Daniel Gebler, Dariusz Kayzer, Anna Budka, Krzysztof Szoszkiewicz

MODELLING VALUES OF RIVER MACROPHYTE METRICS USING ARTIFICIAL NEURAL NETWORKS

Summary

The results of field research at 230 river sections located throughout Poland were used to examine the possibility of predicting values of macrophyte metrics of ecological status. Artificial intelligence methods such as artificial neural networks were used in the modelling. The physicochemical parameters of water (alkalinity, conductivity, nitrate and ammonium nitrogen, reactive and total phosphorus, and biochemical oxygen demand) were used as the explanatory (modelling) variables. The explained (modelled) parameters were the Polish *MIR* (Macrophyte Index for Rivers), the British *MTR* (Mean Trophic Rank) and the French *IBMR* (River Macrophytes Biological Index). The quality of the constructed models was assessed using the normalized root mean square error (*NRMSE*) and the r–Pearson's linear correlation coefficient between variables modelled by the networks and calculated on the basis of the botanical research. These analyses demonstrated that the network modelling *MIR* values had the highest accuracy. The lowest prediction accuracy was obtained for *MTR* and *IBMR* indices. The differences between particular models are likely to result from better adjustment of the Polish method to local rivers (particularly in terms of indicator species used).

Key words: artificial neural networks, biomonitoring, macrophytes, river ecology, water quality

INTRODUCTION

Development of infrastructure in river valleys is associated with the reduction of natural features of watercourses and ecological deterioration of waters. Activities conducted in catchments and aimed at greater economic use of river valleys are important factors in the loss of naturalness. Human developments in the catchments contributed to water pollution. In addition, hydraulic works such as modifying river channels, cutting off meanders and oxbow lakes, constructing embankments and flow regulating structures result in degradation

of fluvial ecosystems involving natural river structures [Żelazo, 2006]. The loss of natural waters causes biocenotic changes among different groups of aquatic organisms, particularly fish, benthic fauna and aquatic vegetation [Herring et al., 2006].

The assessment of the ecological status of surface waters is currently based on biological components of an ecosystem. The obligation of such assessment was imposed on Member States of the European Union in connection with the enactment of the Water Framework Directive (WFD) [Directive 2000]. According to this document, four main groups of aquatic organisms, i.e., fish, macroinvertebrates, phytoplankton and macrophytes are used for the evaluation. Additionally, in order to obtain comprehensive information about the degradation of ecosystem, biological research is supplemented by hydromorphological elements of water bodies and the results of physicochemical evaluation of water samples.

In the aquatic environment biomonitoring, macrophytes, or aquatic plants, are used especially in relation to their tolerance to variable fertility waters. Numerous studies indicate eutrophication as the main factor determining the development of individual species and plant communities. This dependence was used in the development of flowing and stagnant water assessment systems which determine the degree of eutrophication of aquatic ecosystems [e.g. Holmes et al., 1999, Ciecierska, 2008, Szoszkiewicz et al., 2010], or general degradation where eutrophication is the key element of the system [Haury et al., 2006].

Artificial neural networks (ANN) are an alternative tool in the analysis of natural data. Neural networks can be divided according to various criteria and are widely used in various scientific fields and practical applications. The type and structure of the network can be selected depending on the undertaken problem. One of the major advantages of ANNs is significant reduction of data time processing [Tadeusiewicz, 1993]. The relationships and processes observed in ecological research are very often nonlinear, which substantially reduces the possibility of using classical regression methods [Lek et al., 2000]. Boniecki [2008] also shows that neural networks can be the quickest and most convenient way to approximate the nonlinear dependence. A very important feature of the networks is also the possibility of their use in problems where mathematical models of the studied phenomenon is not known.

In the present study we attempted to use artificial neural networks to predict the values of the three river macrophyte metrics: Polish (*MIR*), British (*MTR*) and French (*IBMR*), on the basis of physicochemical analysis of water samples.

MATERIALS AND METHODS

The material for the analysis was derived from observations of 230 river sampling sites located throughout the country where aquatic vegetation survey was conducted. Different river types were considered including lowland, highland and mountain rivers. Analyzed rivers cover the variability level of the hydromorphological degradation.

The botanical research was based on the Polish method used in the national monitoring based on the Macrophyte Index for Rivers (*MIR*) [Szoszkiewicz et al., 2010]. It involves quantitative and qualitative inventory of all species growing in the 100-meter reach of a river. Based on the data collected in the field, in addition to the *MIR* index, two other macrophyte metrics were calculated: the British – Mean Trophic Rank (*MTR*) [Holmes et al., 1999], and French – River Macrophytes Biological Index (*IBMR*) [Haury et al., 2006]. Cover the surface of the river by a single species was conducted using a 9-point scale (Table 1).

Table 1. The scale of the sampling surface coverage by the species used to calculate the various metrics

Surface coverage	*MIR*, *MTR*(P_i) [$i = 1,…,N$]	*IBMR*(K_i) [$i = 1,…,N$]
< 0.1%	1	1
0.1-1%	2	2
1-2.5%	3	3
2.5-5%	4	
5-10%	5	
10-25%	6	4
25-50%	7	
50-75%	8	5
> 75%	9	

On the basis of quantitative and qualitative composition of vegetation at a sampling site, the Macrophyte Index for Rivers (*MIR*) was calculated with the following formula:

$$MIR = \frac{\sum_{i=1}^{N}(L_i \cdot W_i \cdot P_i)}{\sum_{i=1}^{N}(W_i \cdot P_i)} \cdot 10 \tag{1}$$

where:

MIR – value of the Macrophyte Index for Rivers at the sampling site,
N – number of species at the sampling site;
P_i – ratio of coverage for the *i*-th taxon;
W_i – weighting factor for the *i*-th taxon;
L_i – indicator value for the *i*-th taxon [Szoszkiewicz et al., 2010].

The lower the *MIR* value, the more degraded the watercourse in terms of trophic status. *MIR* values range from 10 for eutrophic rivers to up to 100 for rivers with the best ecological status.

Based on botanical observations, it is possible to calculate other macrophyte indices, including the British *MTR* according to the following formula:

$$MTR = \frac{\sum_{i=1}^{N}\left(STR_i \cdot P_i\right)}{\sum_{i=1}^{N} P_i} \cdot 10 \qquad (2)$$

where:

MTR – value of the Mean Trophic Rank at the sampling site,

STR_i – indicator value for the *i*-th taxon [Holmes et al., 1998].

MTR values range from 10 (the most fertile rivers) to 100 (the cleanest rivers) [Dawson et al., 1999].

The *IBMR* index was also calculated on the basis of botanical observation according to the formula:

$$IBMR = \frac{\sum_{i=1}^{N}\left(CS_i \cdot E_i \cdot K_i\right)}{\sum_{i=1}^{N}\left(E_i \cdot K_i\right)} \qquad (3)$$

where:

IBMR – value of the *River Macrophytes Biological Index* at the sampling site,

K_i – ratio of coverage for the *i*-th taxon;

E_i – weighting factor for the *i*-th taxon;

CS_i – indicator value for the *i*-th taxon [Haury et al., 2006].

The higher the *IBMR* value, the less eutrophic the environment.

The botanical research at every sampling site was supplemented by the physicochemical analysis of water (Table 2). All data were used in artificial neural networks modelling. The results of the analyses of water samples were used as input (describing, modelling) variables to the constructed networks, and the macrophyte metrics were used as output (described and modelled) variables. *Automated neural networks* available in STATISTICA 9.1 [StatSoft, Inc., 2010] were used for the statistical analyses.

To model the three macrophyte metrics, a Multilayer Perceptron (MLP) was used. It is a network trained with "a teacher" technique called the delta rule. This type of network is best known and most widely used in the practice of network topologies [Boniecki, 2008]. The collected data consisting of 230 cases were divided into three sets. The first one (training set) contained 160 cases, while the second (validation set) and third (test set) contained over 35 cases. In the process of network learning, the iterative algorithm BFGS (Broyden-Fletcher-Goldfarb-Shanno) was used.

Table 2. Physicochemical parameters of water as the describing data
in the neural network

Parameter	Unit
Alkalinity	mg CaCO$_3$/dm^3
Conductivity	mS/cm
Reactive phosphorus	mg PO$_4$/dm^3
Total phosphorus	mg PO$_4$/dm^3
Nitrate nitro gen	mg N$_{NO3}$/dm^3
Ammonia nitro gen	mg N$_{NH4}$/dm^3
Biochemicaloxygendemand (BOD5)	O$_2$ mg/dm^3

Error of model

Root mean square error (*RMSE*) and normalized root mean square error (*NRMSE*) were calculated to assess the usefulness of the models to estimate *MIR*, *MTR* and *IBMR* based on physico-chemical parameters. The following equations were used:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{oi} - \hat{y}_{mi})^2}{n}}$$

(4)

and

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}}$$

(5)

where

y_o – observed values (*MIR*, *MTR* and *IBMR*);

$\hat{y}_m$ – values calculated by the model (*MIR*, *MTR* and *IBMR*);

y_{min} – minimum observed value (*MIR*, *MTR* and *IBMR*);

y_{max} – maximum observed value (*MIR*, *MTR* and *IBMR*);

n – number of repetition.

RESULTS AND DISCUSSION

The modelling of the three macrophyte metrics was conducted using artificial neural networks. When designing a network, a model describing each macrophyte metrics was developed. The structure of the test models was varied for each index (Table 3). The models had the same number of layers (3), while the number of neurons in individual hidden layers was different. Optimal models for *MIR* and *IBMR* were built of 4 input neurons; the *MTR* network model

had 6 such neurons. In addition, there were differences in the activation functions of hidden neurons (I) and output neurons (II) in each model as shown in Table 3. These differences arise from the essence of the learning process of artificial neural networks in which, e.g., by using iterative algorithms, the network structure is refined in order to minimize the error [Boniecki, 2008].

Table 3. Artificial neural network models structure

	ANN-structure	Activation function I	Activation function II
MIR	MLP 7:4:1	hyperbolic tangent	logistic
MTR	MLP 7:6:1	logistic	exponential
IBMR	MLP 7:4:1	logistic	exponential

Parameters defining the quality of models in various stages of the network structure optimization were calculated for the obtained neural networks (Table 4, Figs. 1 to 3). Both the Pearson's correlation coefficient (r) between the observed and modelled values, and the normalized rootmean square error indicate the best modelling network performance for *MIR*, and then for *MTR* and *IBMR*. The correlation coefficient for the Macrophyte Index for Rivers is 0.929 (empirical significance levelis lower than 0.01) and the normalized root mean square error is 10.4%, which indicates the accurate prediction of this index value. Other studies on artificial neural networks modelling of hydrobiological parameters generated the similar range of error (Soyupak et al., 2003, Dogan et al., 2009).The statistics of the other two networks show higher differences between the values obtained from the models, and the values calculated on the basis of the botanical research.

High correlation for the *MIR* index means that it is adequate for the ecological conditions of rivers in Poland. The Polish *MIR* was created based on the *MTR* and *IBMR* metrics, but has been adapted to local rivers [Szoszkiewicz et al., 2010]. *MIR* was adapted on the basis of verification of indicator values for individual species of aquatic plants. As a result, studies carried out on rivers in Poland showed strong correlation of *MIR* with different physicochemical parameters of water, and trophic pollution in particular [Szoszkiewicz et al., 2010; Gebler and Szoszkiewicz, 2011].

The *MIR* index is so well adapted to the evaluation of rivers in Poland also because the list of indicator species used to calculate this index was adequately prepared. There were 252 species found in the rivers studied, and 133 of them were used to calculate the *MIR* (87% of all *MIR* indicator species). Among the taxa used to calculate the *MTR* index, only 60% of the plants were found, while in the case of *IBMR* only 55%. Adapting the list of indices to the conditions of Polish rivers is very important for the practical use of *MIR*, because the prevalence of a species in an area is one of the main criteria determining a good indicator [Zimny, 2006].

Table 4. Performance parameters of the artificial neural networks for computation of the three macrophytemetrics

Parameter	*r-value*			*RMSE* (*NRMSE*)		
	Training	Validation	Testing	Training	Validation	Testing
MIR	0.840	0.891	0.929	9.61 (11.5%)	9.03 (11.9%)	7.04 (10.4%)
MTR	0.817	0.867	0.916	10.9 (12.1%)	8.05 (12.5%)	10.4 (12.8%)
IBMR	0.766	0.743	0.895	1.74 (13.7%)	2.11 (19.8%)	1.33 (12.2%)

Figures 1 to 3 show the scatterplots (correlation diagrams) depicting a relation between the observed and modelled values of the three considered macrophyte metrics. The observed and predicted values of the river macrophyte index are distributed symmetrically with respect to the line $MIR_m=MIR_o$ (Fig. 1). The correlation coefficient between these variables (no division into the process of training, validation and testing) is 0.86. The relation between the observed and predicted values of *MTR* and *IBMR* are shown analogically. The correlation coefficient for *MTR* is 0.84, and 0.78 for *IBMR*. The values of the correlation coefficients indicate that using the tools of neural networks, we can predict the values of the considered macrophyte indices, given the values of physicochemical parameters of water.
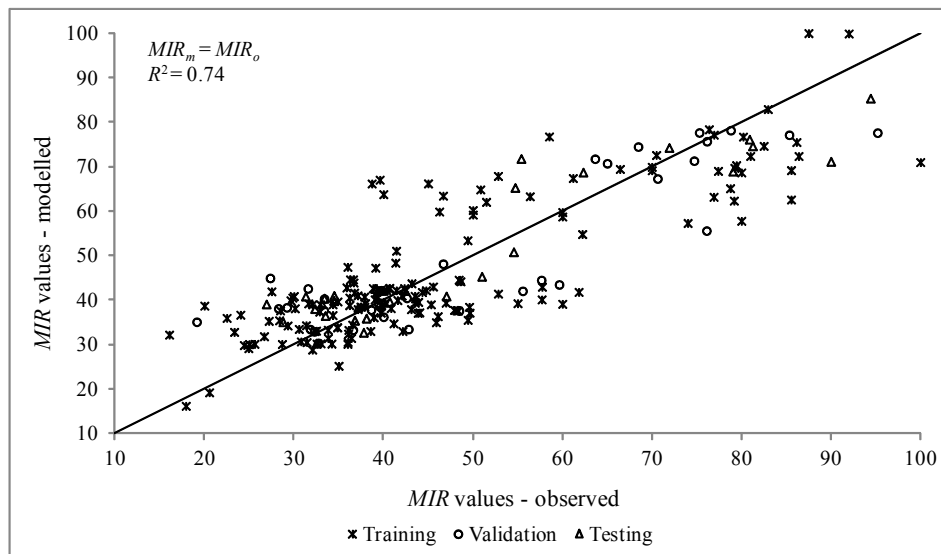
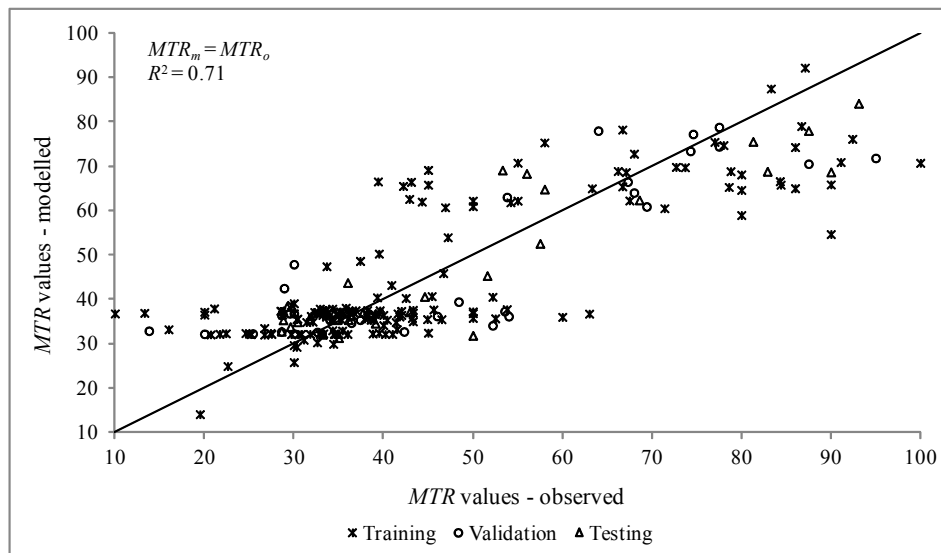**Fig. 1.** Comparison of the modelled and observed *MIR* values

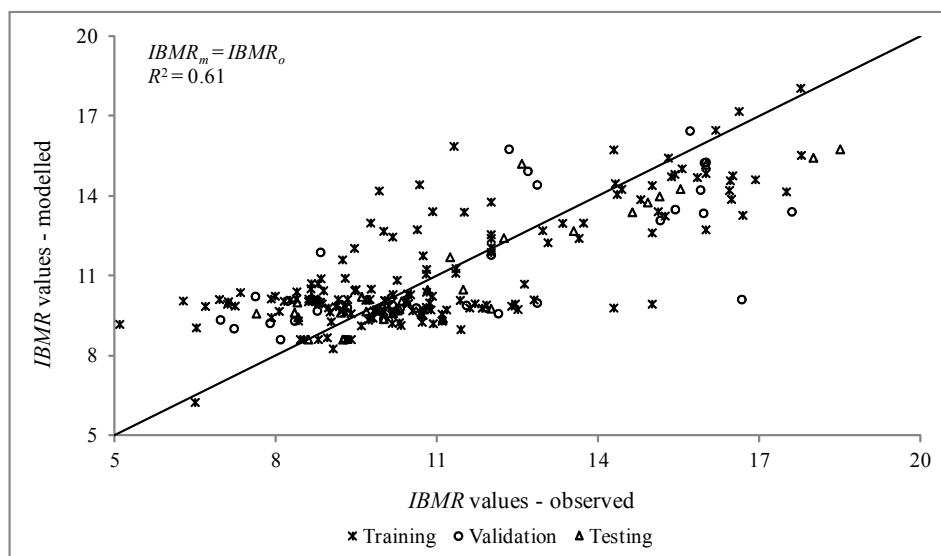**Fig. 2.** Comparison of the modelled and observed *MTR* values

Fig. 3. Comparison of the modelled and observed *IMBR* values

CONCLUSIONS

1. The research results indicate efficiency of using artificial neural networks in the analysis of ecological data, in particular to investigate nonlinear relationships between various components of an aquatic ecosystem.

2. The modelling of macrophyte index values revealed the presence of the relationship between these metrics and physicochemical parameters of water.

3. The neural network model for the Macrophyte Index for Rivers showed lower error (*NRMSE*) and higher correlation coefficient (*r*) in comparison with the network models for *MTR* and *IBMR*. This demonstrates better adaptation of this ecological status assessment system to the conditions of Polish rivers.

REFERENCES

Boniecki J. *Elementy modelowania neuronowego w rolnictwie*. Wydawnictwo UP w Poznaniu 2008, pp 124.

Ciecierska H. *Makrofity jako wskaźniki stanu ekologicznego jezior*. Rozprawy i Monografie, Uniwersytet Warmińsko-Mazurski w Olsztynie. Wydaw. UWM, 2008, nr 139, pp. 202.

Dawson F.H., Newman J.R., Gravelle M.J., Rouen K.J., Henville P. *Assesment of the Trophic Status of Rivers Using Macrophytes. Evaluation of the Mean Trophic Rank*. R&D Technical Report E39. Environment Agency, Bristol 1999, pp 101.

Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy.

Dogan E., Sengorur B., Koklu R. Modeling biochemical oxygen demand of the Melen River in Turkey using an artificial neural network technique. J. Environ. Manag., vol. 90, 2009, p. 1229-1235.

Gebler D., Szoszkiewicz K. *Ecological status assessment of rivers using macrophytes on selected examples (in polish)*. Scientific Review Engineering and Environmental Sciences, vol.20 (2) No 52 2011, p. 75–83.

Haury J., Peltre M.C., Tremolieres M., Barbe J., Thiebaut G., Bernez I., Daniel H., Chatenet P., Haan-Archipof G., Muller S., Dutartre A., Laplace-Treyture C., Cazaubon A., Lambert-Servien E. *A new method to assess water trophy and organic pollutionthe Macrophyte Biological Index for Rivers (IBMR): its application to different types of river and pollution.*Hydrobiologia, vol. 570, 2006, p. 153-158.

Hering D., Johnson R.K., Kramm S., Schmutz S., Szoszkiewicz K. &Vardenschot P.F.M. *Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress*. Freshwater Biology, vol. 51, 2006 p. 1757-1785.

Holmes N.T.H., Newman J.R., Chadd S., Rouen K.J., Saint L., Dawson F.H. *Mean Trophic Rank. A user's manual*. R&D Technical Report E38. Environment Agency, Bristol 1999, pp. 156.

Lek S., Giraudel J.L., Guegan J.F., *Neural Networks:Algorithms and Architectures for Ecologists and Evolutionary Ecologists*, [w:] *Neural Networks. Aplication to Ecology and Evolution*, Ed. S. Lek, J.F. Guegan, Artificial Springer-Verlag Berlin, Heidelberg 2000, p. 3-27.

Soyupak S., Karaer F., Gurbuz H. *A neural network based approach for calculating dissolved oxygen profiles in reservoirs*. Neural Comput. Appl. vol. 12, 2003, p. 166-172.

StatSoft, Inc. (2010). STATISTICA (data analysis software system), version 9.1.

Szoszkiewicz K., Zbierska J., Jusik S., Zgoła T. *Makrofitowa Metoda Oceny Rzek. Podręcznik metodyczny do oceny i klasyfikacji stanu ekologicznego wód płynących w oparciu o rośliny wodne*. Bogucki Wydawnictwo Naukowe, Poznań, 2010, pp. 81.

Tadeusiewicz R. 1993. Sieci neuronowe. Warszawa: Akademicka Oficyna Wydawnicza

Zimny K. Ekologiczna ocena stanu środowiska. Bioindykacja i biomonitoring. Agencja Reklamowo-Wydawnicza A. Grzegorczyk. Warszawa, 2006, pp. 264.

Żelazo J. *River and Valley restoration* (in Polish). Infrastructure and ecology of rural areas. Polska Akademia Nauk, Oddział w Krakowie, Komisja Technicznej Infrastruktury Wsi, Nr 4/1, 2006, p. 11-31

Dr Dariusz Kayzer
(dkayzer@up.poznan.pl)
Dr Anna Budka
(budka@up.poznan.pl)
Department of Mathematical and Statistical Methods,
WojskaPolskiego 28, 60-637 Poznan
Prof. dr hab. inż. Krzysztof Szoszkiewicz
(kszoszk@up.poznan.pl)
Mgr inż. Daniel Gebler
(dgebler@up.poznan.pl)
Department of Ecology and Environmental Protection,
Piątkowska 94c, 60-649 Poznan
Poznan University of Life Sciences