CrossMark

Modeling the species richness and abundance of lotic macroalgae based on habitat characteristics by artificial neural networks: a potentially useful tool for stream biomonitoring programs

José Celso Rocha[1] · Cleto K. Peres[2] · José Leonel L. Buzzo[1] · Vinícius de Souza[1] ·
Eric A. Krause[1] · Pitágoras C. Bispo[1] · Fernando Frei[1] · Lucas S. M. Costa[1] ·
Ciro C. Z. Branco[1]

Abstract One of the major challenges in stream ecology is
the development of computational models that can predict
aspects of the community structure of organisms from these
ecosystems when they are subject to natural or artificial envi-
ronmental fluctuations. To contribute towards this aim, we
conducted a study whose main goal was to evaluate the effi-
ciency and accuracy of different architectures of multilayer
artificial neural networks (ANNs) in predicting the species
richness and abundance of macroalgae based on environmen-
tal variables of tropical streams. We used data from 82 streams
located in southern Brazil, where species richness, macroalgal
abundance, and environmental parameters were measured. A
set of 20 environmental parameters measured directly in the
stream was used as explanatory variables. The performance of
the ANN architectures was assessed using two different pieces
of software (random combinatorial and exhaustive) and the
coefficient of determination (R^2) and mean-squared error
(MSE). For both species richness and macroalgal abundance,
the best ANN architectures were obtained using random com-
bination software and the performance parameters showed a
combination of high R^2 and very low MSE. Our results sug-
gest that computational models that are constructed based on
ANN frameworks can be efficient and accurate in predicting
the species richness and abundance of stream macroalgae
from environmental data. Therefore, considering that models
based on linear relationships have often failed, we recommend
the application of ANNs as a tool to estimate species richness
and abundance of lotic macroalgae from environmental data,
in the management, conservation, and biomonitoring pro-
grams of tropical stream ecosystems.

Keywords Stream macroalgae · Species richness and
abundance · Environmental distribution · Artificial neural
networks · Predictive models · Stream biomonitoring
programs

Introduction

Currently, there is a growing concern about the biomonitoring
and conservation of biodiversity, especially aquatic ecosys-
tems (Ormerod 2014), including lotic ones. Predictive models
can be useful tools to deal with these issues; thus, a huge
challenge for stream ecologists is to construct models that
can predict the richness, abundance, and distribution of taxa
based on environmental and spatial factors. When working
properly, these models have been very useful and can be used
to (1) predict the characteristics of biota only based on envi-
ronmental and spatial factors, which simplifies the effort of
future sampling in ecological investigations (e.g., Oppel et al.
2012; Bucklin et al. 2015; Rovzar et al. 2016), (2) predict
changes in biota due to corresponding changes in environ-
mental and spatial factors (e.g., Pottier et al. 2013; Wisz
et al. 2013), and mainly, (3) predict the features of the biota
before the occurrence of anthropogenic interference (e.g.,
Blois et al. 2013; Catford et al. 2013; Graham et al. 2015).

✉ Ciro C. Z. Branco
czbranco@assis.unesp.br

[1] Department of Biological Sciences, São Paulo State
University—UNESP, Assis, SP 19806-900, Brazil

[2] Federal University for Latin American Integration—UNILA, Foz do
Iguaçu, PR 85867-970, Brazil

The predictive model approach has been used in large biomonitoring programs. Many of these models are built based on Reference Conditions (nonimpacted sites or those with minimal environmental impacts), according to the Reference Conditions Approach (RCA) (Bailey et al. 2004). RCA models aim to build tools capable of estimating community attributes (e.g., taxa richness) under natural conditions. In this context, the impact grade of a given location can be measured as the difference between the observed values and those estimated by the model based on Reference Conditions. RCA supports the theoretical bases of large biomonitoring programs such as River Invertebrate Prediction and Classification (RIVPACS) (Wright 1995), Australian River Assessment System (AUSRIVAS) (Simpson and Norris 2000), and Benthic Assessment of Sediment (BEAST) (Reynoldson et al. 1995).

Although the principles of predictive model construction of lotic ecosystems have the potential to be applied to any aquatic organism, most biomonitoring programs have focused on macroinvertebrates, and models that use primary producers remain scarce (Almeida and Feio 2012; Feio and Dolédec 2012). Among the organisms in lotic environments, benthic macroscopic algae (macroalgae "sensu" Sheath and Cole 1992), offer high potential in biomonitoring programs, mainly when the aim is the trophic state evaluation (Branco and Pereira 2002; Peres et al. 2009; Cantonati et al. 2012; Stancheva et al. 2012). The use of macroalgae in biomonitoring relies on the fact that they are one of the most important primary producers in rivers and streams (Dell'Uomo 1991; Branco et al. 2010), and they are easily recognizable in the field (at least at the genus level), which makes their use quite convenient (Dell'Uomo 1991; Branco et al. 2010). The understanding that composition, richness, and abundance of macroalgae are strongly influenced by both biotic and abiotic processes (Dudley 1992; Branco and Necchi Júnior 1996; Verb and Vis 2001; Necchi Júnior et al. 2003) reinforces the idea that these organisms represent a good model for environmental biomonitoring programs.

In this context, predictive model construction able to efficiently and accurately estimate variations in richness and abundance of stream macroalgae, based on easily measurable environmental variables, could facilitate the progress of stream biomonitoring studies that use these organisms. Richness and abundance are important properties of the community structure, and they are also metrics affected by both natural and anthropogenic factors (Ferreira et al. 2014; Hrivnák et al. 2014; Klose et al. 2015; Teittinen et al. 2016). Taxa richness is one of the most intuitive measures of community diversity level, and it has been universally used as an important descriptor of the community structure. Specifically in stream algae case, studies have shown that taxa richness of these communities is affected by environmental factors (Teittinen et al. 2016), including those of anthropogenic

origins (Verb and Vis 2001, 2005; Schneider et al. 2013; Barinova et al. 2016; Oberholster et al. 2016). Thus, despite the need to use this metric carefully (Stevenson 2014), algae richness is one good attribute for identifying spatial and temporal changes and for biomonitoring at community level.

Algae abundance has also been treated as a very important metric in the identification of anthropogenic perturbations in aquatic ecosystems, since the excessive increase in algal biomass is one of the most usual effects of eutrophication (Yang et al. 2012). Based on this assumption, this parameter of community structure and functioning, used alone or in association with other measures, has been largely applied in biomonitoring studies (e.g., Rakocevic-Nedovic and Hollert 2005; Rier and Stevenson 2006; Stevenson et al. 2006; Peres et al. 2010). Birk et al. (2012) showed, for instance, based on an overview of almost 300 assessment methods, that algal abundance is one of the most used biological metric to evaluate water quality in European lakes. In addition, a great sign of the relevance of this biological metric for water quality program in rivers is the use of algal abundance data, despite the differences in methodological protocols, by several programs of environmental agencies. The European Project for the Standardization of River Classification protocol (STAR, http://www.eu-star.at/frameset.htm), for instance, uses macroscopic benthic algae abundance (estimated according to the percentage of the bottom covered area) as one of the metrics to evaluate changes in benthic algae communities.

Classical ecological studies on lotic macroalgae have shown that it is generally difficult to reliably predict the richness and abundance of species from environmental data using classical linear relationships (Hu and Xie 2006; Branco et al. 2009). Therefore, modern computing techniques, including artificial neural networks (ANNs), have been successfully applied to solve similar complex ecological problems. Examples of the application of ANNs in predictive models for ecological issues include the estimation of tree biomass and organic carbon in forests (Schoeninger et al. 2008), the prediction of the richness of aquatic insects in running water (Park et al. 2003; Goethals et al. 2007), the abundance and species succession of cyanobacteria (Recknagel 1997), and the modeling of macrophyte indices (Gebler et al. 2014), among several others (see examples in Lek and Guégan 1999).

Artificial neural networks are computational techniques (learning algorithms) that are inspired by the operation of typical biological neuronal networks, in which neurons receive stimuli for their dendrites, process them in their cell bodies, and then propagate the signal to other neurons through axons and synaptic endings (Kovács 2002). When performing artificial neural processing, the ANNs use data pairs from a dataset (each input data or independent variable is paired to an output data or dependent/response variable) to perform a large number of calculations, enabling them to learn the patterns of relationship between input and output data (Lippmann 1987).

This ability to learn is defined as the capacity for self-organization or plasticity (Rumelhart et al. 1986). During the learning process, the ANNs progressively modify the interconnection patterns among its elements (neurons), so that the computational properties of the network are also modified, facilitating and improving its learning performance (Rumelhart et al. 1986). After training, the ANNs can be used to estimate the response variable based on independent variables.

An ANN can be composed from between one to hundreds of neurons, usually divided into layers, the so-called multilayer neural networks (Haykin 2008). The arrangement of the layers in an ANN is commonly designated by the term architecture. A specific architecture is defined by the number of layers present in the ANN, the number of neurons in each layer, and the mathematical function in which they propagate the stimuli to the subsequent layers (Haykin 2008). The architecture of an ANN is extremely relevant to their performance, since a network with few neurons might not be able to retain the learning pattern and, conversely, a network with many neurons can enter a state of "overfitting" (Haykin 2008). Considering that multilayer ANNs can work with nonlinear data (Haykin 2008), the application of this type of ANN to ecological data, such as the determination of the richness and abundance of species of stream macroalgae based on environmental variables, is quite adequate.

Thus, considering that ANNs represent an effective predictive modeling technique, especially when applied to nonlinear data (Lek et al. 1996), such as ecological data (Lek and Guégan 1999), this study was conducted to assess the potential application of different architectures of multilayer ANNs in predicting the richness and abundance of macroalgal communities based on the environmental variables of tropical stream ecosystems under reference conditions.

Materials and methods

Study area, sampling, and ecological data

The stream macroalgae and environmental data were collected in eighty-nine 10-m long stream segments (first- to third-orders), in reference conditions (nonimpacted sites or those with minimal environmental impacts) located in ten protected areas for conservation (National or State Parks) of the four major biomes of southern Brazil, between latitudes 22°31′10″ S–33°44′59″ S and longitudes 48°01′08″ W–57°36′05″ W. Stream macroalgal samples were purposely collected in protected areas because we based our sampling program on the Reference Condition Approach (RCA) (Bailey et al. 2004). The central idea of the RCA is to build models based on reference conditions that allow to predict the attributes of the biota under natural or minimally impacted condition

(Bailey et al. 2004; Stancheva and Sheath 2016). In addition, the largest possible data set, including lotic environments with different characteristics and from distinct biomes (always considering the Reference Condition, however), was used in order to allow the maximum possible generalization of the model.

In this study, we used as response variables species richness and abundance of the macroalgal community. The richness and abundance of macroalgal species were investigated by the transect technique, which is a widely used technique in ecological studies involving this group of organisms (Sheath and Cole 1992; Branco et al. 2014). In each stream segment, the presence of each species of macroalgae was recorded and their respective abundance was estimated (in terms of percentage cover) by visual analysis (Sheath and Cole 1992; Branco et al. 2014). From these data, we calculated the total species richness (as the sum of all recorded species) and the total abundance of species (as the sum of the abundances of all recorded species) for each stream segment.

Samples of each macroalga found in the field samples were collected, coded, and preserved in 4% formaldehyde (Branco and Necchi Júnior 1996; Branco et al. 2014). In the laboratory, the preserved samples were observed using a Leica trinocular microscope, model DM1000, and identified at species level whenever feasible. For the identification procedures, at least, ten measurements were randomly taken of diagnostic structures in each sample (Branco and Necchi Júnior 1996). These morphometric analyses were made using an image capture system composed of a Leica video camera model DFC280 coupled to a microcomputer with the software Leica IM-50. The complete list of species identified in this study is given in Online Resource 1.

In addition to the richness and abundance of species, a set of 20 environmental variables was measured directly in the streams. These variables were selected based on their widespread use in ecological studies involving lotic organisms (Allan and Castillo 2007) and because they are recurrently described as being closely related to the spatial and temporal distribution of stream macroalgae (Sheath and Cole 1992; Branco et al. 2014).

In each stream segment, we measured the following water variables, using a Horiba U-10 water analyzer equipped with a multiparameter probe: temperature, turbidity, specific conductance, pH, and dissolved oxygen. We also measured the water content of orthophosphate and total nitrogen (using a Merck, Spectroquant Nov. 60 spectrophotometer). In addition, we recorded the percentage of each substratum type (by visual analysis) that was available for macroalgal attachment (using the particle size classes modified from Gordon et al. 1992), the degree of shading (based on DeNicola et al. 1992), mean current velocity (measured with a mechanical flowmeter, General Oceanics 2030R), and mean depth (measured with a centimeter ruler).

Artificial neural networks

All environmental parameters measured in the field were used to construct a supervised training ANN, totaling 20 input variables, namely water temperature, specific conductance, pH, dissolved oxygen, turbidity, current velocity, depth, shading, orthophosphate, total nitrogen, rock substrate, boulder substrate, pebble substrate, gravel substrate, sand substrate, clay substrate, sand/clay substrate, macrophyte substrate, trunks substrate, and grit substrate. All data were normalized (varying from −1 to 1), to achieve a higher efficiency and speed during the ANN training processes (Zanchettin and Ludermir 2005).

In the network configuration, the environmental variables represent the input data (independent variables), whereas the species richness and abundance of species of macroalgae represent the output data (dependent or response variables), so that together, environmental data and species richness and abundance represent the input and output pairs of the network (Fig. 1). Considering the architecture of the network, the input layer provides environmental data to train the ANN, whereas the intermediate layers provide the understanding of the learning process and the output layer provides the responses of the ANN (Fig. 1). Therefore, the output layers are those that are linked to the results of the desired prediction, which in this study, are species richness and abundance of stream macroalgae.

The performance of an ANN architecture can be evaluated based on various performance measures, including the percentage of correctly classified instances (CCI) (Fielding and Bell 1997; Manel et al. 1999) or Cohen's kappa values (Cohen 1960; Randin et al. 2006). However, when the output of an ANN are data such as the richness and abundance of species, Goethals et al. (2007) reported that the performance measures used are mainly the correlation (r) or determination (R^2) coefficient and the root-mean-squared error (RMSE). In this study, we developed two softwares, which were used to evaluate the best ANN architecture, considering the coefficient of determination (R^2) and mean-squared error (MSE) as the performance measures. The R^2 was calculated from real species richness and abundance data (real data) and the results provided by the ANNs (predicted data), with the ANNs that had the highest R^2 and the lowest MSE (Olaya-Marín et al. 2013) being considered as the best.

The first software, which was the random combinatorial software, created, trained, validated, and tested several networks by random variation in the ANN architecture throughout successive iterations and, at the end of each iteration, retained the architecture that showed the best performance. In contrast to the first software, the exhaustive software, in addition to creating, training, validating, and testing different ANN architectures, also promoted a nonrandom but exhaustive variation of all architectural parameters, to explore all possible combinations of ANN architectures. For training the ANNs, the data from 89 stream segments were divided into a training set (containing data from 63 stream segments), a validation set (13 stream segments), and a test set (13 stream segments).

The learning algorithm used in the ANNs was "backpropagation" (Lek and Guégan 2000; Park et al. 2003). This algorithm is characterized by two phases. In the first stage (feed-forward phase), the input variables are presented to the ANN and their effect propagates to the output

Fig. 1 Schematic illustration of the three-layered neural networks used to predict the species richness and abundance of macroalgae in streams of southern Brazil
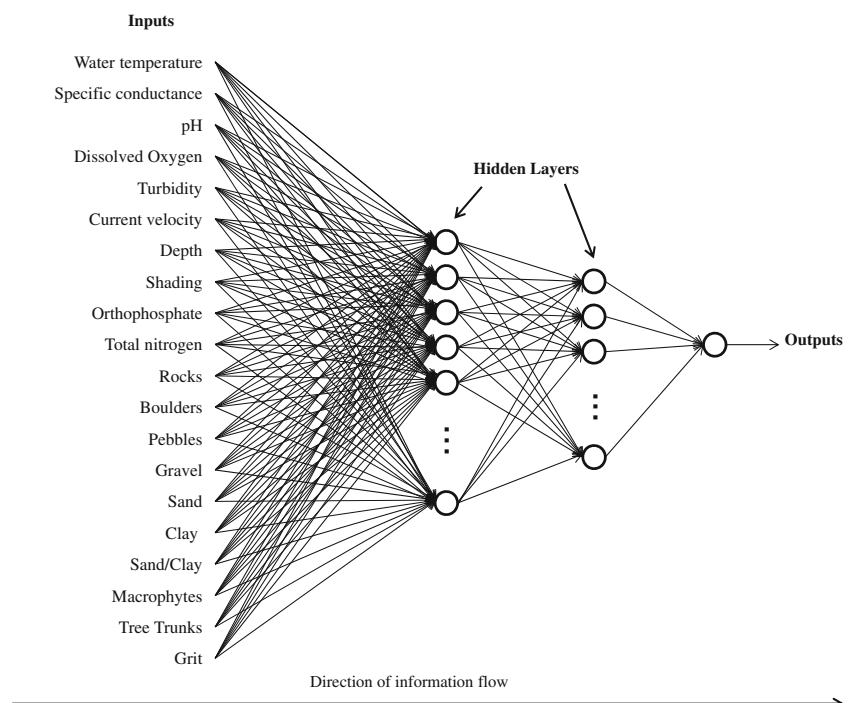
Table 1 Comparison of architectural parameters, correlation values (R^2), and mean square error (MSE) between the best artificial neural networks obtained by Random Combinatorial and Exhaustive software built to predict the species richness of macroalgae in lotic environments of southern Brazil

Architectural parameters	Random Combinatorial Software			Exhaustive Software		
	RCS1	RCS2	RCS3	ES1	ES2	ES3
NN first layer	28	12	7	26	28	21
NN second layer	13	7	4	10	11	10
TFF first and second	tansig	tansig	logsig	tansig	logsig	tansig
TFF second and output	logsig	tansig	logsig	logsig	tansig	logsig
TRF	trainrp	trainrp	trainrp	trainrp	trainlm	trainrp
R^2 - Trainning set	0.85	0.79	0.71	0.98	0.92	0.79
R^2 - Validation set	0.74	0.72	0.64	0.87	0.74	0.72
R^2 - Testing set	0.83	0.74	0.69	0.29	0.59	0.69
MSE	0.018	0.026	0.044	0.040	0.060	0.050

NN numbers of neurons, *TFF* transfer function, *TRF* training function

layer (response layer), and then, the variation responses are compared with the real values previously offered to the model. If the difference between the predicted responses and the real values previously provided to the model is greater than a preestablished error, the ANN initiates the second stage of the learning algorithm (feed-backward phase), where the weights are adjusted according to a rule of error correction (in this study, the "delta rule"). This error signal is then propagated back into the ANN and their weights are recalculated. This process is performed repeatedly until the ANN responses are within the predetermined error (Rumelhart et al. 1986; Kung 1993; Lek and Guégan 1999, 2000).

All ANNs were constructed using the MATLAB modeling environment (Beale et al. 2015).

Results

The predictive models of macroalgal species richness revealed three possible architectures of ANNs for each software used to process the data (Table 1). Among these six possible networks, RCS1, obtained using random combinatorial software,

showed the best results, since this network produced a high determination coefficient ($R^2 = 0.83$) associated with a very low value of the mean squared error (MSE = 0.018) (Table 1). Additionally, a comparison of the values predicted by RCS1 with the real values from the field showed that this particular network was extremely accurate in predicting the richness of macroalgal species (Figs. 2 and 3). Considering the sampling error of RCS1, measured by the comparison of the species richness values predicted by the model and from the values observed in each sampling point, the macroalgal species richness was precisely estimated for 39 sampling sites (representing 44% of all streams), whereas for 82% sampling sites (73 of the 89 streams), the maximum error was only one species and for 96.6% of sites (86 of the 89 streams), the maximum error was two species.

Considering the prediction of macroalgal abundance values (in terms of percentage coverage), we also identified three possible architectures of ANNs for each software used (Table 2). Among the six possible network architectures, RCS2, which was also obtained using random combinatorial software, showed the best results, since it recorded the highest coefficient of determination ($R^2 = 0.76$), associated with an

Fig. 2 A comparison of modeled (by RCS1) and observed values of number of species of stream macroalgae in lotic environments in southern Brazil
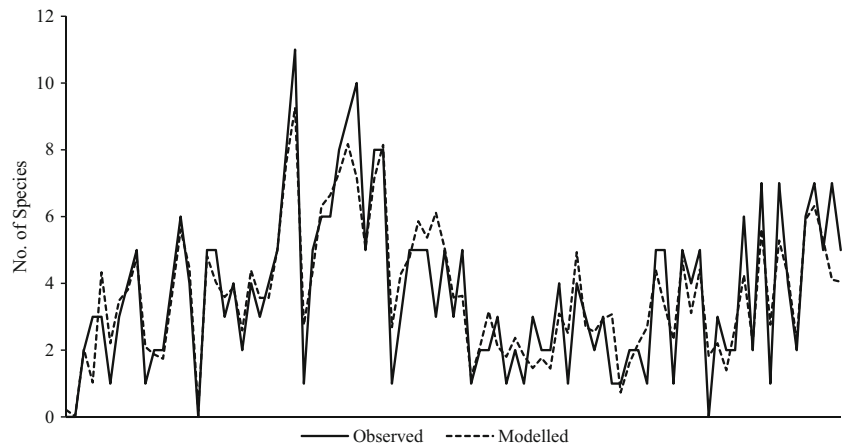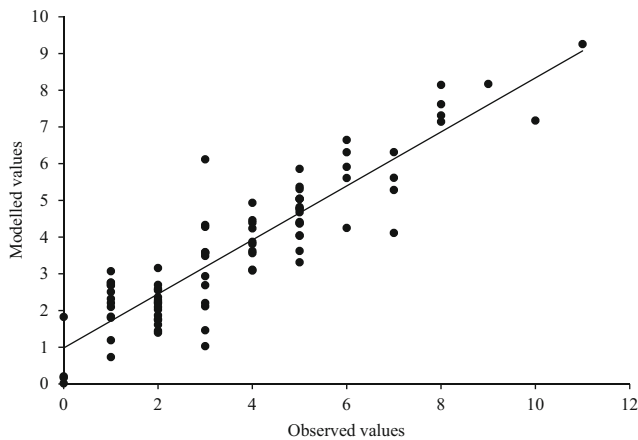
Fig. 3 Correlation plot of modeled (by RCS1) and observed values of number of species of stream macroalgae in lotic environments in southern Brazil

extremely low value of the mean squared error (MSE = 1.2×10^{-11}) (Table 2). The RCS2 recorded a strong correlation among the predicted and the observed values, showing a huge and consistent similarity between the curves of abundance produced by the model and the real data (Figs. 4 and 5). In this context, for 52.8% of the sampling sites (47 of the 89 streams), the error in predicting abundance was less than 3% of the percentage cover. If we consider the maximum acceptable error as being 5%, RCS2 showed an even higher performance, with 71.9% of the sampling sites (64 of the 89 streams) showing predicted values below this limit.

Discussion

Streams are highly dynamic ecosystems where biotic and environmental interactions are complex and natural disturbances are relatively common (Hart 1992; Lake 2000). In these environments, macroalgae often have an erratic distribution and it is very difficult to identify reliable relationships among the

richness and abundance of species and environmental factors by statistical or mathematical methods based on linear models (Necchi Júnior et al. 2000; Hu and Xie 2006; Branco et al. 2009). Considering this limitation, the construction of predictive models based on abiotic features and that provides good prediction of structure of lotic macroalgal communities has become a huge challenge and linear approaches can be inefficient.

In this context, ANNs have been treated as a powerful computational tool, since they can deal with nonlinear relationships, which are typical for most ecological relationships (Gevrey et al. 2003; Gebler et al. 2014). Artificial neural networks have been used in studies on organisms from freshwater ecosystems (e.g., Park et al. 2003; Penczak et al. 2012; Gebler et al. 2014; Lopez-Exposito et al. 2016); however, no further study has yet been conducted to confirm the performance of this approach in predicting the structure (richness and abundance) of lotic macroalgae communities, which are among the most important primary producers and are promising bioindicators of the trophic state in continental running water (Branco and Pereira 2002; Peres et al. 2010; Cantonati et al. 2012; Stancheva et al. 2012). In this study, we showed that some architectures of multilayer ANNs were relatively efficient in predicting the species richness and abundance of macroalgae, even in high dynamic systems such as streams.

The most effective ANNs in this study reported relatively high coefficients of determination (R^2) between the predicted and observed values for both species richness (RCS1) and abundance (RCS2). In general, the R^2 values recorded for stream macroalgae were similar to or higher than those reported in other studies with other organisms (e.g., Park et al. 2003; Olaya-Marín et al. 2012; Gebler et al. 2014). Additionally, the MSE values observed for both species richness and abundance were very low (see Tables 1 and 2), suggesting that the errors of the models were significantly reduced (Brosse et al. 1999). Brey (2012), for instance, accepted as reasonable the accuracy and precision of a neural network to estimate the

Table 2 Comparison of architectural parameters, correlation values (R^2) and mean square error (MSE) between the best artificial neural networks obtained by Random Combinatorial and Exhaustive software built to predict the abundance of macroalgae in lotic environments of southern Brazil

Architectural parameters	Random Combinatorial Software randômico			Exhaustive Software		
	RCS1	RCS2	RCS3	ES1	ES2	ES3
NN first layer	20	19	8	24	23	20
NN second layer	7	9	3	10	9	11
TFF first and second	tansig	tansig	logsig	logsig	tansig	tansig
TFF second and output	tansig	tansig	purelin	logsig	logsig	tansig
TRF	trainlm	trainlm	trainrp	trainlm	trainrp	trainlm
R^2 - Training set	0.87	0.92	0.72	0.94	0.74	0.98
R^2 - Validation set	0.79	0.74	0.86	0.46	0.67	0.62
R^2 - Testing set	0.69	0.76	0.69	0.58	0.56	0.13
MSE	2.2×10^{-4}	1.2×10^{-11}	0.046	0.042	0.062	0.043

NN numbers of neurons, *TFF* transfer function, *TRF* training function

Fig. 4 A comparison of modeled (by RCS2) and observed values of abundance of stream macroalgae in lotic environments in southern Brazil
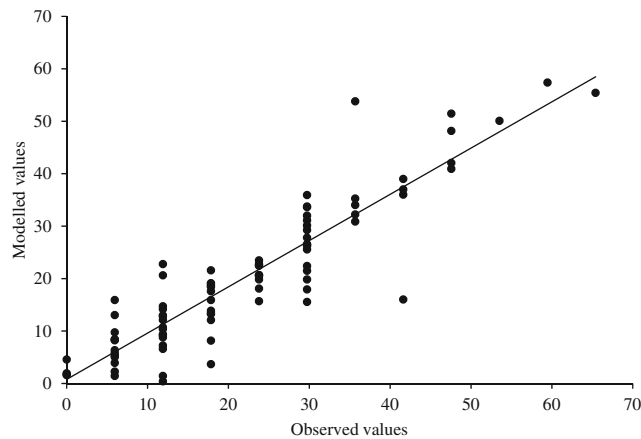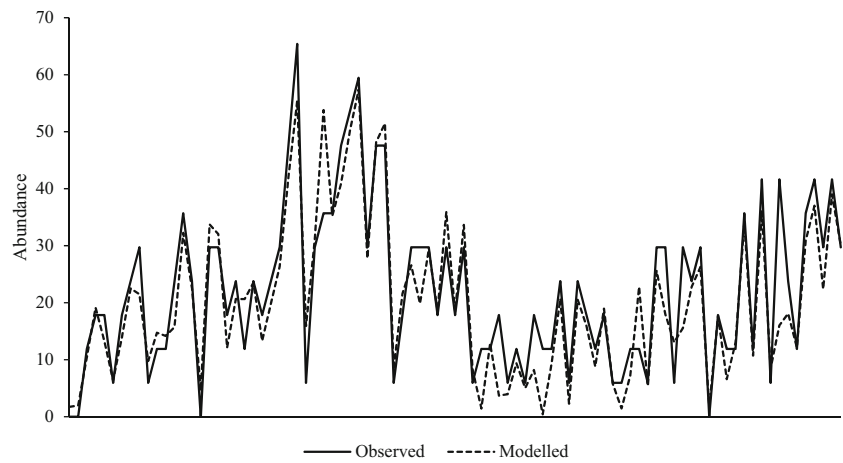

Fig. 5 Correlation plot of modeled (by RCS2) and observed values of abundance of stream macroalgae in lotic environments in southern Brazil

productivity and production of benthic macroinvertebrates with $R^2 = 0.801$ and RMS = 0.083. Similarly, Yoo et al. (2013) recognized a high predictive ability of a neural network constructed to estimate the biodiversity of tidal flat habitats with $R = 0.84$ (i.e., $R^2 = 0.71$). Olaya-Marín et al. (2013) reported variation in R^2 between 0.52 and 0.77 for ANN models constructed to predict the richness of fish species in Mediterranean rivers. Therefore, our results showed that the ANN technique was able to produce models to predict richness and abundance of stream macroalgae with uncertainties within absolutely acceptable levels.

In summary, our results reveal that the use of ANNs offers great potential in predicting the richness and abundance of stream macroalgae based on environmental data, since the models constructed using this technique demonstrated a very high accuracy. Therefore, we strongly recommend the application of ANNs as a tool to estimate the species richness and abundance of macroalgae in lotic environments in the context of stream biomonitoring programs from environmental data. In addition, we also suggest the use of ANNs in investigations that require the evaluation of the effects of environmental

scenarios on the community structure of these organisms, whether in purely ecological studies or as part of the management and conservation of tropical stream ecosystems.

References

Allan JD, Castillo MM (2007) Stream ecology: structure and function of running water. Springer, Dordrecht

Almeida SFP, Feio MJ (2012) DIATMOD: diatom predictive model for quality assessment of Portuguese running waters. Hydrobiologia 695:185–197

Bailey RC, Norris RH, Reynolds TB (2004) Bioassessment of freshwater ecosystems - using the reference condition approach. Springer, New York

Barinova S, Liu N, Ding J, An Y, Qin X, Wu C (2016) Ecological assessment of water quality of the Songhua River upper reaches by algal communities. Acta Ecol Sinica 36:126–132

Beale MH, Hagan MT, Demuth HB (2015) Neural network toolbox: MATLAB user's guide. The MathWorks Inc, Natick

Birk S, Bonne W, Borja A, Brucet S, Courrat A, Poikane S, Solimini A, van de Bund W, Zampoukas N, Hering D (2012) Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the water framework directive. Ecol Indic 18:31–41

Blois JL, Willians JW, Fitzpatrick MC, Jackson ST, Ferrier S (2013) Space can substitute for time in predicting climate-change effects on biodiversity. Proc Natl Acad Sci U S A 110:9374–9379

Branco CCZ, Necchi Júnior O (1996) Distribution of stream macroalgae in the eastern Atlantic rainforest of São Paulo state, southeastern Brazil. Algol Stud 333:139–150

Branco LHZ, Pereira JL (2002) Evaluation of seasonal dynamics and bioindication potential of macroalgal communities in a polluted tropical stream. Arch Hydrobiol 155:147–161

Branco CCZ, Krupek RA, Peres CK (2009) Distribution of stream macroalgal communities from the mid-western region of Paraná

state, southern Brazil: importance of local-scale variation. Braz Arch Biol Techn 52:379–386

Branco CCZ, Necchi Júnior O, Peres CK (2010) Effects of artificial substratum types and exposure time on macroalgal colonization in a tropical stream. Fund Appl Limnol 178:17–27

Branco CCZ, Bispo PC, Peres CK, Tonetto AF, Branco LHZ (2014) The roles of environmental conditions and spatial factors in controlling stream macroalgal communities. Hydrobiologia 732:123–132

Brey T (2012) A multi-parameter artificial neural network model to estimate macrobenthic invertebrate productivity and production. Limnol Oceanogr-Meth 10:581–589

Brosse S, Guegan JF, Tourenq JN, Lek S (1999) The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. Ecol Model 120:299–311

Bucklin DN, Basille M, Benscoter AM, Brandt LA, Mazzotti FJ, Romañach SS, Speroterra C, Watling JI (2015) Comparing species distribution models constructed with different subset of environmental predictors. Divers Distrib 21:23–35

Cantonati M, Rott E, Spitale D, Angeli N, Komárek J (2012) Are benthic algae related to spring types? Freshw Sci 31:481–498

Catford JA, Naiman RJ, Chambers LE, Roberts J, Douglas M, Davies P (2013) Predicting novel riparian ecosystems in a changing climate. Ecosystems 16:382–400

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psycol Meas 20:37–46

Dell'Uomo A (1991) Use of benthic macroalgae for monitoring rivers in Italy. In: Whitton BA, Rott E, Friedrich G (eds) Use of algae for monitoring rivers. Universität Innsbruck, Innsbruck, pp 129–137

DeNicola DM, Hogland KD, Roemer SC (1992) Influence of canopy cover on spectral irradiance and periphyton assemblages in a prairie stream. J N Am Benthol Soc 11:391–404

Dudley TL (1992) Beneficial effects of herbivores on stream macroalgae via epiphyte removal. Oikos 65:121–127

Feio MJ, Dolédec S (2012) Integration of invertebrate traits into predictive models for indirect assessment of stream functional integrity: a case study in Portugal. Ecol Indic 15:236–247

Ferreira WR, Ligeiro R, Macedo DR, Hughes RM, Kaufmann PR, Oliveira LG, Callisto M (2014) Importance of environmental factors for the richness and distribution of benthic macroinvertebrates in tropical headwater streams. Freshw Sci 33:860–871

Fielding HG, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24:38–49

Gebler D, Kayzer D, Szoszkiewicz K, Budka A (2014) Artificial neural network modeling indices based on physical-chemical characteristics of water. Hydrobiologia 737:215–224

Gevrey M, Dimopolous I, Lek S (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol Model 160:249–264

Goethals PLM, Dedecker AP, Gabriels W, Lek S, Pauw N (2007) Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquat Ecol 41:491–508

Gordon ND, McMahon TA, Finlayson BL (1992) Stream hydrology: an introduction for ecologists. Wiley, Chichester

Graham NAJ, Jennings S, MacNeil MA, Mouillot D, Wilsno SK (2015) Predicting climate-driven regime shifts versus rebound potential in coral reefs. Nature 518:94–97

Hart DD (1992) Community organization in streams: the importance of species interactions, physical factors, and chance. Oecologia 91:220–228

Haykin S (2008) Neural networks and learning machines. Prentice Hall, New York

Hrivnák R, Kochjarová J, Ot'ahel'ová H, Pal'ove-Balang P, Slezák M, Slezák P (2014) Environmental drivers of macrophyte species richness in artificial and natural aquatic water bodies – comparative

approach from two central European regions. Ann Limnol 50:269–278

Hu BF, Xie SL (2006) Effect of seasonality on distribution of macroalgae in a stream system (Xin'an spring) in Shanxi Province, North China. J Integr Plant Biol 48:889–896

Klose K, Cooper SD, Bennett DM (2015) Effects of wildfire on stream algal abundance, community structure, and nutrient limitation. Freshw Sci 34:1494–1509

Kovács ZL (2002) Redes neurais artificiais: Fundamentos e Aplicações. Livraria da Física, São Paulo

Kung SY (1993) Digital neural networks. Prentice Hall, Englewood Cliffs

Lake PS (2000) Distubance, patchiness, and diversity in streams. J N Am Benthol Soc 19:573–592

Lek S, Guégan JF (1999) Artificial neural networks as a tool in ecological modeling, an introduction. Ecol Model 120:65–73

Lek S, Guégan JF (2000) Artificial neuronal networks: application to ecology and evolution. Springer, Berlin

Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S (1996) Application of neural networks to modeling nonlinear relationships in ecology. Ecol Model 90:39–52

Lippmann RP (1987) An introduction to computing with neural nets. IEEE a.S.S.P. Magazine 2:4–22

Lopez-Exposito P, Suarez AB, Negro C (2016) Estimation of *Chlamydomonas reinhardtii* biomass concentration from chord length distribution data. J Appl Phycol 28:2315–2322

Manel S, Dias JM, Buckton ST, Ormerod SJ (1999) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. J Appl Ecol 36:734–747

Necchi Júnior O, Branco CCZ, Branco LHZ (2000) Distribution of stream macroalgae in São Paulo state, southeastern Brazil. Algol Stud 97:43–57

Necchi Júnior O, Branco LHZ, Branco CCZ (2003) Ecological distribution of stream macroalgal communities from a drainage basin in the Serra da Canastra National Park, Minas Gerais, southeastern Brazil. Braz J Biol 63:1–12

Oberholster PJ, De Klek AR, De Klek L, Chamier J, Botha A-M (2016) Algal assemblage responses to acid mine drainage and steel plant wastewater effluent up and downstream of pre and post wetland rehabilitation. Ecol Indic 62:106–116

Olaya-Marín EJ, Martínez-Capel F, Soares Costa RM, Alcaraz-Hernandez JD (2012) Modeling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River basin, Spain). Sci Total Environ 440:95–105

Olaya-Marín EJ, Martínez-Capel F, Vezza P (2013) A comparison of artificial neural networks and random forests to predict native fish species richness in Mediterranean rivers. Knowl Manag Aquat Ecosys 409:1–19

Oppel S, Meirinhyo A, Ramírez I, Gardner B, O'Connell AF, Miller PI, Louzao M (2012) Comparison of five modeling techniques to predict the spatial distribution and abundance of seabirds. Biol Conserv 156:94–104

Ormerod SJ (2014) Rebalancing the philosophy of river conservation. Aquat Conserv Mar Freshwat Ecosys 24:147–152

Park YS, Céréghino R, Compin A, Lek S (2003) Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. Ecol Model 160:265–280

Penczak T, Glowacki L, Kruk A, Galicka W (2012) Implementation of a self-organizing map for investigation of impoundment impact on fish assemblages in a large, lowland river: long-term study. Ecol Model 227:64–71

Peres CK, Branco CCZ, Krupek RA (2009) Distribuição ambiental e temporal das comunidades de macroalgas de riachos da Serra da Prata, Estado do Paraná, sul do Brasil. Braz J Bot 32:625–633

Peres CK, Branco CCZ, Krupek RA, Rocha JC (2010) Longitudinal distribution and seasonality of macroalgae in a subtropical stream impacted by organic pollution. Acta Limnol Bras 22:199–207

Pottier J, Dubuis A, Pellisier L, Maiorano L, Rossier L, Randin CF, Vittoz P, Guisan A (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. Glob Ecol Biogeogr 22:52–63

Rakocevic-Nedovic J, Hollert H (2005) Phytoplankton community and chlorophyll *a* as trophic state indices of Lake Skadar (Montenegro, Balkan). Environ Sci Pollut Res 12:146–152

Randin CF, Dirnböck T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in space? J Biogeogr 33:1689–1703

Recknagel F (1997) ANNA – artificial neural network model for predicting species abundance and succession of blue-green algae. Hydrobiologia 349:47–57

Reynoldson TB, Bailey RC, Day KE, Norris RH (1995) Biological guidelines for freshwater sediment based on benthic assessment of sediment (the BEAST) using a multivariate approach for predicting biological state. Aust J Ecol 20:198–219

Rier ST, Stevenson RJ (2006) Response of periphytic algae to gradients in nitrogen and phosphorus in streamside mesocosms. Hydrobiologia 561:131–147

Rovzar C, Gillespie TW, Kawelo K (2016) Landscape to site variations in species distribution models for endangered plants. Forest Ecol Manag 369:20–28

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323:533–536

Schneider SC, Kahlert MK, Kelly MG (2013) Interactions between pH and nutrients on benthic algae in streams and consequences for ecological status assessment and species richness patterns. Sci Total Environ 444:73–84

Schoeninger ER, Koehler HS, Botelho MF, Watzlawick LF, Oliveira PC (2008) Uso de redes neurais artificiais para mapeamento de biomassa e carbono orgânico no componente arbóreo de uma floresta ombrófila densa. Ambiência 4:179–195

Sheath RG, Cole KM (1992) Biogeography of stream macroalgae in North America. J Phycol 28:448–460

Simpson J, Norris RH (2000) Biological assessment of water quality: development of AUSRIVAS models and outputs. In: Wright JF, Sutcliffe DM, Furse MT (eds) RIVPACS and similar techniques for assessing the biological quality of freshwaters. Freshwater Biological Association and Environment Agency, Ambleside, pp 125–142

Stancheva R, Sheath RG (2016) Benthic spft-body algae as bioindicators of stream water quality. Knowl Manag Aquat Ecosyst 417:15

Stancheva R, Fetscher AE, Sheath RG (2012) A novel quantification method for stream-inhabiting, non-diatom benthic algae, and its application in bioassessment. Hydrobiologia 684:225–239

Stevenson RJ (2014) Ecological assessments with algae: a review and synthesis. J Phycol 50:437–461

Stevenson RJ, Rier ST, Riseng CM, Schultz RE, Wiley MJ (2006) Comparing effects of nutrients on algal biomass in streams in two regions with different disturbance regimes and with applications for developing nutrient criteria. Hydrobiologia 561:149–165

Teittinen A, Kallajoki L, Meier S, Stigzelius T, Soininen J (2016) The roles of evaluation and local environmental factors as drivers of diatom diversity in subarctic streams. Freshw Biol 61:1509–1521

Verb RG, Vis ML (2001) Macroalgal communities from an acid mine drainage impacted watershed. Aquat Bot 71:93–107

Verb RG, Vis ML (2005) Periphyton assemblage as bioindicator of mine-drainage in unglaciated western allegheny plateau lotic systems. Water Air Soil Pollut 161:227–265

Wisz MS, Pottier J, Kissling D, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Prachhammer MC, Grytnes JA, Guisan A, Heikkinen RK, Hoyer TT, Kühn I, Luoto M, Maiorano L, Nilsson MC, Normand S, Öckinger E, Schmidt NM, Termanssen M, Timmermann A, Wardle DA, Aastrup P, Svenning JC (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modeling. Biol Rev 88:15–30

Wright JF (1995) Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. Aust J Ecol 20:181–197

Yang J, Yu X, Liu L, Zhang W, Guo P (2012) Algae community and trophic state of subtropical reservoirs in Southeast Fujian, China. Environ Sci Pollut Res Int 19:1432–1442

Yoo JW, Lee Y, Lee CG, Kim CS (2013) Effective prediction of biodiversity in tidal flat habitats using an artificial neural network. Mar Environ Res 83:1–9

Zanchettin C, Ludermir TB (2005) Sistemas neurais híbridos para reconhecimento de padrões em narizes artificiais. SBA Controle & Automação 2:159–172