) raises coleopteran diversity. A similar explanation can be applied to the turbidity profiles.

In relation to pond area, several authors have shown that this is one of most important factors related to macroinvertebrate diversity in ponds (Lassen, 1975; Driver, 1977; Fryer, 1985). Bilton et al. (2006, 2009) reported that the pond area had a significant impact on macroinvertebrate assemblage structure in Lizard and New Forest sites. In the present study, the mean area of ponds contained in cluster 1 for magnesium concentration was significantly higher than the mean area of ponds grouped in clusters 2 and 3. Additionally, the effect of the pond area seemed to be one of the most important factors behind the patterns observed for turbidity and conductivity. The highest diversity levels in relation to turbidity and conductivity were found in cluster 2 and cluster 3, respectively, which had the highest mean pond area.

Apparently, the ponds grouped in the three depth profiles did not have proxy associations with any of the other variables, indicating an *a priori* direct functional relationship between diversity and depth. In this way, the ANN predicted a *quasi* linear increase of diversity for depths greater than 20 cm in ponds included in the clusters 1 and 3, which is in agreement with several earlier studies on temporary ponds (Schneider and Frost, 1996; Brooks, 2000; Batzer et al., 2004), where depth is often a surrogate of hydroperiod. However, a very different pattern was found for the ponds contained in cluster 2. In these ponds, the mean diversity was significant lower than in the ponds of clusters 1 and 3, and a diversity decrease was observed between depths of 5 and 40 cm, with an increase again above 40 cm. Ponds in this cluster had the highest average depth, but were also deeper in proportion to their area than ponds contained in the other two clusters (area/depth ratio ANOVA:  $F(2,73) = 9.109$ ,  $p < 0.05$ ; Tukey HSD test:  $p_{\text{cluster1-cluster2}} < 0.05$ ,  $p_{\text{cluster1-cluster3}} = 0.409$ ,  $p_{\text{cluster2-cluster3}} < 0.05$ ), both of which point to a longer wet phase in these sites. Larger, deeper sites within this cluster show the least water level fluctuation of any of the ponds included in this study, and have a relatively species poor beetle fauna. This finding is in accordance with other studies which report a greater number of coleopteran species in

temporary ponds than in more permanent sites (Nicolet, 2001; Nicolet et al., 2004; Della Bella et al., 2005).

This study has highlighted the non-linear nature of the relationship between water beetle diversity in lowland pond systems and environmental variables, and illustrates the superior performance of artificial neural networks in modelling diversity–environment interactions. The ANNs generalisation capacity (explaining more than 80% of variance in diversity) suggests that realistic simulations of coleopteran diversity can be made under different environmental scenarios. Also, the model may be used to detect changes in coleopteran diversity and therefore assess changes in macroinvertebrate assemblages over time. Comparison of contribution profiles after the incorporation of new data could illuminate the mechanisms behind observed changes in diversity, which would be useful in evaluating the risks of different management options. In short, these relatively simple models have high predictive power, and should be employed to forecast how pond communities may respond to environmental changes in the future.

### Acknowledgments

The authors wish to express their gratitude to the Spanish Ministry of Education and Science for providing financial support through the 'José Castillejo' Grant Program to Dr. J.C. Gutiérrez-Estrada for a research stay at the Marine Biology and Ecology Research Centre of the University of Plymouth (UK) (Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional I+D+I 2008–2011). Much of the fieldwork which generated these data was conducted under a PhD studentship undertaken by Louise McAbendroth. We are grateful to Natural England and the University of Plymouth for financial support for this project.

### References

- Amiaud, B., Pervanchon, F., Plantureux, S., 2005. An expert model for predicting species richness in grasslands: flora-predict. *Grassland Science in Europe* 10, 103–106.
- Ancil, F., Rat, A., 2005. Evaluation of neural network stream flow forecasting on 47 watersheds. *Journal of Hydrologic Engineering* 10, 85–88.
- Batzer, D.P., Palik, B.J., Buech, R., 2004. Relationships between environmental characteristics and macroinvertebrate communities in seasonal woodland ponds of Minnesota. *Journal of the North American Benthological Society* 23, 50–68.
- Bechara, J.A., 1996. The relative importance of water quality, sediment composition and floating vegetation in explaining the macrobenthic community structure of floodplain lakes (Parana river, Argentina). *Hydrobiologie* 333, 95–109.
- Bervoets, L., Knaepkens, G., Eens, M., Blust, R., 2005. Fish community response to metal pollution. *Environmental Pollution* 138, 338–349.
- Biggs, J., Williams, P., Whitfield, P., Nicolet, P., Weatherby, A., 2005. 15 years of pond assessment in Britain: results and lessons learned from the work of Pond Conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* 15, 693–714.
- Bilton, D.T., Foggo, A., Rundle, S.D., 2001. Size, permanence and the proportion of predators in ponds. *Archiv für Hydrobiologie* 151, 451–458.
- Bilton, D.T., McAbendroth, L.C., Nicolet, P., Bedford, A., Rundle, S.D., Foggo, A., Ramsay, P.M., 2009. Ecology and conservation status of temporary and fluctuating ponds in two areas of southern England. *Aquatic Conservation: Marine and Freshwater Ecosystems* 19, 134–146.
- Bilton, D.T., McAbendroth, L.M., Bedford, A., Ramsay, P.M., 2006. How wide to cast the net? Cross-taxon congruence of species richness, community similarity and indicator taxa in ponds. *Freshwater Biology* 51, 578–590.
- Blasius, B.J., Merritt, R.W., 2002. Field and laboratory investigations on the effects of road salt (NaCl) on streams macroinvertebrate communities. *Environmental Pollution* 120, 219–231.
- Blockson, K.A., Johnson, B.R., 2009. Development of a regional macroinvertebrate index for large river bioassessment. *Ecological Indicators* 9, 313–328.
- Brooks, T.S., 2000. Annual and seasonal variation and the effects of hydroperiod on benthic macroinvertebrates of seasonal forest ('vernal') ponds in central Massachusetts, USA. *Wetland* 20, 707–715.
- Brower, J.E., Zar, J.H., von Ende, C.N., 1998. *Field and Laboratory Methods for General Ecology*. McGraw and Hill, Dubuque, IA.
- Céréghino, R., Giraudel, J.L., Compin, A., 2001. A spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen selforganizing maps. *Ecological Modelling* 146, 167–180.
- Collison, N.H., Biggs, J., Corfield, A., Hodson, M.J., Walker, D., Whitfield, M., Williams, P.J., 1995. Temporary and permanent ponds: an assessment of the effects of

- drying out on the conservation value of aquatic macroinvertebrate communities. *Biological Conservation* 74, 125–134.
- Cottenie, K., 2005. Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters* 8, 1175–1182.
- Cowell, B.C., Remley, A.H., Lynch, D.M., 2004. Seasonal changes in the distribution and abundance of benthic invertebrates in six headwater streams in central Florida. *Hydrobiologia* 522, 99–115.
- De Jonge, M., Van de Vijver, B., Blust, R., Bervoets, L., 2008. Responses of aquatic organisms to metal pollution in a lowland river in Flanders: a comparison of diatoms and macroinvertebrates. *Science of the Total Environment* 407, 615–629.
- Dedecker, A.P., Goethals, P.L.M., D'Heygere, T., Gevrey, M., Lek, S., De Pauw, N., 2005. Application of artificial neural network models to analyse the relationships between *Gammarus pulex* L. (Crustacea, Amphipoda) and river characteristics. *Environmental Monitoring and Assessment* 111, 223–241.
- Della Bella, V., Bazzanti, M., Chiarotti, F., 2005. Macroinvertebrate diversity and conservation status of Mediterranean ponds in Italy: water permanence and mesohabitat influence. *Aquatic Conservation: Marine and Freshwater Ecosystems* 15, 583–600.
- Driscoll, C.T., Baker, J.P., Bisogni, J.J., Schofield, C.L., 1980. Effect of aluminium speciation on fish in diluted acidified waters. *Nature* 284, 161–164.
- Driver, E.A., 1977. Chironomid communities in small prairie ponds: some characteristics and controls. *Freshwater Biology* 7, 121–133.
- Fernández-Aláez, C., de-Soto, J., Fernández-Aláez, M., García-Criado, F., 2002. Spatial structure of the caddisfly (Insecta, Trichoptera) communities in a river basin in NW Spain affected by coal mining. *Hydrobiologia* 487, 193–205.
- Foggo, A., Rundle, S.D., Bilton, D.T., 2003. The net result: evaluating species richness extrapolations techniques for pond invertebrates. *Freshwater Biology* 48, 1–9.
- Friday, L.E., 1987. The diversity of macroinvertebrate and macrophyte communities in ponds. *Freshwater Biology* 18, 87–104.
- Fryer, G., 1985. Crustacean diversity in relation to the size of water bodies: some facts and problems. *Freshwater Biology* 15, 347–361.
- Goethals, P.L.M., Dedecker, A.P., Gabriels, W., Lek, S., De Pauw, N., 2007. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. *Aquatic Ecology* 41, 491–508.
- 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 computation neural networks. *Aquatic Sciences* 62, 350–363.
- Gutiérrez-Estrada, J.C., Vasconcelos, R., Costa, M.J., 2008. Estimating fish community diversity from environmental features in the Tagus estuary (Portugal): multiple linear regression and artificial neural network approaches. *Journal of Applied Ichthyology* 24, 150–162.
- Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. *Multivariate Data Analysis*, Fifth ed. Prentice Hall International, Inc., London, UK.
- Hartigan, J.A., 1975. *Clustering Algorithms*. Wiley, New York.
- Hastie, T.J., Tibshirani, R.J., 1990. *Generalized Additive Models*. Chapman & Hall/CRC, Boca Raton.
- Haidekker, A., Hering, D., 2008. Relationship between benthic insects (Ephemeroptera, Plecoptera, Coleoptera, Trichoptera) and temperature in small and medium-sized streams in Germany: a multivariate study. *Aquatic Ecology* 42, 463–481.
- Hunter, A., Kennedy, L., Henry, J., Ferguson, I., 2000. Application of neural networks and sensibility analysis to improved prediction of trauma survival. *Computer Methods and Programs Biomedicine* 62, 11–19.
- Irmiler, U., Heller, K., Meyer, H., Hans-Dieter, R., 2002. Zonation of round beetles (Coleoptera: Carabidae) and spiders (Araneida) in salt marshes at the North and the Baltic Sea and the impact of the predicted sea level increase. *Biodiversity and Conservation* 11, 1129–1147.
- Iyer, M.S., Rhinehart, R.R., 1999. A method to determine the required number of neural-network training repetitions. *IEEE Transactions on Neural Networks* 10 (2), 427–432.
- James, F.C., McCulloch, C.E., 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review in Ecology and Systematics* 21, 129–166.
- Jeffries, M., 2005. Local-scale turnover of pond insects: intra-pond habitat quality and inter-pond geometry are both important. *Hydrobiologia* 543, 207–220.
- Kapoor, N.N., 1978. Effect of salinity on the osmoregulatory cells in the tracheal gills of the stonefly nymph, *Paragnetia media* (Plecoptera: Perlidae). *Canadian Journal of Zoology* 56, 2608–2613.
- Kasangaki, A., Chapman, L.J., Balirwa, J., 2008. Land use and the ecology of benthic macroinvertebrate assemblages of high-altitude rainforest streams in Uganda. *Freshwater Biology* 53, 681–697.
- Kiflawi, M., Eitam, A., Blaustein, L., 2003. The relative impact of local and regional processes on macro-invertebrate species richness in temporary ponds. *Journal of Animal Ecology* 72, 447–452.
- Kitanidis, P.K., Bras, R.L., 1980. Real time forecasting with a conceptual hydrological model. 2. Applications and results. *Water Resources Research* 16 (6), 1034–1044.
- Laë, R., Lek, S., Moreau, J., 1999. Predicting fish yield of African lakes using neural networks. *Ecological Modelling* 120, 325–335.
- Lahr, J., Diallo, A.O., Ndour, K.B., Badji, A., Diouf, P.S., 1999. Phenology of invertebrates living in a sahelian temporary pond. *Hydrobiologia* 405, 189–205.
- Lassen, H.H., 1975. The diversity of freshwater snails in view of the equilibrium theory of island biogeography. *Oecologia* 19, 1–8.
- Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water Resources Research* 35 (1), 233–241.
- Lek, S., Guegan, J.F., 1999. Artificial neural networks as a tool in ecological modeling. An introduction. *Ecological Modelling* 120, 65–73.
- Lek, S., Belaud, A., Baran, P., Dimopoulos, I., Delacoste, M., 1996a. Role of some environmental variables in trout abundance using neural networks. *Aquatic Living Resource* 9, 23–29.
- Lek, S., Delacoste, M., Baran, P., Lauga, J., Aulagnier, S., 1996b. Application of neural network for modeling in ecology. *Ecological Modelling* 90, 39–52.
- Lek-Ang, S., Deharverng, L., Lek, S., 1999. Predictive models of collemolan diversity and abundance in a riparian habitat. *Ecological Modelling* 120, 247–260.
- Lek-Ang, S., Park, Y.S., Ait-Mouloud, S., Deharverng, L., 2007. Collemolan communities in a peat bog versus surrounding forest analysed by using self-organizing map. *Ecological Modelling* 203, 9–17.
- Lemly, D.A., 1982. Modification of benthic insect communities in polluted streams: combined effects of sedimentation and nutrient enrichment. *Hydrobiologia* 87, 229–245.
- Lencioni, V., Maiolini, B., Marziali, L., Lek, S., Rossaro, B., 2007. Macroinvertebrate assemblages in glacial stream systems: a comparison of linear multivariate methods with artificial neural networks. *Ecological Modelling* 203, 119–131.
- Macan, T.T., 1974. *Freshwater Ecology*, Second ed. Longman Group, London.
- Marchant, R., Ryan, D., Metzeling, L., 2006. Regional and local species diversity patterns for lotic invertebrates across multiple drainage basins in Victoria. *Marine and Freshwater Research* 57, 675–684.
- Miserendino, M.L., Archangelsky, M., 2006. Aquatic Coleoptera distribution and environmental relationships in a large Patagonian river. *International Review in Hydrobiology* 91, 423–437.
- Moss, B., 1998. *Ecology of Freshwater, Man and Medium, Past to Future*. Blackwell Scientific, Oxford.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. I. A discussion of principles. *Journal of Hydrology* 10, 282–290.
- Nicolet, P., 2001. Temporary ponds in the UK: a critical biodiversity resource for freshwater plants and animals. *Freshwater Forum* 17, 16–25.
- Nicolet, P., Biggs, J., Hodson, M.J., Reynolds, C., Whitfield, M., Williams, P., 2004. The wetland plant and macroinvertebrate assemblages of temporary pond in England and Wales. *Biological Conservation* 120, 261–278.
- Park, Y.S., Céréghino, R., Compin, A., Lek, S., 2003. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. *Ecological Modelling* 160, 265–280.
- 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. *Journal of Fish Biology* 38, 281–293.
- Pulido-Calvo, I., Portela, M.M., 2007. Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds. *Journal of Hydrology* 332, 1–15.
- Ripley, B.J., Simovich, M.A., 2009. Species richness on islands in time: variation in ephemeral pond crustacean communities in relation to habitat duration and size. *Hydrobiologia* 617, 181–196.
- Rohasliney, H., Jackson, D.C., 2008. Lignite mining and stream channelization influences on aquatic macroinvertebrate assemblages along the Natchez Trace Parkway, Mississippi, USA. *Hydrobiologia* 598, 149–162.
- Rundle, S.D., Foggo, A., Choisel, V., Bilton, D.T., 2002. Are distribution patterns linked to dispersal mechanism? An investigation using pond invertebrate assemblages. *Freshwater Biology* 4, 1571–1581.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. 'Learning' representations by back-propagation errors. *Nature* 323, 533–536.
- Sánchez-Fernández, D., Abellán, P., Mellado, A., Velasco, J., Millán, A., 2006. Are water beetles good indicators of biodiversity in Mediterranean aquatic ecosystems? The case of the Segura river basin (SE Spain). *Biodiversity and Conservation* 15, 4507–4520.
- Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M., Doyle, M., 2003. A rigorous inter-comparison of ground-level ozone predictions. *Atmospheric Environment* 37, 3237–3253.
- Schneider, D.W., Frost, T.M., 1996. Habitat duration and community structure in temporary ponds. *Journal of the North American Benthological Society* 15, 64–86.
- Shepherd, A.J., 1997. *Second-order Methods for Neural Networks*. Springer, NewYork.
- Stenseth, N.C., Lekve, K., Gjøsaeter, J., 2002. Modeling species richness controlled by community-intrinsic and community-extrinsic processes: coastal fish communities as an example. *Population Ecology* 44, 165–178.
- Tan, Y., Van Cauwenbergh, A., 1999. Neural network based d-step-ahead predictors for nonlinear systems with time delay. *Engineering Applications of Artificial Intelligence* 12, 21–25.
- Thomaes, A., Kervyn, T., Maes, D., 2008. Applying species distribution modeling for the conservation of the threatened sproxylic Stag Beetle (*Lucanus cervus*). *Biological Conservation* 141, 1400–1410.
- Tsoukalas, L.H., Uhrig, R.E., 1997. *Fuzzy and Neural Approaches in Engineering*. Wiley Interscience, New York.
- Valladares, L.F., Garrido, J., García-Criado, F., 2002. The assemblages of aquatic Coleoptera from shallow lakes in the northern Iberian Meseta: influence of environmental variables. *European Journal of Entomology* 99 (3), 289–298.
- Ventura, S., Silva, M., Pérez-Bendito, D., Hervás, C., 1995. Artificial neural networks for estimation of kinetic analytical parameters. *Analytical Chemistry* 67 (9), 1521–1525.

- Watts, M.J., Worner, S.P., 2008. Comparing ensemble and cascaded neural networks that combine biotic and abiotic variables to predict insect species distribution. *Ecological Modelling* 3, 354–366.
- Wellborn, G.A., Skelly, D.K., Werner, E.E., 1996. Mechanisms creating community structure across a freshwater habitat gradient. *Annual Review of Ecology and Systematics* 27, 337–363.
- Williams, D.D., 1996. Environmental constraints in temporary fresh waters and their consequences for the insect fauna. *Journal of the North American Benthological Society* 15, 634–650.
- Williams, D.D., 2005. *The Ecology of Temporary Waters*, Second ed. Blackwell Science, Oxford.
- Williams, D.D., Williams, N.E., 1998. Aquatic insects in an estuarine environment: densities, distribution and salinity tolerance. *Freshwater Biology* 39, 411–421.
- Williams, D.D., Williams, N.E., Cao, Y., 1997. Spatial differences in macroinvertebrates community structure in springs in southeastern Ontario in relation to their chemical and physical environments. *Canadian Journal of Zoology* 75, 1404–1414.