Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network

Young-Seuk Park[a,*], Piet F.M. Verdonschot[b], Tae-Soo Chon[c], Sovan Lek[a]

[a] *CESAC, UMR 5576, CNRS—Université Paul Sabatier, 118 Route de Narbonne, Toulouse, Cedex 31062, France*
[b] *Alterra, Green World Research, Department of Freshwater Ecosystems, P.O. Box 47, AA Wageningen 6700, The Netherlands*
[c] *Division of Biological Sciences, Pusan National University, Geumjeong-gu, Pusan 609-735, South Korea*

Abstract

A counterpropagation neural network (CPN) was applied to predict species richness (SR) and Shannon diversity index (SH) of benthic macroinvertebrate communities using 34 environmental variables. The data were collected at 664 sites at 23 different water types such as springs, streams, rivers, canals, ditches, lakes, and pools in The Netherlands. By training the CPN, the sampling sites were classified into five groups and the classification was mainly related to pollution status and habitat type of the sampling sites. By visualizing environmental variables and diversity indices on the map of the trained model, the relationships between variables were evaluated. The trained CPN serves as a 'look-up table' for finding the corresponding values between environmental variables and community indices. The output of the model fitted SH and SR well showing a high accuracy of the prediction ($r > 0.90$ and 0.67 for learning and testing process, respectively) for both SH and SR. Finally, the results of this study, which uses the capability of the CPN for patterning and predicting ecological data, suggest that the CPN can be effectively used as a tool for assessing ecological status and predicting water quality of target ecosystems.
© 2002 Elsevier Science Ltd. All rights reserved.

1. Introduction

Understanding community patterns is important to manage target ecosystems. Especially in aquatic ecosystems, communities of benthic macroinvertebrates are important to monitor changes of the target system. Benthic macroinvertebrates constitute a heterogeneous assemblage of animal phyla and consequently some members respond to stresses placed upon them, and provide both a facility for examining temporal changes and integrating the effects of prolonged exposure to intermittent discharges or variable concentrations of pollutants [1]. Therefore, it is promising to characterize the changes occurring in communities to assess target ecosystems exposed to environmental disturbances [1,2].

It is obvious that biological communities are affected by man-made alterations of nature [3,4]. To evaluate changes of communities in space and/or time, diversity indices are commonly used [1,5]. Species richness (SR) is an integrative descriptor of the community, as it is influenced by a large number of natural environmental factors as well as anthropogenic disturbances [2]. The disturbances of environmental factors lead to spatial discontinuities of predictable gradients and losses of taxa [6]. Therefore, SR is used as a biological indicator of disturbance. As with SR, diversity indices decrease under increasing disturbance and stress on the ecosystem. The Shannon diversity index (SH) is commonly

---

*Corresponding author. Tel.: +33-5-61-55-86-87; fax: +33-5-61-55-60-96.

E-mail addresses: park@cict.fr (Y.-S. Park), p.f.m.verdonschot@alterra.wag-ur.nl (P.F.M. Verdonschot), tschon@pusan.ac.kr (T.-S. Chon), lek@cict.fr (S. Lek).

used to describe the diversity of a particular community and as an ecological indicator for the assessments of ecosystems [7].

Development of methods for patterning spatial and/ or temporal changes in communities has currently become an important issue in ecosystem management. Traditionally, conventional multivariate analyses have been applied to solve these problems [5]. This task, however, is not easy to achieve as nonlinear, complex interactions occur in the dataset consisting of many species and sampling areas. To respect the natural nonlinearity of ecological data, artificial intelligence methods could be preferred [8]. An artificial neural network is a versatile tool for dealing with problems to extract information out of complex, nonlinear data, and it is more and more used in modelling aquatic ecosystems [8–10]. Most of these models have used two popular artificial neural networks: a multilayer perceptron with a backpropagation algorithm (BP) [11] and a Kohonen's self-organizing map (SOM) [12,13]. The networks are mainly used to predict target values or to classify input vectors in a model. It is not easy to conduct both classification and prediction in such networks at the same time.

However, patterning and predicting could effectively be carried out in a network. One example is a counter-propagation network (CPN) [14], which consists of unsupervised and supervised learning algorithms. It classifies input vectors and predicts output values. This study aims to apply a CPN for patterning and for predicting the ecological data consisting of benthic macroinvertebrate communities and environmental variables.

2. Materials and methods

2.1. Modelling procedure

The CPN [14] is a hybrid neural network combining the SOM [12] and the Grossberg outstar [15]. The network is designed to approximate continuous functional associations between variables, and serves as a statistically optimal self-programming look-up table [14]. In this study, we used a forward-only CPN which is a specific type of CPN without counterflow (Fig. 1).

In the modelling process, initially the data vectors **x** (explanatory variables) and **y** (dependent variables) are given to the SOM and the Grossberg layers, respectively. Then, the weights are updated for a given set of data vectors **x** and **y**. For the CPN this occurs in two phases. First, the SOM layer is trained. When the input vector x is sent through the network, each neuron (computation unit) k of the network computes the distance between the weight vector v and the input vector **x**. Among all N output neurons in two dimensions, the best matching
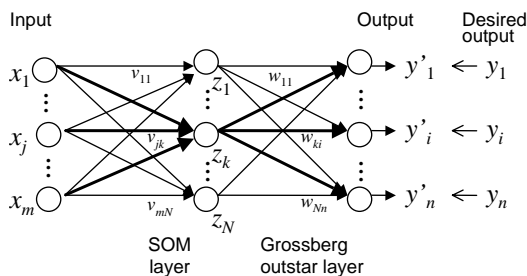

Fig. 1. Schematic diagram of a forward-only CPN.

neuron (BMN) which has minimum distance becomes the winner. The BMN and its neighboring neurons are allowed to learn by changing their weights so as to further reduce the distance between the weight and the input vectors as follows [13]:

$$v_{jk} = v_{jk} + h_{ck}(x_j - v_{jk}), \tag{1}$$

where v_{jk} is the weight between neuron j of the input layer and neuron k of the SOM layer, and h_{ck} is a neighborhood function and a smoothing kernel for location vectors of BMN c and k defined over the lattice of the output layer. This can be written in terms of the Gaussian function:

$$h_{ck}(t) = \alpha \exp\left(-\frac{\|r_c - r_k\|^2}{2\sigma^2}\right), \tag{2}$$

where r_c and r_k are location vectors of neuron c and k, respectively, in the output layer, α and σ are, respectively, a learning rate factor and the width of the kernel, and monotonically decreasing functions of time. This results in training the layer to classify the input vectors by the weight vector **v** they are closest to.

Once the SOM layer is trained, the Grossberg layer is trained. This is done in a supervised mode according to the following procedure. An input vector **x** is applied to the CPN, the output of the SOM layer is established, and the Grossberg layer outputs are calculated. In this process, the Grossberg layer receives z vector signals from the SOM layer. If the difference between the desired and the estimated output values is greater than an acceptable error, the weights are updated as follows:

$$w_{ki} = w_{ki} + \beta(y_i - w_{ki})z_k, \tag{3}$$

where w_{ki} is the weight between neuron k of the SOM layer and neuron i of the Grossberg layer, β is the learning rate, and z_k is assigned to 1 for the BMN while set to 0 for all other neurons of the SOM layer. The weights correspond to the averages of the desired outputs **y** associated to the inputs **x** according to the equiprobability of the winning neurons of the SOM layer. The trained CPN actually functions as a statistically self-programming look-up table.

After training the CPN in this study, a unified-matrix algorithm (U-matrix) [16] was applied to detect the

cluster boundaries on the map of the SOM layer. The algorithm is commonly used to find clusters in the SOM units.

2.2. Relationships between biological and environmental variables

The values calculated for each input variable during the learning process were visualized on the trained SOM map with a gray scale to represent the relationships between the input variables and the clusters of the input vectors. Furthermore, to understand relationships between input (environmental) variables and output (biological) variables, mean values of output variables were calculated in corresponding units of the trained SOM, visualized with a gray scale [17], and compared with maps of environmental variables. The environmental variables were classified into several groups based on their distribution patterns on the trained SOM map with weight vectors of the trained SOM to estimate relationships between variables.

2.3. Ecological data

To implement the capability of the CPN, benthic macroinvertebrate communities and the corresponding environmental variables were used. The datasets were extracted from the database EKOO in The Netherlands [18]. The data were collected at 664 sites (Fig. 2) of 23 different water types (Table 1) in the province Overijssel, The Netherlands. A total of 854 species were recorded

and Chironomidae, Coleoptera, and Oligochaeta were the most abundant taxa in the dataset. From the community matrix, two community indices; SR (number

Table 1
Water types of sampling sites and number of samples collected in each water type

| Acronym	Water type	No. of samples
BB	Lower watercourses	24
BK	Springs sources	21
BO	Upper watercourses	63
BP	Remaining stream pools	17
BR	Springs	22
BV	Spring ponds	1
DW	Temporary water	25
KA	Canals	35
KB	Regulated small rivers	34
KO	Deep ponds	27
LS	Peat ditches	29
ML	Middle watercourses	29
MM	Small lakes	24
PE	Peat pits	26
PO	Shallow pools	24
RM	Large lakes	10
RR	Rivers	33
SB	Regulated streams	24
SG	Spring gutter	1
SL	Ditches	97
VA	Peat canals	42
VE	Moorland pools	32
ZW	Sand and clay pits	24
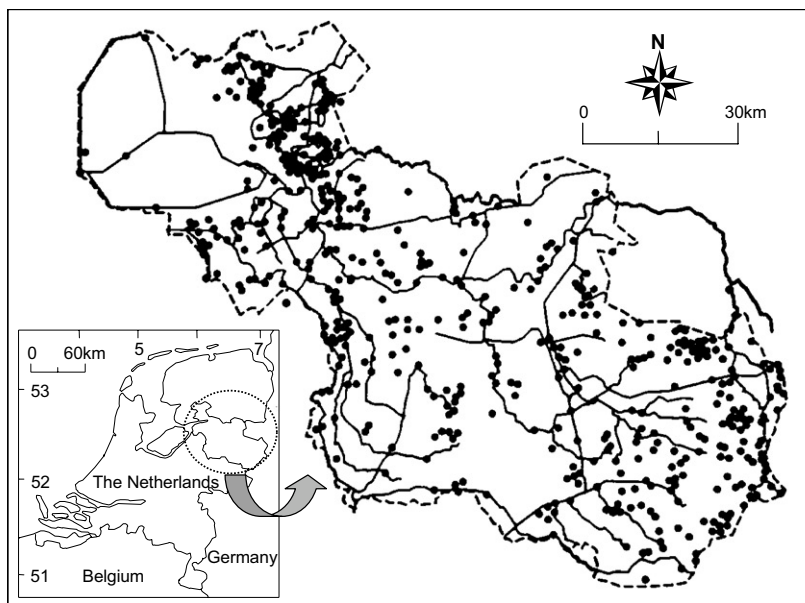
Fig. 2. Sampling sites in the province of Overijssel, The Netherlands. Each sampling site is marked with a spot.

Table 2
Thirty-four quantitative environmental variables used in the model

Variables	Acronym	Unit	Mean (SE)
Percentage cover emergent vegetation	BOVE%	%	6.77 (0.54)
Percentage cover floating vegetation	DRIJ%	%	11.67 (0.89)
Percentage cover floating algae	FLAL%	%	3.82 (0.56)
Percentage sampled habitat: emergent vegetation	MMBO%	%	16.16 (0.86)
Percentage sampled habitat: detritus	MMDE%	%	9.01 (0.69)
Percentage sampled habitat: floating vegetation	MMDR%	%	12.96 (0.79)
Percentage sampled habitat: gravel	MMGR%	%	1.36 (0.20)
Percentage sampled habitat: clay	MMKL%	%	0.51 (0.14)
Percentage sampled habitat: bank	MMOE%	%	18.24 (0.91)
Percentage sampled habitat: submerged vegetation	MMON%	%	12.05 (0.76)
Percentage sampled habitat: silt	MMSL%	%	15.67 (0.73)
Percentage sampled habitat: stones	MMST%	%	0.72 (0.13)
Percentage sampled habitat: peat	MMVE%	%	2.20 (0.26)
Percentage sampled habitat: sand	MMZA%	%	10.51 (0.65)
Dissolved oxygen percent saturation	O_2%	%	90.70 (1.66)
Percentage cover bank vegetation	OEVE%	%	6.10 (0.57)
Percentage cover submerged vegetation	ONDE%	%	11.23 (0.92)
Percentage cover all vegetation	TOTB%	%	33.14 (1.38)
Width of stream	WIDTH	m	64.24 (18.16)
Ratio width/depth	WD/DP		28.51 (4.54)
Calcium	Ca^{++}	mg/l	51.21 (1.01)
Chloride	Cl^-	mg/l	52.79 (1.98)
Depth	DEPTH	m	1.13 (0.06)
Silt thickness	DSAPR	m	0.11 (0.01)
Electric conductivity	ECOND	μs/cm	427.95 (9.18)
Ammonium	NH_4^+	mg N/l	1.46 (0.14)
Nitrate	NO_3^-	mg N/l	3.87 (0.32)
Oxygen concentration	O_2	mg/l	9.71 (0.16)
Ortho-phosphate	O–P	mg P/l	0.29 (0.03)
Acidity	pH		7.13 (0.04)
Flow velocity	VELOC	m/s	0.07 (0.01)
Water temperature	TEMP	°C	13.26 (0.24)
Total-phosphate	T–P	mg P/l	0.51 (0.05)
Slope	VERVA	m/km	5.91 (0.81)

of species collected at each sample) and SH were extracted to evaluate the benthic macroinvertebrate community structure at each sampling site. The mean SR was 54.46 ($\pm$0.94 SE) ranging from 2 to 132, and mean diversity index was 5.29 ($\pm$0.03 SE) ranging from 0.49 to 6.77.

Thirty-four environmental variables (Table 2) were also measured at each sampling site, and showed a wide range in environmental conditions. Verdonschot and Nijboer [18] have reported the general ecological characteristics in the EKOO database. The environmental variables were used to predict SR and SH of benthic macroinvertebrate communities using the CPN. Out of 664 sites 500 were used to train the network; while the remaining 164 were applied to test the performance of the trained network. The input data; both environmental variables and biological attributes; were proportionally scaled between 0 and 1 in the range of the minimum and maximum values. Before scaling

data; the environmental variables were transformed by natural logarithm to reduce skewed distributions.

3. Results

3.1. Patterning input variables

The CPN patterned the dataset in the SOM layer, and a U-matrix method clustered the units of the trained SOM map. The results showed five clusters (I–V) of sampling sites according to environmental gradients, and two subclusters Va and Vb were observed in cluster V (Fig. 3). The acronyms of the water types are given in Table 1. Each cluster was mainly associated with the characteristics of the water types. For instance, cluster I mainly consisted of sites of moorland pools (VE), cluster II of ditches (SL), cluster III of stagnant water bodies (VA, PE, PO, and KA), cluster IV of large rivers and
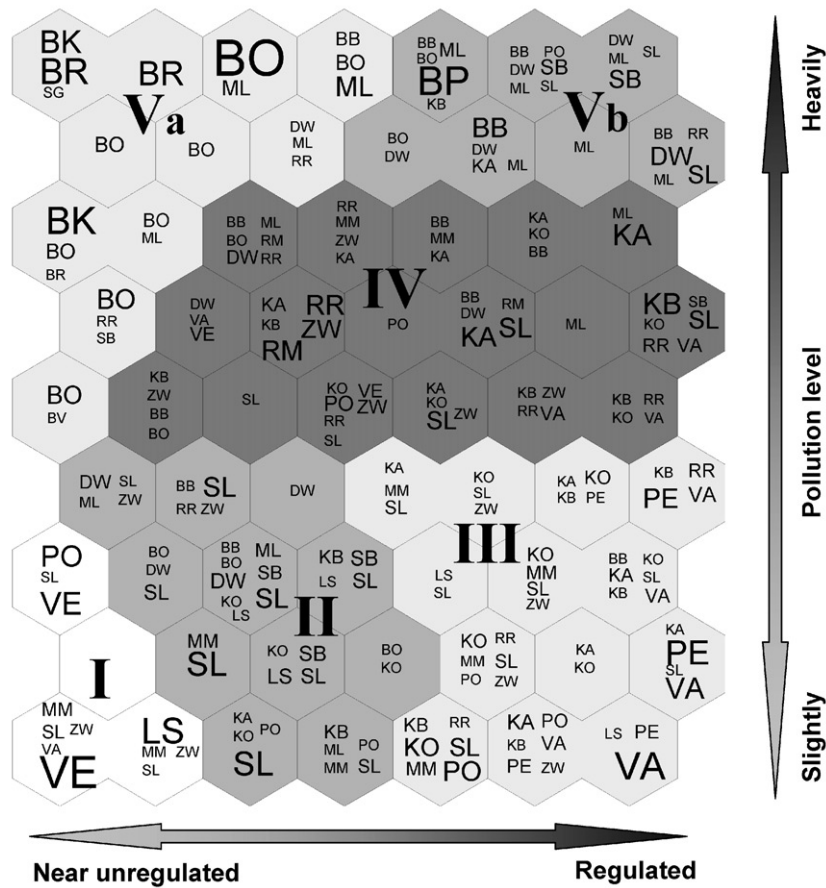
Fig. 3. Classification of sampling sites with environmental variables in the SOM layer of the CPN. The U-matrix algorithm was applied to cluster the SOM units. The Latin numbers (I–V) represent different clusters. The acronyms in the hexagonal units represent different water types, and are shown in Table 1. The font size of the acronym is proportional to the number of sampling sites in the water types in the range of 1–18 samples.

lakes (RR, RM, KA, and ZW) and ditches (SL). Finally, clusters Va and Vb were characterized, respectively, by springs and upper watercourses (BK, BO and BR) and intermittent or regulated streams (BP, DW and SB). These distribution patterns show the characteristics of natural key conditions of water systems. The sampling sites located on the left areas of the SOM map were mainly from unregulated water systems, whereas sites on the right were from regulated areas (Fig. 3).

Fig. 4 displays the contribution of each input variable for the classification of sampling sites on the SOM map. Dark areas represent high contribution of each input variable, while light ones display low values. The values were calculated during the learning process of the network. Acronyms of environmental variables are shown in Table 2. Each variable displays a high-gradient distribution on the SOM map. In the environmental variables, nine groups were observed according to their distribution similarities (A-I). The groups of variables

show different aspects of environment. For example, group B is related to electric conductivity and group F is characterized by inorganic nutrients (NH_4^+, T–P, and O–P). The groups also show different local habitat characteristics. Groups A and D are concerned with percentages of vegetation cover, whereas group H typically represents the characteristics of upper water course habitats showing high percentages of detritus, stones, sands, and gravels with high current velocities and strong slopes. The morphological characters of streams (width and depth) were grouped together in group E.

The next step is to compare the relationship between clusters of sampling sites and groups of environmental variables. Clusters I and II are related to low values of group B and high values of group D, and cluster III is represented by high values of groups D and G, and low values of group H (Fig. 3). Similarly, cluster IV is displayed by high values of groups B and E and
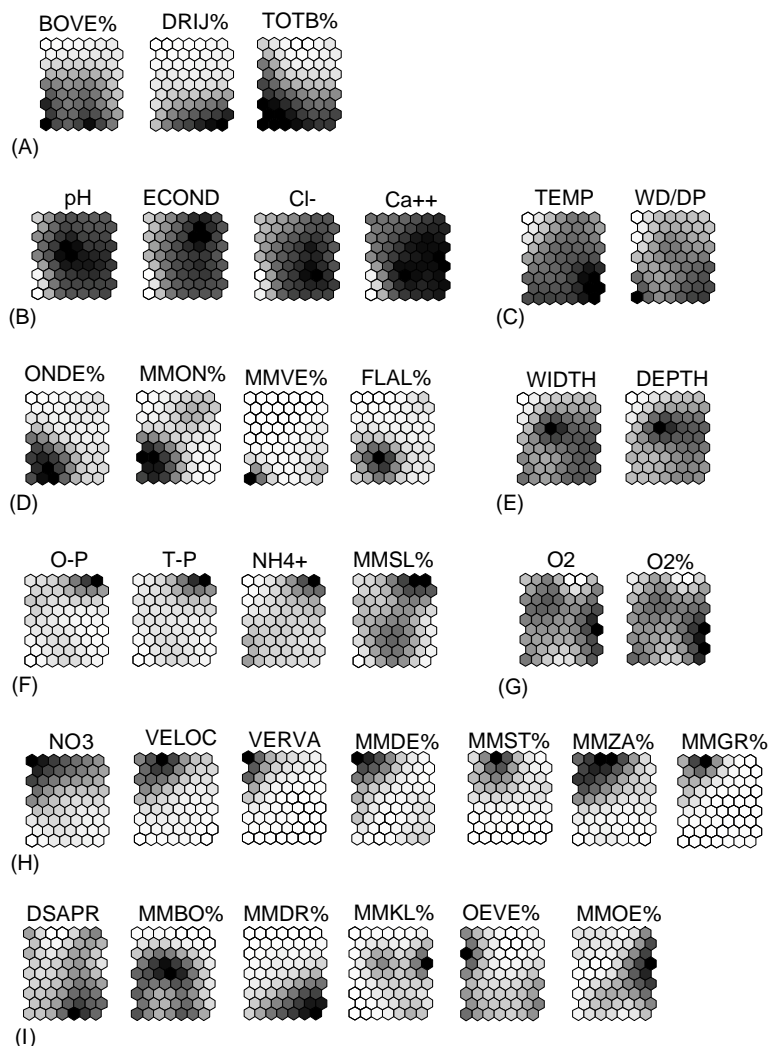
Fig. 4. Component planes displaying the contribution of each environmental variable to classification of sampling sites. Based on the similarity of the distribution pattern, nine groups (A–I) were identified. The names of the environmental variables are given in Table 2. Dark represents high values of each variable, whereas light is for low values. The values were calculated during the learning process of the network.

variables MMBO%, MMKL%, and MMOE% of group I, and subclusters Va and Vb are strongly related to high values of groups H and F, respectively. Nitrogen and phosphorus compounds, which were mainly considered as pollutants at high concentrations, represent the groups F and H. Furthermore, the sampling sites in the left areas of the SOM map (clusters I, II, Va) display mainly unregulated water systems, while the sites in the right areas (clusters III, IV, Vb) reveal regulated aquatic systems like canals. Overall, Fig. 3 shows that sites of clusters I and II in the lower areas of the SOM map are not disturbed and contain well-developed vegetation, whereas the sites of cluster Vb in the upper area are disturbed by regulation and nutrients (e.g. nitrate,

ammonium, ortho-phosphate, and total phosphate) which are presumably due to increased amounts of dissolved ions entering the water through agricultural activities.

3.2. Relationship between environmental variables and community indices

To evaluate the relationships between environmental variables and diversity indices (SR and SH), the mean values of the SR and SH were visualized on the trained SOM map in gray scale (Fig. 5). The results show that SH and SR are higher in the lower areas of the SOM map than in the upper areas, and higher in the right
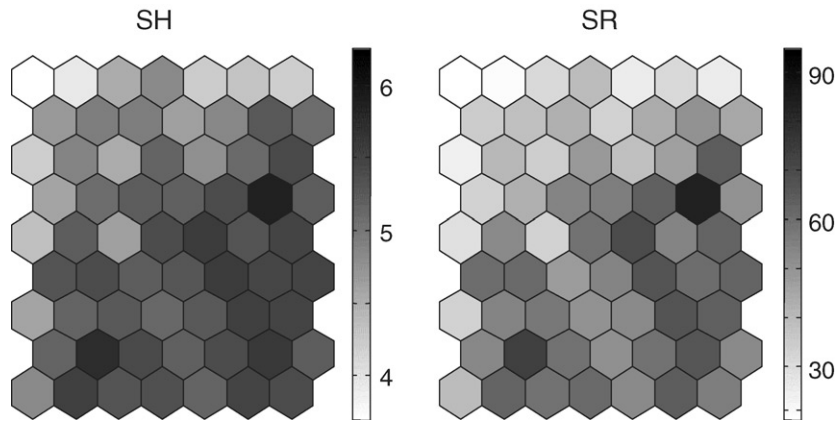
Fig. 5. Distribution of SH and SR on the SOM map trained with environmental variables. Dark represents high values of each variable, whereas light displays low values. The mean values of each variable were calculated in each unit of the SOM map.

areas than in the left areas. The low values in the upper areas (cluster V) are mainly influenced by high concentration of nitrogen and phosphorus compounds in groups F and H (Figs. 3–5). They are also affected by substrate conditions of their habitats with high percentages of stone, gravel, sand, and detritus in substrates. Sampling sites in these areas are characterized by water types of springs and upper courses (cluster Va) and intermittent and regulated water systems (cluster Vb). SR and SH were also related to dissolved oxygen (group G). Thus, both community indices are higher at the samples assigned in the lower right areas, which were slightly polluted by nutrients and morphologically, physically regulated by water managers, while they are lower at samples in the upper areas, which represent upper watercourses and highly influenced by nutrients.

3.3. Prediction of community indices

The trained CPN serves as a 'look-up table' for finding the corresponding values between the input and output variables. The Grossberg layer of the trained network showed a high predictability in the learning process (Figs. 6a and b). Correlation coefficients between observed and estimated values were 0.90 ($P < 0.01$) for both SH and SR. In both cases, overestimations were observed at low values, while underestimations were observed at high values. This is caused by the structural characteristics of data. There are few cases with low values in both SH and SR. The frequency histogram of error values showed that most error values lie around zero (Figs. 6c and d). The residuals between observed and estimated values averaged 0.03 (± 0.02 SE) and 2.51 (± 0.52 SE) for SH and SR, respectively.

The data not used in the learning process were applied to test the feasibility of the trained network. The results showed a high predictability of the network. The correlation coefficients between observed and predicted values were 0.70 and 0.67 for SH and SR, respectively ($P < 0.001$) (Figs. 7a and b). A majority of frequencies of the error terms also appeared around zero (Figs. 7c and d). The residuals between observed and predicted values were located around zero showing averages of 0.11 (± 0.05 SE) and 4.71 (± 1.62 SE) for SH and SR, respectively. Thus, the results show that the trained CPN corresponded well to the reality of SH and SR.

4. Discussion

The CPN was implemented to pattern sampling sites and to predict SR and SH with the environmental variables available in this study. In the first step, the network classified sampling sites into five clusters based on environmental variables in the SOM layer, and afterwards the diversity indices (SR and SH) were predicted in the output layer of the network. Thus, the CPN shows to be a general approach to explain the variation of ecological data in two steps: ordination methods to summarize the variability of the data as a first step, and exploration for possible relationships between biological and environmental variables as a second step [19].

The SOM layer showed the ability to produce a classification of input vectors as well as visualization of relationships among input variables in their contribution to the classification. The analysis using visualization of component planes is comparable to principal component analysis, but more directly describes the discriminatory power of the input variables in the mapping procedure [13]. A clear distribution gradient of a variable represents a high contribution to the classification of input vectors. In this study, the sampling sites were classified into five clusters and input variables were divided into nine groups. Each cluster was explained very well by environmental groups (Figs. 3 and 4).
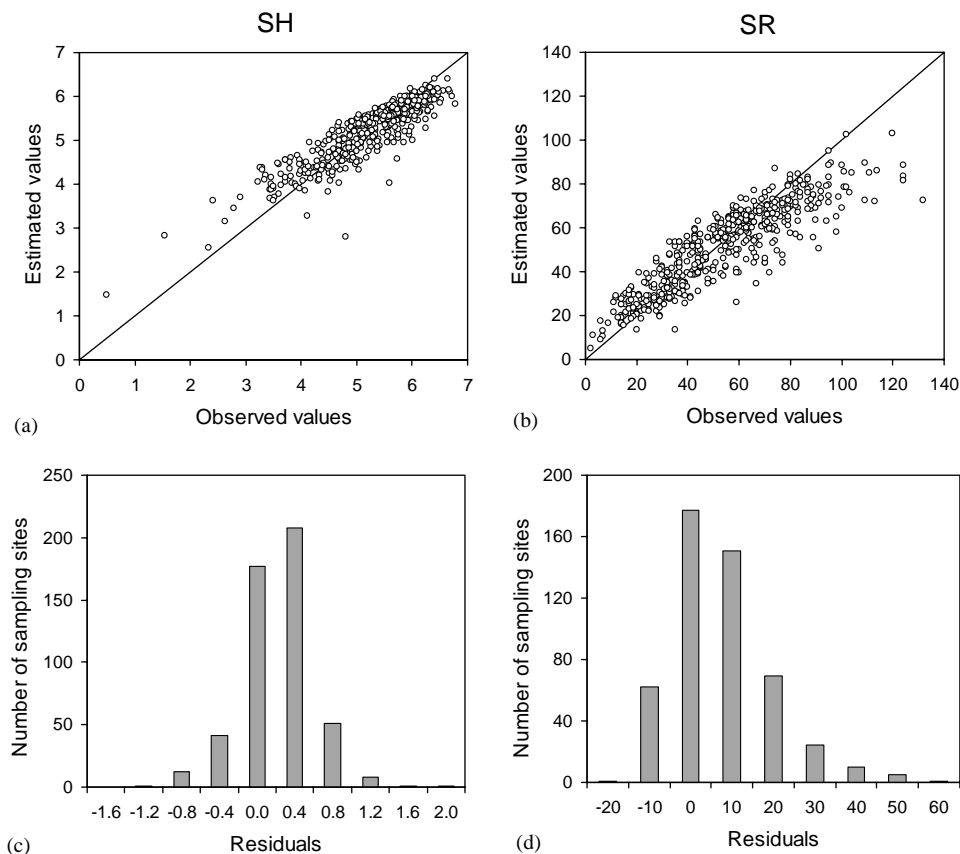
Fig. 6. Training results of the model to predict diversity index (SH) and SR with environmental variables. Scatter plots represent correlations between observed values and estimated values of the model trained with 34 environmental variables (a), (b); and distribution of residuals in the learning phase (c), (d).

Furthermore, by overlapping the distribution of both input variables and mean values of diversity indices on the SOM map, the relationships between explanatory (input) variables and dependent (output) variables could be analyzed. When there are strong relationships between input and output variables, the component planes show clear gradients and similar patterns of their distribution on the trained SOM map. However, it is necessary to quantify the distribution gradient of each variable as well as the relationships between biological and environmental variables.

The structure of the CPN is similar to a combination of two networks; SOM and multilayer perceptron with BP. Especially when prediction output values are considered, the CPN is related to the BP. It is considered that the BP is relatively better than the CPN, although there is still debate on this point [20]. In contrast, the CPN is more effective in noise sensitivity, and perform well without being influenced by the increase in data size. Recently, these characteristics were successfully applied for patterning hierarchical relationships among taxonomic groups of benthic macroinvertebrates [21].

Since information extraction and noise sensitivity are equally important in adaptive learning processes with ecological data, it is difficult to decide which algorithm should be better suited for patterning communities at the present time. Further, comparative research may be required with various ecosystem data in the future.

According to the distribution gradients of the environmental variables on the SOM map, influence of environmental variables on the classification of the sampling sites as well as on diversity indices could be assessed effectively. The low values of SR and diversity index were mainly affected by high values of nutrients concentration such as nitrogen and phosphorus compounds, and substrate conditions of their habitats. Thus, both diversity indices are higher at the slightly polluted and regulated samples, while lower at samples highly influenced by nutrients. Both nitrogen and phosphorus compounds are essential for living organisms and the limiting nutrients for algal growth and, therefore, control the primary productivity of a water body [22]. The eutrophication due to the artificial increase in concentration of these nutrients affects on energy flow of
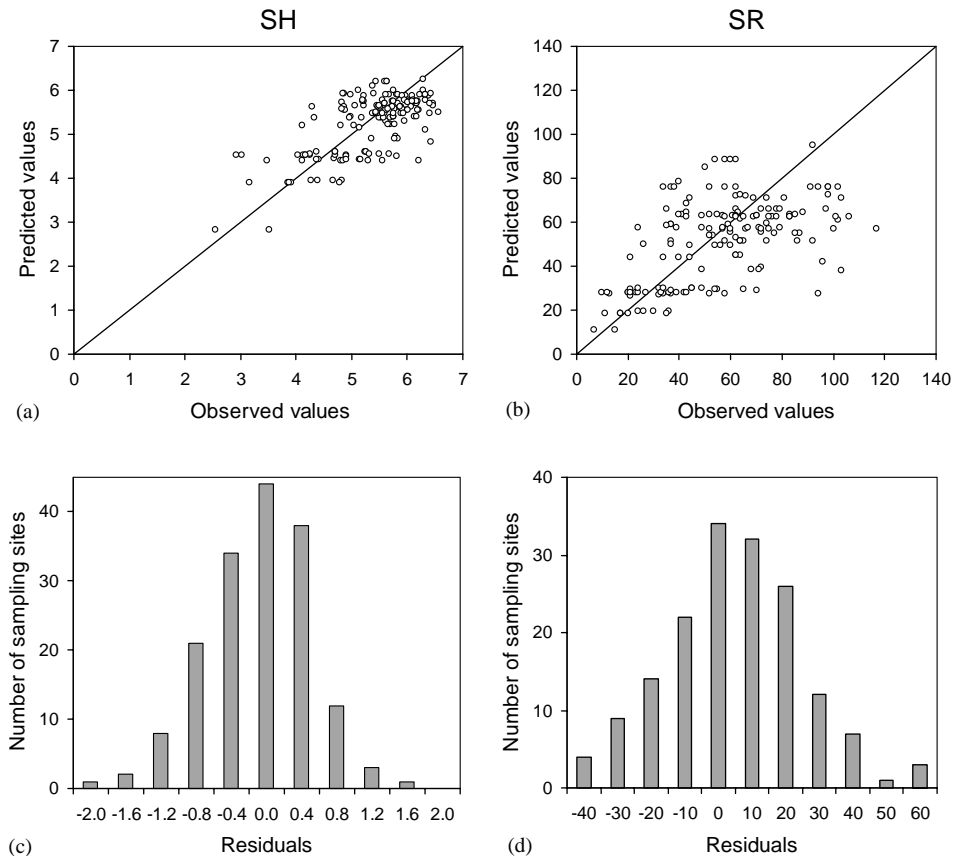
Fig. 7. Results of the model tested with the datasets not used in the learning process. Scatter plots represent correlations between observed and predicted values for both diversity index (SH) and SR (a), (b); and distribution of residuals (c), (d).

aquatic ecosystems and cause decline of biodiversity [23,24]. Furthermore, sampling sites in the low-diversity areas are also characterized by water types of springs and upper courses. This is supported by the intermediate disturbance hypothesis [25] assuming that high species diversity is a result of intermediate frequency of disturbance, while either too low or too high frequency of disturbance will result in a low biodiversity [26].

The community structure is changed by perturbations in the environment and the degree of the structure change is used to assess the intensity of the environmental stress [1]. The SR is a function of the stability of the environment [5]. A stable environment contains more species and more niches, because a more stable environment involves a higher degree of organization and complexity of the food web [27]. The niche of a species is the set of environmental conditions that the species does not share with any other sympatric species, so SR is concerned with the number of niches [28]. Diversity index further accommodates the evenness concepts in addition to the taxon richness, and represents heterogeneity of species composition, char-

acterizing the ecological status of communities at a given site and a given time [1]. Based on these facts, SR and diversity indices are frequently used as biological indicators of target ecosystems in combination. It is worth predicting these indices with their explanatory variables, and they can be used as a tool for the assessment of disturbance in a given ecosystem.

5. Conclusion

By combining two different neural network models, aquatic ecological data were patterned and predicted with concerning descriptor variables. At first, the sampling sites were classified into several clusters in the SOM layer, and the classification was mainly related with pollution status and habitat types of sampling sites. According to the distribution gradients of the environmental variables on the SOM map, their influence on the classification of the sampling sites could be assessed effectively. Furthermore, by visualizing variables on the trained SOM map, we could evaluate the relationships

between environmental variables and community indices showing that SR and diversity indices were strongly influenced by concentration of nutrients, dissolved oxygen, and percentages of vegetation cover as well as by different water types. This method, classifying sampling sites and visualizing environmental and biological variables on the trained same SOM map, is useful to understand complex ecological data. Furthermore, the trained CPN serves as a 'look-up table' for finding the corresponding values between the explanatory and dependant variables displaying a high accuracy of the prediction. Finally, these results suggest that the capability of the CPN for patterning and predicting ecological data can be effectively used as a tool for assessing ecological status and for predicting water quality of target ecosystems in managing aquatic ecosystems according to the EU Water Framework Directive.

Acknowledgements

References

[1] Hellawell JM. Biological indicators of freshwater pollution and environmental management. London: Elevier, 1986.

[2] Rosenberg DM, Resh VH. (Eds.). Freshwater biomonitoring and benthic macroinvertebrates. London: Chapman & Hall, 1993.

[3] Rosenzweig ML. Species diversity in space and time. Cambridge: Cambridge University Press, 1995.

[4] Wilson EO. The diversity of life. New York: Norton, 1999.

[5] Legendre P, Legendre L. Numerical ecology. Amsterdam: Elsevier, 1998.

[6] Ward JV, Stanford JA. Ecological factors controlling stream zoobenthos with emphasis on thermal modification of regulated streams. In: Ward JV, Stanford JA, editors. The ecology of regulated streams. New York: Plenum Press, 1979. p. 35–55.

[7] Bahls LR, Burkantis R, Tralles S. Benchmark biology of Montana reference streams. Department of Health and Environmental Science, Water Quality Bureau, Helena, Montana, 1992.

[8] Lek S, Guégan JF. (Eds.). Artificial neuronal networks: Application to ecology and evolution. Berlin: Springer, 2000.

[9] Huang W, Foo S. Neural network modeling of salinity variation in Apalachicola River. Water Res 2002;36: 356–62.

[10] Recknagel F. (Ed.). Ecological informatics: understanding ecology by biologically-inspired computation. Berlin: Springer, 2002.

[11] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McCelland JL, editors. Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1 Foundations. Cambridge: MIT Press, 1986. p. 318–62.

[12] Kohonen T. Self-organized formation of topologically correct feature maps. Biol Cybernet 1982;43:59–69.

[13] Kohonen T. Self-organizing maps, 3rd ed.. Berlin: Springer, 2001.

[14] Hecht-Nielsen R. Neurocomputing. Reading, MA: Addison-Wesley, 1990.

[15] Grossberg S. On the production and release of chemical transmitters and related topics in the cellular control. J Theoret Biol 1969;22:325–64.

[16] Ultsch A. Self-organizing neural networks for visualization and classification. In: Opitz O, Lausen B, Klar R, editors. Information and classification. Berlin: Springer, 1993. p. 307–13.

[17] Park Y, Céréghino R, Compin A, Lek S. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. Ecol Modell, 2003;160(3):265–80.

[18] Verdonschot PFM, Nijboer RC. Typology of macrofaunal assemblages applied to water and nature management: a dutch approach. In: Wright JF, Sutcliffe DW, Furse MT, editors. Assessing the biological quality of fresh waters: RIVPACS and other techniques. Ambleside Cumbria: The Freshwater Biological Association, 2000. p. 241–62.

[19] Jongman RHG, ter Braak CJF, van Tongerenm OFR. (Eds.). Data analysis in community and landscape ecology. Cambridge: Cambridge University Press, 1995.

[20] Ruiz ME, Srinivasan P. Automatic text categorization using neural networks. In: Efthimiadis E, editor. Proceedings of the Eighth ASIS/SIGCR Workshop on Classification Research. Washington: American Society for Information Science, 1997. p. 59–72.

[21] Park Y, Kwak I, Cha E, Lek S, Chon T. Relational patterning on different hierarchical levels in communities of benthic macroinvertebrates in an urbanized steam using an artificial neural network. J Asia-Pacific Entomol 2001;4:131–41.

[22] Chapman D. (Ed.). Water quality assessments. London: Chapman & Hall, 1992.

[23] Lods-Crozet B, Lachavanne J. Changes in the chironomid communities in Lake Geneva in relation with eutrophication over a period of 60 years. Arch Hydrobiol 1994;130(4):453–71.

[24] Schindler DW. Experimental perturbations of whole lakes as tests of hypotheses concerning ecosystem structure and function. Oikos 1990;57:25–41.

[25] Connell J. Diversity in tropical rain forests and coral reefs. Science 1978;199:1304–10.

[26] Jørgensen SE, Padisak J. Does the intermediate disturbance hypothesis comply with thermodynamics? Hydrobiologia 1996;323:9–21.

[27] Margalef R. Information theory in ecology. Gen Syst 1958;3:36–71.

[28] Hutchinson GE. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 1957;22:415–27.