Limnologica 65 (2017) 46–54

Contents lists available at ScienceDirect

Limnologica

journal homepage: www.elsevier.com/locate/limno

# Modeling of the river ecological status with macrophytes using artificial neural networks

D. Gebler[a,*], K. Szoszkiewicz[a], K. Pietruczuk[b]

[a] Department of Ecology and Environmental Protection, Poznan University of Life Sciences, Poznan, Poland
[b] Environmental Protection Inspectorate, Poznan, Poland

ARTICLE INFO

*Keywords:*
Water Framework Directive
Macrophytes
Water quality
River monitoring
Artificial neural networks

ABSTRACT

Biomonitoring methods based on macrophytes have been used mandatorily in the assessment of freshwaters since the implementation of the Water Framework Directive (WFD). The Macrophyte Index for Rivers (MIR) was developed in Poland for the monitoring of running waters under the WFD requirements. This index shows the degree of river degradation under the influence of water pollutants, especially nutrients. The aim of the present study was to determine the relationship between the MIR and various hydrochemical parameters using artificial neural networks (ANNs). Physico-chemical parameters of water (monthly results for the whole year), which were derived from 147 lowland river survey sites, all located in Poland, were applied to model the MIR values. Water quality variables were determined over three timeframes: the annual average; the average for the vegetation period; and the average for the summer period. Quality of the networks was assessed using coefficient of determination ($R^2$), Nash-Sutcliffe efficiency (NSE) and root mean square error (RMSE). The best modeling quality was obtained for yearly average values of water quality parameters. The quality statistics were: $R^2 = 0.722$, NSE = 0.721 and RMSE = 0.056 (training dataset); $R^2 = 0.555$, NSE = 0.533 and RMSE = 0.101 (validation dataset); $R^2 = 0.650$. NSE = 0.600 and RMSE = 0.089 (testing dataset). This indicates that macrophytes reflect the whole year impact of pollution, whereas summer.

## 1. Introduction

The implementation of the Water Framework Directive (WFD) (Directive 2000/60/EC) by EU member countries created a new chapter in research on various water ecosystems. A new approach to evaluate the water quality based on biological quality elements has been introduced. Numerous systems for the evaluation of ecological conditions have been developed and applied for the needs of this new type of monitoring approach throughout the whole European Union (Birk et al., 2012). Despite some skepticism arising from, e.g., multiple factors influencing macrophytes and difficulties in the statistical detection of important ecological gradients due to the collinearity of ecological gradients (Demars et al., 2012), biological methods have been widely developed throughout the whole EU. Thereby, widely implemented biological methods in monitoring have in recent years enabled the collection of a large volume of data. Nevertheless, despite the extensive application of the ecological monitoring approach and methods established under the WFD, there are still many challenges and problems to achieve environmental objectives in natural rivers (Moss, 2008; Hering et al., 2010; Aguiar et al., 2014).

Aquatic plants are one of the groups of water organisms used to evaluate the ecological quality of surface waters (Directive 2000/60/EC; Birk et al., 2012). The development of macrophytes is strongly determined by various habitat elements. The most significant is their response to nutrients, and thus these organisms are regarded as useful indicators of water and/or sediment eutrophication (Westlake, 1975; Birk and Willby, 2010; Fabris et al., 2009; Haury et al., 2006; Szoszkiewicz et al., 2006; Brabec and Szoszkiewicz, 2006; Kelly and Whitton, 1998). Macrophytes are also sensitive to acidification (Roelofs et al., 1984; Tremp and Kohler, 1995), alkalinity and hardness (Vestergaard and Sand-Jensen, 2000; Triest, 2006), and herbicides (Hellawell, 1986). Apart from water quality, the development of aquatic plants is also influenced by flow rate (Westlake, 1975; Dawson, 1988), hydrological regime (Westlake, 1975; Haslam, 1987), light availability (Westlake, 1975; Dawson and Kern-Hansen, 1979; Triest, 2006; Brabec and Szoszkiewicz, 2006) and hydromorphological alterations (O'Hare et al., 2006; Szoszkiewicz et al., 2006; Carey et al., 2011).

The ecological responses of macrophytes are used in long-term monitoring of water quality, and numerous studies have proved their

* Corresponding author.
*E-mail addresses:* dgebler@up.poznan.pl (D. Gebler), kszoszk@up.poznan.pl (K. Szoszkiewicz), karlo14@op.pl (K. Pietruczuk).

http://dx.doi.org/10.1016/j.limno.2017.07.004
Received 27 February 2017; Received in revised form 8 July 2017; Accepted 10 July 2017
Available online 15 July 2017
0075-9511/ 

high value in the evaluation of aquatic environments (Haury et al., 2006; Staniszewski et al., 2006). The long lifespan of aquatic plants and their high level of resistance to degradation compared to other groups of organisms allow for a highly integrative bioindication statement (Tremp and Kohler, 1995). On the other hand, some other studies have highlighted several problems inherent in macrophyte monitoring, since the trophic gradient is often not detectable with statistical methods due to its collinearity with major ecological gradients − altitude and alkalinity (Demars and Edwards, 2009; Demars et al., 2012; Wiegleb et al., 2014). Thus, more research is required on plant-environment relationships in the surface water environment.

The high level of complexity of interactions occurring in ecosystems (James and McCulloch, 1990) requires advanced methods for data analysis (Hobbs and Hilborn, 2006), including artificial neural networks (ANNs), which are increasingly used in ecological research. ANNs are based on iterative and repeated data analysis. The insufficient knowledge of the mathematical structure of the constructed model resulting in a lack of transparency − so-called 'black boxes' (Lek and Guegan, 1999; Gevrey et al., 2006). Even with this limitation, and thanks to their ease of implementation for solving complex relations, ANNs have become a very popular tool for analysis of ecological data (Lek et al., 1996; Lek and Guegan, 1999; Lek and Park, 2008). In water ecosystem studies, ANNs are widely used for modeling water quality parameters (Singh et al., 2009; Chen and Liu, 2014), both in relation to all main groups of water organisms and for different types of ecosystems (Park et al., 2008; Wenger and Olden, 2012; Millie et al., 2012; Choi et al., 2014; Gebler et al., 2014).

The main goal of our study was to reveal plant-environment relationships based on artificial neural networks. The relationship between the macrophyte index for ecological status assessment and water quality was verified.

## 2. Materials and methods

### 2.1. Site selection

The research was carried out at 147 river sites located in the Polish lowlands (below 200 m a.s.l.). Each of the survey sites was included in the national environmental monitoring program (Fig. 1).

### 2.2. Macrophyte survey

A qualitative and quantitative survey of macrophyte species was performed on a 100-m stretch of river. The macrophyte surveys were conducted in summer, when the vegetation was well developed. Based on the floristic data, the Polish index of the ecological assessment of rivers was calculated − the Macrophyte Index for Rivers (Szoszkiewicz et al., 2010). This index is widely used by scientists investigating fluvial systems in Poland (e.g., Lewin and Szoszkiewicz, 2012; Szoszkiewicz et al., 2014; Jusik et al., 2015). It has been applied in international comparisons of macrophyte methods, including the EU inter-calibration exercise (Birk and Willby, 2010; Wiegleb et al., 2016). The MIR is based on the coverage of certain species in a river stretch and two ecological indicator values. Each of the indicator taxa obtains a stenoecy coefficient showing the response of individual taxa along a eutrophication gradient. The list of indicator macrophytes in this method includes over 150 taxa (Szoszkiewicz et al., 2010).

### 2.3. Physico-chemical analysis

Within the national environmental monitoring program, all selected sites were subjected to monthly sample collections for complex hydrochemical analyses. The selected sites cover a broad trophic spectrum in terms of phosphorus and nitrogen concentration. Both the cleanest (mesotrophic) and most heavily eutrophicated river sites in Poland were included. The eleven determined water parameters are listed in Table 1, and their basic statistics are presented in Appendix A (Tables A1–A3).

Twelve water samples were sampled, but three time periods were considered for analysis: the summer period (June–September), the vegetation period (April–September) and the whole year (January–December). For each of these periods, models of artificial neural networks predicting the values of the Macrophyte Index for Rivers were constructed. The physico-chemical parameters of water were used in the models as explanatory variables.

### 2.4. Artificial neural networks

In the study artificial neural networks available in STATISTICA 12.1 (StatSoft, Inc., 2016) were used. To achieve the objectives the supervised type of network − a MultiLayer Perceptron (MLP) − was utilized. This type of network is widely used to model non-linear relationships (Lek and Park, 2008). The networks had the same number of layers (3) and the same number of neurons in input and output layers that correspond to input and output variables, respectively. The number of neurons in hidden layers can vary in different networks. This is the result of the essence of the learning process of artificial neural networks, in which the structure is refined using iterative algorithms leading to minimization of errors in the networks (Amirikian, 2009). However, according to Fletcher and Goss's (1993) recommendations, in our study the number of hidden neurons ranged from ($2n^{1/2} + m$) to ($2n + m$), where $n$ is the number of input neurons and $m$ the number of output neurons in each of the respective variants.

The process of learning by that network is based on iterative selection and modification of values of weights for all neurons constituting the topology of the network. Neural network learning consists of three steps. The first is the process of network training − during which the weights of neurons are chosen, to minimize the error of the model. After training, validation of the network follows − in this process, on the basis of new data, the quality of the network is evaluated and weight values are verified to minimize error. The last step is the testing process, where the quality of the final model is assessed. According to the three stages of ANN learning, the data set is divided into three separate groups: training, validation, and testing sets. In this study, the training set included 103 river sites (70% of the dataset), the validation set contained 22 sites (15%), and the testing set also contained 22 sites.

Due to the broad range of input variables, and pursuant to recommendations (Lek et al., 2000), the creation of networks was preceded by data standardization using auto scaling (2), whereas the values of the modeled variable were standardized to a range of < 0.1 to 0.9 > . To perform this, the min-max standardization method presented in Eq. (3) was used.

$$z_i = \frac{x_i - \mu}{\sigma} \tag{2}$$

where: $x_i$ is the $i$-th value of each physico-chemical variable at the i-th site (i = 1,…,147); $z_i$ is the $i$-th standardized value of the variable; $\mu$ is the mean of the variable; and $\sigma$ is the standard deviation of the variable.

$$SMIR_i = (SMIR_{\max} - SMIR_{\min})\frac{(MIR_i - MIR_{\min})}{MIR_{\max} - MIR_{\min}} + SMIR_{\min} \tag{3}$$

where: $MIR_i$ & $SMIR_i$ are the absolute value and standardized value of the Macrophyte Index for Rivers; $MIR_{\max}$ & $SMIR_{\max}$ are the maximum value and maximum standardized value of the Macrophyte Index for Rivers; and $MIR_{\min}$ & $SMIR_{\min}$ are the minimum value and minimum standardized value of the Macrophyte Index for Rivers.

The modeling accuracy was demonstrated by $R^2$ (calculated as the square of Pearson's correlation coefficient), which indicates the percentage of explained variance of a dependent variable, the Nash-Sutcliffe efficiency (NSE) (4), which shows the fitting of the modeled data, and the error of models calculated as the root mean square error (RMSE) (5).