Egyetemi doktori (PhD) értekezés

NEURAL NETWORKS IN ECOLOGICAL DATA ANALYSIS

NEURÁLIS HÁLÓZATOK ALKALMAZÁSA AZ ÖKOLÓGIAI ADATOK ÉRTÉKELÉSÉBEN

Várbíró Gábor

Témavezetők: Dr. Tóthmérész Béla
Dr. Borics Gábor

DEBRECENI EGYETEM
Juhász-Nagy Pál Doktori Iskola
Debrecen, 2009

A doktori értekezés betétlapja

Ezen értekezést a Debreceni Egyetem TTK ....Doktori Iskola … **Kvantitatív és Terresztris Ökológia** …programja keretében készítettem a Debreceni Egyetem TTK doktori (PhD) fokozatának elnyerése céljából.
Debrecen, 2009. június. 25.

…………………………
a jelölt aláírása

Tanúsítom, hogy ....Várbíró Gábor......................... doktorjelölt 2000- 2009. között, a fent megnevezett Doktori Iskola **Kvantitatív és Terresztris Ökológia** programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javasolom.
Debrecen, 2009. június 25.

…………………………..…
a témavezető aláírása

Tanúsítom, hogy ....Várbíró Gábor......................... doktorjelölt 2000- 2009. között, a fent megnevezett Doktori Iskola **Kvantitatív és Terresztris Ökológia** programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javasolom.
Debrecen, 2009 június 25.

……………………………..……
a témavezető aláírása

NEURAL NETWORKS IN ECOLOGICAL DATA ANALYSIS

NEURÁLIS HÁLÓZATOK ALKALMAZÁSA AZ ÖKOLÓGIAI ADATOK ÉRTÉKELÉSÉBEN

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében készült,
a Természettudományi tudományágban.

Írta: …**Várbíró Gábor...** okleveles Biológus-ökológus
Készült a Debreceni Egyetem **… Juhász-Nagy Pál Doktori Iskola**
Kvantitatív és Terresztris Ökológia programja keretében

Témavezetők:
Dr. Tóthmérész Béla …………………………
Dr. Borics Gábor …………………………

A doktori szigorlati bizottság:
elnök: Dr. Nagy Sándor Alex …………………………
tagok: Dr. Magura Tibor …………………………
Dr. Nagy Miklós …………………………
A doktori szigorlat időpontja: 2007. június 14.

Az értekezés bírálói:

………………………… …………………………
………………………… …………………………

A bírálóbizottság:
elnök: ………………………… …………………………
tagok: ………………………… …………………………
………………………… …………………………
………………………… …………………………
………………………… …………………………

Az értekezés védésének időpontja: 2009… . ……………… … .

CONTENTS

Papers included

1. **Várbíró, G.,** Ács, É**.,** Borics, G., Érces, K., Fehér, G., Grigorszky, I., Japport, T., Kocsis, G., Krasznai, E., Nagy, K., Nagy-László, Zs., Pilinszky, Zs., Kiss, K.T. (2007): Use of Self-Organizing Maps (SOM) for characterisation of riverine phytoplankton associations in Hungary. - Arch. Hydrobiol. Suppl. Large Rivers 17: 383-394. *(Printed with kind permission from the Journal of Arch. Hydrobiol)*

2. **Várbíró, G.**, Borics, G., Kiss, K.T., Szabó, K. É., Ács, É. (2007): Use of Kohonen Self Organizing Maps (SOM) for the characterisation of benthic diatom associations of the River Danube and its tributaries. – Arch. Hydrobiol. Suppl. Large Rivers 17: 395-403. *(Printed with kind permission from the Journal of Arch. Hydrobiol)*

3. Ács, É., Borsodi, A.K., Kiss, É., Kiss, K.T., Szabó, K.É., Vladár, P., **Várbíró, G.,** Záray, Gy. (2008): Comparative algological and bacteriological examinations on biofilms developed on different substrata in a shallow soda lake – Aquat Ecol (2008) 42:521–531. *(Printed with kind permission from the Journal of Aquatic Ecology)*

4. Borics, G., **Várbiró, G.**, Grigorszky, I., Krasznai, E., Szabó, S., Kiss, K. T. (2007): A new evaluation technique of potamoplankton for the assessment of the ecological status of rivers. Arch. Hydrobiol. Suppl. Large Rivers 17: (3-4) 465-486. *(Printed with kind permission from the Journal of Arch. Hydrobiol)*

5. Van Dam, H., Stenger-Kovács, Cs., Ács, É., Borics, G., Buczkó, K., Hajnal, É., Soróczki-Pintér, É., **Várbíró, G**., Tóthmérész, B., Padisák, J. (2007): Implementation of the European Water Framework Directive: Development of a system for water quality assessment of Hungarian running waters with diatoms. - Arch. Hydrobiol. Suppl. Large Rivers 17: 339-364. *(Printed with kind permission from the Journal of Arch. Hydrobiol)*

Introduction and Objectives

"in the 19 century the data speak for itself,
at present its completely different …"

The need for better techniques, tools and practices to analyse ecological and economic systems within an integrated framework at wider scales has never been so great. Currently, individuals of different professions consider scientific predictions that are based on highly complicated principles and hypotheses as beyond their comprehension and often ignore them (Clark et al., 2001). Scientists are becoming more focused on science and research of course than improving their communication with the general public (Buckeridge, 2001). The transformation of large quantities of disparate data into simple, useable information is seen as a major challenge in many current environmental monitoring programmes. Vant (1999) identified two key aspects of this, namely, the identification of robust methods to summarise large volumes of data without losing the useful information in it, and presenting this information to a largely non-technical audience in an understandable and compelling manner.

The task of transforming data into useful information is a major challenge to environmental scientists, who were probably attracted to the study of ecosystems because of the complexity involved, and now increasingly find themselves having to move beyond the security of complicate conceptual models and the jargon of the highly technical journals, to communicate with a wider community.

Over the last few years, the use of Artificial Neural Network (ANN) techniques has provided ecologists with a powerful method for modelling large volumes of ecological data (Paisley et al., 2003).Neural networks are powerful computational tools that can be used for classification, pattern recognition, empirical modelling and for many other tasks. Neural networks can be "trained" to provide the right output (binary, fuzzy, quantitative) if enough input-output patterns are available and if these patterns effectively describe the system that is to be modelled.

The earliest work in neural computing goes back to the 1940's when McCulloch and Pitts (McCulloch and Pitts, 1943) introduced the first neural network computing model. Reinforcing this concept of neurons and how they work was a book written by Donald Hebb. The *Organization of Behavior* was written in 1949 (Hebb, 1949). It pointed out that neural pathways are strengthened each time that they are used. In the 1950's, Rosenblatt's (Rosenblatt, 1956, 1957) work resulted in a two-layer network, the perceptron, which was capable of learning certain classifications by adjusting connection weights. The architecture inspired by the human brain which contains neuron connected to each other by synapses. The artificial analogue of the neuron is the perceptron. (Fig. 1.)

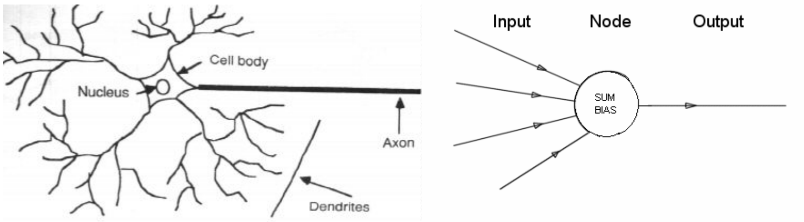


Fig. 1. The structure of the natural neuron and the artificial perceptron

Although the perceptron was successful in classifying certain patterns, it had a number of limitations, which the decline of the field of neural networks. However, the perceptron had laid foundations for later work in neural computing. In the early 1980's, researchers showed renewed interest in neural networks.

These studies show promising results in modelling complex relationships between biological and environmental variables of natural habitats to support sustainable environment development (Shanmuganathan, 2005). Bayesian belief networks and ANN techniques, namely *back propagation* and *self organizing maps* (SOM), were found to be successful in calculating a set of biotic indices for monitoring of river quality based on the distribution and abundance of macroinvertebrate taxa and environmental parameters of 1995 national survey data (Walley et al., 1998). The study concluded that the methods tested to have shown considerable potential for use in river quality classification and river management. Modelling Patterns in Environmental Data (MOPED) uses SOM techniques for modelling patterns within fish species distribution and elevation of sampled freshwater systems in a region (Jowett, 2001). MOPED also was found to be capable of predicting the biological assemblages that should be present in certain streams. Onkal-Engin et al. (2005) successfully applied a back-propagation ANN algorithm to develop an electronic nose for the purpose of controlling sewage odours in wastewater treatment plants.

Ecosystem analysis and prediction with empirical statistical and analytical methods are often limited by the spatial complexity and temporal dynamic character of ecological processes. This is one reason for the typically non-linear interrelations of variables and species with data being not normally distributed. Artificial neural networks provide an attractive alternative tool for analysing ecological data and for modelling due to their specific features such as non-linearity, adaptivity, generalisation and model independence (no a-priori model needed). ANNs have been applied to various fields of aquatic sciences and engineering, such as modelling water quality (e.g. Daniell and Wundke, 1993; Maier and Dandy, 1996a,b; 1997; Lachtermacher and Fuller, 1994; Schizas et al., 1994; Maier, 1995; Winkler and Voigtlander, 1995; Kaluli et al., 1998; Wen and Lee, 1998) and relating community characteristics with environmental variables (e.g. Chon et al., 1996; Lek et al., 1996; Recknagel, 1997; Recknagel et al., 1997, 1998; Guegan et al., 1998; Lee et al., 1998; Maier et al., 1998).

An important aspect of an ANN model is whether it needs guidance in learning or not. Based on the way they learn, all artificial neural networks can be divided into two learning categories - *supervised* and *unsupervised*. In *supervised* learning, a desired output result for each input vector is required when the network is trained.

An ANN of the supervised learning type, such as the multi-layer perceptron, uses the target result to guide the formation of the neural parameters. It is thus possible to make the neural network learn the behaviour of the ecological system. In *unsupervised* learning, the training of the network is entirely data-driven and no target results for the input data vectors are provided. An ANN of the *unsupervised* learning type, such as the self-organizing map, can be used for clustering the input data and find features inherent to the problem. (Burke, and Ignizio, 1992)

The biological monitoring of the water currents in several countries provided large volume datasets during the last decades. In case of large volume of data, the traditional multivariate statistical methods like cluster analysis and ordination is difficult to interpret and present the information. Self Organizing Map is a novel approach for the visualization of high-dimensional data. Self-organizing Map is a data visualization technique invented by Professor Teuvo Kohonen which reduces the dimensions of data through the use of self-organizing neural networks. SOM converts complex statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display. Thus, it compresses information while preserving the most important topological and metric relationships of the primary data items (Kohonen 2001). The main advantages of the SOM are the better data visualisation and noise reduction (Vesanto and Alhoniemi 2000). Mangiameli et al. (1996) compared SOM and several hierarchical clustering methods, and found SOM superior to hierarchical clustering in both robustness and accuracy. In addition, the two-level clustering approach (SOM neural network followed by K-means clustering) was developed and successfully applied to cluster data (Vesanto and Alhoniemi 2000; Beccali et al. 2004). SOMs are increasingly popular tools in ecology, and have been used to describe bentic algal assemblages in France and Luxembourg (Gosselain et al. 2005, Rimet et al. 2005a, 2005b). The self-organizing map (SOM) is an efficient tool for mining non-linear data and has been extensively used for patterning community data since 1990s (e.g., Chon et al., 1996, Giraudel et al., 2000, Kwak et al., 2000; Levine et al., 1996; Park et al., 2003, 2005). Chon et al. (1996) classified benthic macroinvertebrate communities in polluted streams with the SOM and elucidated community patterning according to anthropogenic disturbances and locality of the sample sites. Chon et al. (2002) further implemented the SOM to a large-scale data collected in different river systems in the Korean Peninsula for 16 years. The large-scale data were accordingly arranged to reveal the impact of environmental disturbances. In this study, we further elaborated to show the trained SOM as a means of providing a comprehensive view on ecological states of the communities and to use the SOM as a tool for assessing biological water quality. Further articles which summarise the application of the SOM in ecological application(Kaski et al., 1996, Kalteh et al.,2008, Ceregrino and Park, 2009)

The Hungarian routine monitoring program for chemical parameters has been monitored in running and standing waters for more than 40 years already. Unfortunately in the early years the only two biological components which were stored in a database were the index value and the chlorophyll-a content of the given samples. In 2000 the author of the theses started to build a database for storing the taxa list in digital form. Since then the system became the part of the Hungarian monitoring program. In 2001 there has been an initial start in the qualitative monitoring of macroinvertebrates, as a first step towards implementation of the Water Framework Directive (European Union, 2000) requirements. In 2005 other biological elements of the WFD were involved in the national monitoring program such as phytobenton and macrophytes. At present all of these biological quality elements have a relevant database to store the taxalist and sampling event features.

From the presented publications the author’s main interests and responsibilities, as well as the main goals of the PhD thesis were the followings:

Characterize selected Hungarian running waters according to their characteristic phytoplankton assemblages with the help of SOM techniques (Paper 1)

Construction of useable biological database for storing relevant ecological information on phytoplankton and phytobenton biological elements (Paper 4, 5)

Characterize the upper Danube section based on their characteristic diatom communities by SOM methods (Paper 2)

Describe the differences and similarities of the periphyton of Lake Velencei on different substrate using SOM statistics (Paper 3)

.

Material and methods

"Gagnants and Perdants"

1.1 Statistical methods

SOM algorithm

For statistical analysis in Paper 1, 4 and 5 we used the SOM algorithm which can make a projection of the database into a two dimensional hexagonal map. Closely related communities were placed into neighbouring hexagons by their similarities, while samples with different communities were in distant hexagons. The SOM can display the groupings of samples and species together; therefore, each species can be evaluated by its importance (Kohonen, 2001). The SOM Toolbox (http://www.cis.hut.fi/projects/somtoolbox) was used to implement the SOM under a MATLAB™s environment.
In this chapter I would describe the algorithm in details. We followed the abbreviations of the publication Giraduel (Giraduel and Lek , 2001).
The SOM consists of two layers: input and output layers connected by connection intensities (weights).(Fig. 2.)

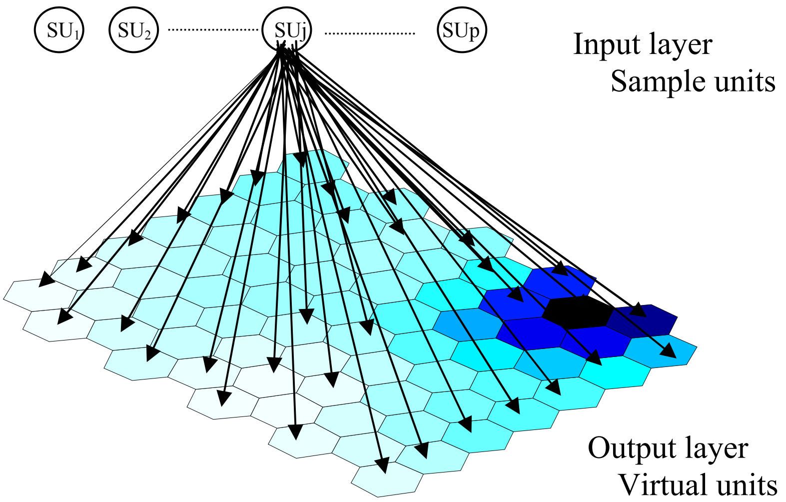


Fig. 2. The structure of the SOM neural network

Input layer gets information from data matrix, while output layer visualizes the computational results. When an input vector is sent through the network, each neuron of the network computes the distance between the weight vector and the input vector. The output layer consists of output neurons, which are usually arranged into a two-dimensional grid for better visualization. The best arrangement for the output layer is a hexagonal lattice, as it does not favour horizontal and vertical directions as much as a rectangular array (Kohonen, 2001).

The ecological data usually contains a matrix of sample units (SU) and species composition data as like in table 1.

Table 1.An ecological data matrix, i species ($sp_1,\ldots, sp_i$) are observed in SUp sample units (SU1,…, SUp), X_{ip} species weight (like abundance etc.)

		Species / variables/		
		Sp_1	-------------------	Sp_i
Sample units	SU_1	X_{11}	-------------------	X_{i1}
	SU_2	X_{12}	-------------------	X_{i2}
	⋮	⋮		⋮
	SU_p	X_{1p}	-------------------	X_{ip}

This is the input layer of the SOM. The goal of the SOM is to make the output layer consist of virtual units (VU) with the virtual species weight (prototype vector or codebook vector) as seen in table .2.

Table 2. .The SOM data matrix, i species ($sp_1,\ldots, sp_i$) are observed in VUp virtual units (VU1,…, VUp), W_{ip} weight vector (species abundance etc)

		Species / variables/		
		Sp_1	-------------------	Sp_i
Virtual units	VU_1	W_{11}	-------------------	W_{i1}
	VU_2	W_{12}	-------------------	W_{i2}
	⋮	⋮		⋮
	VU_p	W_{1p}	-------------------	W_{ip}

The algorithm of the SOM process can be described as:

- Step 1: Epoch $t=0$, the virtual units VU_p are initialized, weights drawn from the input dataset.
- Step 2: A sample unit SU_j is randomly chosen as an input unit.
- Step 3: Using some metric, the distances between SU and each virtual unit are computed.
- Step 4: The virtual unit VU_c closest to the input SU_j is chosen as the winning neuron. VU_c is called the Best Matching Unit (BMU).
- Step 5: The virtual units VU_p are updated with the learning rule:
- Step 6: Increase time t to t+1. If $t<t_{max}$ then go to step 2 else stop the training.

In step 1, the user decide the size of the map, this is an important question because if the map is too big there would be empty cells in the final map, whereas if the size is too small there would be less distinct groups as expected. The optimal numbers of nodes according to Vesanto (Vesanto, 2000) can be calculated as:

$$Nn = 5\sqrt{SU_d}$$

Where Nn is the number of nodes in the final map and SU_d is the number of sample units.

There are two ways to initialize the weight vectors. The first is to give each weight vector a random value based on the SU, the second is to make a linear gradient from the weights.

In step 3, the distance or the measure of dissimilarity between stations can be chosen. Usually the conventional Euclidean distance is widely usable but is not the only possibility. Some measurements that take into account the difficult problems of double zero or scale (e.g. binary data) can also be used. However, the use of the learning rule has to be made very prudently. If Euclidean distance is not used, the learning equation has to be adapted in order to be compatible with the chosen distance or the measure of dissimilarity (Kaski, 1997). For instance Giraudel and Lek (2001) used the Whittaker's relative transformation for absolute distance (Whittaker, 1952), as suggested by Orlóci (1978) in this way the learning rule should also change as seen in table 3.

Table 3. The equation for selecting the BMUs and the learning rules

Distance	Learning rule
$d(x,w_i) = \left[\sum_{j=1}^{p}(x_j - w_{ij})^2\right]^{\frac{1}{2}}$ Euclidean	$w_{ik}(t+1) = w_{ik}(t) + h_{ck}(t)[x_{ij}(t) - w_{ik}(t)]$ Euclidean learning rule
$D(SU_i, SU_j) = \sum_{l=1}^{n}\left\lvert \frac{x_{li}}{\sum_{l-1}^{n} x_{li}} - \frac{x_{li}}{\sum_{l-1}^{n} x_{lj}} \right\rvert$ Whittaker	$w_{ik}(t+1) = \frac{w_{ik}(t+1) = w_{ik}(t) + h_{ck}(t)[x_{ij}(t) - w_{ik}(t)]}{\sum_{l-1}^{n}(w_{lk}(t+1) = w_{lk}(t) + h_{ck}(t)[x_{lj}(t) - w_{lk}(t)])}$ Modified learning rule

In step 5, in the Euclidean learning rule the function $h_{ck}(t)$ is the neighbourhood function. During the learning process, the BMU defined in step 4 is not the only updated unit. In the grid, a neighbourhood is defined around the BMU and all units within this neighbourhood are also updated. Several choices can be made for the definition of the neighbourhood function (see Kohonen, 2001). The most often neighbourhood is the Gaussian function:

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

$\|\mathbf{r}_k - \mathbf{r}_c\|$ is the Euclidean distance on the map between the winning unit VU_c and each virtual unit VU_k.

$\boldsymbol{\sigma}$ is a decreasing function of time, which defines the width of the part of the map, affected by the learning.

$\boldsymbol{\alpha}$ is the 'learning-rate factor',

Both factors are decreasing functions of the time and converge towards 0. The learning is broken down into two parts:

1. the ordering phase: during this phase, the virtual stations are widely modified in a large neighbourhood of the BMU with large values of $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$
2. the tuning phase: when this second phase takes place, only the virtual units adjacent of the Best Matching Unit are modified. This phase is much longer than the former one and $\boldsymbol{\sigma}$ is decreasing very slowly towards 0.

It is recommended that the number of steps in the tuning phase has to be 500 times the number of units in the output layer.

When the learning process is finished, a map with S hexagons is obtained and in each hexagon, there is a virtual station in which the weights/species composition has been computed.

Then, this map can be visualised in different ways:

- The sample units can be mapped. The BMU is computed for each sample unit and label the corresponding hexagon.
- The species composition of each virtual unit can display the distribution of each weights/species. The SOM map of the variables called *component planes*
- The two presentations can be mixed to display the sample units and the weights/species distribution on the same map.

Structuring index (SI)

In order to characterise the VU groups we used the Structuring Index (SI) which was originally developed to define the species that show the strongest influence on the organization of the SOM map (Park et al., 2005). Tison et al. (2004, 2005) used the SI to evaluate relevant diatom species in the classification of diatom communities. The set of species showing high SI can be considered as the indicator species. Taxa showing strong gradient display have high SI values, whereas species showing weak gradient present low SI values. Thus, the higher the value of SI, the more relevant the variable is to the structure of the map. The index calculation based on the weight vectors of the SOM see Fig. 3.

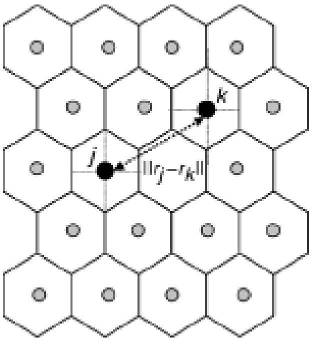


Fig. 3. topological distance of the SOM output units used in the SI calculation

$$SI_i = \sum_{j=1}^{S} \sum_{k=1}^{j-1} \frac{| W_{ij} - W_{ik} |}{\| r_j - r_k \|}$$

This equation calculates the SI value, where
w_{ij} and w_{ik} are the connection weights of the ith species in the SOM units,
$||rj-rk||$ is the topological distance between units j and k,
S is the total number of SOM output unit.

Then based on the SI value of each variable it is possible to order them by these properties and select the ones with high values. These variables are responsible mostly for the topology of the SOM therefore can regard as indicators. This is a very useful function if we are working with large number of variables or species. An example of the visualization can be seen in Fig. 4. On this example the data of paper 1 were used. The two SOM component plane represent two variables one (TIB) shows high gradient which expressed as colour gradient from red to blue, red means high abundance of the given variable, whereas blue means low value. The second variable (M) shows very small gradient distribution. These distributions are also expressed in the SI index values.

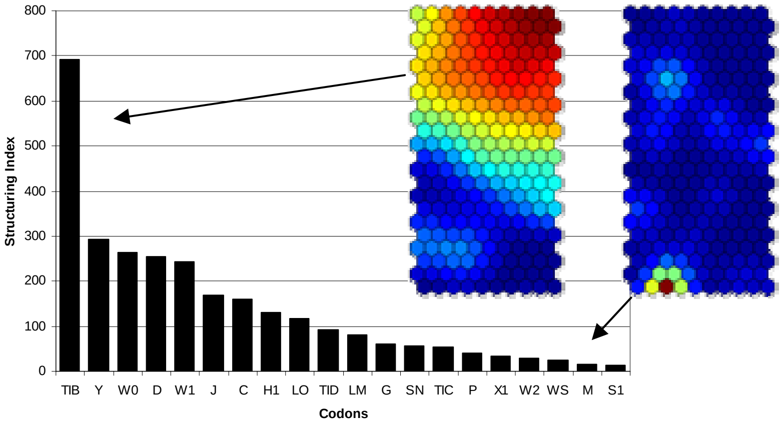


Fig. 4. The SI profile of a SOM map and the relevant component planes showing high or low gradient distribution, values of the component planes are abundances where red means high and blue express low values.

Clustering of the SOM

The map obtained after the learning process of the SOM represents all the SU assigned to neurons so that similar samples are located close to each other and far from those dissimilar. However, it is not easy to distinguish subsets because there are still no boundaries between possible clusters.

Therefore, it is necessary to subdivide the output neurons into different groups according to their similarity.

One frequent clustering is based on the U-matrix algorithm (Ultsch, 2003), which calculates the distance of a weight vector (w) to its neighbours in the SOM, and displays the cluster structure of the map units. Supposing the map has a size of m columns and n rows, the following value ($M_{x,y}$; U-matrix) is calculated for all positions as

$$M_{x,y} = \frac{1}{H} \sum_{a=x-1}^{x+1} \sum_{b=y-1}^{y+1} \left\| w_{x,y} - w_{a,b} \right\|$$

where *H* is the number of neighbour units ranging from 2 to 6, depending upon the location of the virtual unit. The values are rescaled between 0 and 1 for visual comparison. The matrix presented as a grey-scaled picture based on the calculated values: bright areas with low values depict short distances while dark areas with high values represent long distances to the surrounding neighbours. High values of the U-matrix indicate groups' boundaries, while low values reveal groups themselves as seen in Fig. 5.

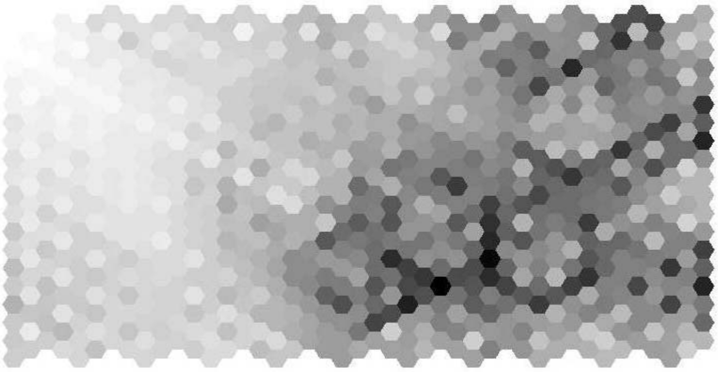


Fig. 5. U-matrix of the SOM, darker cells indicate more distance between neighbouring cells.

For clustering the SOM we can use several conventional clustering techniques however in our work we used the K-means clustering technique. K-means clustering is an algorithm to classify or to group objects based on attributes (in this case species composition) into K number of group. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid as the square error of each data point is calculated and clusters reformed such that the sum of square errors is made to be minimum. Details of the methodological background can be found in Vesanto and Alhoniem (2000) or Beccaleti al. (2004).

For determining the suitable cluster size we used Davies-Bouldin Validity Index. (Davies and Bouldin, 1979) This index is a function of the ratio of the sum of within-cluster scatter to between-cluster separation, it uses both the clusters and their sample means. The index calculated as follows:

$$DB = \frac{1}{n}\sum_{i=1}^{n}\max_{i\neq j}\left\{\frac{S_n(Q_i)+S_n(Q_j)}{S(Q_i,Q_j)}\right\},$$

where n - number of clusters, S_n - average distance of all objects from the cluster to their cluster centre, $S(Q_iQ_j)$ - distance between clusters centres. The ratio is small if the clusters are compact and far from each other, consequently, Davies-Bouldin index will have a small value for a good clustering.

1.2 Ecological Databases

A database is a structured collection of records or data. The most commonly used database model today is the relational model. The relational model was introduced by E. F. Codd in 1970 (Codd, 1970) and make database management systems more independent of any particular application. A relational database contains multiple tables, which are connected to each others by relations.

The basic data structure of the relational model is the table, where information is represented in columns and rows. Thus, the "relation" in "relational database" refers to the various tables in the database; a relation is a set of tables. The columns can be expressed as various attributes of the entity (sampling date, site name etc.), and a row is an actual instance of the entity (like 2009.05.27., Concó, Ács).

One of the strengths of the relational model is that any value occurring in two different records implies a relationship among those two records. It also minimizes the redundant information.

There are three type of identifying parent-child relationships (Fig. 6.):

- 1:1 when the two table has the same unique key,
- 1:∞ when the first table is connected to the second by a foreign key, thus one table has many instances on the other, practically it occurs such in Fig. 6. when one *Sampling_location* table has multiple *Sampling_event_phytobenton* instance
- ∞:∞ when both tables have multiple instance in each other, like the connection between sampling events and taxonlists

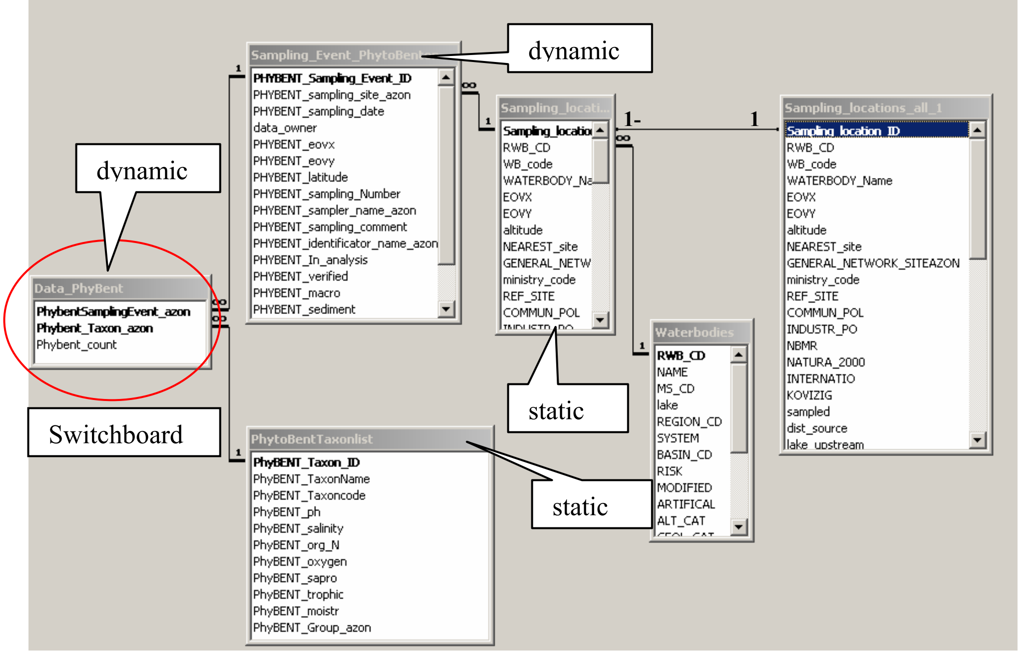
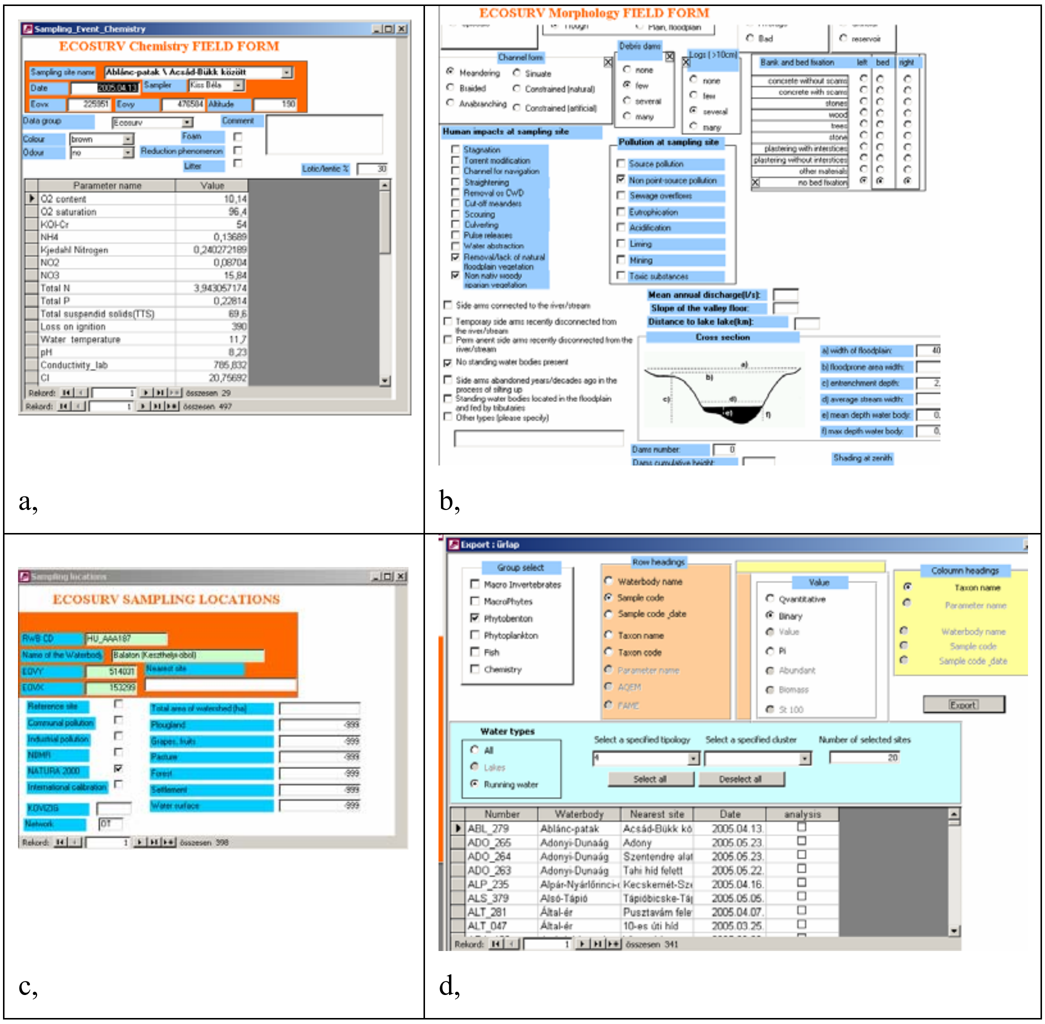


Fig 6. The phytobenton database structure of the ECOSURV Project, which represents the different connection types among tables and static and dynamic variables.

Practically the ∞:∞ or “many to many” connection is not exist in the most of the database management software such as Access. So these features are solved by tables called switchboard which contains foreign keys from both tables.

According to variable types there can be static and dynamic ones. Static variables are like attributes of sampling sites (eg. site name GIS coordinates) or taxa names, whereas dynamic ones are like attributes of sampling events (eg. sampling date, substrate type) or species abundances which could change to survey to survey. (Fig. 6.)

The author produces several ecological databases in connection to surface water monitoring. One of the biggest achievements was the ECOSURV database, which were used in Paper 5. The ECOSURV was a PHARE project co-funded by the EU and Hungary’s Ministry of Environment and Water, implemented by an international consortium led by ARCADIS Euroconsult, was the first in the EU to provide a nation-wide overview of the ecological status of the water bodies. Sampling and assessment methods have been selected for all biological elements including fish, macrophyte, phytoplankton, phytobenton and macroinvertebrates. (Arcadis, 2005) The final database contains 70 tables and thousands of individual records. Some form of the database can be seen in Fig. 7.



Sampling_Event_Chemistry

ECOSURV Chemistry FIELD FORM

Sampling site name: Ablánc-patak \ Acsád-Bükk között
Date: 2005.04.13
Sampler: Kiss Béla
Eovx: 225951
Eovy: 476504
Altitude: 150
Data group: Ecosurv
Comment:
Colour: brown
Odour: no
Foam ☐
Reduction phenomenon ☐
Litter ☐
Lotic/lentic %: 30

Parameter name	Value
O2 content	10,14
O2 saturation	96,4
KOI-Cr	54
NH4	0,13689
Kjedahl Nitrogen	0,240272189
NO2	0,08704
NO3	15,84
Total N	3,943057174
Total P	0,22814
Total suspendid solids(TTS)	69,6
Loss on ignition	390
Water temperature	11,7
pH	8,23
Conductivity_lab	785,832
Cl	20,75692

Rekord: 1 összesen 29
Rekord: 1 összesen 497

a,

ECOSURV Morphology FIELD FORM

○ Plain, floodplain
○ Bad
○ reservoir

Channel form
● Meandering ○ Sinuate
○ Braided ○ Constrained (natural)
○ Anabranching ○ Constrained (artificial)

Debris dams
○ none
● few
○ several
○ many

Logs (>10cm)
○ none
○ few
● several
○ many

Bank and bed fixation	left	bed	right
concrete without scams	○	○	○
concrete with scams	○	○	○
stones	○	○	○
wood	○	○	○
trees	○	○	○
stone	○	○	○
plastering with interstices	○	○	○
plastering without interstices	○	○	○
other materials	○	○	○
☒ no bed fixation	●	●	●

Human impacts at sampling site
☐ Stagnation
☐ Torrent modification
☐ Channel for navigation
☐ Straightening
☐ Removal os CWD
☐ Cut-off meanders
☐ Scouring
☐ Culverting
☐ Pulse releases
☐ Water abstraction
☑ Removal/lack of natural floodplain vegetation
☑ Non nativ woody riparian vegetation

Pollution at sampling site
☐ Source pollution
☑ Non point-source pollution
☐ Sewage overflows
☐ Eutrophication
☐ Acidification
☐ Liming
☐ Mining
☐ Toxic substances

☐ Side arms connected to the river/stream
☐ Temporary side arms recently disconnected from the river/stream
☐ Perm anent side arms recently disconnected from the river/stream
☑ No standing water bodies present
☐ Side arms abandoned years/decades ago in the process of silting up
☐ Standing water bodies located in the floodplain and fed by tributaries
☐ Other types (please specify)

Mean annual discharge(l/s):
Slope of the valley floor:
Distance to lake lake(km):

Cross section
a) width of floodplain: 40
b) floodprone area width:
c) entrenchment depth: 2.
d) average stream width:
e) mean depth water body: 0.
f) max depth water body: 0.

Dams number: 0
Dams cumulative height:
Shading at zenith

b,

Sampling locations

ECOSURV SAMPLING LOCATIONS

RWB CD: HU_AAA187
Name of the Waterbody: Balaton (Keszthelyi-öböl)
EOVY: 514031
EOVX: 153299
Nearest site:
Reference site ☐
Communal pollution ☐
Industrial pollution ☐
NBMR ☐
NATURA 2000 ☑
International calibration ☐
KOVIZIG:
Network: 01
Total area of watershed (ha):
Plougland: -999
Grapes, fruits: -999
Pasture: -999
Forest: -999
Settlement: -999
Water surface: -999

Rekord: 1 összesen 398

c,

Export : űrlap

Group select
☐ Macro Invertebrates
☐ MacroPhytes
☑ Phytobenton
☐ Phytoplankton
☐ Fish
☐ Chemistry

Row headings
○ Waterbody name
● Sample code
○ Sample code _date
○ Taxon name
○ Taxon code
○ Parameter name
○ AQEM
○ FAME

Value
○ Qvantitative
● Binary
○ Value
○ Pi
○ Abundant
○ Biomass
○ St 100

Coloumn headings
● Taxon name
○ Parameter name
○ Waterbody name
○ Sample code
○ Sample code _date

Water types
○ All
○ Lakes
● Running water

Select a specified tipology: 4
Select a specified cluster:
Number of selected sites: 20

Number	Waterbody	Nearest site	Date	analysis
ABL_279	Ablánc-patak	Acsád-Bükk kö	2005.04.13.	☐
ADO_265	Adonyi-Dunaág	Adony	2005.05.23.	☐
ADO_264	Adonyi-Dunaág	Szentendre alat	2005.05.23.	☐
ADO_263	Adonyi-Dunaág	Tahi híd felett	2005.05.22.	☐
ALP_235	Alpár-Nyárlőrinci-c	Kecskemét-Sze	2005.04.16.	☐
ALS_379	Alsó-Tápió	Tápióbicske-Tá	2005.05.05.	☐
ALT_281	Által-ér	Pusztavám fele	2005.04.07.	☐
ALT_047	Által-ér	10-es úti híd	2005.03.25.	☐

Rekord: 1 összesen 341

d,

Fig. 7. Selected forms from the ECOSURV database, a, Chemical sampling events form; b, hydromorphological information; c, Sampling locations form; d, Data export form.

Results

> *“although our results are in contradiction with the hypotheses and statistically not significant, but from the figures its clearly visible …”*

Paper 1.

The phytoplankton database of the Middle Danube Basin was analysed and evaluated in order to describe the characteristics algal assemblages of the rivers. The dataset were extracted from the database of the Hungarian monitoring network and academic institutions. The database which has been included in the analysis contains the results of 1897 phytoplankton investigation from 189 locations. Implementing the SOM method we can identify the different algal communities which characterise different river types.

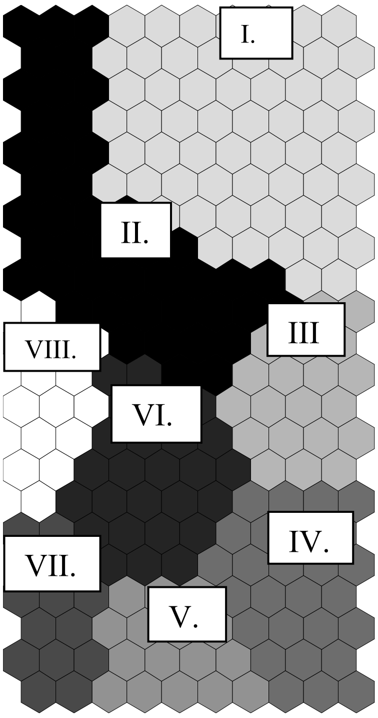


Fig. 8. Clusters of the SOM based on codon abundances

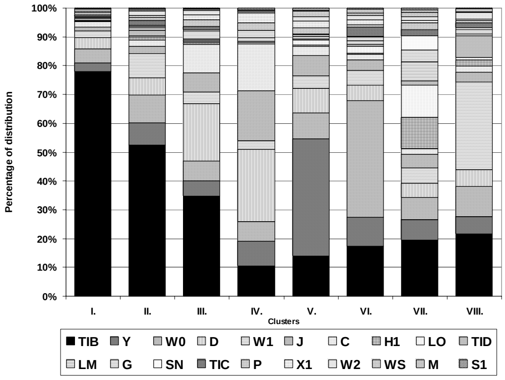


Fig. 9. The distribution of the different codons according to the SOM clusters, in percentage of the total biomass.

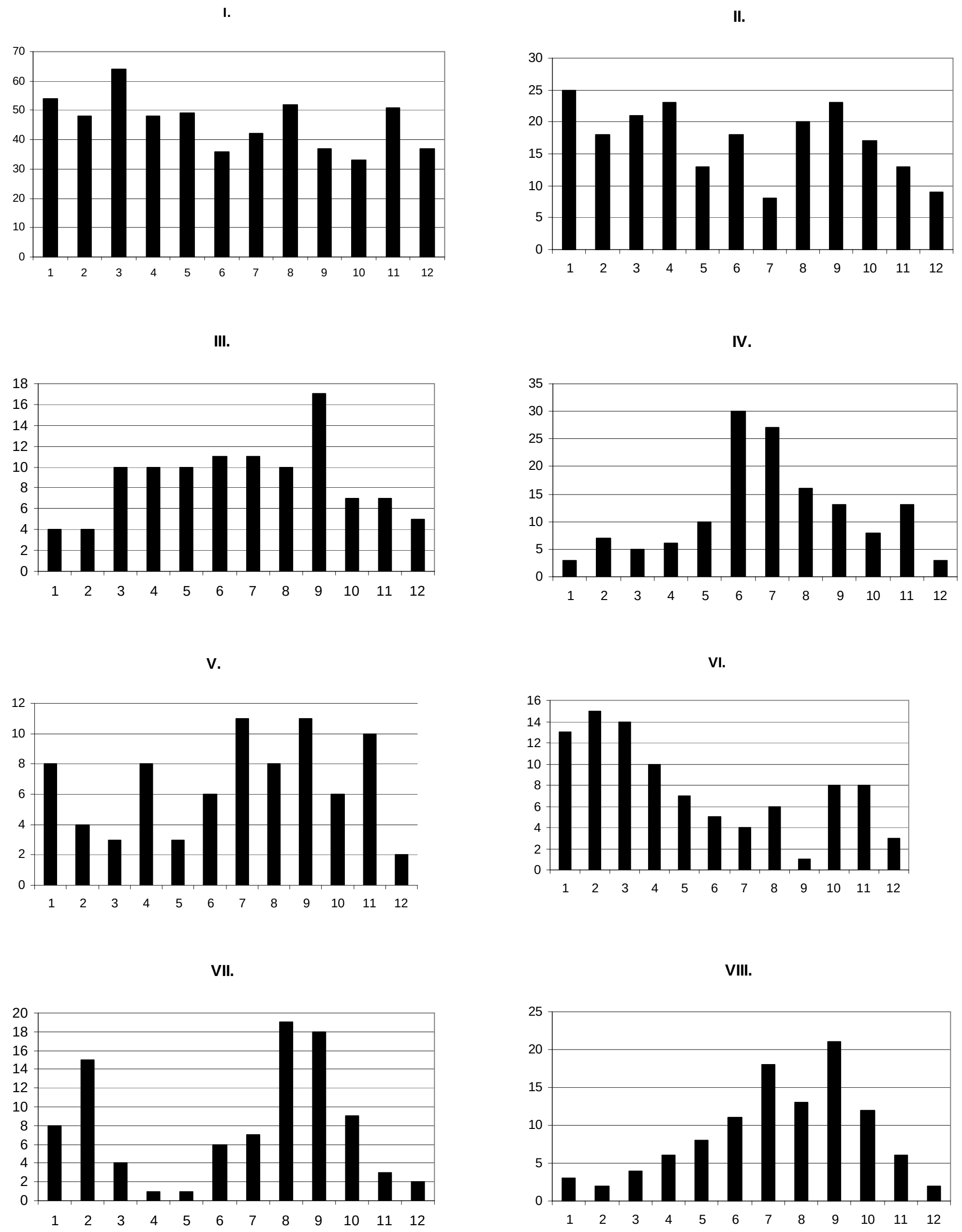


Fig. 10. Seasonal distribution of the selected clusters. Roman letters means the cluster numbers, vertical axis represents the months, while horizontal axis the number of samples belonging to that given month.

The algal communities were described as different ratios of algal functional groups. Since some of the groups are in close relation with certain types of environmental pressure it is also possible to highlight those rivers or river sections (or those periods) which are far from the expected good ecological status. We identified eight clusters or river types. (Fig. 8.)

Type I. This is the most frequent phytoplankton type which is representative for the upper section of rivers. The dominant codon is the TB (bentic diatoms) with more than 70 % of the total biomass abundance (Fig. 9*.)*. This association has no seasonal preference. The most frequent taxa are *Nitzschia palea*, *Nitzschia fonticola, Navicula capitatoradiata, Surirella brebissonii, Diatoma vulgaris*.

Type II. Members of the TB group are dominating type I1 (Fig. 9*.)* but the elements of the Wo and W1 functional groups (*Chlamydomonas reinhardtii, Euglena spp*.) indicate organic pollution. This type is also characteristic for upstream sections of rivers. Only a slight seasonal preference for early spring and autumn can be found.

Type III. This is typical plankton of small rivers with the dominance of TB and D (Fig. 9*.)* codons. This assemblage occurs on those river sections where the retention time due to hydro-morphological changes or upper stream reservoirs enables real phytoplankton to develop.

Type IV. This association can be referred to as "Danube type" summer plankton. The most important codons are D, J and C (Fig. 9*.)*. Dominant species of this association are *Stephanodiscus hantzschii, Cyclotella comta Skeletonema potamos* and *Nitzschia acicularis* from codon D, *Cyclotella meneghiniana* from codon *C,* and *Scenedesmus* species from codon J. However, the favourable season of this type is the early summer (Fig. 10). Certain species such as *Stephanodiscus invisitatus* could bloom however, in the Danube during winter (Kiss and Genkal, 1993).

Type V. This type could be mainly found on the lower section of the river Tisza with the dominance of Y codon (Fig. 9.). *Cryptomonas reflexa, C. marssonii, C. rostratiformis, Rhodomonas minuta* become dominant in this section. Development of this association is absolutely independent from the seasons. (Fig. 10.)

Type VI. The characteristic functional group of type VI is Wo (Fig. 9.). The dominance of this group is due to very strong organic pollution. *Chlamydomonas spp.*, *Euglena viridis, Polytoma uvella, Spermatozopsis exultans* are typical elements of this assemblage. This type of plankton is usually dominant in winter and early spring (Fig. 10.)

Type VII. This type is a mixture of a very divers association with the presence of relatively rare groups like Lo, HI, L„ S, (Fig. 9.). The occurrence of this type is expected in slow flowing channels and small rivulets in late winter and summer. The bimodal character of the type VII is cause by the Lo functional group. This plankton type frequently occurs in the summer epilimnion of lakes **(**Reynolds et al., 2002). Several dinophytes, however, can be important members of the winter phytoplankton **(**Grigorszky et al., 1998). Therefore it would be necessary to separate certain oligotherm *Peridinium* tax

Type VIII. The dominant codon of this type is the W1 (Fig. 9*.)* which is according to Reynolds characteristic for small organic pools. Frequent taxa are the metaphytic *Phacus* and *Trachelomonas* spp. This type is typical in slow-flowing rivulets and channels which are under the risk of organic pollution and have rich macrophyte vegetation. Development of this type is expected in summer.

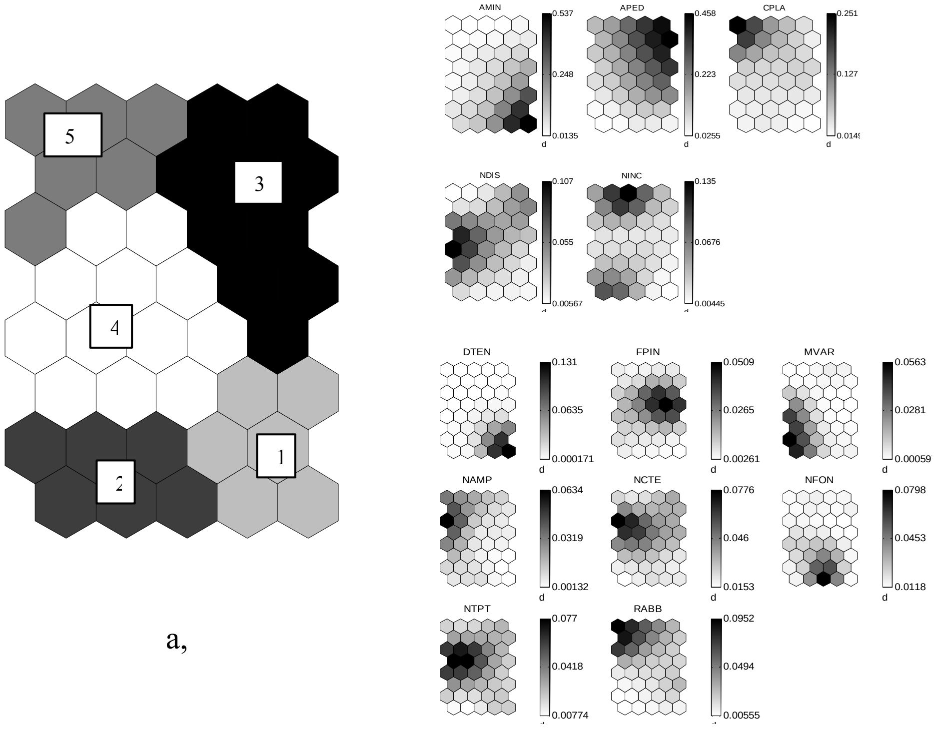
Paper 2.

In this paper the investigation of the benthic diatom flora of the River Danube is presented. The investigations include samples from the source streams to the end of the Hungarian stretch during a three years period. Characteristic diatom assemblages were established using SOM algorithm.

K-means clustering of the SOM allocated 5 different sample groups (clusters) (Fig. 11a). With few exemptions, the first, second and third clusters fell in the German-Austrian stretch, whereas the fourth and fifth clusters fell in the Slovakian-Hungarian stretch. Temporal differences were less pronounced: the five clusters were mainly organized by the sampling localities. The first cluster was formed by some sampling points scattered along the German and Austrian stretches, including the tributaries Enns and Lech. The second cluster comprised the sampling points around the source of the Danube such as the Breg and Brigach sources and streams, the first stretch of the River Danube after the fusion of Breg and Brigach, furthermore the tributaries Ipel, Morava and Garam. The third cluster was formed by the sampling sites on the lower part of the Austrian stretch. The fourth cluster was mostly built up by Slovakian, and the fifth by the Hungarian sampling sites, although the clustering is not perfectly unambiguous.

The character species of the five sample clusters were defined (Fig.11a.). We found that the clusters 1 and 2 were characterized by *Achnanthidium minutisimum* (AMIN), *Denticula tenuis* (DTEN), *Nitzschia fonticola* (NFON), *Psammothidium bioretii* (ABIO), *Melosira varians* (MVAR), *Planothidium subatomoides* (ASAT). The character species of cluster 3 is *Amphora pediculus*. (APED), while character species of the clusters 4 and 5 are *Nitzschia dissipata* (NDIS), *Navicula cryptotenella* (NCTE), *Cocconeis placentula* var. *euglypta* (CPLA), *Rhoicosphenia abbreviata* (RABB), *Nitzschia inconspicua* (NINC)(Fig. 11b.). However, not all of these species had equally high SI values. Thirteen species had higher SI values, these were *Amphora pediculus, Achanthidium minutissimum, Cocconeis placentula* *var. euglypta, Nitzschia inconspicua, N. dissipata, Denticula tenuis, Rhoicosphenia abbreviata, Navicula tripunctata* (NTPT)*, Navicula cryptotenella* (NCTE)*, Staurosirella pinnata* (FPIN)*, Nitzschia fonticola* (NFON)*, Melosira varians* and *Nitzschia amphibia* (NAMP).

Considering the average relative abundance values of the character species of the five cluster groups it is obvious that the main difference between the cluster groups is not so much the species composition, but rather the different relative abundance ratios of the different species in the different clusters. This arrangement is also well in agreement with the poorer water quality of the lower, and the better water quality of the upper stretch.



a,

b,

Fig. 11. a, Clusters of the sampling site by K-means clustering of the SOM. b, Individual SOM maps of the characteristic species. Scale bars showing the species calculated abundances by the learning process of the SOM in percentage. Abbreviations of the species can be found in the text.

Paper 3.

The periphyton of the Lake Velencei which is a diverse and dense community were investigated in this article. The main interest was to distinguish the differences of periphyton communities according to the different substrate type.

More than 120 algal species were found in this study, most of them are diatoms.

According to the SOM analysis based on algological data (species composition and their abundance), samples of different substrata formed two distinctive groups. Old reed (OR), andesite (A) and granite (G) were grouped together, respectively, and polycarbonate (PC) and green reed (GR) clearly separated from the other three in May (Fig. 12.). *Achnanthidium minutissimum* and *Navicula cryptotenella*, a pioneer colonist diatom species were found to be vigorously dominant on all substrata.

In November, based on the SOM analysis, granite and andesite come together with green and old reeds comprised one group, plexi and polycarbonate the other. These similarity was due to the presence of *A. minutissimum,* strongly dominating both plexi and polycarbonate substrata. Characteristic species were also *Fragilaria pulchella* and *Nitzschia perminuta* for this type of substrates, while for green and old reed a clear gradient of increasing of *Nitzshia palea* and *Gomphonema olivaceum var. olivaceum* can be viewed from the SOM component planes. (Fig. 13.)

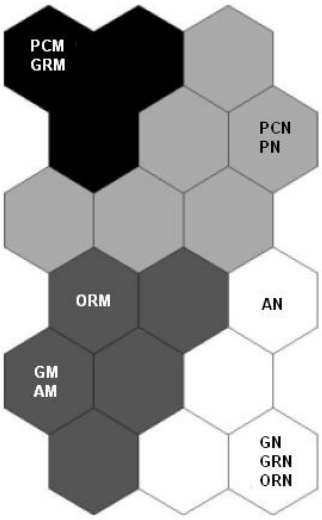


Fig. 12. Clustered SOM based on species composition, the last letter means the month(may=M, November=N) The first letter represents the substrate (old reed=OR, green reed=GR, granite=G, andezi t=A, polycarbonate=PC)

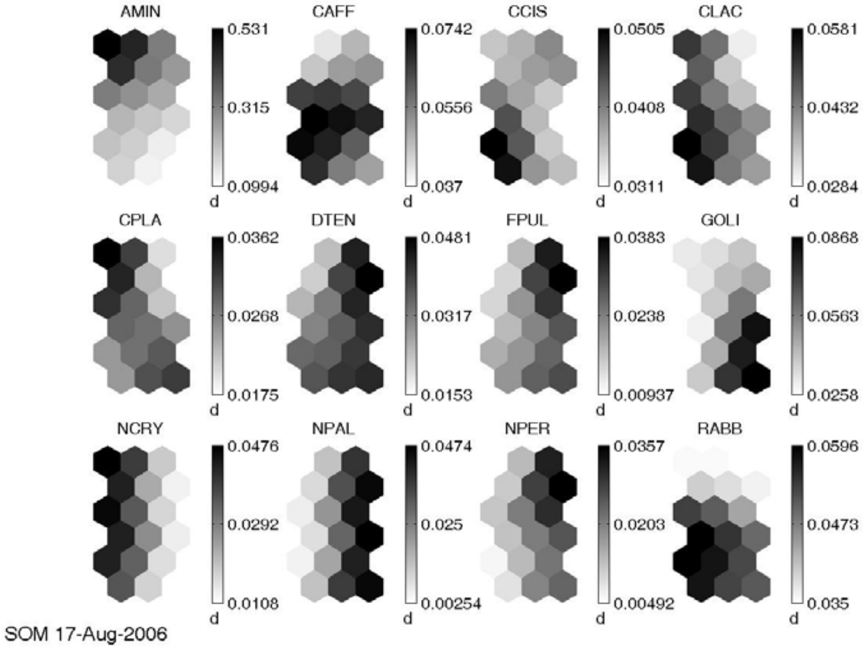


Fig. 13. The component planes of the SOM, the vertical bars represent species abundances in percentage. *Achnanthidium minutisimum* (AMIN), *Cymbella affinis var. affinis* (CAFF), *Cymbella cistula* (CCIS), *Cymbella lacustris* (CLAC), *Cocconeis placentula var. euglypta* (CPLA), *Denticula tenuis* (DTEN), *Fragilaria pulchella* (FPUL), *Gomphonema olivaceum var. olivaceum* (GOLI), *Navicula cryptocephala* (NCRY), *Nitzschia palea* (NPAL), *Navicula permitis* (NPER), *Rhoicosphenia abbreviate* (RABB).

Paper 4.

Monitoring of the naturally occurring algal communities in rivers provide data on species composition, species number, diversity, or quantitative occurrence of the algae. Experts of administrative institutions who are responsible for water quality management need simple numerical values rather than species lists or scientific evaluation of the assemblages. Since up to now, little attention has been attended to the application of the phytoplankton for evaluation of ecological state of the rivers, the present study aims to show a new assessment method. The method is based on the functional group of algae, described for lakes by Reynolds (Reynolds , 2002), represented in the potamoplankton and provides a single index number (Q). The index has been tested on phytoplankton data of different rivers, and proved to be more sensitive than the earlier used saprobic index.

To achieve an index, each species in the sample must be assigned to the appropriate functional group. Then the relative share of each functional groups are calculated. Relative shares are then multiplied by the factor number. The sum of these scores is the index. The reference values of the upper river sections are close to 5, while those of the lower river stretches are approximately 4. The method has been tested with hundreds of phytoplankton samples, it is simple, and after applying to a phytoplankton database can be computerised easily.

The new assessment method has been tested by using different datasets from Hungarian rivers that contain phytoplankton data against the Pantle-Buck index (Pantle & Buck, 1955), which is the officially accepted qualification method in Hungary and several other countries in Europe.

Paper 5.

In this study the results of the first nation wide survey of benthic river diatoms is presented. The dataset is robust as the primary sources of uncertainty in diatom training sets (choice of site and substrate for sampling, inter-operator differences in diatom taxonomy and counting techniques (Besse-Lototska et al., 2006) were eliminated as much as possible. Moreover, parallel with the diatoms, investigations were made on other biological quality elements, including macro-invertebrates and fish. In total, nearly 500 taxa were found in the 339 sites. This certainly does not include all diatom taxa in the inland waters of a country like Hungary, which probably has number between 1000 and 2000 'conservative' taxa. The present study provided a relatively homogeneous diatom data set for the whole country, which is lacking in most of the other European countries.

Many of the diatom taxa were very rare and occurred only in one or in a few samples. The commonest genera are *Navicula* (104 taxa), followed by *Nitzsclzia* (go), *Fragilaria* (43), *Achnanthes* (41), *Gomphonema* (30) and *Cymbella* (26). The most common taxa are listed in Table 2. The most common species, *Achnanthidium minutisimum,* is also the most common freshwater diatom species in the world. The majority of the other common taxa are characteristic of eutrophic, alkaline waters. Some of these are tolerant to organic pollution (e. g. *Gomphonema parvulum, Nitzschia paleacea),* others require cleaner waters (e. g. *Achnanthes biasolettiana, Gomphonema micropus). Meridion circulare* is one of the very few diatom species which are characteristic for running waters (Van Dam et al., 1994). Species from acidic or oligotrophic waters are rare in the dataset: the most common one is *Eunotia bilunaris.* The study shows that despite the intensive human use of over 90 % of the area of the country, natural variation is still the predominant factor responsible for the composition of river diatom assemblages.

The diatom database used in the paper was part of the ECOSURV database. The author of the theses was responsible for building the database structure as well as providing the relevant data matrixes. The structure and the connection between the tables involved in the analysis can be seen in Fig. 14.

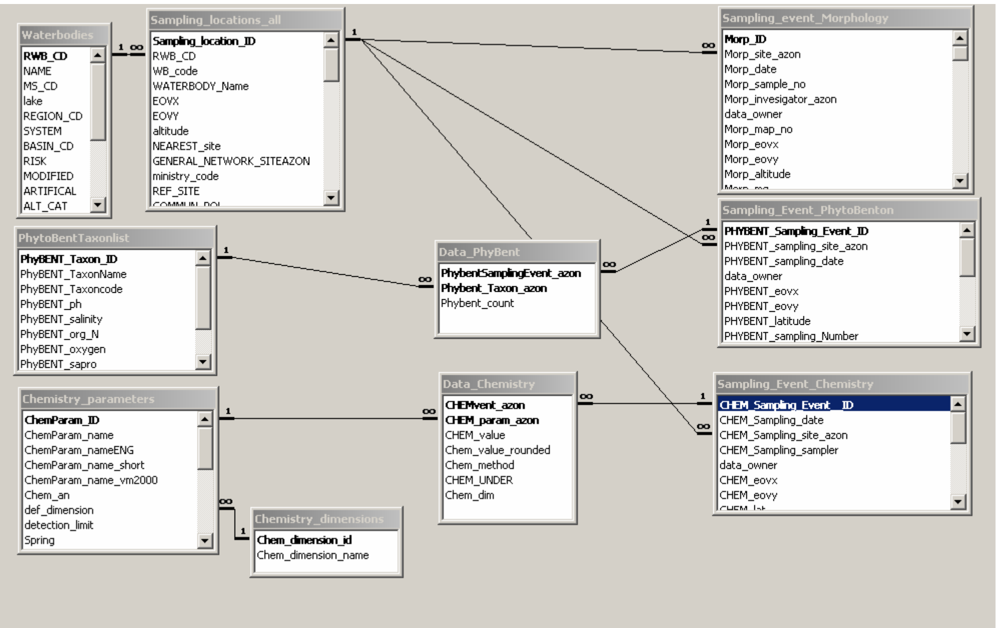


Fig. 14. The table structure of the diatom database used in the analysis in Paper 5. The variables on the left side considered as static whereas on the rights side the dynamic variables can be seen.

A better understanding of the relationships between diatom assemblages and environmental factors (including stressors) will be gained by a more time-intensive monitoring of selected stations in the present network.

Summary

The need for better techniques, tools and practices to analyse ecological systems within an integrated framework at wider scales has never been so great. The biological monitoring of the water currents in Hungary provided large volume datasets during the last years thanks to the EU Water Framework Directive (2000).

For easy handling the data of the monitoring it is essential to store them in an appropriate ecological database. This database should be capable of the storage of relevant static and dynamic variables from the ecosystems and should also allow the production of matrix inputs for ecosystem analysis. In case of large volume of data, the traditional multivariate statistical methods like cluster analysis and ordination are difficult to interpret and present the information. Self Organizing Map is a novel approach for the visualization of high-dimensional data.

In the theses I described in details the SOM algorithm as well as the basic structure of an ecological database. By combining the database structures and the SOM algorithm we were able to characterize frequent Hungarian running water types according to their characteristic phytoplankton assemblages.

I constructed the phytoplankton and phytobenton database which is presently in use in the environmental authorities, and is concerned as National Database.

By the application of SOM methods we characterized the upper Danube section based on their characteristic diatom communities, and described the differences and similarities of the periphyton of the Lake Velencei on different substrates.

Összefoglalás

A doktori értekezés előzményei és célkitűzései;

A 2001-ben bevezettet EU Vizkeretirányelv, melynek célja a felszíni és felszín alatti vizek jó ökológiai állapotának elérése 2015-re, több élőlénycsoport rendszeres monitorozását irja elő, mely tevékenység nagy mennyiségű ökológiai adatot eredményez. A rendszeres adatgyűjtés és az így keletkező adatok adatbázisba rendezése nemcsak a strukturált adatfeldolgozást teszi lehetővé, hanem az adatok kiértékelésében is nagy szerepet kap. A benne tárolt adatok élettartama nemcsak a vizsgálat idejére korlátozódik, így azok a későbbi elemzésekben is részt vehetnek.

Az ökológiai folyamatok tér és időbeli összetettsége, a fajok és a környezeti változók közötti nem lineáris kapcsolat és a változók normalitásának hiánya komoly kihívást jelent az adatfeldolgozásban és az ökológiai folyamatok feltárásában. A mesterséges neurális hálózatok vonzó alternatív eszközzzé váltak az ökológiai adatelemzés során, mert számos előnyös tulajdonsággal rendelkeznek, mint például a nem-linearitás, adaptivitás, általánosítás és modell függetlenség (nem szükséges a-priori modell). A neurális hálózatok alkalmazása, - bár már az 1950-s években megtörténtek a kezdeti lépések - csak az utóbbi évtizedekben terjedt el és vált az ökológusok által is használt statisztikai eszköztár részévé. A neurális hálózatok erős és hathatós segítséget nyújtottak az ökológiai modellezésben, osztályozási feladatok megoldásában és a mintázatelemzések során (Paisley et al., 2003).

A neurális hálózatokat számos hidrológiai és hidrobiológiai kutatásban sikerrel alkalmazták; így például a vízminőségi paraméterek modellezésében (e.g. Daniell and Wundke, 1993; Lachtermacher and Fuller, 1994; Schizas et al., 1994; Kaluli et al., 1998; Wen and Lee, 1998) vagy az élőlényközösségek és a környezeti változók közötti összefüggések vizsgálatában (e.g. Chon et al., 1996; Lek et al., 1996; Recknagel et al., 1998; Guegan et al., 1998; Lee et al., 1998; Maier et al., 1998).

Nagymennyiségű adat elemzésénél a hagyományos többváltozós elemzések számos esetben nehezen értelmezhetők és eredményik nem jeleníthetők meg közérthető formában. A Kohonen-féle Önszervező Térkép Módszer egy olyan neurális hálózati típus, amely lehetővé teszi nagymennyiségű, összetett, (több dimenziós), adatok elemzését és értékelését. A Teuvo Kohonen által kifejlesztett módszer a dimenzió csökkentése révén a kiindulási adatmátrixot kétdimenziós hexagonális térképre képezi le, megőrizve a kiindulási adatok topológiai és metrikai sajátságait (Kohonen 2001). A módszer legfőbb erénye az adatok szemléletes megjelenítése (Vesanto and Alhoniemi 2000).

Anyag és módszer

A statisztikai elemzések elvégzéséhez az önszervező térkép (SOM) algoritmust használtuk. Az algoritmus lehetővé teszi nagy adatbázisok sokdimenziós adatainak leképezését kétdimenziós hexagonális térképpé. Az egymáshoz közeli térkép egységek (neuronok) úgy helyezkednek el a végső térképen, hogy az egymáshoz hasonló tulajdonsággal rendelkező neuronok egymás közelében, míg a különbözőek a térkép távolabbi pontjain helyezkednek el. A SOM a mintákat és a változókat egyidejüleg is képes megjeleníteni, s ugyanakkor a változók fontosságát is meghatározhatjuk. (Kohonen, 2001). Az értékelés során a SOM Toolbox modult (http://www.cis.hut.fi/projects/somtoolbox) és a MATLAB™ fejlesztői környezetet használtuk.

A SOM neurális hálózat a nem-ellenőrzött tanulási típusú hálozatok kategóriájába tartozik, ami azt jelenti, hogy nem adunk meg kimeneti paramétert, a csoportok képzését az algoritmus önnállóan végzi. Maga a hálózat egy bemeneti (input) és egy kimeneti (output) rétegből áll. A bemeneti réteg tartalmazza a mintaegységeket (SU), a kimeneti réteg a virtuális egységekből (VU) vagy neuronokból áll. A kimeneti réteg neuronjai a bemeneti réteg súlyvektorainak megfelelő mintázatot vesznek fel egy tanulási folyamat során. Ez ökológiai analógiára lefordítva annyit tesz, hogy a kiindulási mintavételekhez (SU) tartozó faj-abundancia viszonyokra jellemző lesz a végső SOM térkép neuronjainak súlyfaktora, azaz minden egyes neuron egy egy virtuális faj abundancia listával fog rendelkezni. Az eljárás kellően robosztus, lehetővé teszi több ezer mintavételi egység sűrítését néhány virtuális neuronba. A végső SOM térkép leggyakoribb ábrázolási módja a hexagonális rács (Fig. 2.).

A SOM algoritmusát az alábbik szerint adhatjuk meg:

- 1. lépés: Ciklus t=0, a virtuális egységet (VU_p) inicializáljuk, a súlyokat (*w*) az input adatokból töltjük fel.
- 2. lépés: Egy mintaegységet kiválasztunk (SUj) véletlenszerűen
- 3. lépés: Egy megfelelő metrika segítségével meghatározzuk a SU_j és az összes VU_p közötti távolságot.
- 4. lépés: A SU_j-hez legközelebbi VU_c lesz a győztes neuron, ez a neuront Legjobban hasonlító egységnek (BMU) nevezzük
- 5. lépés: A virtuális egységek VU_p súlyfaktorai a tanulási szabály alapján módosulnak, úgy, hogy az új faktorok a SU_j bemeneti vektorához hasonlítsanak.
- 6. lépés: Újabb ciklus kezdődik t+1, amíg $t < t_{max}$ addig 2. lépésre lép; a ciklus akkor fejeződik be, ha eléri a kívánt számú ismétlést .

Az első lépésben a felhasználó dönti el a térkép méretét. A kimeneti réteg neuronjainak oiptimális száma az alábbiak szerint számolható (Vesanto, 2000)

$$Nn = 5\sqrt{SU_d}$$

Ahol *Nn* a neuronok száma a végső térképen és SU_d a mintavételi egységek száma. A 3. lépésben a leggyakrabban használt távolság metrika az Euklideszi.

Az 5. lépésben Euklideszi tanulási szabály $h_{ck}(t)$ változója a szomszédsági függvény. A 4. lépésben kiválasztott BMU neuron nem az egyetlen, amelynek vektorait módosítjuk, hanem a szomszédos neuronok is módosulnak. A leggyakrabban használt szomszédsági függvény a Gauss féle:

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

$||r_k - r_c||$: az Euklideszi távolság a győztes neuron VU_c és a térkép többi neuronja VU_k.között,

σ: időben csökkenő faktor azt határozza meg, hogy a térkép mekkora része érintett a tanulási folyamatban,

α: a tanulási faktor. Mindkét faktor idővel csökken, és nullához konvergál.

A SOM jellemzésére és a karakterfaj elemzések során a strukturális indexet (SI) használtuk. Az index azt mutatja meg, hogy az egyes változók (súlyfaktorok) milyen mérétékben felelősek a SOM térkép topológiájáért (Park et al., 2005). Azok a változók (esetünkben fajok) amelyek nagyon erős grádiens eloszlást mutatnak, nagy SI index értékkel rendelkeznek, karakterfajoknak tekinthetők. Az SI értékét az alábbiak szerint számolhatjuk:

$$SI_i = \sum_{j=1}^{S} \sum_{k=1}^{j-1} \frac{|\ W_{ij} - W_{ik}\ |}{\|r_j - r_k\ \|}$$

ahol w_{ij} és w_{ik} az ith faj súlyfaktorai,

$||rj-rk||$ a topológiai távolság a térkép két j és k neuronja között,

S a SOM térkép neuronjainak száma.

Az algoritmus eredményeképpen kapott térkép neuronjait a kiindulási mintaegységekkel megfeleltethetjük és ábrázolhatjuk. Bár az egymáshoz hasonló minták egymás közelében helyezkeden el, a mintacsoportok elkülönítése nem minden esetben egyértelmű.

A mintacsoportok elkülönítésére számos lehetőségünk van. Használhatunk hagyományos klaszterezési technikát. Mi az elemzések során a szakirodalom által ajánlott *k*-átlag módszert használtuk. A megfelelő clusterszám meghatározása a Davies-Bouldin indexet felhasználásával történt (Davies and Bouldin, 1979)

Az **ökológiai adatbázis** az adatok illetve rekordok stukturált összessége. Napjainkban a leggyakrabban használt adatbázis modell a relációs adatmodell. A reláció nem más, mint egy táblázat, a táblázat soraiban tárolt adatokkal együtt. A relációs adatbázis, pedig a relációk összessége. Ennek alapstruktúrája az adattábla, amely sorokból (rekordok) és oszlopokból (mezők) áll. A relációk vagy kapcsolatok az adatbázis táblái között állnak fenn. A táblák oszlopai vagy mezői az adatok tulajdonságai, pl. mintavételi hely neve, mintavétel dátuma, a sorok vagy adatok pedig a konkrét értékeket taralmazzák pl. 2009.05.27., Concó, Ács

A relációs adatbázis előnye, hogy minimalizálja a tárolt redundáns információt, ezáltal csökkenti az adatbázis méretét és az egyszerre kezelt adatok mennyiségét, valamint a kapcsolatok révén bármely mező szerinti lekérdezét ill. csoportosítást lehetővé tesz.

Eredmények

A disszetáció öt tudományos közleményt tartalmaz. A disszertációban részletesen bemutattam a SOM algoritmusát, valamint az ökológiai adatbázis szerkezetét. A megfelelő adatszerkezet és a SOM statisztikai módszer segítségével lehetővé vált magyarországi folyóvíz típusok jellemzése a fitoplakton közösségek alapján.

Létrehoztam egy fitoplakton és egy fitobenton adatbázist amelyet jelenleg is használnak a Környezetvéfelmi Felügyelőségek és ezek részei az országos monitoring hálózatnak is.

A SOM módszer segítségével jellemeztem a felső Duna szakasz diatóma közösségeit és a Velencei tavi, különböző aljzaton előforduló perifitonjának sajátosságait.

Folyóvízi fitoplankton közösségek

Use of Self-Organizing Maps (SOM) for characterisation of riverine phytoplankton associations in Hungary, (Várbíró et al., 2007) 1.

A célkitűzésünk az volt, hogy a magyarországi folyóvizek jellemző fitoplakton közösségeit a meghatározzuk és jellemezzük. Az adatsor a magyarországi monitoring hálózat adatait tartalmazta és ezen adatok adatbázisba történő rendezése után vált értékelhetővé.

Az elemzésben használt adatbázis 1897 mintavétel adatait tartalmazta 189 mintavételi hely esetében. A SOM módszer segítségével nyolc folyóvíztípust lehetett elkülöníteni. Minden típusra meghatároztuk a domináns fitoplakton közösséget és jellemeztük az egyes csoportokat. A fitoplakton taxonokat a Reynolds féle kodonbesorolás alapján elemeztük (Reynolds, 2002). Néhány csoport szoros összefüggést mutatott a környezeti változókkal és lehetővé vált azon folyószakaszok illetve fitoplakton közösségek meghatározása, amelyek nem megfelelő ökológiai állapotot jeleznek.

Az általunk meghatározott folyóvízi fitoplankton típusok (2. ábra):

I. típus A folyók felső szakaszaira legjellemzőbb típus a TIB kodon (bentikus kovaalgák), gyakran 100% abundanciaértékkel. Leggyakoribb taxonok a *Nitzschia palea*, *Nitzschia fonticola, Navicula capitatoradiata, Surirella brebissonii, Diatoma vulgaris*.

II. típus Ebben a típusban is a TIB csoport dominál, de jelentős arányban jelen vannak a szerves szennyezésre utaló W0 és W1 funkcionális csoportok fajai (*Chlamydomonas reinhardtii, Euglena spp.*).

III. típus Ez a típus a kisebb folyókra jellemző, ahol a víz tartózkodási ideje, hidromorfológiai beavatkozások miatt megnő és lehetővé teszi valódi plankton kialakulását. Jellemző kodonok a TIB és a D csoportok.

IV. típus Ez a típus a Dunára jellemző plaktontípus. Jellemző kodonok a D, J és a C. Domináns fajok a D kodonba tartozó *Stephanodiscus hantzschii, Cyclotella comta Skeletonema potamos* és *Nitzschia acicularis*, a C kodonba tartozó *Cyclotella meneghiniana* és a J kodonba tartozó *Scenedesmus* fajok.

V. típus Ebbe a típusba a Tisza alsó szakasza tartozik. A domináns kodon az Y, jellemző fajok a *Cryptomonas reflexa, C. marssonii, C. rostratiformis.*

VI. típus Ebbe a típusba a szerves szennyezéssel terhelt vizek tartoznak. A terhelést jellző W0 kodon dominál, jellemző fajok *Chlamydomonas spp.*, *Euglena spp., (Polytoma uvella, Spermatozopsis exultans.)*

VII. típus A lassú folyású kis csatornák és erek jellemző asszociációja. Több relatívan ritka kodon fordul elő benne, például L0, HI, L, S. Szezonálisan bimodalis eloszlás jellemzi, azaz a nyári és a téli vegetációs periódusban is dominánsá válhatnak a H1 másrészt az L0 (Dinofita) funkcionális csoportok fajai.

VIII. típus A W1 kodon dominanciája jellemzi ezt a típust, amely Reynolds szerint a kis szerves tavak jellemző plantkonja. Gyakori taxonok a metafitikus *Phacus* és *Trachelomonas* fajok. Ez a típus a dús növényzettel benőtt csatornákra jellemző, tipikusan a nyári vegetációs periódusban előforduló asszociáció.

A Duna bentikus kovaalga flórája

> *Use of Kohonen Self Organizing Maps (SOM) for the characterisation of benthic diatom associations of the River Danube and its tributaries. (Várbíró et al., 2007)*

Ebben a cikkben a Duna felső szakasza, bentikus kovaalga flórájának bemutatása történt meg.

A mintavételi helyek a Duna forrásától a Magyarországi szakaszának végig terjedtek és három éves periódust öleltek fel. A jellemző kovaalga közösségeket SOM módszer segítségével határoztuk meg. Az elkészült SOM térképet *k*-átlag klaszterezési eljárással 5 klaszterre különítettük el. Néhány kivételtől eltekitve az első három klasztercsoportba a Német-Osztrák Duna-szakasz került, a 4. és 5. csoportok a Szlovák-Magyar szakaszra estek. A mintavételek időbeni különbsége nem volt számottevő. A klaszterek a mintavételi helyek szerint alakultak ki. Az 1-2. klaszter karakterfajai, melyet a strukturális index segítségével határoztunk meg: *Achnanthidium minutisimum*, *Denticula tenuis*, *Nitzschia fonticola*, *Psammothidium bioretii*, *Melosira varians*, *Planothidium subatomoides*. A 3. klaszter karakterfaja az *Amphora pediculus*, míg a 4-5. klaszteré a *Nitzschia dissipata*, *Navicula cryptotenella*; *Cocconeis placentula var. euglypta, Rhoicosphenia abbreviata, Nitzschia inconspicua* voltak.

Aljzat szerepe kovaalga közösségek kialakulásában

Comparative algological and bacteriological examinations on biofilms developed on different substrata in a shallow soda lake (Ács et al.,2008)

A felmérés a Velencei tó gazdag perifiton közösséinek vizgálatára irányult. Arra a kérdésre kerestük a választ, hogy meghatározzuk az eltérű aljzatokon előforduló kovalaga közösségek összetételét.

A SOM analízis alapján a kovalga közzösségek abundancia viszonyai a májusi mintavételek esetén két csoportra különíthetők el. Az avas nád, az andezit és a gránit aljzatok alkottak egy csoportot, a zöld nád és a polikarbonát pedig egy másikat. Az *Achnanthidium minutissimum* és *Navicula cryptotenella*, mint pionír kovaalga kolonista fajok, erőteljesen meghatározók voltak valamennyi aljzat esetén.

Novemberben a SOM analízis alapján az avas és a zöld nád az andezit és gránit aljzatal alkotott egy csoportot, míg a polikarnonát a plexi aljzattal. Ez a hasonlósága az A*chnanthidium minutissimum,* fajnak volt köszönhető, ami az utóbbi két szbsztráton dominált. További jellemző fajok voltak még a *Fragilaria pulchella* és a *Nitzschia perminuta.* A zöld és avas nád jellemző bevonatalgái a SOM komponens térképek alapján a *Nitzshia palea* és a *Gomphonema olivaceum var. olivaceum* voltak.

Folyóvizek fitoplakton alapú minősítése

A new evaluation technique of potamoplankton for the assessment of the ecological status of rivers.(Borics et al., 2007)

A folyóvizek fitoplaktonjának monitorozása során a természetes algaközösségek fajszámáról, fajösszetételéről, diverzitásáról nyerünk információt. Azon szakemberektől akik, a vízminőség védelem területén dolgoznak nem várható el, hogy tudományos részletességű eredményekből és fajlistákból hámozzák ki a megfelelő információt. Számukra jól értellmezhető, egyszerű indexeket kell kidolgozni.

Egészen napjainkig kevés sikeres próbálkozás irányult arra, hogy a folyóvizek ökológiai állapotát a fitoplankton alapján értékeljük. Jelen tanulmányukban egy olyan értékelő rendszert mutatunk be, amely a folyóvizek algaflóráján alapul.

A módszer az eredetileg állóvizekre kidolgozott fitoplakton funkcionális csoprtokon alapul (Reynolds, 2002). Az értékelés eredményeként egyetlen számmal, Q index, jellemezzük a vizek ökológiai állapotát. Az indexet különböző magyarországi víztesten teszteltük és jóval érzékenyebbnek bizonyult, mint a korábban használt szaprobitási index.

Magyarországi felszíni vizei ökológiai állapotának áttekintése bentikus kovaalga közösségek alapján

Implementation of the European Water Framework Directive: Development of a system for water quality assessment of Hungarian running waters with diatoms (Van Dam et al., 2007)

Ebben a cikben az első országos kovalga felmérés eredményei kerültek bemutatásra. Az adatsor megfelelően robosztus volt, mert mind a mintavételi, mind a mintafeldolgozási, határozáési és számolási hibákat a lehető legminimálisabra csökkentettük.

Ezen felül a diatoma vizgálatokkal párhuzamosan más, a Vízkeretirányelv hatálya alá tartozó biológiai minpőségi elem (halak, makroszkípikus vízi gerinctelenek, fitoplakton, makrofita) vizsgálatára is sor került, valamint kémiai és hidromorfológiai felmérés is készült. Közel 359 kovalga taxon került elő a vizsgált 339 mintavételi helyről. A leggyakoribb genus a *Navicula* (104 db faj) volt , amit a követett a *Nitzsclzia* (80), *Fragilaria* (43), *Achnanthes* (41), *Gomphonema* (30) és *Cymbella* (26). A leggyakrabban előforduló taxon az *Achnanthidium minutisimum* volt, amely a leggyakroibb diatoma taxon a világon. A legtöbb előkerült taxon a meszes, eutróf vizekre jellemző. Számos a szennyezésre toleráns faj is előfordult (pl. *Gomphonema parvulum, Nitzschia paleacea),* de a szennyezéstől mentes vizeket jelző fajok is előkerültek (pl. *Achnanthes biasolettiana, Gomphonema micropus).* A *Meridion circulare* a kevés tipikusan folyóvízre jellemző kovaalga egyike, szintén jellemző volt a mintákban (Van Dam et al., 1994). A savas és oligotróf vizekre jellemző fajok ritkán kerülte elő pl. *Eunotia bilunaris.* Az adatelemzésben felhasznált kovalaga adatbázis az ECOSURV project adatbázisának része.

Ezen adatbázis képezi az alapját a jelenleg az országos monitoring hálózat által használt minősítő és adattároló rendszereknek.

Acknowledgements

First of all I am deeply grateful to my wife Anikó and children Nándi and Danó, my Parents and Family, thanks for your neverending support, patience and love, these would have never been done without you.

I am indebted to a number of people for their invaluable help and support, most of all, I am grateful to my supervisors Professor Bela Tóthmérész for his professional guidance and Gabor Borics for his great help, valuable scientific discussions and friendship.

I have to thank in depth to Janos Attila Tóth, whom I started my scientific carrier with, at fisrt as a university, then as a PhD student. By his guidance I met a lot of inspiring people. Thanks to his help I was involved in the scientific community and to be participated in international projects and scholarships.

I wish to thank Éva Ács for introducing me the SOM method and for providing the neverending amount of data for analysis.

I have to thank to Jean Pierre Garrec for his inspiring cooperation and friendship and for the warm welcome during my short but Determinative visit in Nancy, which was one of a turning point in my life and Cristine for the great culinary experiences.

Special thanks to István Grigorszky, Csaba Balogh, Enikő Krasznai, Csaba Deák and Muci for their support, friendship and inspiring discussions.

In the end I would like to thank to Ágnes Boronkai for her guidance in my life to remain and the right path …

Köszönetnyílvánítás

Mindenekelőtt hatalmas köszönettel tartozom feleségemnek Anikónak, gyermekeimnek Nándinak és Danónak, Szüleimnek és Családomnak végtelen Türelmükért, Támogatásukért és Szeretetükért, mert nélkülük soha nem született volna meg ez a dolgozat.

Köszönöm témavezetőimnek Tóthmérész Bélának a kiváló szakmai és tudományos iránymutatásért, Borics Gábornak, az óriási segítségért, tanácsaiért és barátságáért, mely hozzájárult ahhoz, hogy a dolgozat és a szakmai cikkek megszülessenek.

Hálás köszönettel tartozom Tóth János Attillának, akinél elkezdtem tudományos pályafutásomat először, mint egyetemi diák, majd szakdolgozó, később mint PhD hallgató. Általa sok olyan embert ismertem meg, akik a pályám során meghatározó szerepet játszottak. Segítségével sikerült bekapcsolódnom a nemzetközi tudományos közéletbe, részese lehettem számos pályázatnak, ösztöndíjnak.

Szerettel és tisztelettel szeretném megköszönni Ács Évának, hogy ráirányította figyelmemet a SOM módszerre, és folyamatosan ellátott elemezésre váró adatok tömegével.

Külön szeretném megköszönni Jean Pierre Garrec szakmai iránymutatásait, és azt a szívélyes fogadtatást amiben rövid de annál meghatározóbb Nancy-i látogatásaim során részem volt.

Szintén külön köszönet illeti Grigorszky Istvánt, Krasznai Enikőt, Deák Csabát és Koncz Csabánét (Mucit) a támogatásukért , barátságukért és az inspiráló vitákért.

Köszönöm munkatársaimnak és vezetőimnek Kelemen Bélának, Balogh Csabának és Dr. Pádár Istvánnak, hogy a dolgozat elkészítéséhez lehetőséget biztosítottak.

Végül, pedig szeretném megköszönni Boronkay Ferencnének, hogy életem és pályám során mindig segített az úton maradni…

References

ARCADIS 2005. Ecosurv: final technical report, part A. Europe Aid 114951/D/SV/2002- 000-180-04-01-02-02. - 48pp. Ministry of Environment and Water, Budapest.

Beccali, M., Cellura, M. L., Brano, V., Marvuglia, A. 2004. Forecasting daily urban electric load profiles using artificial neural networks. - Energy Conversion and Management **45:** 2879-2900.

Besse-Lototskaaya, V., Verdonschop, T. F. M., Sinkeldam, J. A. 2006. Uncertainty in diatom assessment: sampling, identification and counting variation. – Hydrobiologia 566: 247-260.

Buckeridge, J.S. 2001. Ethics, Environment and Culture: Their Significance within Engineering Education. International Ecological Engineering Conference International Ecological Engineering Society (IEES) and New Zealand Engineers for Social Responsibility (ESR), Lincoln University, 26e29 November 2001

Burke, L.I., Ignizio, J.P., 1992. Neural networks and operations research: an overview. Computer and Operations Research 19 (3/4), 179–189.

Cereghino, R., Park, Y. S. 2009. Review of the Self-Organizing Map (SOM) approach in water resources: Commentary Environmental Modelling & Software 24 945–947

Chon, T.-S., Park, Y.S., Moon, K.H., Cha, E.Y. 1996. Patternizing communities by using an artificial neural network. Ecol. Model. 90, 69–78.

Clark, J.S., Carpenter, S.R., Barber, N., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke Jr., R., Pizer, W., Pringle, C., Reid, W.V., Orse, K.A., Sala, O., Schlesinger, W.H., Wall, D.H., Wear, D. 2001. Ecological forecasts: an emerging imperative. Science 293 (5530), 657e660. The American Association for the Advancement of Science.

Codd, E.F. 1970. A Relational Model of Data for Large Shared Data Banks. Communications of the ACM 13 (6): 377–387

Davies, D.L., Bouldin, D.W. 1979. A cluster separation measure. 1979. IEEE Trans. Pattern Anal. Machine Intell. 1 (4). 224-227.

Daniell, T.M., Wundke, A.D. 1993. Neural networks—assisting in water quality modelling. Preprints, Watercomp, Melbourne, Australia, pp. 51–57.

European Union 2000. Directive 2000160/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. - PE-CONS 363911100, REV 1, ENV 221, CODEC 513: 1-152. Brussels.

Giraudel, J.L., Aurelle, D., Lek, S. 2000. Application of the self-organizing mapping and fuzzy clustering microsatellite data: how to detect genetic structure in brown trout (*Salmo trutta*) populations. In: Lek, S., Gueguan, J.F. (Eds.), Artificial Neuronal Networks: Application to Ecology and Evolution, Environmental Science. Springer, Berlin.

Giraudel, J.L., Lek, S. 2001. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. Ecological Modelling 146 (1e3), 329e339.

Gosselain, V., Champeau, S. Gevrey, M., Coste, M., Ector, L., Park, Y. S., Lek, S., Descy, J. P. 2005. Diatom typology of reference situations at a large multi-regional scale: combined results of multivariate analysis and SOM. - In: Lek, S., Scardi, M., Verdonscho, P., Park, Y. S., Descy, J. P. (eds.) Modelling Community Structure in Freshwater Ecosystems. Springer-Verl.

Grigorszky, I., Padisak, J. 1998. Jég alatti *Peridinium aciculiferum Lemmermann* (Dinophyta) populáció a Balatonban [*Peridinium aciculiferum Lemmermann* (Dinophyta) under-ice population from the shallow lake Balaton, Hungary]. – Hidrológiai Kozlony 78: 282-284. [in Hungarian with English summary]

Guegan, J.-F., Lek, S., Oberdorff, T. 1998. Energy availability and habitat heterogeneity global riverine fish diversity. Nature 391, 382–384.

Hebb, D. O. 1949 The Organization of Behavior Wiley, New York

Giraudel, J.L., Lek, S. 2001. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination Ecological Modelling 146, 329–339

Jowett, I. 2001. Artificial intelligence and river biomonitoring. Water & Atmosphere 9 (1), 6.

Kalteh, A.M., Hjorth, P., Berndtsson, R. 2008. Review of the self-organizing map (SOM) approach in water resources: analysis, modelling and application. Environmental Modelling and Software 23, 835–845.

Kaluli, J.W., Madramootoo, C.A., Djebbar, Y. 1998. Modelling nitrate leaching using neural networks. Water Sci. Technol. 38 (7), 127–134.

Kaski, S., Kangas, J., & Kohonen, T. 1998. Bibliography of Self-Organizing Map (SOM) papers: 1981–1997. Neural Computing Surveys, 1, 102–350.

Kohonen, T. 2001. Self-organizing Maps. 3rd ed. - Springer, Berlin

Kiss, K.T. & Genkals, I. 1993. Winter blooms of centric diatoms in the River Danube and in its side arms near Budapest. -In: H. Van Dam (ed.): Twelfth International Diatom Symposium. Hydrobiologia 2691270: 317-325. Kluwer Academic Publ.

Lachtermacher, G., Fuller, J.D. 1994. Backpropagation in hydrological time series forcasting, in stochastic and statistical methods. In: Hipel, K.W., McLeod, A.I., Panu, U.S., Singh, V.P. (Eds.): Hydrology and Environmental Engineering. Kluwer Academy, Norwell, MA, pp. 229–242.

Lee, H.-L., DeAngelis, D., Koh, H.-L. 1998. Modelling spatial distribution of the unionid mussels and the core-satellite hypothesis. Water Sci. Technol. 38 (7), 73–79.

Lek, S., Belaud, A., Baran, P., Dimopoulos, I., Delacoste, M. 1996. Role of some environmental variables in trout abundance models using neural networks. Aquat. Living Resour. 9, 23–29.

Maier, H.R. 1995. A review of artificial neural networks. Research Report No. R131. Department of Civil and Environmental Engineering, The University of Adelaide, Adelaide, pp. 94.

Maier, H.R., Dandy, G.C. 1996a. Neural network models for forecasting univariate time series. Neural Network World 5, 747–771.

Maier, H.R., Dandy, G.C. 1996b. The use of artificial neural networks for the prediction of water quality parameters. Water Resources Research 32 (4), 1013–1022.

Maier, H.R., Dandy, G.C. 1997. Modelling cyanobacteria (blue-green algae) in the River Murray using artificial neural networks. Mathematics and Computers in Simulation 43, 377–386.

Maier, H.R., Dandy, G.C., Burch, M.D. 1998. Use of artificial neural networks for modelling cyanobacteria *Anabena* spp. in the River Murray, South Australia. Ecol. Model. 105, 257–272.

Mangiamelien, P., Chen, S. H. K., West, D. 1996. A comparison of SOM neural network and hierarchical clustering methods. - Europ. J. Operational Res. 93: 402-417.

McCulloch, W.S. Pitts, W (1943): A logical calculus of the ideas immanent in nervous activity Bull. Math. Phys. 5: 115-133.

Onkal-Engin, G., Demir, I., Engin, S.N., 7 July 2005. Determination of the relationship between sewage odour and BOD by neural networks. Environmental Modelling & Software 20, 843e850.

Orloci, L. 1978. Multivariate Analysis in Vegetation Research. W. Junk, The Hague.

Paisley, M.F., Walley, W.J., Nikhade, J., Dils, R. 2003. Identification of the key biological indictors of nutrient enrichment in rivers for use in predictive/diagnostic models. Diffuse Pollution Conference, Dublin 2003 8A Ecology.

Pantle, R., Buck, H. 1955. Die biologische ijberwachung der Gewasser und die Darstellung der Ergebnisse. - Gas- u. Wasserfach 96: 604.

Park, Y.-S., Cereghino, R., Compin, A., Lek, S. 2003. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. Ecological Modelling 160, 165–280.

Park, Y.-S., Gevrey, M. Lek, S., Giraudel, J. 2005. Evaluation of relevant species in communities: Development of structuring indices for the classification of commuuities using a self- organizing map. - In: Lek, S., Scardi, M., Verdonscho, P., Park, Y. S., Descy, J. P. (eds.): Modelling Community Structure in Freshwater Ecosystems: 369-380. Springer, Berlin.

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

Recknagel, F., French, M., Harkonen, P., Yabunaka, K.-I. 1997. Artificial neural network approach for modelling and prediction of algal blooms. Ecol. Model. 96, 11–28.

Recknagel, F., Fukushima, T., Hanazato, T., Takamura, N., Wilson, H. 1998. Modelling and prediction of phyto- and zooplankton dynamics in Lake Kasumigaura by artificial neural networks. Lakes Reserv. Res. Manag. 3, 123–133.

Reynolds, C., Huszar, V., Kruk, C., Nasseli-Florel, S., Melo, S. 2002. Towards functional classification of the freshwater phytoplankton. - J. Plankton Res. 24: 417- 428.

Rosenblatt, F. 1957. The perceptron a perceiving and recognizing automation project Cornell Aeronautical Laboratory Report 85-400-1. 1957

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics, 27:832–835.

Schizas, C.N., Patticheς, C.S, Michaelides, S.C. 1994. Forecasting minimum temperature with short time-length data using artificial neural networks. Neural Netw. World 4 (2), 219–230.

Shanmuganathan, S. 2005. Applications of an unsupervised neural network to support sustainable development by modelling environmental, social and economic conditions. Proceedings of the Fourth International Symposium on Digital Earth-Japan (ISDEJapan), March 28e31, 2005, A05.

Tison, J., Park, Y.S., Coste, M., Delmas, F., Giraudel, J.L. 2004. Use of unsupervised neural networks for eco-regional zonation of hydrosystems through diatom communities: case study of Adour-Garonne watershed (France). Archiv fur Hydrobiologie 159, 409–422.

Tison, J., Park, Y.-S., Coste, M., Wasson, J.G., Ector, L., Rimet, F., Delmas, F. 2005. Typology of diatom communities and the influence of hydro-ecoregions: A study on the French hydrosystem scale. Wat. Res. 39, 3177–3188.

Ultsch, A. 2003. U-matrix: a tool to visualize clusters in high dimensional data. Tech. rep. Department of Mathematics and Computer Science, Philipps-University Marburg.

Van Dam, H., Mertens, A., Sinkeldam, J. 1994. A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands. - Netherl. J. Aquat. Ecol. 28: 117-131.

Vant, B. 1999. Analysing and reporting SoE monitoring results for rivers in the Waikato region. In: Proceedings of the Environmental Indicators Symposium. University of Otago.

Vesanto, J . 2000. Neural network tool for data mining: SOM Toolbox. - Proc. Symp. On Tool Environments and Development Methods for Intelligent Systems (TOOLMET2000): 184.196. Oulun yliopistopaino, Oulu, Finland.

Vesanto, J., Alhonieme, I. 2000. Clustering of the self-organizing map. - IEEE Transact. on Neural Networks 11**:** 586-600.

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J. 2000. SOM Toolbox for Matlab 5. Technical Report A57. Neural Networks Research Centre, Helsinki University of Technology, Helsinki, Finland.

Walley, W.J., Fontama, V.N., Martin, R.W. 1998. Applications of Artificial Intelligence for Biological Surveillance of River Quality, Environment Agency R&D Technical Report E52, Environment Agency R&D Dissemination Centre, c/o WRc, Frankland Road, Swindon, Wiltshire, SN5 8YF.

Wen, C.-G., Lee, C.-S. 1998. A neural network approach to multiobjective optimization for water quality management in a river basin. Water Resour. Res. 34, 427–436.

Whittaker, R.H. 1952. A study of summer foliage insect communities in the Great Smoky Mountains. Ecological Monographs 22, 1–44.

Winkler, U., Voigtlander, G. 1995. Anwendung neuronaler Netze fur die Simulation von Prozeßablaufen auf vorhandenen Klaranlagen. Korrespond. Abwasser 10, 1784–1792.

Publications of Gábor Várbíró

1.1 Publications connected to the theses

1. **Várbíró, G**., Borics, G., Kiss, K.T., Szabó, K. É., Ács, É. (2007): Use of Kohonen Self Organizing Maps (SOM) for the characterisation of benthic diatom associations of the River Danube and its tributaries. – Arch. Hydrobiol. Suppl. Large Rivers 17: 395-403. **IF: 1.409**
2. **Várbíró, G.,** Ács, É., Borics, G., Érces, K., Fehér, G., Grigorszky, I., Japport, T., Kocsis, G., Krasznai, E., Nagy, K., Nagy-László, Zs., Pilinszky, Zs., Kiss, K.T. (2007): Use of Self-Organizing Maps (SOM) for characterisation of riverine phytoplankton associations in Hungary. - Arch. Hydrobiol. Suppl. Large Rivers 17: 383-394. **IF: 1.409**
3. Borics, G, **Várbiró, G.,** Grigorszky, I, Krasznai, E., Szabó, S., Kiss, K. T. (2007): A new evaluation technique of potamoplankton for the assessment of the ecological status of rivers. Arch. Hydrobiol. Suppl. Large Rivers 17: (3-4) *465-486* . **IF: 1.409**
4. Ács, É., Borsodi, A.K., Kiss, É., Kiss, K.T., Szabó, K.É., Vladár, P., **Várbíró, G**., Záray, Gy. (2008): Comparative algological and bacteriological examinations on biofilms developed on different substrata in a shallow soda lake – Aquat Ecol (2008) 42:521–531. **IF: 0.925**
5. Van Dam, H., Stenger-Kovács, Cs., Ács, É., Borics, G., Buczkó, K., Hajnal, É., Sorόczki-Pintér, É., **Várbíró, G.,** Tóthmérész, B., Padisák, J. (2007): Implementation of the European Water Framework Directive: Development of a system for water quality assessment of Hungarian running waters with diatoms. - Arch. Hydrobiol. Suppl. Large Rivers 17: 339-364**. IF: 1.409**

1.2 Other publications

6. Borics, G., Tóthmérész, B., Grigorszky, I., Padisák, J., **Várbíró, G.,** Szabó, S. (2003): Algal assemblage types of bog-lakes in Hungary and their relation to chemistry, hydrological conditions and habitat diversity. Hydrobiologia 502: *145-155.* **IF: 0.720**
7. Krasznai, E., Fehér, G., Borics, G., **Várbíró, G.**, Grigorszky, I. & Tóthmérész, B. (2008): Use of desmids to assess the natural conservation value of a Hungarian oxbow (Malom-Tisza, NE-Hungary). Biologia Bratislava 63/6: 924—931, **IF: 0,207**

8. Ács, É., Borics, G., Fehér, G., Kiss, K.T., Reskóné, N. M., Stenger-Kovács, Cs., **Várbíró, G.** (2009): Implementation of the European Water Framework Directive to assessmment the water quality in Hungarian running waters with diatoms. – Diatomededelingen 33: 29-32. ISSN1872-9673

9. Borics, G., Krasznai, E., **Várbíró, G.**, Abonyi, A., Grigorszky, I., Szabó, S. (2008): Néhány Tisza-menti holtág jellegezetes fitopkankton asszociációi. Hidrológiai Közlöny 88: (6) *34-36*

10. Grigorszky, I., Borics G., **Várbíró, G.**, Grigorszkyné Budai, T., Béres, V., Krasznai, E., Schnitchen, Cs., (2008): A nitrát és nitrit ion hatása az üledékből történő foszfor felszabadulásra két Körös-vidéki holtmederben Hidrológiai Közlöny 88: (6) *61-63*

11. Ács, É., Borics, G., Kiss, G., Reskóné, Nagy, M., **Várbíró, G**., Kiss, K.T. (2007): Fitobenton vizsgálatok tanulságai a Balaton és vízgyűjtője patakjainak példáján. – MHT XXV. Országos Vándorgyűlés. Tata, 2007. július 4-5. pp. 1-27. CD. - ISBN 978-963-8172-20-4

1.3 Oral presentations and posters

12. Ács, É., Bolla, B., Kiss, K.T., Reskóné Nagy, M., **Várbíró, G**. (2009): A bevonatlakó kovaalgák szerepe a Balaton ökológiai állapotának megítélésében. – leadott absztrakt a 8. Magyar Ökológus Kongresszusra.

13. Krasznai, E., Borics, G., **Várbíró, G**., Abonyi, A., Padisák, J., Tóthmérész, B. (2008): Spatial and temporal pattern of vertical distribution of phytoplankton in a shallow oxbow (Hungary). – The 15th workshop of the International Association of phytoplankton taxonomy and ecology (IAP), Israel

14. Borics, G., **Várbíró, G.,** Krasznai, E., Grigorszky, I. (2008): Milyen mértékben határozza meg a meder morfológiája sekély tavaink trofitását? – L. Hidrobiológus Napok, Hungary,

15. Krasznai, E., Borics, G., **Várbíró, G.,** Tóthmérész, B. (2008): Cianobaktérium dinamika egy sekély holtágban. – L. Hidrobiológus Napok, Hungary,

16. Krasznai, E., Fehér, G., Borics, G., **Várbíró, G**., Grigorszky, I. & Tóthmérész, B. (2007): Use of desmids in the assessment of the natural value of a Hungarian oxbow – The International Symposium on Biology and taxonomy of green algae V. Smolenice Castle, June 25-29, 2007, Slovakia, Programme & Abstracts

17. Ács, É., Szabó, K.É., **Várbíró, G.,** Kiss, K.T., Borics G. (2007): Use of Kohonen Self Organizing Maps (SOM) for the characterisation of benthic

diatom associations of the River Danube and its tributaries. 6th International Symposium Use of Algae for Monitoring Rivers p. 20.

18. Ács, É., Borics, G., Kiss, G., Reskóné, Nagy, M., **Várbíró, G.,** Kiss, K.T. (2007): Fitobenton vizsgálatok tanulságai a Balaton és vízgyűjtője patakjainak példáján. – MHT XXV. Országos Vándorgyűlés. Tata, 2007. július 4-5. pp. 1-27. http://www.hidrologia.hu/ovgytata/25/7szekcio/250700.htm
19. Borics, G., Krasznai, E., Abonyi, A., **Várbíró, G.**, Grigorszky, I. (2007): Néhány Tisza menti holtág jellegzetes fitoplankton asszociációi – IL. Hidrobiológus Napok, 2007, Hungary,
20. Grigorszky, I., Borics, G., **Várbíró, G**., Grigorszkyné Budai, T., Béres, V., Krasznai, E., Schnitchen, Cs. (2007): A nitrát és nitrit hatása a holtágak üledékéből történő foszfor pótlására – IL. Hidrobiológus Napok, Hungary,
21. Borics, G., **Várbíró, G.,** Kiss, K. T., Grigorszky, I., Krasznai, E. (2006): Possible evaluation of the rheo-plankton for assessing th ecological status of rivers. 6th International Symposium Use of Algae for Monitoring Rivers p. 29.
22. **Várbíró, G.,** Ács, É., Borics, G., Érces, K., Fehér, G., Grigorszky, I., Japport, T., Kocsis, G., Krasznai, E., Nagy, K., Nagy-László, Zs., Pilinszky, Zs., Kiss, K. T. (2006): Use of Self Organizing Maps (SOM) for characterisation of riverine phytoplankton associations in Hungary. 6th International Symposium Use of Algae for Monitoring Rivers p. 172.
23. Grigorszky, I., Kiss, K. T., Dévai Gy., Nagy, A.S., Borics, G., Gligora, M., Kraj K., Plenkovich, A., Béres, V., Owsianny, P.,**Várbíró G.** (2006): The main charecteristics of phytoplankton populations in the hungarian section of the River Tisza in connection with the WFD. 6th International Symposium Use of Algae for Monitoring Rivers p. 53.
24. Krasznai, E., **Várbíró, G**., Borics, G., Grigorszky, I. Some interesting algological characteristics of the Tiszadob-oxbow. 16th Hungarian Algological Meeting 2005. Abstract book. 19.
25. **Várbíró, G.**, Borics, G. (2004): Establishment of the East Hungarian Biomonitoring Network (EHBN), In Klumpp, A., Ansel, W., Klumpp, G., (Eds,) Urban air Pollution Bioindication and environmental awareness. 321-327; Cuvillier Verlag, Göttingen.
26. **Várbíró, G**., Borics, G. (2002): Establishment of the East Hungarian Biomonitoring Network (EHBN), 2002. Oradea
27. **Várbíró, G**., Borics, G. (2001): The establishment of the East Hungarian Biomonitoring Network. EURASAP 41. Newsletter. April 2001. p.36-37. ISSN-1026-2172. Sofia.

28. Borics, G., Grigorszky, I., Padisák, J. and **Varbiró, G.** (2001): Two morphologically and ecologically different types of *Peridinium bipes* Stein in Hungarian waters. Plankton symposium 2001. Abstract book. p. 222.

29. **Várbíró, G**., O'Heix, B.-C., Tóth, J. A., Dizengremel, P., Garrec, J.P., (1998): Air quality mapping of the city Debrecen with epiphytic green algae as biomonitors.. Microflora et Pollution. Egregia, Nyiregyhaza-Nancy, pp. 93-102.

30. O'Heix, B.-C., **Várbíró, G.**, Tóth, J. A., Dizengremel, P., Garrec, J.P., Balazsy S. (1997): Use of epiphytic green algae growing on tree trunks as biomonitors of air quality. International Regional Seminar Environmental Protection: Modern Studies in Ecology and Microbiology. Ukraine Uzhgorod., Vol., 1. pp. 32-39.

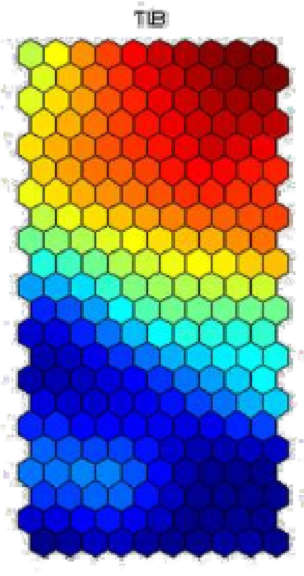
31. O'Heix, B.-C., **Várbíró, G.**, Tóth, J. A., Dizengremel, P., Garrec, J.P. (1997): Method for the air quality mapping of a city thanks the use of biomonitoring example of Debrecen - Hungary.. International Regional Seminar Environmental Protection: Modern Studies in Ecology and Microbiology. Ukraine Uzhgorod., Vol., 2. pp. 50-55.

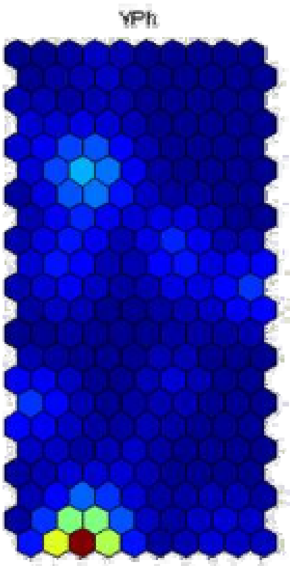
Paper 1.

Use of Self-organizing Maps (SOM) for characterization of riverine phytoplankton associations in Hungary

Várbíró, G., Borics, G., Kiss, K.T., Szabó, K. É., Ács, É.

Large Rivers Vol. **17,** No. 3-4
Arch. Hydrobiol. Suppl. 16113-4, p. 383-394,ok;ober 2007





Large Rivers Vol. 17, No. 3–4
Arch. Hydrobiol. Suppl. 161/3–4, p. 383–394, Oktober 2007

Use of Self-Organizing Maps (SOM) for characterization of riverine phytoplankton associations in Hungary

G. Várbíró[1]*, É. Ács[2], G. Borics[1], K. Érces[3], G. Fehér[4], I. Grigorszky[5], T. Japport[6], G. Kocsis[7], E. Krasznai[8], K. Nagy[9], Zs. Nagy-László[10], Zs. Pilinszky[11] and **K.T. Kiss[2]**

With 6 figures in the text

Abstract: The phytoplankton database of the Middle Danube Basin was analysed and evaluated in order to describe the characteristic algal assemblages of the rivers. The dataset were extracted from the database of the Hungarian monitoring network and academic institutions. We implemented the Kohonen Self Organizing Map (SOM) method by which we can visualize the assemblages in topology-preserving projection of two-dimensional space. The method is capable of evaluating large datasets (more than 1800 samples in the present investigation). As a result, we can identify the different algal communities which characterize different river types. The algal communities were described as different ratios of algal functional groups. Since some of the groups are in close relation with certain types of environmental pressure it is also possible to highlight those rivers or river sections (or those periods) which are far from the expected good ecological status.

Key words: Self Organizing Map, algal functional groups, algal communities.

Authors' addresses: [1] Environmental Protection, Nature Conservation and Water Authority, Trans-Tiszanian Region, Hatvan u. 16, H-4025 Debrecen, Hungary. [2] Institute of Ecology and Botany, Hungarian Danube Research Station of the Hungarian Academy of Sciences, Jávorka S. u. 14, H-2131 Göd, Hungary. [3] Environmental Protection, Nature Conservation and Water Authority, North-Danubian Region, Árpád u. 28–32, H-9021 Győr, Hungary. [4] Environmental Protection, Nature Conservation and Water Authority, Lower-Danube Region, Pf. 113, H-6501 Baja, Hungary. [5] Debrecen University, Department of Hydrobiology, Egyetem tér 1, H-4010 Debrecen, Hungary. [6] Environmental Protection, Nature Conservation and Water Authority, Körös Region, Pf. 99, H-5701 Gyula, Hungary. [7] Environmental Protection, Nature Conservation and Water Authority, Upper-Tiszanian Region, Pf. 246, H-4401 Nyíregyháza, Hungary. [8] Debrecen University, Department of Ecology, Egyetem tér 1, H-4010 Debrecen, Hungary. [9] Environmental Protection, Nature Conservation and Water Authority, North-Hungarian Region, Pf. 246, H-3501 Miskolc, Hungary. [10] Environmental Protection, Nature Conservation and Water Authority, Lower-Tiszanian Region, Pf. 1048, H-6712 Szeged, Hungary. [11] Environmental Protection, Nature Conservation and Water Authority, South-Danubian Region, Pf. 412, H-7602 Pécs, Hungary.

* corresponding author's E-mail address: varbirog@gmail.com

0945-3784/07/0161-0383 $ 3.00

Introduction

The traditional multivariate statistical methods, like cluster analysis and ordination, are difficult to interpret and cannot well present the information of very large data sets. Self Organizing Map (SOM) is a novel approach for the visualization of high-dimensional data. SOM converts complex statistical relationships between data sets into simple geometric relationships on a low-dimensional display. Thus, it compresses information while preserving the most important topological and metric relationships of the primary data (KOHONEN 2001). The main advantages of the SOM are the better data visualization and noise reduction (VESANTO & ALHONIEMI 2000). MANGIAMELI et al. (1996) compared SOM and several hierarchical clustering methods, and found SOM superior to hierarchical clustering in both robustness and accuracy. In addition, the two-level clustering approach (SOM neural network followed by K-means clustering) was developed and successfully applied to cluster data (VESANTO & ALHONIEMI 2000; BECCALI et al. 2004). SOMs are increasingly popular tools in diatom ecology, and have been used to describe bentic algal assemblages in France and Luxembourg (GOSSELAIN et al. 2005; RIMET et al. 2005a, 2005b). The method however, has not been used so far to analyse riverine phytoplankton assemblages, although the biological monitoring of the waterways provided large datasets in several countries during the last decades. The EC Water Framework Directive (2000) defines the minimal level for the ecological monitoring of surface waters in Europe. The importance of the use of macroscopic invertebrates and bentic diatoms as well-known indicator groups in rivers is emphasized, but it also indicates that the most relevant elements should be applied in the given situation. In the case of eutrophic rivers, the investigation of the phytoplankton is at least as important and informative as that of the other elements. The problem is that phytoplankton investigations alone are not equal with water quality assessment. In most cases, the phytoplankton of rivers is a mixed assemblage memorizing the ecological conditions of the upper river segments. The phytoplankton assessment has to be based on the evaluation of the occurring species, or groups of algae, which groups can be created on the basis of their taxonomic relationship (genera, divisions), or on the basis of their similar ecological behaviour. The assessment method proposed by BORICS et al. (2007) is based on the evaluation of the functional groups of algae (REYNOLDS et al. 2002; PADISÁK et al. 2005).

The aim of this study was to show the applicability of the functional groups approach to describe characteristic phytoplankton assemblages in riverine ecosystems and to extract the most frequent assemblage types.

Material and methods

Data from 1897 phytoplankton investigations from 189 locations have been included in the analysis. The investigations were carried out in Hungarian rivers and rivulets by regional authorities and research institutions during the period 2000–2004.

The original database contained the relative abundance of the species. Since the analysis here is based on the relative biomass of the species, cell volume data from earlier measurements were used.

The species were sorted into different algal functional groups. The original list of algal functional groups of Reynolds (REYNOLDS et al. 2002) was supplemented with some new groups such as T_B benthic diatoms (most of the species that belong to the *Navicula*, *Nitzschia*, *Gomphonema* etc.); T_D benthic desmids; T_C benthic Cyanobacteria (species such as *Lyngbia*, *Oscillatoria*). Smaller alterations were made on the species pool of the W_1 and W_2 groups, because species that prefer waters of very high organic content including *Chlamydomonas reinhardtii, C. ehrenbergii*, *Euglena viridis*, *Spermatozopsis exultans* were assembled into a new W_0 codon. Detailed explanation of the new functional groups and the total list of the codons and their relevant species can be found in BORICS et al. (2007).

For statistical analysis, we used KOHONEN's Self Organizing Map method. The detailed algorithm of the SOM can be found in KOHONEN (2001). CHON et al. (1996), PARK et al. (2003) and LEK & GUEGAN (2000) described ecological applications. The numbers of nodes were determined as 5x sqrt (number of samples) according to VESANTO (2000).

For clustering the SOM we used the K-means clustering technique. K-means clustering is an algorithm to classify or to group objects based on attributes (in this case species composition) into K number of group. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid as the square error of each data point is calculated and clusters reformed such that the sum of square errors is made to be minimum. Details of the the methodological background can be found in VESANTO & ALHONIEMI (2000) or BECCALI et al. (2004).

The Structuring Index (SI) was originally developed to define species showing the strongest influence on the organization of the SOM map (see PARK et al. 2005). TISON et al. (2004) used the SI to evaluate relevant diatom species in the classification of diatom communities. The SI is the value indicating the relative importance of each species in determining the distribution patterns of the samples in the SOM. Therefore, the set of species showing high SI can be considered as the indicator species.

Results

The application of SOM and the use of K-means clustering resulted in 8 different types of riverine phytoplankton association. (Fig. 1a–c). These types were determined by high structuring indexed codons such as T_B, Y, W_0, D, J, C and W_1 (Figs. 2, 3a). Other codons, important in the riverine phytoplankton associations

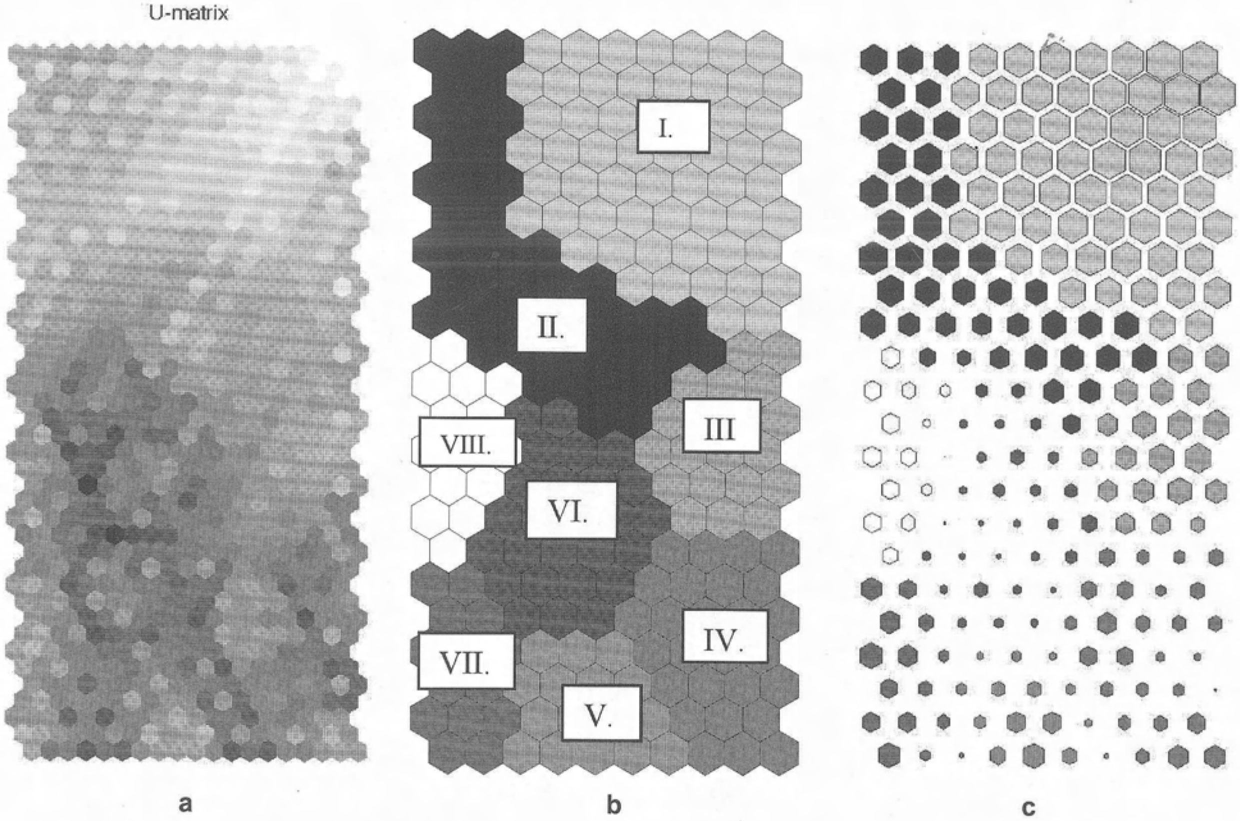


Fig. 1a. U-matrix of the SOM, darker cells indicate more distance between neighbouring cells.
Fig. 1b. Clusters of the SOM resulted by the K-means clustering. Roman letters mean the different clusters.
Fig. 1c. SOM histogram of samples – the larger the hexagons are the larger numbers of samples belong to that given virtual map unit, colours are according to the clusters indicated in Fig. 1b.

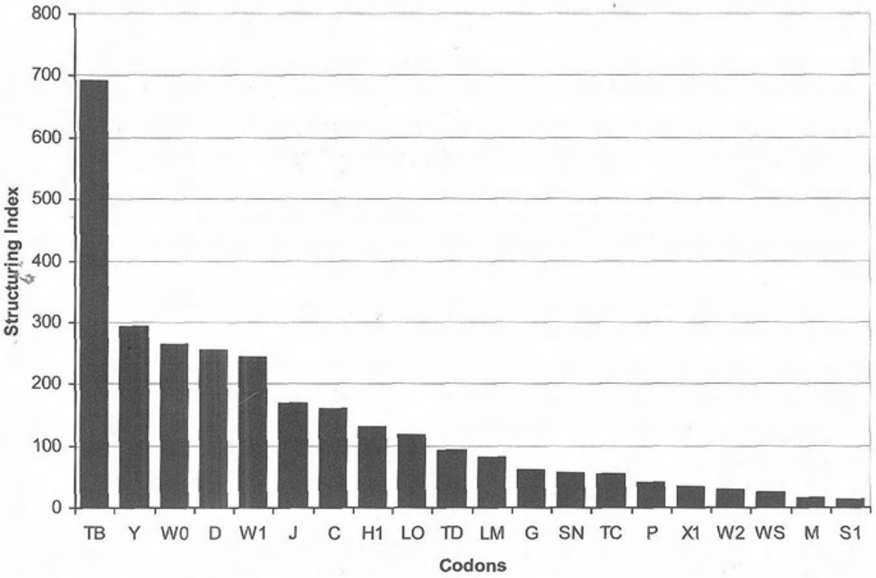


Fig. 2. Codons with high structuring index. The values indicate the relative importance of each codon in determining the SOM patterns.

with their abundance of occurrence on the SOM map can be found in Fig. 3b. After creating the SOM, we were able to count the number of samples belonging to a given virtual hexagonal map unit (Fig. 4), in this way the most frequent was the type I, in which 931 samples occurred. The eight plankton associations can be defined as follows:

Type I. This is the most frequent phytoplankton type which is representative for the upper section of rivers. The dominant codon is the T_B (bentic diatoms) with more than 70 % of the total biomass abundance (Fig. 5). This association has no seasonal preference. The most frequent taxa are *Nitzschia palea, Nitzschia fonticola, Navicula capitatoradiata, Surirella brebissonii, Diatoma vulgaris.*

Type II. Members of the T_B group are dominating type II (Fig. 5) but the elements of the W_0 and W_1 functional groups (*Chlamydomonas reinhardtii, Euglena* spp.) indicate organic pollution. This type is also characteristic for upstream sections of rivers. Only a slight seasonal preference for early spring and autumn can be found.

Type III. This is typical plankton of small rivers with the dominance of T_B and D (Fig. 5) codons. This assemblage occurs on those river sections where the retention time due to hydro-morphological changes or upper stream reservoirs enables real phytoplankton to develop.

Type IV. This association can be referred to as "Danube type" summer plankton. The most important codons are D, J and C (Fig. 5). Dominant species of this

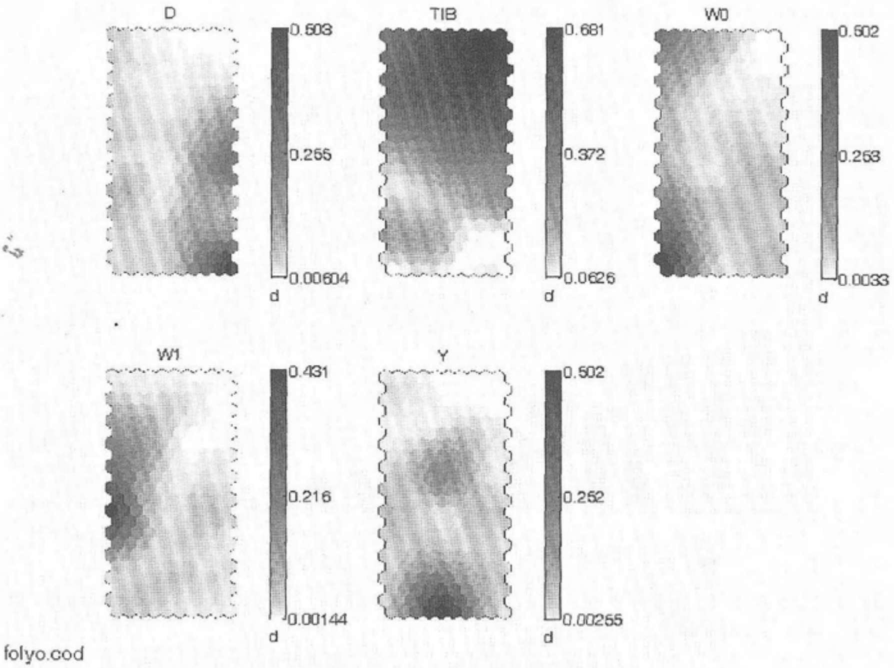


Fig. 3a. Gradient distribution of the most important codons on the SOM. Darker cells mean higher abundance of the codon in the hexagonal unit.

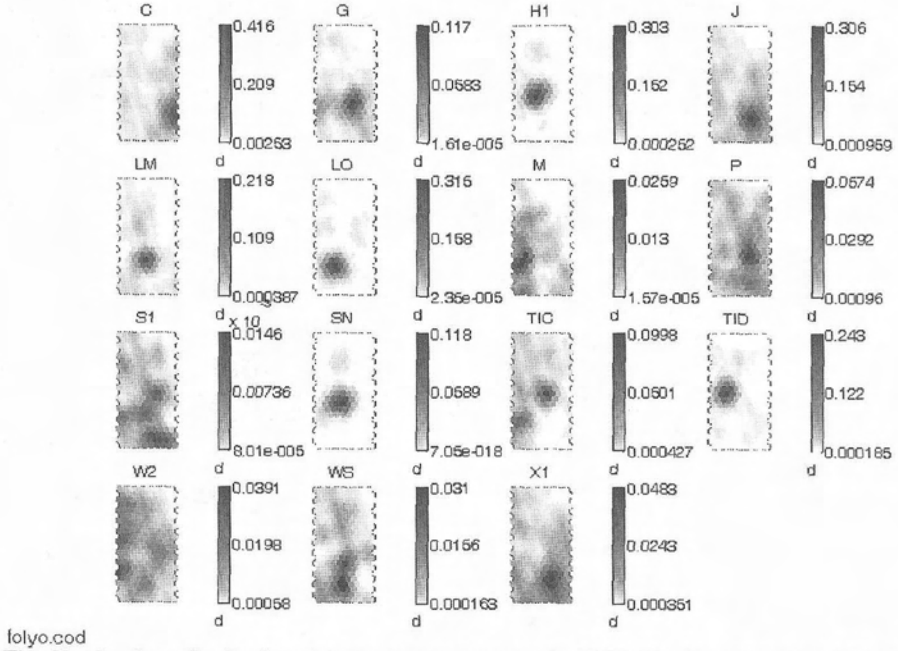


Fig. 3b. Gradient distribution of important codons on the SOM. Darker cells mean higher abundance of the codon in the hexagonal unit. Scale bar refers to relative percentage.

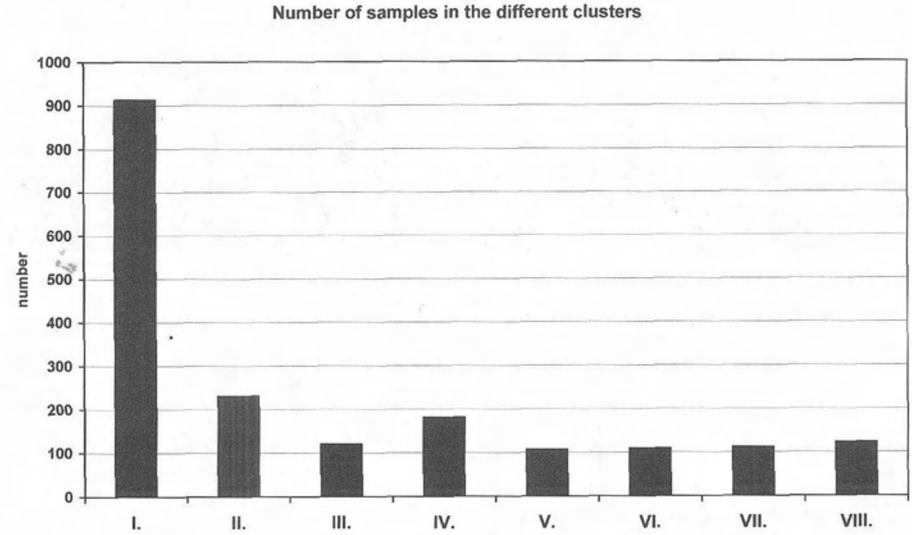


Fig. 4. The number of the samples belongs to the different cluster groups. Roman letters refer to the cluster numbers.

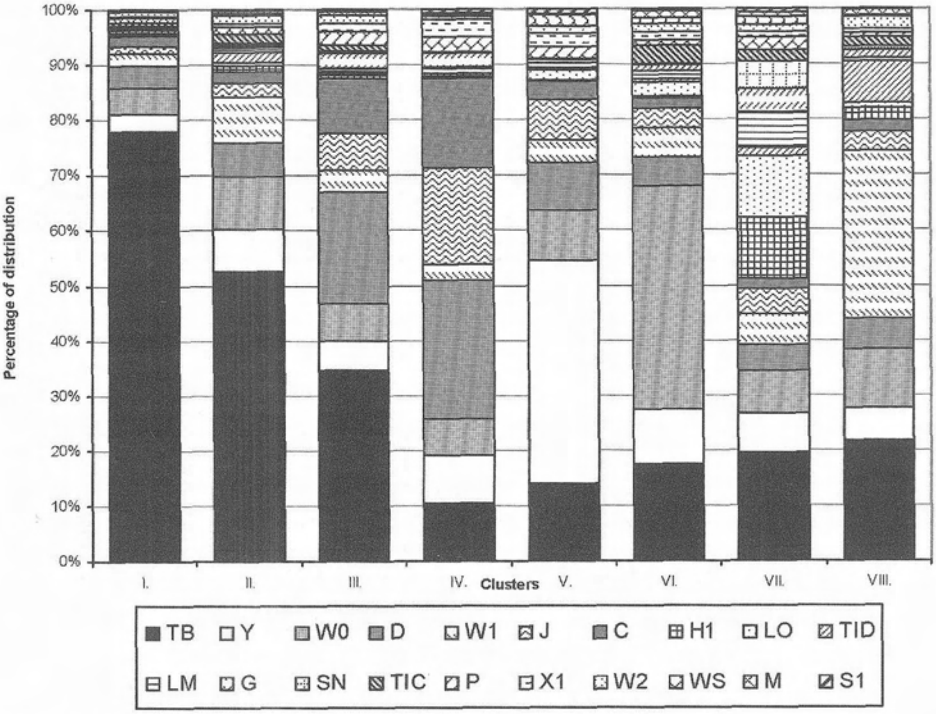


Fig. 5. Percentage distribution of the different codons according to the SOM clusters. Roman letters refer to the cluster numbers.

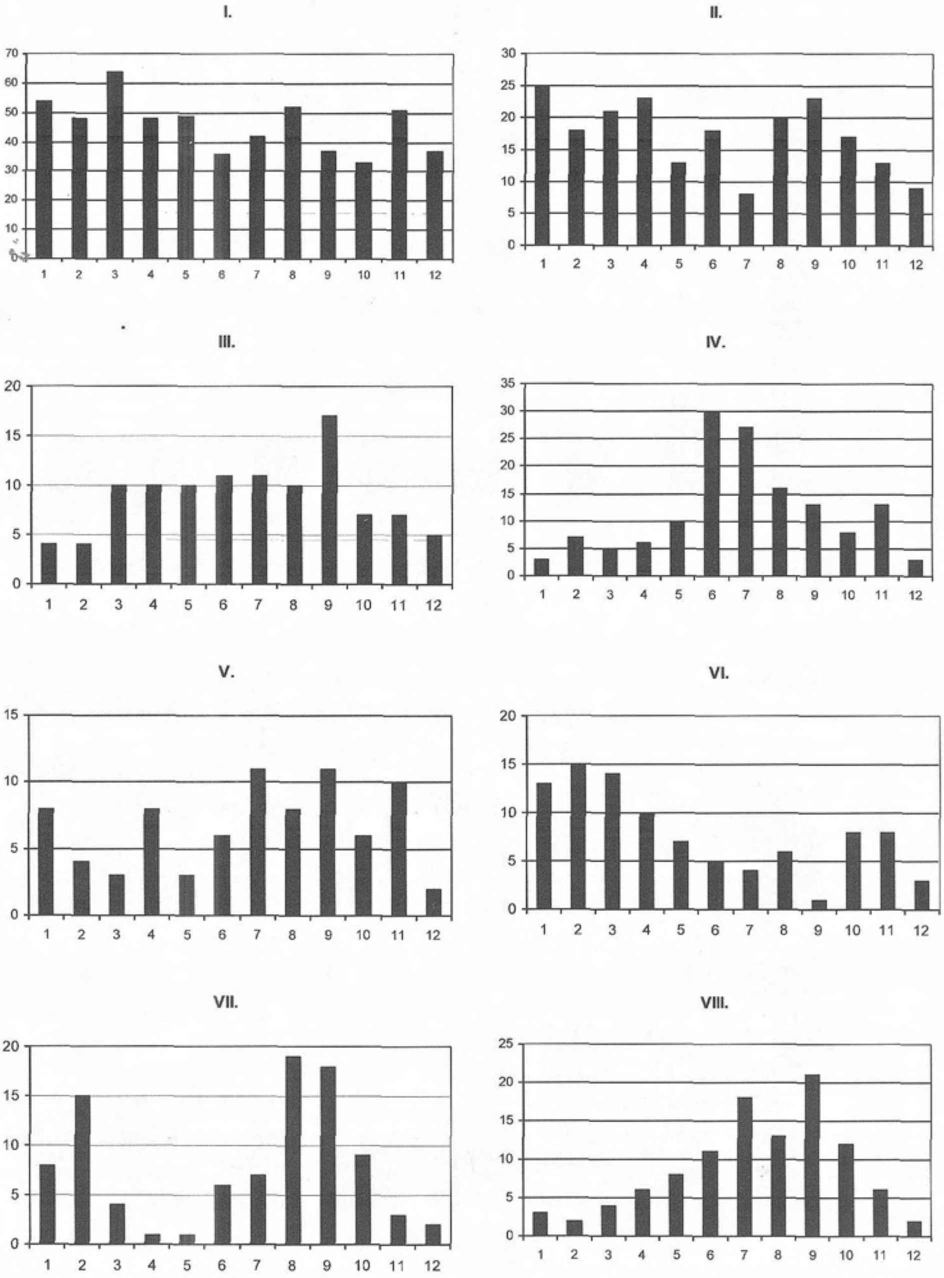


Fig. 6. Seasonal distribution of the algal assemblage types. Roman letters mean the cluster numbers, horizontal axis represents the months, while vertical axis the number of samples belonging to that given month.

association are *Stephanodiscus hantzschii, Cyclotella comta, Skeletonema potamos* and *Nitzschia acicularis* from codon D, *Cyclotella meneghiniana* from codon C, and *Scenedesmus* species from codon J. However, the favourable season of this type is the early summer (Fig. 6). Certain species such as *Stephanodiscus invisitatus* could bloom however, in the Danube during winter (KISS & GENKAL 1993).
Type V. This type could be mainly found on the lower section of the river Tisza with the dominance of Y codon (Fig. 5). *Cryptomonas reflexa, C. marssonii, C. rostratiformis, Rhodomonas minuta* become dominant in this section. Development of this association is absolutely independent from the seasons (Fig. 6).
Type VI. The characteristic functional group of type VI is W_0 (Fig. 5). The dominance of this group is due to very strong organic pollution. *Chlamydomonas* spp., *Euglena viridis, Polytoma uvella, Spermatozopsis exultans* are typical elements of this assemblage. This type of plankton is usually dominant in winter and early spring (Fig. 6.)
Type VII. This type is a mixture of a very divers association with the presence of relatively rare groups like L_0, H_1, L_m, S_n (Fig. 5). The occurrence of this type is expected in slow flowing channels and small rivulets in late winter and summer.
Type VIII. The dominant codon of this type is the W1 (Fig. 5) which is according to REYNOLDS is characteristic for small organic pools. Frequent taxa are the metaphytic *Phacus* and *Trachelomonas* spp. This type is typical in slow-flowing rivulets and channels which are under the risk of organic pollution and have rich macrophyte vegetation. Development of this type is expected in summer.

The seasonal distribution of the defined plankton types (Fig. 6) shows, that several of them have affinity to a certain period of the year. Types IV and VIII develop usually in summer. Type VI occurs in late winter and early spring. Type VII has got a special bimodal character with a late winter and mid summer peak.

Discussion

The Danube river basin is the second largest system in Europe. Flow regulation and river fragmentation has crated hundreds of impoundments in its catchment. These lentic habitats, depending on their altitude, depth, trophic state and residence time, provide different ecological conditions for the development of specific algal assemblages. It is no exaggeration to say that the diversity of habitats (lakes, impoundments) enables the development of almost all of the functional groups of algae, but survival and further development downstream is rather different. With the implementation of Self Organizing Map and K-means clustering eight significant algal assemblages were defined and described by their relative contribution of algal functional groups (REYNOLDS et al. 2002). Since some of the functional groups are closely related to certain types of environmental pressure (organic pollution W_0,W_1; high nutrient status Y, H_1, J; impounding L_0, L_m) it is possible to highlight those river sections or periods which are impacted by human activities. Evaluation of the dataset validated the separation of the W_0 from the Reynolds' W_1 group. The establishment of the W_0 group enables the description of the worst status of the rivers with serious organic pollution. This status usually occurs in late winter and early spring, because mineralization is slow due to low temperatures. Our results demonstrate that the abundance of the W_0 decreases with elevated water temperature, but the occurrence of the W_1 shows an increasing tendency (Fig. 6, VI).

In our dataset, the most frequent algal assemblage was the type I, dominated by bentic diatoms (Fig. 3). This can be explained by the short residence time of the water bodies in question.

The bimodal character of the type VII is cause by the L_0 functional group. This plankton type frequently occurs in the summer epilimnion of lakes (Reynolds et al. 2002). Several dinophytes, however, can be important members of the winter phytoplankton (Grigorszky et al. 1998). Therefore it would be necessary to separate certain oligotherm *Peridinium* taxa from the L_0 codon.

The application of functional groups in lake quality assessment has been proved by Padisák et al. (2005). Further studies have been started for the application of this approach on riverine phytoplankton quality assessment, see Borics et al. (2006).

Acknowledgements

This study was supported by the Ministry of Environment and Water, by the Bolyai János fellowship of the Hungarian Academy of Sciences, the Hungarian National Research Fund (OTKA, K60452) and by GVOP-3.2.1.-2004-04-0151/3.0. The authors would like to thank Mr. Jean Giraudel for help in calculating of the Structuring Index. We wish to thank Martin Dokulil and the unknown referee for their critical comments that improved this manuscript.

References

Beccali, M., Cellura, M., L. Brano, V. & Marvuglia, A. (2004): Forecasting daily urban electric load profiles using artificial neural networks. – Energy Conversion and Management **45**: 2879–2900.

Borics, G., Várbíró, G., Kiss, K.T., Grigorszky, I., Krasznai, E. & Szabó, S. (2007): Possible evaluation of the rheo-plankton for assessing the ecological status of rivers. – Arch. Hydrobiol. Suppl. **161** (Large Rivers **17**): 465–486.

Chon, T.S., Park, Y.S., Moon, K.H. & Cha, E.Y. (1996): Patternizing communities by using an artificial neural network. – Ecol. Modelling **90**, Issue 1.

Gosselain, V., Campeau, S., Gevrey, M., Coste, M., Ector, L., Park, Y.S., Lek, S. & Descy, J.-P. (2005): Diatom typology of reference situations at a large multi-regional scale: combined results of multivariate analysis and SOM. – In: Lek, S., Scardi, M., Verdonschot, P., Park, Y.S. & Descy, J.-P. (eds.): Modelling Community Structure in Freshwater Ecosystems. Springer-Verl.

Grigorszky, I. & Padisák, J. (1998): Jég alatti *Peridinium aciculiferum* Lemmermann (Dinophyta) populáció a Balatonban [*Peridinium aciculiferum* Lemmermann (Dinophyta) under-ice population from the shallow lake Balaton, Hungary]. – Hidrológiai Közlöny **78**: 282–284. [in Hungarian with English summary]

Kiss, K.T. & Genkal, S.I. (1993): Winter blooms of centric diatoms in the River Danube and in its side arms near Budapest. – In: H. van Dam (ed.): Twelfth International Diatom Symposium. Hydrobiologia **269/270**: 317–325. Kluwer Academic Publ.

Kohonen, T. (2001): Self-organizing maps. 3rd ed. – Springer, Berlin.

Lek, S. & Guégan, J.F. (2000): Artificial neuronal networks: application to ecology and evolution. – Springer, Berlin.

Mangiamelien, P., Chen, S.H.K. & West, D. (1996): A comparison of SOM neural network and hierarchical clustering methods. – Europ. J. Operational Res. **93**: 402–417.

Padisák, J., Grigorszky, I., Borics, G. & Soróczki-Pintér, É. (2005): Use of phytoplankton assemblages for monitoring ecological status of lakes within the Water Framework Directive: the assemblage index. – Hydrobiologia **553**: 1–14.

Park, Y.-S., Cereghino, R., Compin, A. & Lek, S. (2003): Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. – Ecol. Modelling **160**: 265–280.

Park, Y.-S., Gevrey, M., Lek, S. & Giraudel, J.L. (2005): Evaluation of relevant species in communities: Development of structuring indices for the classification of communities using a self- organizing map. – In: Lek, S., Scardi, M., Verdonschot, P., Descy, J.P., Park, Y.-S. (eds.): Modelling Community Structure in Freshwater Ecosystems: 369–380. Springer, Berlin.

Reynolds, C.S., Descy, J.P. & Padisák, J. (1994): Are phytoplankton dynamics in rivers so different from those in shallow lakes? – In: J.P. Descy, C.S. Reynolds & J. Padisák (eds.): Phytoplankton in turbid environments: Rivers and shallow lakes. Hydrobiologia **289**: 1–7. Kluwer Academic Publ.

Reynolds, C.S., Huszar, V., Kruk, C., Naselli-Flores, L. & Melo, S. (2002): Towards functional classification of the freshwater phytoplankton. – J. Plankton Res. **24**: 417–428.

Rimet, F., Cauchie, H.M., Tudesque, L. & Ector, L. (2005a): Use of artificial intelligence (MIR-max) and chemical index to define type diatom assemblages in Rhône basin and Mediterranean region. – In: Lek, S., Scardi, M., Verdonschot, P.F.M., Descy, J.P. & Park, Y.S. (eds.): Modelling community structure in freshwater ecosystems: 288–303. Springer.

Rimet, F., Ector, L., Hoffman, L., Gevrey, M., Giraudel, J.L., Park, Y.S. & Lek, S. (2005b): Prediction with artificial neural networks of diatom assemblages in headwater streams of Luxembourg. – In: Lek, S., Scardi, M., Verdonschot, P.F.M., Descy, J.P. & Park, Y.S. (eds.): Modelling community structure in freshwater ecosystems. Springer.

Tison, J., Giraudel, J.L., Coste, M., Park, Y.S. & Delmas, F. (2004): Use of unsupervised neural networks for eco-regional zonation of hydrosystems through diatom communities: case study of Adour Garonne watershed (France). – Arch. Hydrobiol. **159**: 409–422.

VESANTO, J. (2000): Neural network tool for data mining: SOM Toolbox. – Proc. of Symp. on Tool Environments and Development Methods for Intelligent Systems (TOOLMET2000). Oulun yliopistopaino, Oulu, Finland: 184–196.

VESANTO, J. & ALHONIEMI, E. (2000): Clustering of the self-organizing map. – IEEE Transact. on Neural Networks **11**: 586–600.

WFD (2000): Directive 2000/60/ec of the European Parliament and of the Council 22.12.2000. – Off. J. Europ. Comm. L327/1-72.

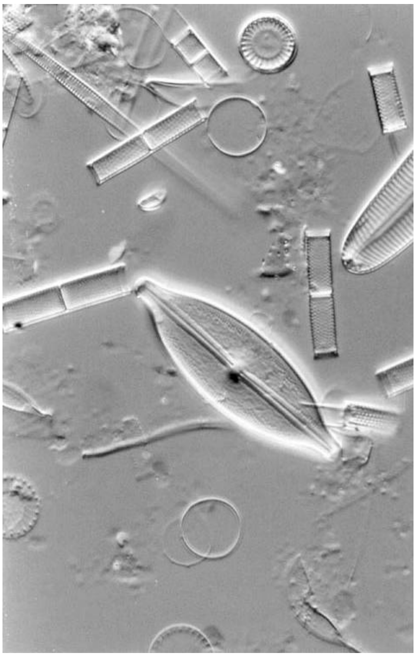
Accepted: 14 February 2007

Paper 2.

Use of Kohonen Self Organizing Maps (SOM) for the characterization of benthic diatom associations of the River Danube and its tributaries

Várbíró, G., Borics, G., Kiss, K.T., Szabó, K. É., Ács, É.

Large Rivers Vol. **17,** No. 3-4
Arch. Hvdrobiol. SUDDI1. 6113-4. D. 395-403. Oktober 2007



Large Rivers Vol. **17**, No. 3–4
Arch. Hydrobiol. Suppl. 161/3–4, p. 395–403, Oktober 2007

Use of Kohonen Self Organizing Maps (SOM) for the characterization of benthic diatom associations of the River Danube and its tributaries

Gábor Várbíró[1], Gábor Borics[1], Keve T. Kiss[2], Katalin É. Szabó[2], Anđelka Plenković-Moraj[3] and **Éva Ács[2*]**

With 7 figures and 1 table in the text

Abstract: Benthic diatom flora of the River Danube was investigated along the river from the source streams to the end of the Hungarian stretch during a four-year period. Characteristic diatom community assemblages were established using KOHONEN Self-Organizing Map (SOM) algorithm. We found five sample clusters, arranged according to the location. The first three groups are located in the German-Austrian stretch, while the fourth and fifth groups are located in the Slovakian-Hungarian stretch. Characteristic species of the clusters 1 and 2 are *Achnanthidium minutisimum* (KÜTZ.) CZARNECKI, *Denticula tenuis* KÜTZ., *Nitzschia fonticola* GRUN., *Psammothidium bioretii* (GERMAIN) BUKHT. & ROUND, *Melosira varians* AG., *Planothidium subatomoides* (HUST.) BUKHT. & ROUND, character species of cluster 3 is *Amphora pediculus* (KÜTZ.) GRUN., while character species of the clusters 4 and 5 are *Nitzschia dissipata* (KÜTZ.) GRUN., *Navicula cryptotenella* LANGE-BERT., *Cocconeis placentula* var. *euglypta* (EHR.) GRUN., *Rhoicosphenia abbreviata* (AG.) LANGE-BERT., *Nitzschia inconspicua* GRUN. This arrangement is also well in agreement with the poorer water quality of the lower, and the better water quality of the upper stretch.

Introduction

The River Danube is the largest (2,780 km long, 817,000 km^2 catchment area) and one of the most important rivers in Central Europe. During a four-year project, epilithic algae of the River Danube and its main tributaries were investigated along the river from the source streams Breg and Brigach up to the end of the Hungarian stretch (Fig. 1, Table 1). These samples were analysed from a taxonomical perspective, furthermore, water quality analysis using diatom indices was carried

---

Authors' addresses: [1] Environmental Protection, Nature Conservation and Water Inspectorate, Transtiszanian Region, Hatvan u. 16, H-4025 Debrecen, Hungary. [2] Institute of Ecology and Botany, Hungarian Danube Research Station of the Hungarian Academy of Sciences, Jávorka S. u. 14, H-2131 Göd, Hungary. [3] University of Zagreb, Faculty of Science, Division of Biology, Rooseveltov trg 6, HR-10000 Zagreb, Croatia.
[*] corresponding author's E-mail address: acs@ludens.elte.hu, acseva@freemail.hu

0945-3784/07/0161-0395 $ 2.25

Fig. 1. Map of the River Danube with our sampling points (numbers refer to the sampling sites in Table 1).

out. These results have been published and discussed elsewhere (Ács et al. 2006).

In the current study, we analysed and evaluated the diatom database of our previous study in order to describe the characteristic algal assemblages of the River Danube and some of its tributaries using Kohonen Self-Organizing Map.

Self Organizing Map is a novel approach for the visualization of high-dimensional data. SOM converts complex statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display. Thus, it compresses information while preserving the most important topological and metric relationships of the primary data items (Kohonen 1995). SOMs are an increasingly popular tool in diatom ecology, and have been used to describe algal assemblages in France and Luxembourg (Gosselain et al. 2005; Rimet et al. 2005a, 2005b).

Material and methods

Epilithon samples were taken in August 2000, 2001, 2002, and in October 2001 along the Danube and some of its tributaries (close to their mouths) from the source streams to the end of the Hungarian stretch (Table 1, Fig. 1). Samples were scraped off from stones in three replicates, washed into tap water, homogenized and fixed with formaldehyde. Diatoms were treated with H_2O_2 and washed three times in distilled water, afterwards mounted in Naphrax. In each sample, 400 valves were counted and IPS diatom index was calculated using Omnidia Version 4 (Lecointe et al. 1993, 1999).

Table 1. Sampling points and sampling times.

No.		Aug. 2000	Aug. 2001	Oct. 2002	Aug. 2002
	MAIN ARM:				
1	Breg-Brigach join at Donaueschingen				+
2	Immendingen before disappearing				+
3	Immendingen after disappearing				+
4	Nasgenstadt		+		
5	Öpfingen		+		
6	Elchingen		+		
7	Ingolstadt	+			
8	Kehlheim		+		+
9	Bad Abbach	+			
10	Deggendorf		+		
11	Passau	+	+		+
12	Mauthausen	+	+		
13	Melk	+	+		+
14	Bad Deutsch	+	+		
15	Bratislava			+	
16	Gabcikovo			+	
17	Medved			+	
18	Nagybajcs	+			
19	Komárom	+	+		
20	Esztergom		+		
21	Göd	+	+	+	+
22	Dunaföldvár			+	
23	Baja			+	
24	Mohács			+	
	TRIBUTARIES:				
25	Breg source				+
26	Breg at Furtwangen				+
27	Brigach source				+
28	Brigach				+
29	Iller		+		
30	Lech		+		
31	Danube-Main channel	+	+		+
32	Isar	+	+		
33	Enns	+	+		
34	Morava (Morva)	+			
35	Mosoni Danube	+	+		+
36	Vah (Vág)	+	+		+
37	Garam	+	+		
38	Ipel (Ipoly)	+	+		

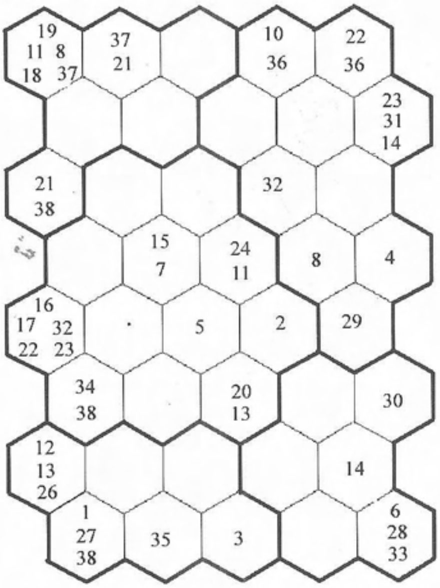
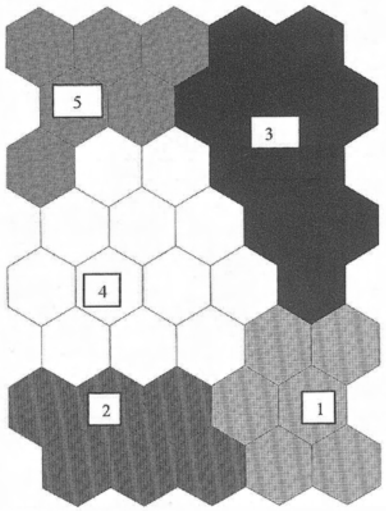


Fig. 2. The localization of the sampling sites on the SOM map. Numbers are according to the site numbers in Table 1.



SOM 21-Sep-2006

Fig. 3. Clusters of the sampling sites by K-means clustering of the SOM. Numbers represent the different clusters.

For statistical analysis, we used the Kohonen Self Organizing Map (SOM) which can make a projection of the database into a two-dimensional hexagonal map. Closely related communities were placed into neighbouring hexagons by their similarities, while samples with different communities were in distant hexagons. The SOM can display the groupings of samples and species together; therefore, each species can be evaluated by its importance (Kohonen 2001). With this method, we were able to distinguish different algal communities which characterize different river sections. It is also possible to select the characteristic species for these assemblages. The number of output neurons is an important factor in the analysis, we used the suggestion of Vesenato which determines the output number as:

$$Noutput = 5x\sqrt{Nsamples}$$

(Vesenato 2000). The rough tuning phase includes 5000, while the fine tuning phase lasted contains 15,000 iterations.

The Structuring Index (SI) was originally developed to define the species that show the strongest influence on the organization of the SOM map (Park et al. 2005). Tison et al. (2004) used the SI to evaluate relevant diatom species in the classification of diatom communities. The set of species showing high SI can be considered as the indicator species. Taxa showing strong gradient display high SI values, whereas species showing weak gradient present low SI values. Thus, the higher the value of SI, the more relevant the variable is to the structure of the map.

Results and discussion

K-means clustering of the SOM allocated 5 different sample groups (clusters) (Fig. 2). With few exemptions, the first, second and third clusters fell in the German-Austrian stretch, whereas the fourth and fifth clusters fell in the Slovakian-Hungarian stretch. Temporal differences were less pronounced: the five clusters were mainly organized by the sampling localities. The first cluster was formed by some sampling points scattered along the German and Austrian stretches, including the tributaries of Enns and Lech. The second cluster comprised the sampling points around the source of the Danube such as the Breg and Brigach sources and streams, the first stretch of the River Danube after the fusion of Breg and Brigach, furthermore the tributaries of Ipel, Morava and Garam. The third cluster was formed by the sampling sites on the lower part of the Austrian stretch. The fourth cluster was mostly built up by Slovakian, and the fifth by the Hungarian sampling sites, although the clustering is not perfectly unambiguous.

The character species of the five sample clusters were defined (Fig. 3). We found that the clusters 1 and 2 were characterized by *Achnanthidium minutisimum* (KÜTZ.) CZARNECKI (AMIN), *Denticula tenuis* KÜTZ. (DTEN), *Nitzschia fonticola* GRUN. (NFON), *Psammothidium bioretii* (GERMAIN) BUKHT. & ROUND (ABIO), *Melosira varians* AG. (MVAR), *Planothidium subatomoides* (HUST.) BUKHT. & ROUND (ASAT). The character species of cluster 3 is *Amphora pediculus* (KÜTZ.) GRUN. (APED), while character species of the clusters 4 and 5 are *Nitzschia dissipata* (KÜTZ.) GRUN. (NDIS), *Navicula cryptotenella* LANGE-BERT. (NCTE), *Cocconeis placentula* var. *euglypta* (EHR.) GRUN. (CPLA), *Rhoicosphenia abbreviata* (AG.) LANGE-BERT. (RABB), *Nitzschia inconspicua* GRUN. (NINC). However, not all of these species had equally high SI values (the higher the SI value, the better indicator is the species) (Fig. 4). Thirteen species had higher SI values, these were *Amphora pediculus, Achanthidium minutissimum, Cocconeis placentula* var. *euglypta, Nitzschia inconspicua, N. dissipata, Denticula tenuis, Rhoicosphenia abbreviata, Navicula tripunctata* (O.F. MÜLLER) BORY (NTPT), *Navicula cryptotenella* LANGE-BERT. (NCTE), *Staurosirella pinnata* (EHR.) WILLIAMS & ROUND (FPIN), *Nitzschia fonticola* GRUN. (NFON), *Melosira varians* and *Nitzschia amphibia* GRUN. (NAMP).

Considering the average relative abundance values of the character species of the five cluster groups (Fig. 5), it is obvious that the main difference between the cluster groups is not so much the species composition, but rather the different relative abundance ratios of the different species in the different clusters. The source area of the Danube (streams Breg and Brigach and the uppermost stretch of the River) are better separated (cluster 2) than the rest of the clusters.

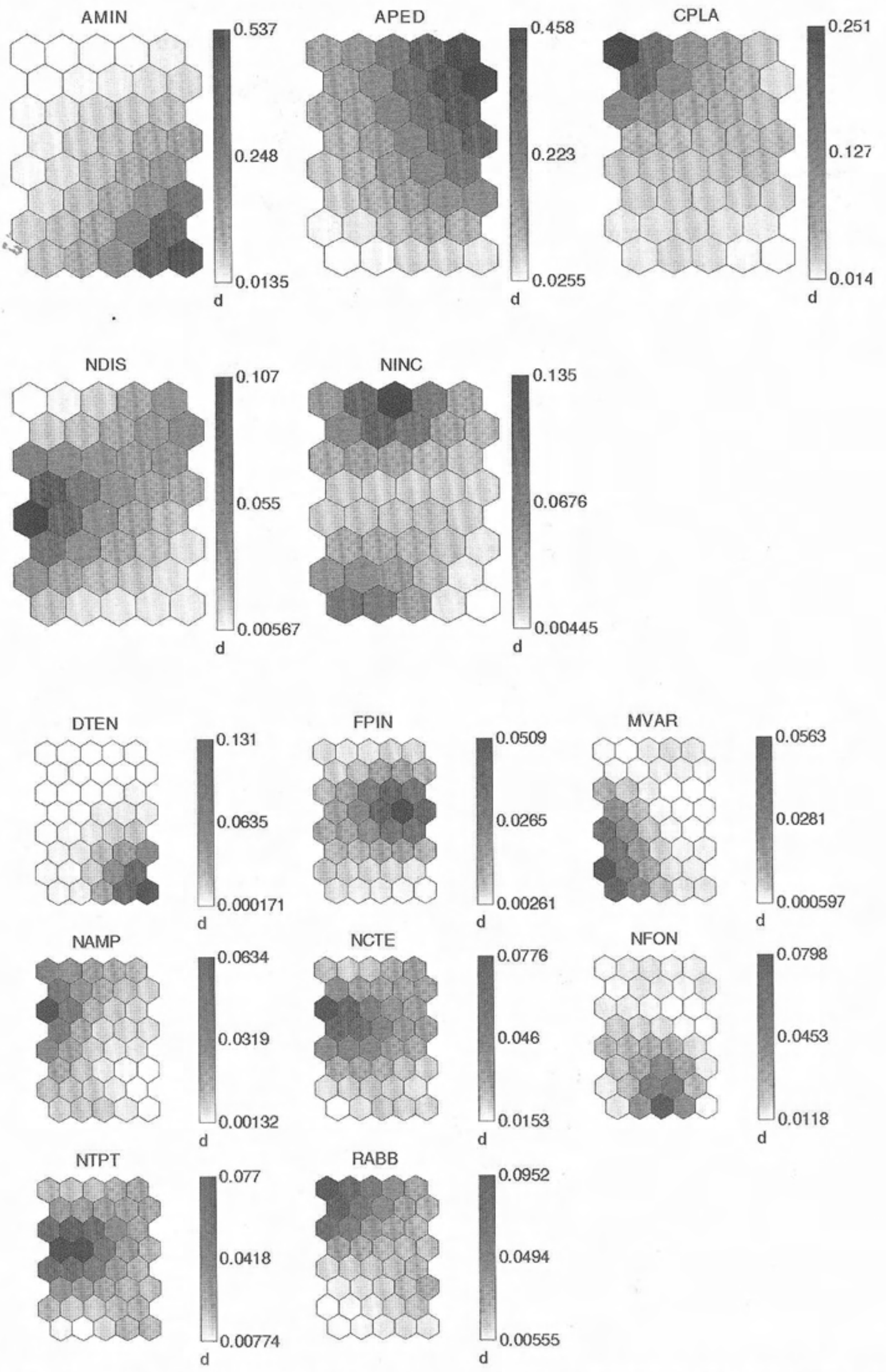


Fig. 4. Individual SOM maps of the characteristic species. Scale bars showing the species calculated abundances by the learning process of the SOM in percentage (abbreviation of species can be found in the text).

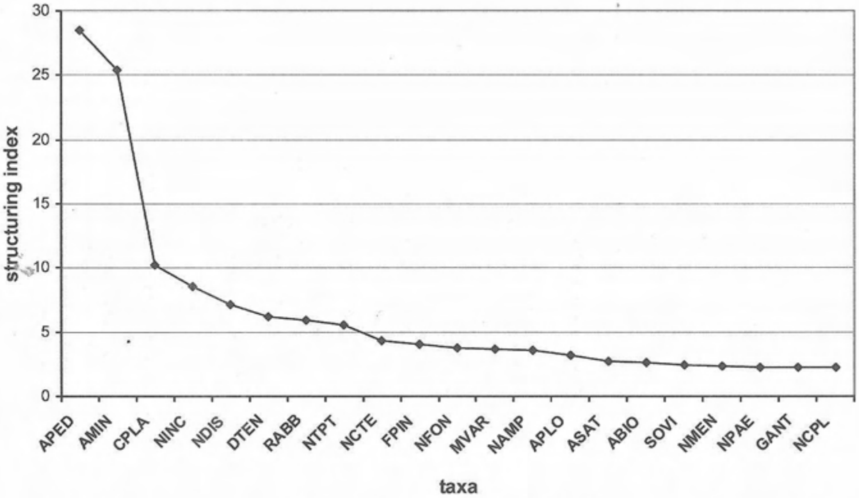


Fig. 5. The highest structuring indexed taxa. APLO = *Kolbesia ploenensis* (HUST.) KINGSTON, SOVI = *Surirella ovalis* BRÉB., NMEN = *Navicula menisculus* SCHUMANN, NPAE = *Nitzschia paleacea* GRUN., GANT = *Gomphonema angustum* AGH., NCPL = *Nitzschia capitellata* HUST. (see more abbreviations in the text).

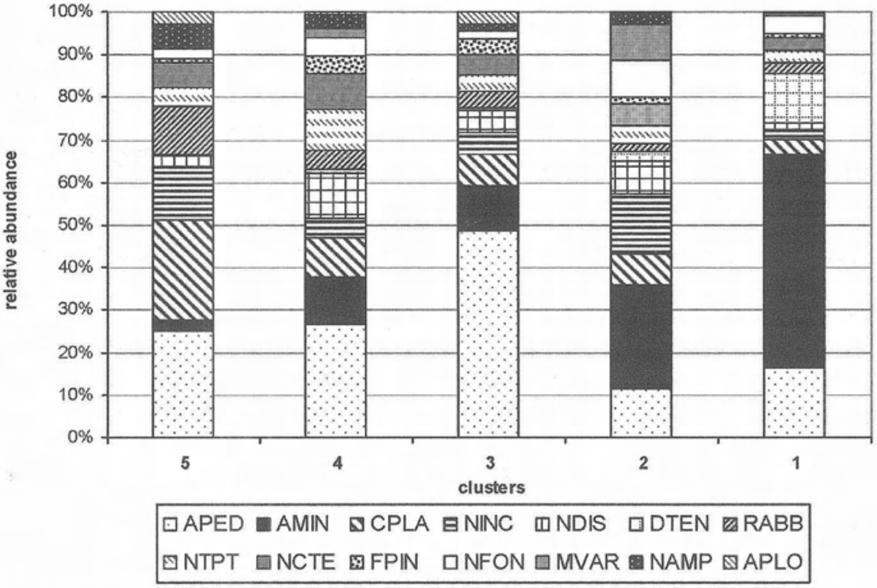


Fig. 6. Average taxonomic composition of the clusters (see abbreviations in the text).

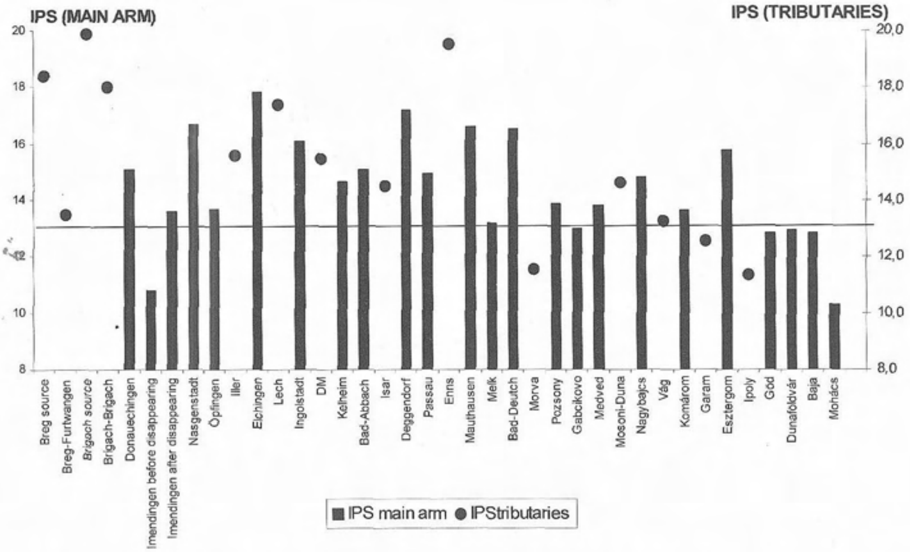


Fig. 7. IPS index values of River Danube and its tributaries (close to their mouth).

According to the diatom indices, the water quality of Danube river is generally good on the German-Austrian-Slovakian part and moderate on the Hungarian part and the water quality of the tributaries (close to their mouth) are good or excellent around the sources and, through gradual deterioration, it becomes moderate in Slovakia and Hungary (Fig. 7).

Conclusions

In conclusion, we found that the SOM was a powerful tool in resolving benthic diatom associations of the River Danube. SOM algorithms are a more sensitive tool for resolving species assemblage types than conventional clustering methods. We found five cluster groups, arranged primarily according to the sampling site; the clustering was but little influenced by the sampling time. We have shown character species for all five cluster groups, however, at the same time, we have shown that the difference between the clusters is more a result of the differences of relative abundance ratios between the cluster groups than a true difference of the species composition, although the uppermost stretch of the river and the source streams was well separated from the lower stretches. The clusters 1, 2 and 3 are in the German-Austrian stretch, the clusters 4 and 5 in the Slovakian-Hungarian stretch. This separation is also well in accordance with differences in water quality between these stretches.

Acknowledgements

This study was supported by the Hungarian National Research Fund (OTKA, K60452) and by the Bolyai János fellowship of the Hungarian Academy of Sciences.

References

ÁCS, É., SZABÓ, K., KISS, Á.K., TÓTH, B., ZÁRAY, GY. & KISS, K.T. (2006): Investigation of epilithic algae on the River Danube from Germany to Hungary and the effect of a very dry year on the algae of the River Danube. – Arch. Hydrobiol. Suppl. **158** (Large Rivers **16**): 389–417.

GOSSELAIN, V., CAMPEAU, S., GEVREY, M., COSTE, M., ECTOR, L., RIMET, F., TISON, J., DELMAS, F., PARK, Y.S., LEK, S. & DESCY, J.P. (2005): Diatom typology of low-impacted conditions at a multi-regional scale: combined results of multivariate analyses and SOM. – In: LEK, S., SCARDI, M., VERDONSCHOT, P.F.M., DESCY, J.P. & PARK, Y.S. (eds.): Modelling community structure in freshwater ecosystems: 317–342. Springer.

KOHONEN, T. (1995): Self-Organizing Maps. – Series in Information Sciences **30**. Springer, Heidelberg. Second edition 1997.

– (2001): Self-organizing maps, 3rd ed. – Springer, Berlin.

LECOINTE, C., COSTE, M. & PRYGIEL, J. (1999): Omnidia version 3: notice d'utilisation. – CLCI Monbazillac, Mai 1999, 46 pp.

– (1993): 'Omnidia': a software for taxonomy, calculation of diatom indices and inventories management. – Hydrobiologia **269/270**: 509–513.

PARK, Y.-S., GEVREY, M., LEK, S. & GIRAUDEL, J.L. (2005): Evaluation of relevant species in communities: Development of structuring indices for the classification of communities using a self- organizing map. – In: LEK, S., SCARDI, M., VERDONSCHOT, P., DESCY, J. P. & PARK, Y.-S (eds.): Modelling Community Structure in Freshwater Ecosystems: 369–380. Springer, Berlin.

RIMET, F., CAUCHIE, H.M., TUDESQUE, L. & ECTOR, L. (2005a): Use of artificial intelligence (MIR-max) and chemical index to define type diatom assemblages in Rhône basin and Mediterranean region. pp. 288-303. In.: LEK, S., SCARDI, M., VERDONSCHOT, P.F.M., DESCY, J.P. & PARK, Y.S. (eds.) (2005): Modelling community structure in freshwater ecosystems. Springer.

RIMET, F., ECTOR, L., HOFFMAN, L., GEVREY, M., GIRAUDEL, J.L., PARK, Y.S. & LEK, S. (2005b): Prediction with artificial neural networks of diatom assemblages in headwater streams of Luxembourg. – In: LEK, S., SCARDI, M., VERDONSCHOT, P.F.M., DESCY, J.P. & PARK, Y.S. (eds.): Modelling community structure in freshwater ecosystems. Springer.

TISON, J., GIRAUDEL, J.L., COSTE, M., PARK, Y.S. & DELMAS, F. (2004): Use of unsupervised neural networks for eco-regional zonation of hydrosystems through diatom communities: case study of Adour Garonne watershed (France). – Arch. Hydrobiol. **159**: 409–422.

VESANTO, J. (2000): Neural network tool for data mining: SOM Toolbox. – Proc. Symp. on Tool Environments and Development Methods for Intelligent Systems (TOOLMET2000): 184-196. Oulun yliopistopaino, Oulu, Finland.

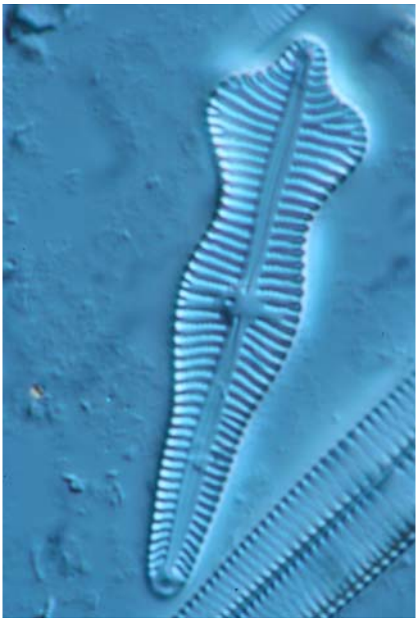
Accepted: 13 February 2007

Paper 3.

Comparative algological and bacteriological examinations on biofilms developed on different substrata in a shallow soda lake

Ács, É., Borsodi, A.K., Kiss, É., Kiss, K.T., Szabó, K.É., Vladár, P., **Várbíró, G.,** Záray, Gy.

Aquatic Ecology (2008) 42521-531



Aquat Ecol (2008) 42:521–531
DOI 10.1007/s10452-007-9132-0

Comparative algological and bacteriological examinations on biofilms developed on different substrata in a shallow soda lake

Éva Ács · A. K. Borsodi · É. Kiss · K. T. Kiss ·
K. É. Szabó · P. Vladár · G. Várbíró · Gy. Záray

Received: 15 February 2007 / Accepted: 9 July 2007 / Published online: 6 September 2007

Abstract According to the European Water Framework Directives, benthic diatoms of lakes are a tool for ecological status assessment. In this study, we followed an integrative sample analysis approach, in order to find an appropriate substratum for the water qualification-oriented biomonitoring of a shallow soda lake, Lake Velencei. Six types of substrata (five artificial and one natural), i.e., andesite, granite, polycarbonate, old reed stems, Plexiglass discs and green reed, were sampled in May and in November. We analysed total alga and diatom composition, chlorophyll *a* content of the periphyton, surface tension and roughness of the substrata and carbon source utilisation of microbial communities. Water quality index was calculated based on diatom composition. Moreover, using a novel statistical tool, a self-organising map, we related algal composition to substratum types. Biofilms on plastic substrates deviated to a great extent from the stone and reed substrata, with regard to the parameters measured, whereas the biofilms developing on reed and stone substrata were quite similar. We conclude that for water quality monitoring purposes, sampling from green reed during springtime is not recommended, since this is the colonization time of periphyton on the newly growing reed, but it may be appropriate from the second half of the vegetation period. Stone and artificially placed old reed substrata may be appropriate for biomonitoring of shallow soda lakes in both spring and autumn since they showed in both seasons similar results regarding all measured features.

Keywords Biomonitoring · Natural and artificial substrata · Water qualification

É. Ács (✉) · K. T. Kiss · K. É. Szabó
Hungarian Danube Research Station of HAS, Jávorka S. U. 14, God 2131, Hungary
e-mail: acs@ludens.elte.hu

A. K. Borsodi · P. Vladár
Department of Microbiology, Eötvös Loránd University, Budapest, Hungary

É. Kiss
Interfacial and Nanostructures Laboratory, Eötvös Loránd University, Budapest, Hungary

G. Várbíró
Environmetal Inspectorate of Transtiszanian Region, Debrecen, Hungary

Gy. Záray
Department of Analytical Chemistry, Eötvös Loránd University, Budapest, Hungary

Introduction

Surfaces exposed to adequate moisture, mainly solid–liquid phase interfaces, provide a niche for microbial colonisation, where the majority of microorganisms are thought to exist naturally as biofilms (Sutherland 2001). Biofilms provide both a spatially and temporally heterogeneous milieu, comprising an organised, three-dimensional structure of microbial communities embedded in an exopolysaccharide matrix (EPS), with a network of water-filled channels (Jenkinson and Lappin–Scott 2001). Bacteria are observed to be the primary colonisers, followed by the adhesion of other bacteria and different microorganisms, among others algae, with each cell potentially serving as a new surface for the wide variety of further colonising cells (Ács et al. 2000). The composition and the architecture of the biofilm are subject to both intrinsic and extrinsic factors, the latter including several physico–chemical attributes of the surrounding environment and the nature of the substratum.

Periphytic algal studies in lentic environments have lagged behind phytoplankton investigations. Nowadays, in several countries of the EU, in studies connected to the EU Water Framework Directive (WFD, European Parliament 2000, directive 2000/60/EC), benthic algae are considered key organisms for the assessment of the ecological quality of watercourses. For a long time, palaeolimnology has been using transfer functions based on weighted averaging which are developed on the basis of the composition of the recent flora and fauna as well as on their inherent measurable chemical parameters and the strength of the model; knowing the composition of the fossil flora and fauna, one can infer the environmental parameters of that time (e.g., Christie and Smol 1993; Schnitchen et al. 2006). It is clearly visible from these results that the epilithic algal community of the littoral zone shows good correlation with the water quality parameters and can be used well in the assessment of trophic condition of lakes in biomonitoring.

In Germany, water qualification investigations based on the benthos of lakes have a long history. Hofmann (1994) described the basis of the use of benthic algae in lakes as the indicator species of trophic and saprobic condition, when she noted that the difference in species composition on stone and plants did not cause significant change to the indices. Later, she (Hofmann 1999) developed a new index (the trophic index: TI) for more precise assessments, which was used subsequently by others also in the country (e.g., Seele et al. 2000). Later it was built into the national monitoring program, too, and developed further (Schaumburg et al. 2005). A lake diatom index was created (DI_{Seen}). The lake diatom index (DI_{Seen}) is calculated from two elements. One element (module) of which is the TI, while the other module is a reference index (reference species ratio, RAQ) . Schaumburg et al. suggested stone and the surface of the sediment as substrate where the water level was at least 30 cm away from the macrophyton.

In the work of Blanco et al. (2004), the indicator role of diatoms was investigated in six Spanish lakes of different water qualities for the determination of the trophic and saprobic condition of the lakes. The indices calculated by the OMNIDIA software were used in the qualification. Their results showed that the investigation of epiphytic algae and the diatom indices calculated by OMNIDIA software could be used well in the biological monitoring of lake ecosystems, too, for the estimation of the ecological condition, although the indices were basically developed for flowing waters.

Poulickova et al. (2004) investigated the indicator role of the diatoms of the littoral zone in Austrian lakes near the Alps; samples were collected from the surface of live reed, stone and sediment. The results showed that the species composition of the various substrates was significantly different that influencing the values of the indices. According to her, the most suitable substrate in the lake is the live reed, because dead reed, stone and sediment accumulate the dead diatoms from the previous years. Live reed grows anew every year, however, and the community developed on it reflects the actual condition. The investigations of Grimes et al. (1980) showed unambiguously that there was not a significant difference between the diatom composition of the periphyton of dead and live reed.

By comparing the element content of the periphyton of live and dead reed in Lake Balaton and Lake Velence with the element content of the reed, Lakatos (1983) found significant positive correlation in the case of the periphyton growing on live reed, while in the case of the dead reed there was not any significant correlation. This is an important statement, because it indicates that the dead reed serves only as substrate for the developing periphyton on it. So until the beginning of considerable decomposition, the algal composition of the periphyton developing on it probably reflects more the chemical components of the water than the periphyton on a live reed, which is in considerable metabolistic connection with the reed.

The introduction of an artificial substratum, through which more standardised and reproducible results were gained, has circumvented this problem to a certain extent. Furthermore, artificial substrata have been extensively used for the study of freshwater periphyton (reviewed e.g., in Kralj et al. 2006).

A proposal was formulated by King et al. (2006) concerning the sampling of benthos of lakes for the assessment of ecological conditions. For the proposal, she surveyed numerous literatures and summarized the most important statements regarding the topic. In her opinion, however, the type of the substrate should be selected in a way that it is the most characteristic substrate of the littoral region of the lake. In the question of the substrate, the available data in the literature are not sufficient and are very contradictory.

The community level BIOLOG (Hayward, California, USA) assay is a rapid and cultivation-independent method for distinguishing between microbial communities based on temporal, spatial and physicochemical parameters (Garland 1997). Community level physiological profiling (CLPP) of bacterial biofilms has proved an effective measure of biofilm condition changes due to copper-induced toxicity (Massieux et al. 2004).

Shallow soda lakes are characteristic for the Hungarian plane, but also occur elsewhere in Europe (e.g., Spain). Up to now, there have been no existing guidelines as to the periphyton-based water quality analysis of this lake type. The aim of our study was to find an appropriate substratum for the biomonitoring of shallow, sodic lakes, and to compare the biofilm development in an early and late phase of the vegetational period. Based on our previous studies, Lake Velencei was chosen as a model system for gaining new experience on the subject of substratum preference. A multidisciplinary approach was followed using physical parameters of substrata, community-level physiological profiling and microscopic alga investigations along with novel statistical tools.

Materials and methods

Study site and sample collection

Lake Velencei is the third largest lake of Hungary (24.5 km^2) with a 615 km^2 water catchment region. The lake is situated at the foot of the Velencei Mountains (which consists mainly of granite). Its proximity to the capital, 45 km SW, makes it an important recreation centre. It is a shallow soda lake, rich in dissolved salts and generally mesotrophic. The water level of the lake is regulated by two reservoirs situated in its catchment area and generally fluctuates between 140 cm and 160 cm measured at Agárd near our sampling point (N°47°11′21″, E°018°34′41″). On the SW part of the lake, a large reed-belt can be found and the water colour is dark brown. This part belongs to a nature conservation area. The middle and NE part of the lake is the "open water" area with grey water colour caused by mud stirred up by the wind. This part belongs to the holiday area. Since the larger, open water area of the lake is used for bathing, field experiments were conducted here.

Four different artificial substrata (granite [G], andesite [A], Plexiglass [P] discs (made by Satunaplex Ltd. Balatonfűzfő, Hungary), and polycarbonate [PC] slides (made by Arla-plast Ltd. Sweden)) and two natural substrata (old reed [OR] stems and living green reed [GR] stems) were used for biofilm development. The common reed (*Phragmites australis* (Cav.) Trin ex Steudel) stems located near the sample holder served as green reed substratum. The additional substrata (with the exception of Plexiglass in May, when it was not included) in 10 replicates were placed into the same Plexiglass holder (Fig. 1a), which was fixed to a rack equipped with a float (Fig. 1b) to keep the samples approximately 30 cm below the water surface. Biofilms developed on different substrata were studied in spring and autumn 2002. Surface sterilized granite, andesite and polycarbonate slides (12 × 3 × 0.7 cm) and Plexiglass discs with a diameter of 3 cm as well as old reed stems (all in ten replicates) placed in the same distance from one another and in vertical position were dipped into the lake water on 15 April and 23 September. Following 6 weeks' colonization, they were collected on 27 May and 4 November, respectively.

At the end of the colonisation periods (in May and November), green reed samples from the same depth were also collected. Samples were transported to the laboratory in a dark cooling box within 2 h of sampling. The biofilm from different substrata was washed into separate flasks filled with known volume of sterilized physiological saline solution (0.89 g NaCl l^{-1}, pH 8.0) with the help of a sterile brush.

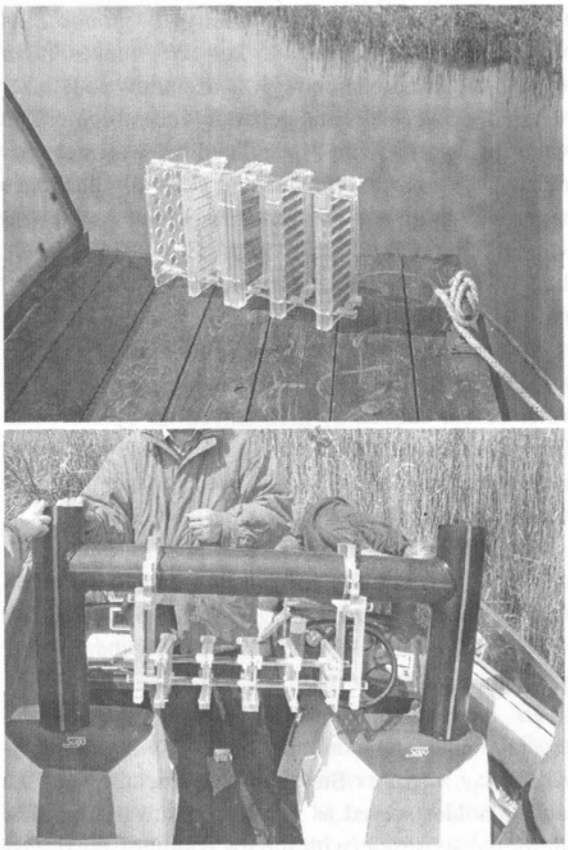


Fig. 1 Frames for the artificial substrata (above) and their floating holder used in Lake Velencei

Composite samples—collected from the same type of substrata—were divided into subsamples for different further analyses.

Surface tension and roughness of substrates

In order to determine the surface tension of the substrata, advancing and receding contact angles of water and formaldehyde droplets on the different substrata were measured by a contact goniometer. Based on the measured values, the polar and dispersion components of the surface tension were calculated according to Fowkes geometrical mean equation. Polarity (P) can be calculated based on the following equation $P = \frac{\gamma_s^{\mathrm{pol}}}{\gamma_s} \cdot 100$, where γ_s is the surface tension and γ_s^{pol} is its polar component of the solid material.

Determination of roughness was carried out using atomic force microscopy (AFM) at a resolution of 10×10 μm. Section analysis provided the roughness information on an arbitrary cross section of the surface. The typical roughness parameter given to characterize the surface roughness is Ra, which represents the mean roughness, which is the mean value of a surface cross section relative to the center line.

Algal abundance and identification

A part of the biofilm samples was preserved in Lugol's iodine solution. The quantitative determination and the identification of algae were completed by Utermöhl's (1958) method, according to the statistic instructions of Lund et al. (1958). For identification of diatoms, samples were treated with H_2O_2 and mounted in Naphrax.

Chlorophyll-a analysis

Another part was prepared on the day of sampling for chlorophyll *a* measurement. Following the method of Goodwin (1976) the chlorophyll *a* content was extracted by cc methanol and measured by a photometer at 653, 666 and 750 nm. The chlorophyll *a* concentration was calculated by the following equations: $C_a = 17.12 \times E_{666} - 8.68 \times E_{653}$, and the chlorophyll *a* concentration (μg cm^{-2}) $= m \times C_a \times 1000 \times M^{-1} \times t^{-1}$, where m was the volume of methanol used for the extraction (ml), M was the volume of the sample filtered onto the glass fiber filter (ml), and t was the area of substrates (cm^2).

Carbon source utilisation of bacterial communities

From the third part of the biofilm samples, a sole-carbon-source-utilisation test of bacterial communities was carried out. In order to inoculate BIOLOG GN2 microplates, an initial 10^{-1} biofilm dilution was prepared by suspending wet biofilm—equivalent to 10 g of dried matter—in 100 ml of sterile saline solution. Serial dilutions were prepared up to a dilution factor of 10^{-3}. A 20 ml aliquot of each dilution was shaken for 10 min and left settling for 10 min, in order to avoid interference in the assay by co-extracted biofilm components, causing unspecific turbidity and absorbance. From the diluted biofilm samples, 150–150 µl volumes were inoculated into the wells of BIOLOG GN2 microplates, providing 95 different carbon sources and a redox indicator (tetrazolium violet). The majority of carbon sources on the BIOLOG GN2 plates were carbohydrates (30), carboxylic acids (24) and amino acids (20). Polymers (5), amines/amides (6) and miscellaneous compounds (10) were represented in lower numbers (Preston-Mafham et al. 2002). The detection of substrate utilization—based on the reduction of the tetrazolium dye and therefore the optical density values—was measured at 595 nm with an ELISA Reader (Labsystems Multiscan PLUS), after incubation at 25°C from 24 h to 96 h. The type and number of the carbon sources utilised by the bacterial communities were evaluated according to Garland and Mills (1991).

Statistical analysis

Diatom diversity and evenness values were calculated using the Shannon–Weaver index. The self-organizing map (SOM) method introduced by Kohonen (2001) was used for grouping the samples and evaluating the structure of diatoms and the epiphyton. SOM can make a projection of the database into a two-dimensional hexagonal map. Closely related communities are placed into neighbouring hexagons by their similarities, while samples with different communities are in distant hexagons. The SOM can display the groupings of samples and species together; therefore, each species can be evaluated by its importance. The number of output neurons is an important factor in the analysis. Using the suggestion of Vesenato (2000), we determine the output number as $N_{\text{output}} = 5x\sqrt{N_{\text{samples}}}$. The rough tuning phase includes 2,000, while the fine tuning phase contains 8,000 iterations.

The structuring index (SI) was originally developed to define the species that show the strongest influence, according to the the organization of the SOM map (Park et al. 2005). Tison et al. (2004) used the SI to evaluate relevant diatom species in the classification of diatom communities. The set of species showing high SI can be considered as the indicator species. Taxa showing strong gradient display high SI values, whereas species showing weak gradient present low SI values. Thus, the higher the value of SI, the more relevant the variable is to the structure of the map. For clustering the SOM we used the K-means clustering technique, an algorithm for classifying or grouping objects based on attributes (in this case species composition) into K number of a group. The grouping is done by minimizing the sum of squares of the distances between data and the corresponding cluster centroid as the square error of each data point is calculated and clusters reformed such that the sum of square errors is made to be minimum. The SOM Toolbox (http://www.cis.hut.fi/projects/somtoolbox) was used to implement the SOM under a MATLABs environment.

The diatom index (IBD) was calculated by using OMNIDIA V4. software.

Comparison of the substrate oxidation results originating from the different samples was accomplished by statistical analysis (principal component analysis [PCA]), using the SYNTAX 2000 software package (Podani 2000).

Results

Among the tested substratum types, granite had the highest surface roughness, polycarbonate and Plexiglass substrata were two orders of magnitude smoother (granite: 510 nm, andesite: 420 nm, polycarbonate: 9 nm, Plexiglass: 4 nm) (Fig. 2). The polarity values were 76.2 for the granite, 68.4 for the andesite, 60.4 for the Plexiglass, 39.2 for the polycarbonate and 22.9 for the old reed substrata. The old reed substratum had the highest dispersion and lowest polar component of the surface tension, while granite had the highest polar component and the lowest dispersion (Fig. 3).

Studying the periphyton communities developed on different substrata in Lake Velencei, we found more than 120 algal species, most of them belonging to diatoms. Comparing the species number, Shannon diversity and evenness values of the different substrata, we found the smallest values on the plastic substrata in almost all cases (Fig. 4). Certain species such as *Navicula cryptocephala* Kütz. and *Oedogonium capitellatum* Wittrock were more dominant in the spring, while others such as e.g., *Nitzschia palea* (Kutzing) W. Smith and *Ctenophora (Fragilaria) pulchella* (Ralfs ex Kutz.) Williams et Round

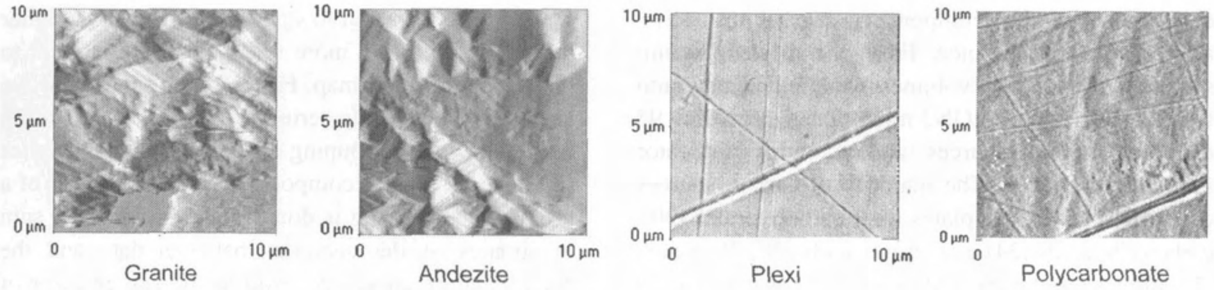


Fig. 2 Atomic force micrographs of the surface of the artificial substrata

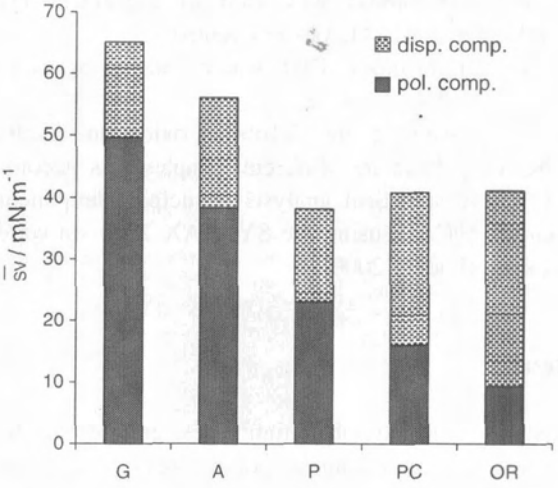


Fig. 3 Polar and dispersal components of the surface tension of the different substrata (G = granite, A = andesite, P = Plexiglass, PC = polycarbonate, OR = old reed)

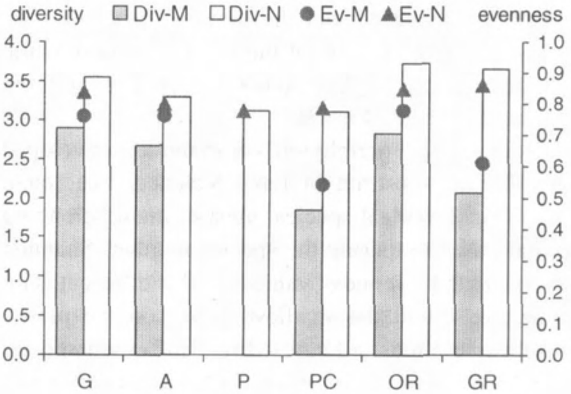


Fig. 4 Diversity and evenness of diatom communities on different substrata (Div = diversity, M = May, N = November, Ev = evenness, G = granite, A = andesite, P = Plexiglass, PC = polycarbonate, OR = old reed, GR = green reed)

dominated in the autumn (Fig. 5). The communities grown on Plexiglass and polycarbonate substrata were strongly dominated by *Achnanthidium minutissimum* (Kütz.) Czarnecki. Other characteristic species were *Microcystis aeruginosa* (Kütz.) Kütz. for granite, *Oedogonium capitellatum* for andezite, *Gomphonema olivaceum* var. *olivaceum* (Hornemann) Brébisson for both granite and andezite, while for green and old reed no significant differences were detected in the dominance of algal species.

According to the SOM analysis based on the algological data (species composition and their abundance), a strong seasonality can be observed (Fig. 6). Substratum types formed five distinct groups. In May granite, andesite and old reed formed a group, while polycarbonate and green reed formed a group, which clearly separating from the ones originated in November. In autumn Plexiglass and polycarbonate as well as granite, green reed and old reed comprised two distinct groups, while andesite was loosely connected to the latter (Fig. 6). IBD indices displayed the most analogous values on the granite, andesite and old reed substrata both in May and November, respectively (Fig. 7).

The chlorophyll *a* content was highest on the granite (326 µg cm^{-2}) and andesite (328 µg cm^{-2}) in May, and on the old reed in November (51 µg cm^{-2}), and it was the lowest on green reed (24 and 0.5 µg cm^{-2}) on both occasions.

Microbial BIOLOG substrate utilization patterns of biofilm communities developed on different substrata were compared on the average well color development value (AWCD; average of absorbance values from 95 wells). With the exception of green reed substratum, the AWCD values were higher in May than in November (Fig. 8). Out of green reed

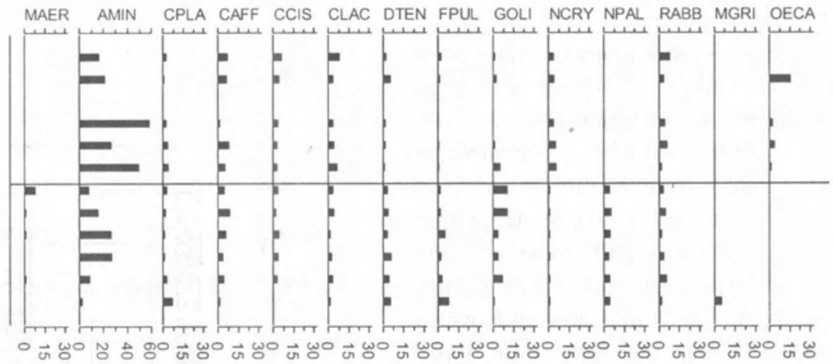


Fig. 5 Dominant alga species in the samples. (MAER = *Microcystis aeruginosa* (Kütz.) Kütz., AMIN = *Achnanthidium minutissimum* (Kütz.) Czarnecki, CPLA = *Cocconeis placentula* Ehrenberg var. *lineata* (Ehr.) Van Heurck, CAFF = *Cymbella affinis* Kutz, CCIS = *Cymbella cistula* (Ehrenberg) Kirchner, CLAC = *Encyonema (Cymbella) lacustre* (Agardh) F.W.Mills, DTEN = *Diatoma tenuis* Agardh, FPUL = *Ctenophora (Fragilaria) pulchella* (Ralfs ex Kutz.) Williams et Round, GOLI = *Gomphonema olivaceum* var. *olivaceum* (Hornemann) Brébisson, NCRY = *Navicula cryptocephala* Kutz., NPAL = *Nitzschia palea* (Kutzing) W.Smith, RABB = *Rhoicosphaenia abbreviata* (C.Agardh) Lange-Bertalot, MGRI = *Monoraphidium griffithii* (Berk.) Kom.-Legn., OECA = *Oedogonium capitellatum* Wittrock. See more abbreviation on Fig. 4)

substratum, the higher substrate utilization values were measured in case of granite and old reed substrata on both occasions. According to the bacteriological investigations, carbohydrates (sucrose, α-D-glucose, maltose, D-mannitol and D-trehalose) were the most preferred among the carbon sources. The number of carbon sources characteristic for certain microbial communities was the largest in the case of granite while this number was lower in the case of the other substrata types. Among the carbon sources the utilization of D-fructose and D-gluconic-acid was characteristic in May, while glycogen, tween 80, D-mannose and D-sorbitol in November. Community level physiological profiles (CLPP) based on the colour development on BIOLOG plates were

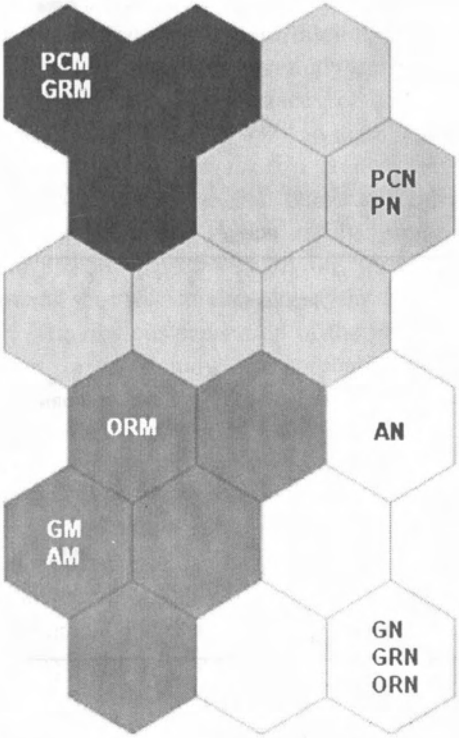


Fig. 6 Classification of the samples through SOM, the SOM virtual units were classified into four clusters by K-means clustering algorithm (see abbreviation on Fig. 4)

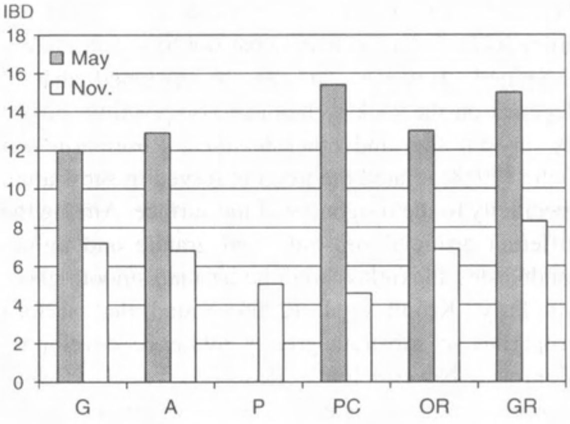


Fig. 7 Values of the indices IBD (Indice Biologique Diatomées) in May and November, based on the benthic diatom composition of the different substrata (see abbreviation on Fig. 4)

examined by PCA. In PCA, microbial communities from different substrata served as objects and the 96 h absorbance values of carbon source utilization as variables. According to the ordination, the first two PC axes accounted for 69% and 63% of the variance in the data in May and November samples, respectively (Fig. 9). In May, samples from andesite, granite and old reed substrata grouped tightly together along PC1, while polycarbonate and green reed clearly separated from them. As Fig. 9 indicates, in November the differences among the substratum types increased as the objects scores diverged along the PC1. Apart from this, in autumn CLPP of bacterial communities from granite and old reed substrata were the most similar to each other with their highest object scores for axis 1.

Discussion

The potential of epilithic microalgae for deriving nutrients from rock substrata has not been extensively examined. However, it can be assumed that it depends on the rock's chemical composition, porosity, crystal size and other features. Rosemarin and Gelin (1978) related the trend observed in substratum specificity to the roughness of the surface. Among the different artificial substrata, red granite and lightly sandblasted Plexiglass were better than smooth glass. Similarly, Kröpfl et al. (2006) found that surface properties of substrata greatly influence developing biofilms. Nevertheless, Rosowski et al. (1986) showed that the organic layer developing on the substrata may mask the substratum effect already after 2 weeks of colonization.

Differences in the polar and dispersal components of the surface tension of the different substrata used in our study seemed to influence the composition of the mature periphyton. Comparing species number, Shannon diversity and evenness values of algae and AWCD values of bacterial communities detected on the different substrata, we found smaller values on the smooth surfaced plastic substrata (Plexiglass and polycarbonate) than on the rougher surfaced ones (andesite, granite and reeds). As can also be seen from the diatom composition, this may have been primarily caused by the dominance of *A. minutissimum* on the plastic substrata.

Based on the results of the SOM analysis, it can be stated that the separation of plastic substrata—both in

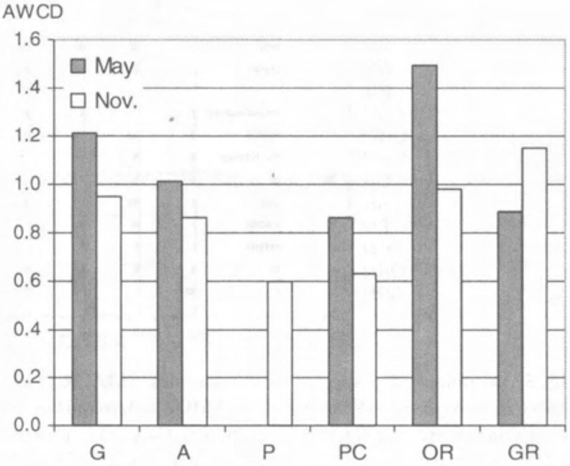


Fig. 8 Differences of the average well color development (AWCD) values of microbial communities detected on BIOLOG GN2 microplates originating from different substrata placed in Lake Velencei (see abbreviation on Fig. 4)

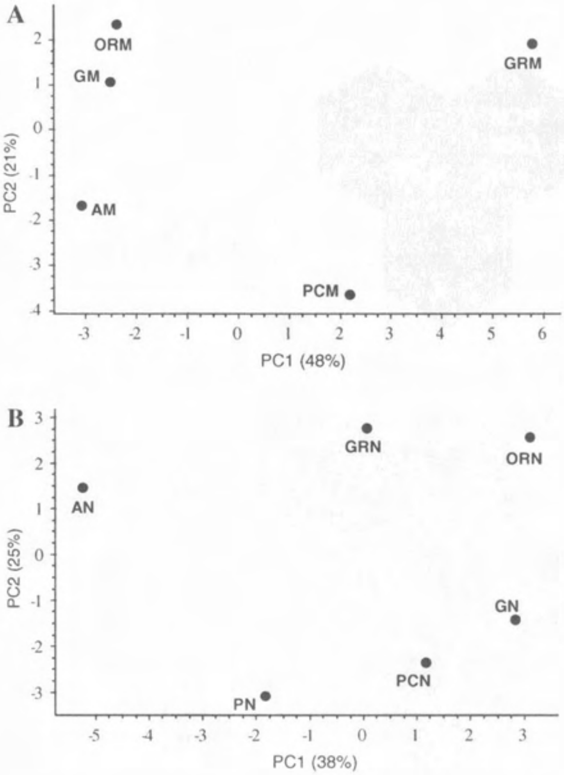


Fig. 9 Scattergram of substrata based on bacterial data (the colour development on BIOLOG plates. A = in May, B = in November. M = May, N = November, see more abbreviations on Fig. 4)

May and November—is in good correspondence with the dominance of certain species. In May, the two rocks (granite, andesite) and old reed showed the greatest similarity to each other and it was reflected in the chlorophyll *a* contents, diatom indices and the carbon source utilisation test of bacterial communities. In November, the most similar periphyton communities were also detected on the granite and old reed substrata, while the Plexiglass was greatly deviant both in algological and in bacteriological respects. The analysis of BIOLOG carbon source utilization patterns revealed a more expressed heterogeneity among the same type of substrata (e.g., rocks, plastics or reeds) compared to the different type of substrata, especially granite and old reed. In both May and November sample series, granite and old reed had the most similar and the most diverse bacterial communities according to their BIOLOG carbon source utilization, respectively.

Comparing the metabolic fingerprints as well as algological data of biofilm samples, we found that the positions of object scores and the results of SOM analysis of natural substrata indicated that bacterial and algal communities developed on green and old reed substrata were more dissimilar in May than in November. In an earlier study performed on green and old reed stem surfaces of Lake Velencei in summer (Ács et al. 2003) no significant differences were found in the species composition algae and metabolic activities of bacterial communities in mature biofilms. These results suggest that the differences between them on green and old reed stems start to diminish already in summer.

The obvious separation of the green reed substratum in this study was probably caused by the different colonization phase of the periphyton (early colonization phase communities differ from mature communities). There are three different hypotheses concerning the interaction between epiphytes and their host plants: (a) some studies showed a positive nutrient interaction (e.g., Wetzel 1983), when the substratum is a second nutrient source for the algae; (b) others indicate a negative interaction, when the macrophyte releases allelopathic substances that inhibit epiphyte growth (e.g., Anthoni et al. 1980); (c) a third hypothesis states that macrophytes are only a neutral site for attachment (e.g., Cattaneo and Kalff 1979). Our results based on the aligned algological and bacteriological studies on biofilms indicated that in May mainly dissimilar or opposite plant-microbe interactions could be dominant on green and old reed substrata, whereas in November these interactions became more complex and related to each other. As Burkholder (1996) also pointed out, each of these hypothesized interactions might be true, depending on the season, water quality (with respect to nutrient content) and plant substratum conditions. Significantly different epiphyte communities have been found on natural as opposed to plastic plants in mesotrophic and oligotrophic lakes (Cattaneo and Kalff 1978), while in eutrophic lakes, they did not exhibit any substratum preference (Eminson and Moss 1980).

Summarising, both the algological and bacteriological investigations showed that periphyton developing on stone and reed material were not significantly different. Especially in the case of old reed material where even the diatom indices showed similar values. However, it has to be noted that in this experiment cut sterilized and artificially placed reed was used as old reed substrata rather than naturally growing, rooted reed from the lake. Using rooted reed might have resulted in greater differences of the community composition between reed and stone periphyton by November, since active microbiological degradation processes would have influenced algal composition to a greater extent.

Conclusions

For water quality monitoring purposes, sampling from green reed during springtime is not recommended, since this is the colonization time of periphyton on the newly growing reed. Consequently, algal and bacterial composition cannot unambiguously be in consonance with water quality, but it may be appropriate from the second half of the vegetation period. Stone and artificially placed old reed substrata may be appropriate for biomonitoring of shallow soda lakes in both (spring and autumn) periods inasmuch as they showed highly similar results regarding all measured features.

Acknowledgements The authors would like to thank for the supports of a Bolyai grant of the Hungarian Academy of Sciences and that of the Hungarian National Fund (OTKA T032444).

References

Anthoni U, Christophersen C, Madsen JO, Wium-Andersen S, Jacobson N (1980) Biologically active sulphur compounds from the green alga *Chara globularis*. Phytochemistry 19:1228–1229

Ács É, Kiss KT, Szabó K, Makk J (2000) Short-term colonization sequence of periphyton on glass slides in a large river (River Danube, near Budapest). Algol Stud Arch Hydrobiol Suppl 136:135–156

Ács É, Borsodi AK, Makk J, Molnár P, Mózes A, Rusznyák A, Reskóné NM, Kiss KT (2003) Algological and bacteriological investigations on reed periphyton in Lake Velencei, Hungary. Hydrobiologia 506–509:549–557

Blanco S, Ector L, Bécares E (2004) Epiphytic diatoms as water quality indicators in Spanish shallow lakes. Vie Milieu 54:71–79

Burkholder JM (1996) Interaction of benthic algae with their substrata. In: Stevenson RJ, Bothwell ML, Lowe RL (eds) Algal ecology. Freshwater benthic ecosystems. Academic Press, San Diego, New York, Boston, London, Sydney, Tokyo, Toronto, pp 253–297

Cattaneo A, Kalff J (1978) Seasonal changes in the epiphyte community of natural and artificial macrophytes in Lake Memphremagog (Que. & Vt.). Hydrobiologia 60:135–144

Cattaneo A, Kalff J (1979) Primary production of algal growing on natural and artificial aquatic plants: a study of interaction between epiphytes and their substrate. Limnol Oceanogr 24:1031–1037

Christie CE, Smol JP (1993) Diatom assemblages as indicator of lake trophic status in Southeastern Ontario lakes. J Phycol 29:575–586

Eminson D, Moss B (1980) The composition and ecology of periphyton communities in freshwaters. I. The influence of host type and external environment on community composition. Br Phycological J 15:429–446

Garland JL, Mills AL (1991) Classification and characterization of heterotrophic microbial communities on the basis of patterns of community level sole-carbon-source utilization. Appl Environ Microbiol 57:2351–2359

Garland JL (1997) Analysis and interpretation of community-level physiological profiles in microbial ecology. FEMS Microbiol Ecol 24:289–300

Goodwin TW (ed) (1976) Chemistry and biochemistry of plant pigments. Academic press, London, p 235

Grimes JA, Larry L, Clair St, Rushforth SR (1980) A comparison of epiphytic diatom assemblages on living and dead stems of the common grass Phragmites australis. Great Basin Nat 40:223–228

European Parliament (2000) Directive 2000/60/EC of the european parliament and of the council establishing a framework for community action in the field of water policy. Official J L327:1–72

Hofmann G (1994) Aufwuchs-Diatomeen in Seen und ihre Eignung als Indikatoren der Trophie. Bibliotheca Diatomologica 30. pp 241. J. Cramer. Berlin, Stuttgart

Hofmann G (1999) Trophiebewertung von Seen anhand von Aufwuchsdiatomeen, pp 319–333. In: Tümpling W, Friedrich G (eds) Methoden der Biologischen Wasseruntersuchung. Band 2. Biologische Gewässeruntersuchung, Gustav Fischer Jena, Stuttgart, Lübeck, Ulm

Jenkinson HF, Lappin-Scott HM (2001) Biofilm adhere to stay. Meeting Report in: TRENDS Microbiol 9:9–10

King L, Clarke G, Bennion H, Kelly M, Yallop M (2006) Recommendation for sampling littoral diatoms in lakes for ecological status assessments. J Appl Phycol 18:15–25

Kohonen T (2001) Self-organizing maps, 3rd edn. Springer, Berlin

Kralj K, Plenković-Moraj A, Gligora M, Primc-Habdija B, Šipoš L (2006) Structure of periphytic community on artificial substrata: influence of depth, slide orientation and colonization time in karstic Lake Visovačko, Croatia. Hydrobiologia 560:249–258

Kröpfl K, Vladár P, Szabó KÉ, Ács É, Borsodi AK, Szikora Sz, Caroli S, Záray Gy (2006) Chemical and biological characterisation of biofilms formed on different substrata in Tisza river (Hungary). Environ Pollut 144:626–631

Lakatos Gy (1983) Accumulation of elements in biotecton forming on reed (*Phragmites australis*) in two shallow lakes in Hungary. In: Proceedings of international symposium on aquatic macrophytes, The Netherlands, Nijmegen, pp 117–122

Lund JWG, Kipling C, Lecren ED (1958) The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. Hydrobiologia 11:143–170

Massieux B, Boivin MEY, van den Ende FP, Langenskiöld J, Marvan P, Barranguet C, Admiraal W, Laanbroek HJ, Zwart G (2004) Analysis of structural and physiological profiles to assess the effects of Cu on biofilm microbial communities. Appl Environ Microbiol 70:4512–4521

Park YS, Gevrey M, Lek S, Giraudel JL (2005) Evaluation of relevant species in communities: development of structuring indices for the classification of communities using a self- organizing map. In: Lek S, Verdonschot P, Descy JP, Park YS (eds) Modelling community structure in freshwater ecosystems. Springer, Berlin, pp 369–380

Podani J (2000) Introduction to the exploration of multivariate biological data. Backhuys, Leiden

Poulicková A, Duchoslav M, Dokulil M (2004) Littoral diatom assemblages as bioindicators of lake trophic status: a case study from perialpine lakes in Austria. Eur J Phycol 39:143–152

Preston-Mafham J, Boddy L, Randerson PF (2002) Analysis of microbial community functional diversity using sole-carbon-source utilisation profiles—a critique. FEMS Microbiol Ecol 42:1–14

Rosemarin AS, Gelin C (1978) Epilithic algal presence and pigment composition on naturally occurring and artificial substrates in Lake Trummen and Fiolen, Sweden. Verh Int Ver Limnol 20:808–813

Rosowski JR, Hoagland KD, Aloi JE (1986) Structural morphology of diatom-dominated stream biofilm communities under the impact of soil erosion. In: Evans LV, Hoagland KD (eds) Algal biofouling, Elsevier, Amsterdam, New York, Oxford, Tokyo Studies in Environmental Sciences 28:247–298

Schaumburg J, Schmedtje U, Schranz C, Köpf B, Schneider S, Stelzer D, Hofmann G (2005) Instruction protocol for the ecological assessment of lakes for implementation of the EU Water Framework Directive: macrophytes and phytobenthos. Manuscript, Bavarian Water Management Agency, München, p 44

Schnitchen C, Magyari E, Charman DJ, Tóthmérész B, Grigorszky I, Béres V, Braun M (2006) The possibility of reconstruction of past hydrological changes in a Sphagnum bog by the use of testate amoebae (Rhizopoda: Testacea) in the Carphatian Basin. Verh Int Ver Limnol 29:1751–1756

Seele J, Mayr M, Staab F, Raeder U (2000) Combination of two indication systems int he pre-alpine lakes—diatom index and macrophyte index. Ecol Modell 130:145–149

Sutherland IW (2001) The biofilm matrix—an immobilized but dynamic microbial environment. Review in: TRENDS in Microbiology 9:222–227

Tison J, Giraudel JL, Coste M, Park YS, Delmas F (2004) Use of unsupervised neural networks for eco-regional zonation of hydrosystems through diatom communities: case study of Adour Garonne watershed (France). Arch Hydrobiol 159:409–422

Utermöhl H (1958) Zur Vervollkommung der quantitativen Phytoplankton-Methodik. Mitt Verh Int Ver Limnol 9:1–38

Vesanto J (2000) Neural network tool for data mining: SOM Toolbox. In: Proceedings of Symposium on tool environments and development methods for intelligent systems (TOOLMET2000), Oulun yliopistopaino, Oulu, Finland, pp 184–196

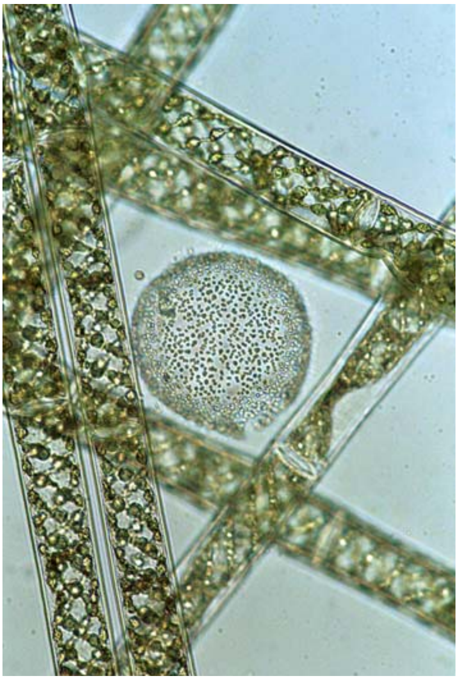
Wetzel RG (1983) Attached algal-substrata interactions: fact or myth, and when and how? In: Wetzel RG (ed) Periphyton of freshwater ecosystems. W. Junk, The Hague/Boston/Lancaster, pp 207–215

Paper 4.

A new evaluation technique of potamo-plankton for the assessment of the ecological status of rivers

Gábor Borics, Gabor Varbiro, Istvan Grigorszky, Eniko
Krasznai~Sándor Szabó and Keve Tihamer Kis

Large Rivers Vol. 17, No. **3-4**
Arch. Hydrobiol. Suppl. 16113-4, **p.** 465-486, Oktober **2007**



Large Rivers Vol. **17**, No. 3–4
Arch. Hydrobiol. Suppl. 161/3–4, p. 465–486, Oktober 2007

A new evaluation technique of potamo-plankton for the assessment of the ecological status of rivers

Gábor Borics[1], **Gábor Várbíró**[1], **István Grigorszky**[2], **Enikő Krasznai**[3], **Sándor Szabó**[4] and **Keve Tihamér Kiss**[5]

With 5 figures and 5 tables in the text

Abstract: Based on the phytoplankton associations described for lakes (REYNOLDS, 2002), an assessment method has been elaborated for rivers. All phytoplankton associations were evaluated and scored by a number between 0-5. As many rivers can be defined as shallow, turbid, mesotrophic ecosystems of short residence time, those associations that prefer this type of environment were given high factor numbers, and those that are typical of stable hypertrophic lakes have got the lowest values. Highest values were given to those assemblages that contain mainly periphytic diatoms. To achieve an index, each species in the sample must be assigned to the appropriate functional group. Then the relative share of each functional groups are calculated. Relative shares are then multiplied by the factor number. The sum of these scores is the index. The reference values of the upper river sections are close to 5, while those of the lower river stretches are approximately 4. The method has been tested with hundreds of phytoplankton samples, it is simple, and after applying to a phytoplankton database can be computerised easily. Another advantage of the method that it is not restricted to a specific geographic region.

Introduction

The increasing demand for developing new assessment methods for evaluating the ecological status of lakes and rivers has mainly been fuelled by the WFD (2000) in the recent years. The Directive does not deem necessary to investigate the phyto-

---

Authors' addresses: [1] Environmental Protection, Nature Conservation and Water Authority, Trans-Tiszanian Region, Hatvan u. 16, H-4025 Debrecen, Hungary; E-mails: boricsg@gmail.com; varbirog@gmail.com [2] Debrecen University, Department of Hydrobiology, Egyetem tér 1, H-4010 Debrecen, Hungary; E-mail: gege@tigris.klte.hu [3] Debrecen University, Department of Ecology, Egyetem tér 1; H-4010 Debrecen, Hungary; E-mail: ekrasznai@gmail.com [4] College of Nyíregyháza, Department of Botany, P.O. Box 166, H-4401 Nyíregyháza, Hungary; E-mail: szabos@zeus.nyf.hu [5] Institute of Ecology and Botany, Hungarian Danube Research Station of the Hungarian Academy of Sciences, Jávorka S. u. 14, H-2131Göd, Hungary; E-mail: kis7972@ella.hu

0945-3784/07/0161-0465 $ 5.50

plankton in rivers which is not surprising, because the streams owing to the predominance of allochthonous organic material over the autochthonous primary production are naturally heterotrophic systems (REYNOLDS 2000). Nevertheless several investigations proved (UHERKOVICH 1971; DESCY et al. 1988; KISS 1994; DOKULIL 1996; SKIDMORE et al. 1998) that the lower sections of the large, sluggish rivers (especially in lowland areas) can be characterized by highly eutrophic phytoplankton. Analysis of the Hungarian water quality database that contains more than 60,000 chlorophyll-*a* data for the Hungarian rivers (VÍGH 2002) has also demonstrated the development of eutrophic, and even worse, hypertrophic situations in most of the lowland rivers in Hungary (Fig. 1). Undesirable increment of the phytoplankton biomass is at least as problematic in riverine ecosystems as it is in lakes, because it may have deleterious effect on the other assemblages of the river, and it impairs water uses.

The basic characteristics of the rivers are, that during their course they change their trophic state from heterotrophy to autotrophy. In line with this, the philosophy of the river monitoring approaches also have to change. Although the water quality monitoring of the streams is usually based on the composition of the macroinvertebrate fauna and the benthic diatoms, because of the above mentioned reasons, the investigation of the riverine phytoplankton for monitoring purposes is unavoidable. Methodology of the rheoplankton investigations (including the sampling, counting etc.) has been worked out in details for decades (UHERKOVICH 1971; KISS et al. 1995, 1996). These are being applied as routine techniques by those organizations which are responsible for the water quality monitoring. Nevertheless phytoplankton-based quality assessment of the rivers has not been elaborated yet, therefore in several countries the old saprobic systems (PANTLE & BUCK 1955) are in current use.

The aim of this study is to present a new phytoplankton-based method, by which the ecological state of the rivers can be assessed. In compliance with the expectations of the WFD, we make an attempt to give the characteristic phytoplankton assemblages of the rivers.

Elaboration of the new assessment method

Earlier assessment methods tried to characterize the water quality on an absolute scale, and gave some kind of "water-goodness" instead of the real water quality. The new procedures have to be based on the degree of deviation from a pristine ecological state. This approach needs a detailed description of the type specific reference conditions for those biological elements which the assessment methods are based on. The main problem with this philosophy is to find enough aquatic ecosystems of pristine state to the "reference conditions" with the required statistical certainty. The probability of finding untouched aquatic ecosystems among the large lakes and rivers of higher order is virtually zero (REVENGA et al. 2000).

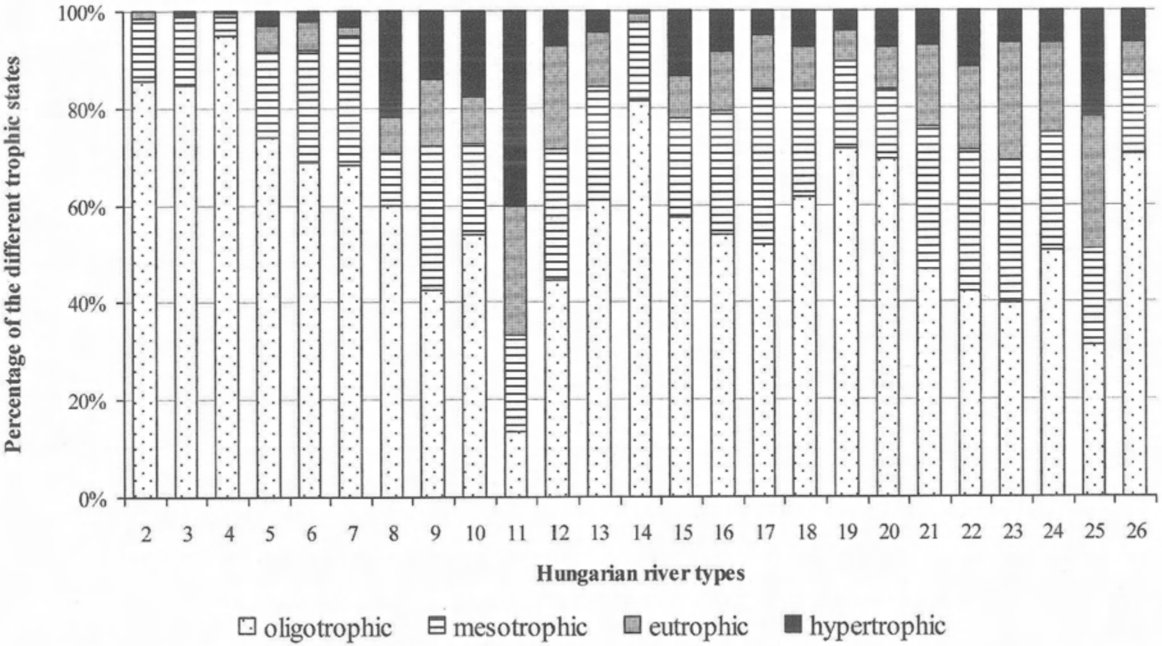


Fig. 1. Trophic status of the Hungarian rivers (brooks and small streams: 2, 3, 4, 11; streams: 5, 8, 9, 12; small rivers: 15, 16, 17, 18, 21, 22; rivers: 6, 7, 10, 13; large rivers: 14, 19; very large rivers: 20, 23, 24, 25; large, artificial canals: 26).

Consequently, simplified mechanistic models might be used, besides historical data and the expert judgement, to describe characteristic algal types in rivers.

Basic hydraulic features of rivers are used in a simple model to differentiate sections where benthic elements are dominant and where considerable populations of euplanktonic elements develop.

Theoretical bases

The premise is that algae will grow wherever and whenever they have the opportunity to do so. For the development of any algal population at a given place three basic criteria need to be present simultaneously:

1. inocula of the species,
2. appropriate environmental variables (temperature, light, nutrients),
3. sufficient time.

If any of these is missing, the population has no chance to develop. In standing waters, with the exeption of those with extremely high flushing rate, the above three criteria are fulfilled for the phytoplankton. In dynamic systems, such as pristine rivers, the probability of the coexistence of all the three factors changes along the river channel.

1. The phytoplankton of rivers can be traced back to the inocula from lentic habitats that are connected with the main river. Natural "alga-cradles" can be the slow flowing or dead arms of the river (Kiss 1987, 1997; Stoyneva 1994; Skidmore et al. 1998), separated oxbows and swampy areas of the floodplain (Uherkovich 1971), or the dead zones of the river channel (Reynolds 2000). The number of such habitats has been reduced as a result of river regulation. Similar lentic habitats are created by human activities as impoundments, cascades, or off-river reservoirs. In many cases these impounded waters mimic natural lakes. Contrary to natural lentic habitats, which can inoculate the rivers with euplanktic elements mainly during low or high discharge periods, several of the artificial water bodies can be considered as continuous chemostats. Their efficiency in inoculation of rivers is supposedly higher than that of the natural ones.
2. As far as the appropriate environmental parameters are concerned, rivers usually contain sufficient inorganic nutrients (Felföldy 1969; Kiss 1994) but phytoplankton can be frequently light limited because of the suspended particles.
3. Long term investigation of potamoplankton (Kiss 1985; Kiss et al. 1998; Kiss & Schmidt 1998; Schmidt 1994; Dokulil 1991) provided corroborative evidence that the hundreds of impoundments built in the Carpathian basin, exerted dramatic impacts on the phytoplankton composition and biomass by improving the light climate (Rákóczi 1989) and by increasing the residence time (Reynolds & Glaister 1993). Except for rivers of higher order, where the

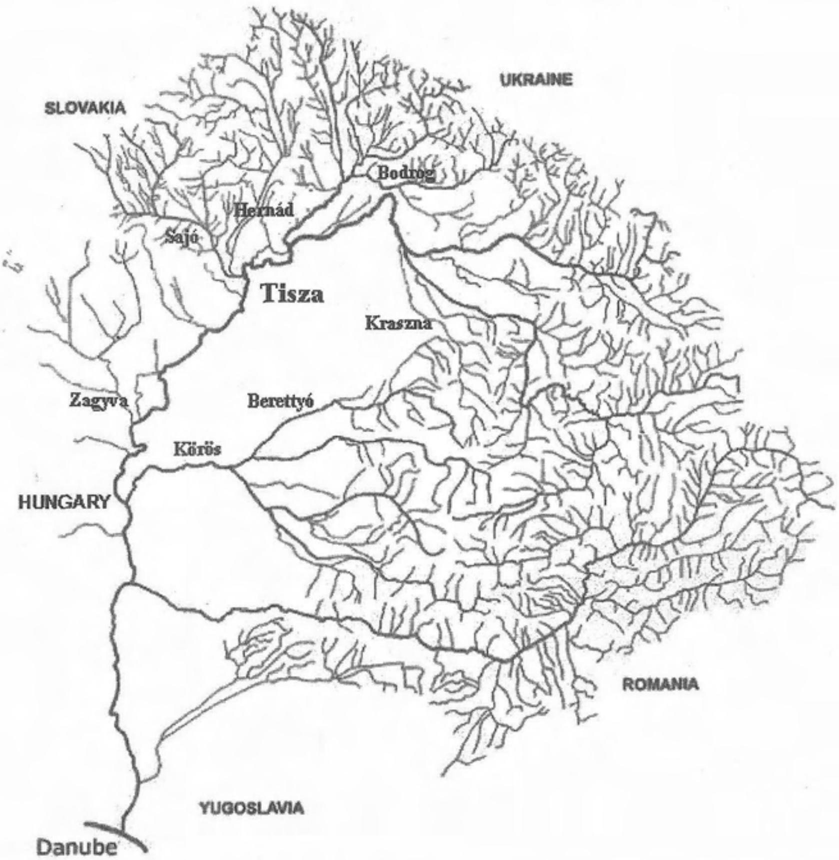


Fig. 2. River Tisza and its main tributaries.

phytoplankton production is curbed primarily by the light conditions, residence time is the key factor in the upper and middle sections of rivers.

Estimation of the residence time for rivers is easier than for lakes, because evapotranspiration which is the most problematic component of the water-balance of lakes is unimportant. In the case of rivers, the residence time depends on the length of the investigated river section and water velocity which can be calculated from the slope, the geometric characteristics of the streambed and from the friction factor. These traits vary considerably along the river. Consequently, accurate calculations can only be accomplished at a given point of a river for a given time. However the residence time can be estimated with high certainty from field measurements. The calculated residence times of the River Tisza and its main tributaries (Fig. 2) are shown in Table 1. The values have been calculated from the

Table 1. Calculated residence time and certain hydrological features of the River Tisza and its main tributaries.

Name of the river	Section	Length (km)	Velocity of the current (m s^{-1})	Covered distance/day (km)	Residence time (day)	*Number of generations
Tisza	Upper section	258	2	172.8	1.5	1.0
Tisza	Middle section	349	1	86.4	4.0	2.7
Tisza	Lower section	338	0.5	43.2	7.8	5.2
Tisza	Whole river section	945			13.4	8.9
Körös	Whole river section	363.4	1	86.4	4.2	2.8
Bodrog	Whole river section	266.9	1	86.4	3.1	2.1
Kraszna	Whole river section	193.4	1	86.4	2.2	1.5
Zagyva	Whole river section	179.4	1	86.4	2.1	1.4
Sajó	Whole river section	229.4	1	86.4	2.7	1.8
Hernád	Whole river section	282	1	86.4	3.3	2.2
Berettyó	Whole river section	204	1	86.4	2.4	1.6

* Estimated number of phytoplankton generations on the given river section, calculated with 1.5 day doubling time.

length of the river section and from velocity measurements (LÁSZLÓFFY 1982). Velocity of the current is approximately 1.5–2 m s^{-1} in mountainous areas, and 0.5–1 ms^{-1} in the lowlands at low discharge. For the tributaries of the River Tisza, current velocity of 1 m s^{-1} was assumed. Since we aimed for characterizing pristine situations, the river-impoundments have not been considered.**

With respect to the trophic state of the rivers, the last two columns in Table 1 are rather surprising. Apart from the lower sections of the Tisza, the residence times of the other rivers are rather short (Table 1). If we calculate generation times setting the doubling time of centric diatoms, a distinct feature of riverine plankton assemblages, to 1.5 days under ideal conditions (KISS 1996; KISS et al. 1997, 1998; SKIDMORE et al. 1998) the number of the generations attainable is also very small. Consequently, the hypertrophic state of these rivers is not explicable. The phytoplankton of natural rivers with short residence time should consist of benthic algae, mainly diatoms (tychoplanktic elements). Nevertheless, euplanktic elements are expected to dominate in the middle or in the lower segments of the large rivers, and the plankton is akin to those of shallow, eutrophic lakes (REYNOLDS et al. 1994).

Due to the anthropogenic modification of the river corridors the phytoplankton species composition and biomass of the rivers in the Carpathian Basin differs considerably from the picture outlined above. The chlorophyll-*a* data of the Hungarian water quality database (VÍGH 2002) and hundreds of algological investigations prove that even rivers of second or third order with residence times as short as one or two days, can be highly eutrophic, and their phytoplankton assemblages mainly consist of euplanktic elements. These characteristics indicate that lake-like conditions prevail at the upper segment of the river basin. The different impoundments, the artificial lakes, ponds, and off-river reservoirs that are in connection with the main river channel exert influence on the natural riverine assemblages (THORNTON et al. 1996). The phytoplankton of the small and medium lowland rivers may entirely loose their riverine character, and comprise an eclectic mixture of the different types of eutrophic lake-phytoplankton. For the water quality and the water-use, however, it is important which types of the limnetic phytoplankton become dominant in rivers, because development of nuisance algal blooms (Cyanobacteria) may have deleterious effect on the aquatic biota and can cause health problems.

** The calculated residence times appear to be rough estimations but it is worth mentioning that the values concerning the River Tisza are in accordance with those that were measured during the cyanide pollution in February 2000. It took the contaminated water 14 days to pass along the entire river section. Although the accident lasted only a few minutes, the contamination extended for 60 hours at the lower section. This furnishes evidence for the existence of the dead zones, which in this case contributed to the enlargement of the residence time by 20 %.

Evaluation of the assemblages

Elaboration of a phytoplankton-based quality assessment method needs the evaluation of phytoplankton associations in rivers. Therefore, we evaluate here the occurrence of the elements of the phytoplankton assemblages (Reynolds et al. 2002; Padisák 2003) by estimating how far, or rather, how close these limnetic assemblages are to those types that can be considered as reference algal assemblages of rivers. We focus on those environmental elements that are specific for rivers (short residence time), or important from the environmental point of view (trophic state, toxic algae). We also consider that the phytoplankton of the rivers is quite similar to those of the shallow, turbid lakes (Reynolds et al. 1994). Finally, the assemblage is characterized numerically by a compound factor (F) estimated from the following components.

Each functional group of algae (Reynolds et al. 2002) was assigned to trophic state, turbulence and residence time on a scale from 0 to 5, on the basis of:

1. trophic state: (hypertrophic 0; eutrophic 1; mesoeutrophic 2; mesotrophic 3; oligo-mesotrophic 4; oligotrophic 5),
2. turbulence character: the habitats of the codons were scored by their turbulence preference (totally standing waters like pools among macrophytes 0, lentic 1, slightly lotic (mixed epilimnion) 2, medium lotic (mixed layers) 3, lotic 4 (well mixed layers), highly lotic 5),
3. residence time, needed for the development of the given assemblage. Lower values were given to those assemblages that need relatively long time for their development. We can conclude to this from the biomass of the algal species belonging to the assemblage. Species of large biovolume usually can be dominating in the late stage of succession, whereas small-celled, invasive species are mostly the members of young, pioneer associations. During the estimation of this number the succession sequence of the codons (Reynolds 2002; Padisák 2003) was also considered (climax assemblages 1, pre-climax assemblages 2, transitional assemblages 3, pioneer assemblages of lakes and rivers 4, benthic codons 5).
4. As a highly subjective element, we also estimated how the occurrence of the given assemblage in the riverine ecosystem is "undesirable", that is, indicates pollution, or may be toxic. Higher values were given to the lower risk.

The designated values of each element were summed and on the basis of this, on a scale ranging from 0–5 the assemblages were classified (Table 2). A high value of the factor (F) indicates that the occurrence of this functional group in the riverine phytoplankton can be considered as natural.

The list of the functional groups (Reynolds et al. 2002) and its updated version (Padisák 2003) has to be augmented with those groups that are characteristic elements for the potamoplankton. Tychoplanktic elements that contain originally benthic species have been inserted into the assemblages. These species can be dominant even in low discharge periods.

Table 2. Calculation of the F numbers.

Sum of the points	Factor number
0–3	0
4–6	1
7–10	2
11–14	3
15–17	4
18–20	5

Three tychoplanktic groups have been established:

- The T_B group contains benthic Bacillariophyceae species.
- The T_D group consists of benthic Desmidiales.
- The T_C group contains benthic Cyanobacteria.

Already REYNOLDS et al. (2002) noted that the W1 and W2 groups need to be refined. In harmony with this, a new functional group W_0 has been created, in which those organisms were collected, that prefer waters of extremely high organic content and are capable of surviving even under septic conditions. The natural representatives of these habitats are the shallow oxbows on the floodplain, while the artificial ones are the sewage treatment ponds. Because of this new functional group some alterations were necessary in the species pool of the W_1 and W_2 groups. Details of the functional groups are shown in Table 3.

For the assessment of the ecological state of the rivers, an assemblage index (Q) has been developed. Calculation of the Q index follows the suggestions by PADISÁK et al. (2005):

$$Q = \sum_{i=1}^{n} p_i F_i$$

where p_i is the relative share of the *i*-th functional group equal to n_i/N, where n_1 is the biomass of the *i*-th group and N is the total biomass, and F is the factor number from Table 3. The theoretical maximum of Q is 5, the minimum is 0.

Relationship between Q and the different river types

River typology based on hydrogeology and geochemistry in Hungary revealed 26 types (KVVM 2005). The results of a nationwide ecological survey of the Hungarian waters (Arcadis Euroconsult 2005) suggest that several of the 26 types can be amalgamated. Surprisingly, this proved to be true even for the macroinvertebrates, although this group is quite sensitive to small differences between the physical characteristics of riverine habitats. It also holds for those biological elements – like phytoplankton – which are more or less independent from the

Table 3. Evaluation of the functional groups of algae. The first three columns were elaborated by Reynolds et al. (2002) and Padisák (2003).

Codon	Characteristic species	Habitat	Nutrient status	Turbulence	Time sufficient for the development of the given assemblage	Risk (inverse scale)	Sum of the points	Factor (F)
A	*Urosolenia (Rhizosolenia), Cyclotella comensis, C. glomerata*	Clear, often well-mixed, base poor, lakes	4	4	4	5	17	**4**
B	*Aulacoseira subarctica, A. islandica, Stephanodiscus neoastraea, S. rotula, Cyclotella comta*	Vertically mixed, mesotrophic, small-medium lakes	3	4	4	5	16	**4**
C	*Asterionella formosa, Aulacoseira ambigua, Stephanodiscus neoastrea, Cyclotella meneghiniana, C. stelligera*	Well mixed, eutrophic small-medium lakes and rivers	1	5	4	5	15	**4**
D	*Synedra acus, Nitzschia* spp., *Stephanodiscus hantzschii, Cyclotella ocellata, C. pseudostelligera*	Shallow, enriched turbid waters, including rivers	1	5	4	5	15	**4**
T_C	*Epiphytic cyanobacteria Oscillatooria, Phormidium, Lyngbya, Rivularia*	Enriched standing waters, or slow-flowing rivers with emergent macrophytes	1	1	5	3	10	**2**
T_D	Benthic (epiphytic) desmids and filamentous green algae, Benthic diatoms	Enriched standing waters, or slow-flowing rivers with emergent macrophytes	2	1	5	4	12	**3**
T_B	*Nitzschia* spp., *Navicula, Gomphonema, Didymosphaenia, Fragilaria, Achnanthes, Surirella*	Highly lotic environments including rivers and rivulets	3	5	5	5	18	**5**

Table 3 (continued).

N	*Tabellaria, Cosmarium, Staurodesmus, Xanthidium*	Mesotrophic epilimnia	3	3	1	4	11	**3**
P	*Fragilaria crotonensis, Aulacoseira granulata, Staurastrum pingue, S. chaetoceras Pediastrum duplex, P. simplex, Coelastrum* spp.	Eutrophic epilimnia						
T	*Geminella, Mougeotia, Tribonema, Planctonema, Closterium aciculare, C. acutum* v. *variabile*	Deep, well-mixed epilimnia	2	4	3	4	13	**3**
S1	*Planktothrix agardhii, Limnothrix redekei, L. planctonica, Pseudanabaena limnetica, Planktolyngbya limnetica, P. contorta*	Mixed layers	0	2	1	0	3	**0**
S2	*Spirulina, Arthrospira, Raphidiopsis*	Shallow, mixed layers	0	2	1	0	3	**0**
S_N	*Cylindrospermopsis, Anabaena minutissima*	Warm slightly mixed layers	0	2	1	0	3	**0**
Z	*Synechococcus, Pseudodictyosph aerium, Choriocystis* Single celled prokaryotic pikoplankton	Clear, mixed layers	4	3	4	3	14	**3**
X3	*Koliella, Chrysococcus,* eukaryotic pikoplankton	Shallow, clear, mixed layers	2	3	4	4	13	**3**
X2	*Plagioselmis (Rhodomonas) Chrysochromulima*	Shallow, clear, well mixed layers in mesoeurophic lakes	3	4	4	4	15	**4**
X1	*Ankyra, Monoraphidium*	Shallow well mixed layers in enriched conditions	1	4	4	4	13	**3**

Table 3 (continued).

Codon	Characteristic species	Habitat	Nutrient status	Turbulence	Time sufficient for the development of the given assemblage	Risk (inverse scale)	Sum of the points	Factor (F)
Y	Large microflagellates, *(Cryptomonas)*	Usually, small, well mixed, enriched lakes	1	4	4	4	13	**3**
$\mathbf{Y_{Ph}}$	*Phacotus*	Small, Ca-rich, alkaline lakes	1	2	3	3	9	**2**
E	*Dinobryon, Mallomonas, Synura*	Usually small, oligotrophic, basepure lakes or heterotrophic ponds	3	2	3	4	12	**3**
F	Colonial Chlorococcaleans *(Botryococcus, Pseudosphaerocystis, Coenochlorys, Oocystis), Elakatothrix*	Clear epilimnia	4	2	2	4	12	**3**
G	*Volvox, Eudorina*	Short, nutrinent-rich water columns, small pools among macrophytes	1	0	1	2	4	**1**
J	*Scenedesmus, Golenkinia, Tetrastrum, Crucigenia, Actinastrum Micractinium*	Shallow, enriched lakes, ponds and rivers	2	3	3	2	10	**2**
K	*Aphanothece, Aphanocapsa*	Short, nutrient-rich columns	2	2	3	2	9	**2**
H1	*Anabaena flos-aquae, Aphanizomenon flos-aquae*	Dinitrogen fixing Nostocaleans	1	2	1	1	5	**1**
H2	*Anabaena lemmermannii, Gloeotrichia echinulata*	Dinitrogen-fixing Nostocaleans of larger mesotrophic lakes	3	1	1	1	6	**1**

Table 3 (continued).

U	*Uroglena*	Summer epilimnia	4	1	1	2	8	**2**
$\mathbf{L_O}$	*Peridinium, Woronichinia, Merismopedia*	Summer epilimnia in mesotrophic lakes	3	1	1	2	7	**1**
$\mathbf{L_M}$	*Ceratium, Microcystis*	Summer epilimnia in eutrophic lakes	1	1	1	2	5	**1**
M	*Microcystis, Sphaerocavum*	Daily mixed layers of small, eutrophic, law latitude lakes	0	2	1	0	3	**0**
R	*Planktothrix rubescens, P. mougeotii*	Metalimnia of mesotrophic, stratified lakes	3	2	1	0	6	**1**
V	*Chromatium, Chlorobium*	Metalimnia of eutrophic, stratified lakes	0	0	3	0	3	**0**
W0	*Chlamydomonas, Spermatozopsis, Pyrobotrys, Chlorella, Polytoma, Oscillatoria chlorina, Beggiatoa alba*	Small ponds of extremely high organic content	0	1	3	0	4	**1**
W1	*Euglena, Phacus, Lepocinclis, Gonium pectorale, G. sociale (Pandorina morum)*	Small, organic ponds	1	2	2	1	6	**1**
W2	*Trachelomonas, Strombomonas, Dysmorphococcus* spp. *Small dinoflagellates Peridinium, Glenodinium, Gymnodinium,* other metaphytic species	Shallow, mesotrophic, well mixed lakes	2	4	3	4	13	**3**
$\mathbf{W_S}$	*Synura*	Small, mesotrophic, mixed lakes, pH neutral	3	3	2	4	12	**3**
Q	*Gynostomum*	Small, humic lakes	2	2	1	1	6	**1**

quality of the substrate. The species composition of the rheoplankton depends primarily on the inocula present at the upper river-segment, the light climate and the residence time. Consequently, the number of the river types, based on the characteristic phytoplankton assemblages, will be much smaller than those made on the basis of hydro-geological and geochemical characteristics.

In case of a pristine river, phytoplankton assemblages that can be considered as reference depend primarily on the residence time. Residence time, however, is not constant at the given point, but changes with the actual current velocity, therefore must be considered when defining the reference assemblages. In general, the T_B functional group can be considered as a reference assemblage in the upper river segments and along the river it is enriched with other phytoplankton elements, in natural rivers mainly with those, that have high factor number ($F = 4$). The value of the Q index decreases downstream.

On basis of the residence time 6 main river types are proposed that are more or less in accordance with MEYBECK's (1996) system. These types and the recommended Ecological Quality Ratio (EQR) values are shown in Table 4. Boundary values would be preferably derived from detailed analysis of large data sets. This approach is not feasable here because of the lack of data and the heterogeneity of the existing datasets. Therefore, the boundary values have been extracted from the theoretical contribution of the assemblages that belong to F = 5, 4, 3 or 2 categories (Table 5) in the phytoplankton. These example distributions explain the relatively high EQR values in certain river types.

Table 4. Proposed river types and Ecological Quality Ratio ($EQR = Q/5$) values for different water quality classes.

River type	Code of the type	Stream order[1]	Residence time (day)	EQR				
				excellent	good	moderate	poor	bad
Brooks and small streams	1	1–5	<2	1	**0.99**	0.97	0.95	<0.05
Streams	2	3–6	2–4	0.99	**0.97**	0.95	0.90	<0.90
Small rivers (lowland streams)	3	4–7	4–8	0.95	**0.9**	0.8	0.7	<0.7
Rivers	4	6–9	8–12	0.9	**0.8**	0.7	0.6	<0.6
Large rivers	5	7–1	12–16	0.8	**0.7**	0.6	0.5	<0.5
Very large rivers	6	>10	16<	0.7	**0.6**	0.5	0.4	<0.4

[1] Depending on local conditions

Table 5. Relative share (%) of the assemblages belong to the functional groups F = 5, 4, 3, and 2, and the calculated Ecological Quality Ratio (EQR) values. This table contains example distribution of the different functional groups.

5	4	3	2	EQR
100	0	0	0	1
95	5	0	0	0.99
85	15	0	0	0.97
75	25	0	0	0.95
50	50	0	0	0.9
0	100	0	0	0.8
0	50	50	0	0.7
0	0	100	0	0.6
0	0	50	50	0.5
0	0	0	100	0.4

Case studies

The new assessment method has been tested by using different datasets from Hungarian rivers that contain phytoplankton data against the Pantle-Buck index (PANTLE & BUCK 1955), which is the officially accepted qualification method in Hungary and several other countries in Europe. Some of the databases had to be supplemented with biovolume data. In such cases cell-volumes from earlier investigations were used.

Data from the following rivers were considered: River Kösely at Hajdúszovát, River Tisza at Tiszalök, and River Danube near Göd. The two methods are in accordance when their values move to opposite direction.

River Kösely at Hajdúszovát, riv.-km 46.0

Lowland, calcareous, small river with yearly average water discharge 1.0 $m^3 s^{-1}$ in East Hungary (type 3 in Table 5). This small streamlet carries the effluent of the sewage treatment plant of the city of Debrecen (average wastewater load 50,000 $m^3 day^{-1}$) and it is one of the most polluted rivers in this region with high organic content, but with relatively small hydromorphological alterations.

Compared to the saprobic index, the Q values exhibit a much wider range (Fig. 3). The Q values indicated better quality in several cases, when the saprobity was α-mesosaprobic or α-β-mesosaprobic. This can be explained by the frequently occurring pennate diatoms, *Navicula rhyncocephala* and *Nitzschia palea*, belonging to the T_B group, which have the highest factor value (F = 5) in the new assessment system while these species were considered as α-mesosaprobic indicators in the saprobic system. The T_B group will have to be refined again on the basis of the pollution tolerance.

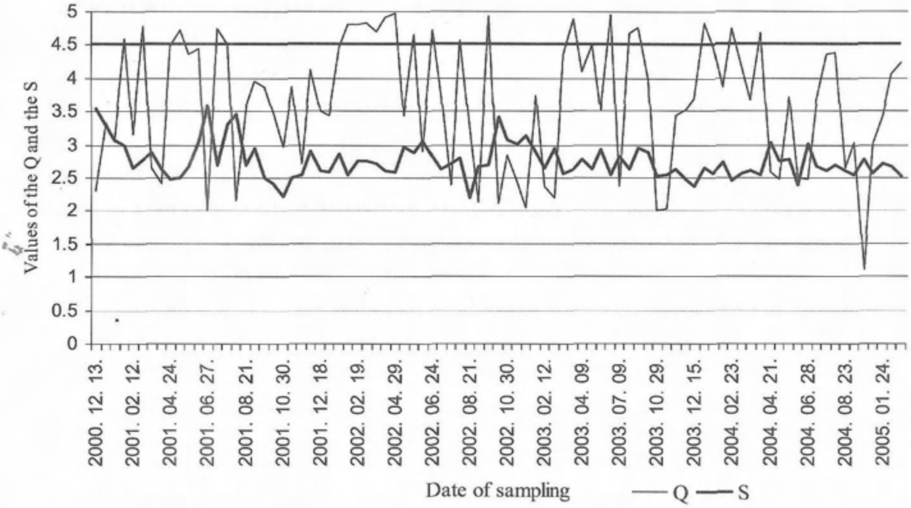


Fig. 3. Quality assessment of the River Kösely on basis of the Q and S (Pantle & Buck 1955). The thick line at 4.5 indicates the good-moderate boundary.

River Tisza at Tiszalök, riv.-km 518.2

This is a lowland, calcareous, large river with yearly average water discharge 540 $m^3 s^{-1}$ (type 4 in Table 5). Sampling point is just upstream of the Tiszalök barrage, which is situated at the upper part of the middle river-segment. Compared to the Pantle-Buck index, the Q values fluctuate at a much larger scale (Fig. 4). In situations when water-quality is worse, both methods coincide, because those phytoplankton groups that have low F numbers have usually high S values in the saprobic system. Since the highest factor number (F = 5) was given exclusively to the representatives of the T_B group, the Q values higher than 4, refer to the dominance of benthic diatoms. The large number of Q values that exceed 4 indicate, that this section has frequently been in good ecological state despite impounding which increases residence time. The Q index, however, indicates bad quality situations in several cases when the saprobic index points to favourable water quality. The reason is that some groups, like Cyanobacteria, have low saprobic values, and consequently, refer to relatively good water quality. In the present method they have low values as well (F = 0, or F = 1), but here they indicate undesirable water quality situations.

River Danube at Göd, riv.-km 1669

Lowland, calcareous, large river with average discharge 2,300 $m^3 s^{-1}$ at Göd (type 5 in Table 5). Sampling was carried out at the middle part of the Hungarian river

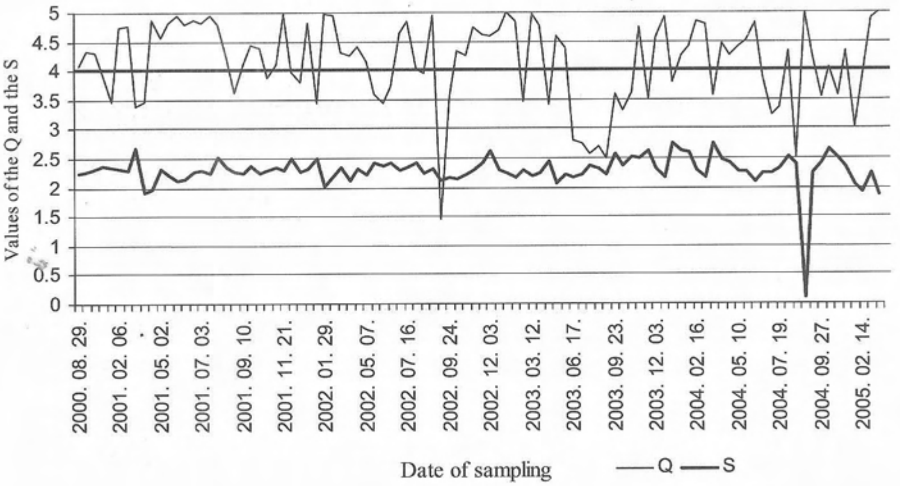


Fig. 4. Quality assessment of the River Tisza on basis of the Q and S (Pantle & Buck 1955). The thick line at 4.0 indicates the good-moderate boundary.

section. The Q values (Fig. 5) indicate that the phytoplankton of the river consists primarily of euplanktic species during the vegetation period. The estimated residence time at the sampling point of >20 days at low discharge allows the development of assemblages B, C, and D, which are dominated mainly by centric diatoms (F = 4). The values lower than 4 indicate the dominance of P and F assemblages. The saprobic index proved to be insensitive during the whole investigation period.

Supposedly, the residence time was long enough for the development of characteristic riverine phytoplankton in the large and very large rivers even in the earlier, pristine situation, but the stream regulations by damming and impounding at the upper river basin considerably reduced flow rate and increased residence time. The composition of the phytoplankton under pristine conditions is therefore difficult. Re-investigation of samples collected from the River Danube in the late 1950s proved, that several taxa that now are considered as "new and dominant" elements of the algal flora (e.g. *Skeletonema potamos*), were present in the plankton (Kiss 1986). Although the ratio of these euplanktic elements must have been lower in the past than nowadays, we have the opinion that the phytoplankton of large pristine rivers consisted of these elements during low discharge periods. The meso-eutrophic state must have been a natural characteristic of the very large pristine rivers. Probably large European rivers were potentially eutrophic in the middle of the last century, but their actual trophic state was mesotrophic as a result of high concentration of the suspended solids (Kiss 1985, 1997).

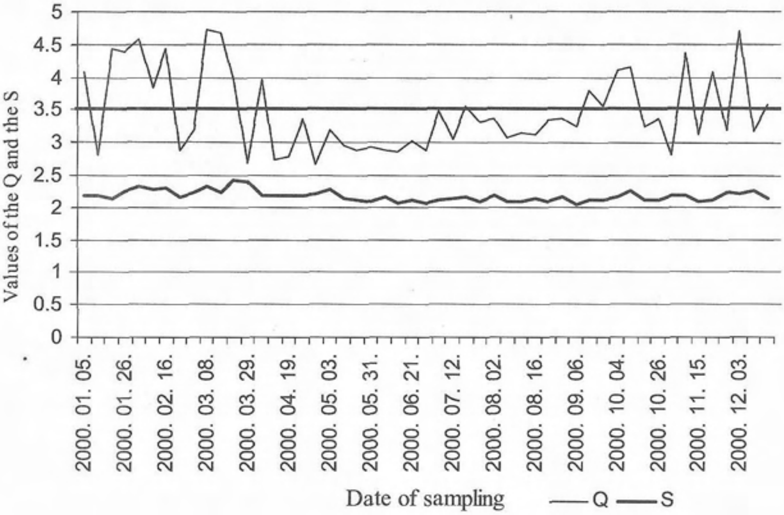


Fig. 5. Quality assessment of the River Danube on basis of the Q and S (PANTLE & BUCK 1955). The thick line at 3.5 indicates the good-moderate boundary.

Discussion

Monitoring of the naturally occurring algal communities in rivers provide data on species composition, species number, diversity, or quantitative occurrence of the algae. Experts of administrative institutions who are responsible for water quality management need simple numerical values rather than species lists or scientific evaluation of the assemblages. Much of the monitoring activities concentrated on bentic Bacillariophyceae species in the recent years, and numerous indeces have been developed for characterizing the quality of the rivers (DESCY 1979; COSTE in CEMAGREF 1982; WATANABE et al. 1988; KELLY & WHITTON 1995; SCHIEFELE & SCHREINER 1991; LENOIR & COSTE 1996; ROTT et al. 1997). These methods became widely used in the last decades, but their applicability in large, deep, slow-flowing rivers (because of the lack of appropriate substrate) is problematic. Nevertheless, this type of the rivers is usually enriched with nutrients and – pending the hydro-meteorological situations – can be characterized by highly eutrophic phytoplankton. High phytoplankton biomass may occur in smaller water currents as well, consequently the phytoplankton as a tool in the water quality assessment should not be ignored. Since up to now, little attention has been attended to the application of the phytoplankton for evaluation of ecological state of the rivers, the present study aims to show a new assessment method. The method is based on the functional group of algae represented in the potamoplankton and provides a single index number (Q). The index has been tested on phytoplankton data of different rivers, and proved to be more sensitive than the earlier used saprobic index. After calculating the Q, the Ecological Quality Ratio (expected by the WFD) could also be given.

Conclusions

As an evaluation it is worth mentioning the strengths and weaknesses of the method.

Strengths of the method:

- It has been elaborated on the basis of the fundamental hydrological characteristics of the rivers.
- Applicability is not restricted to a geographic region.
- It is sensitive.
- It is capable of indicating different anthropogenic impacts (organic pollution, impoundment, occurrence of toxic elements).
- It can be computerized easily.

Weaknesses of the method:

- It is difficult to define those river stretches where the given associations are expected to occur or to be dominant (supposing natural conditions).
- Efficiency of the lentic areas as the "cradles" of the euplanktic species is not known.
- It is also not known how much the "dead zones" (REYNOLDS et al. 1991; REYNOLDS 2000) enlarge the residence time.
- The factor numbers have been given by expert judgement.
- Those species that have not been mentioned in the Reynolds' system (REYNOLDS et al. 2002) are sorted into the groups on the basis of expert opinion.
- During the testing, average biomass data were used.
- In case of a water bloom, the method may indicate good or excellent ecological state.
- The boundaries cannot be set by analysing the variation of the given biological indicator (Q) along a pressure gradient (as it suggested by the ECOSTAT (2005) document), therefore the boundary setting is subjective.

The method has been tested with hundreds of phytoplankton samples, it is simple, and after applying to a phytoplankton database it can be computerized easily.

Acknowledgements

This study was supported by the Ministry of Environment and Water, the Hungarian National Research Fund (OTKA, K60452) and by the Bolyai János fellowship of the Hungarian Academy of Sciences. The authors cordially thank the anonymous referees for their constructive comments, and the comprehensive and detailed review of the paper.

References

Arcadis Euroconsult (2005): Ecosurv – Ecological survey of surface waters in Hungary. – Final Techn. Report, Budapest.

Coste, M. in CEMAGREF (1982): Etude des méthodes biologiques d'appréciation quantitatives de la qualité des eaux. – Rapport Groupement de Lyon-Agence de l'Eau Rhone – Méditerranée – Corse, Pierre – Bénite, 28 pp.

Descy, J.P. (1979): A new approach to water quality estimation using diatoms. – Nova Hedwigia Beih. **64**: 305–323.

Descy, J.P., Everbecq, E. & Smitz, J.S. (1988): Primary production in the River Meuse (Belgium). – Verh. Int. Verein Limnol. **23**: 1287–1293.

Dokulil; M.T. (1991): Review on recent activities, measurements and techniques concerning phytoplankton algae of large rivers in Austria. – In: Whitton, B.A., Rott, E., Friedrich, G. (eds.): Use of algae for monitoring rivers: 53–57. Innsbruck, ISBN 3-95000-0-0.

– (1996): Evaluation of eutrophication potential in rivers: The Danube example, a review. – In: Whitton, B.A. & Rott, E. (eds.): Use of algae for monitoring rivers **II**: 173–178. Innsbruck.

ECOSTAT Working Group A (2005): Template for the development of a boundary setting protocol for the purposes of the intercalibration exercise. – Common Implementation Strategie SCG agenda point 6.1 WGA/26271005/4/d. Development of tools for a thematic eutrophication assessment (HELCOM EUTRO). Second Meeting Riga, Latvia, 30 June – 1 July 2005.

Felföldy, L. (1969): Felszíni vizeink eutrofizálódásáról. [Eutrophication of surface waters in Hungary.] – Vízügyi közlemények **209**: 352–362. [in Hungarian with English summary]

Kelly, M.G., Penny, C.J. & Whitton, B.A. (1995): Comparative performance of benthic diatom indices used to assess river water quality. – Hydrobiologia **302**: 179–188.

Kiss, K.T. (1985): Changes of trophity conditions in the River Danube at Göd. Danubialia Hungarica XCIV. – Ann. Univ. Sci. Budapest Eötvös Sect. Biol. **24**: 47–59.

– (1986): Species of the Thalassiosiraceae in the Budapest section of the Danube. Comparison of samples collected in 1956–63 and 1979–83. – In: Ricard, M. (ed.): Proc. 8th International Diatom Symposium: 23–31. Koeltz, Koenigstein.

– (1987): Phytoplankton studies in the Szigetköz section of the Danube during 1981–1982. – Arch. Hydrobiol., Algol. Studies **47**: 247–273.

– (1994): Trophic level and eutrophication of the River Danube in Hungary. – Verh. Int. Verein. Limnol. **25**: 1688–1691.

– (1996): Diurnal change of planktonic diatoms in the River Danube near Budapest (Hungary). – Arch. Hydrobiol., Algol. Studies **80**: 113–122.

– (1997): The main results of phytoplankton studies on the River Danube and its side arm system at the Szigetköz area during the nineties (Hungary). – In: Dokulil, M.T. (ed.): Limnologische Berichte Donau 1997, **I**: 153–158.

Kiss, K.T. & Genkal, S.I. (1997): Télvégi – koratavaszi Centrales (Bacillariophyceae) vízvirágzás a Dunán. [Late summer – early spring Centrales (Bacillariophyceae) bloom in the river Danube.] – Hidrológiai Közlöny **77**: 57–58. [in Hungarian with English summary].

KISS, K.T., NOSEK, J. & KREMMER, L. (1998): Diurnal change of phytoplankton, some physical and chemical components in the River Danube near Budapest (Hungary) – Verh. Internat. Verein. Limnol. **26**: 1041–1044.

KISS, K.T. & SCHMIDT, A. (1998): Changes of the Chlorophyta species in the phytoplankton of the Hungarian Section of the Danube river during the last decades (1961–1997). – Biologia Bratislava **53**: 509–518.

KISS, K.T., SCHMIDT, A. & ÁCS, É. (1995): Fitoplankton-mintavétel gyakoriságának kérdése egy nagy eutrófikus folyóvízben, a Dunában. [Phytoplankton sampling frequency in case of a large river, Danube.] – In: BÍRÓ, P. (ed.): XXXVII. Hidrobiológus Napok kötete, MTA Veszprémi Akadémiai Bizottság: 79–82. [in Hungarian with English summary].

– (1996): Sampling strategies for phytoplankton investigations in a large river (River Danube, Hungary). – In: WHITTON, B.A., ROTT, E. & FRIEDRICH, G. (eds.): Use of algae for monitoring rivers **II**: 179–185. Inst. f. Botanik, Univ. Innsbruck.

KVVM (2005): Nemzeti Jelentés a Duna vízgyűjtőkerület magyarországi területének jellemzőiről, az emberi tevékenységek környezeti hatásairól és a vízhasználatok gazdasági elemzéséről. 2005 március 22. [EU WFD National Report of Hungary.] – In: http://www.euvki.hu. [in Hungarian]

LÁSZLÓFFY, W. (1982): A Tisza vízi munkálatok és vízgazdálkodás a Tisza vízrendszerében. [Tisa River. Water basin management in the Tisa River basin.] – Akadémiai Kiadó, Budapest, 610 pp.

LENOIR, A. & COSTE, M. (1996): Development of pratical diatomic index of overall water quality applicable to the French National Water Board Network. – In: WHITTON, B.A. & ROTT, E. (eds.): Use of Algae for monitoring Rivers **II**: 29–43. Inst. f. Botanik, Univ. Innsbruck.

MEYBECK, M., FRIEDRICH, G., THOMAS, R. & CHAPMAN, D. (1996): Rivers. – In: CHAPMAN, D. (ed.): Water Quality Assessments. UNESCO/WHO/UNEP Second ed., Cambridge Univ. Press, 626 pp.

PADISÁK, J. (2003): Javaslat vízterek biológiai állapotának jellemzésére alkalmas biológiai indikátorokra. Az ökológiai minősítés kérdései, fitoplankton. [Proposal for ecological qualification of surface waters, phytoplankton. Project report.] – Szakértői jelentés. BMGE Vízi Közmű és Környezetmérnöki Tanszék Budapest. [in Hungarian]

PADISÁK, J., GRIGORSZKY, I., BORICS, G. & SORÓCZKI-PINTÉR, É. (2005): Use of phytoplankton assemblages for monitoring ecological status of lakes within the Water Framework Directive: the assemblage index. – Hydrobiologia **553**: 1–14.

PANTLE, R. & BUCK, H. (1955): Die biologische Überwachung der Gewässer und die Darstellung der Ergebnisse. – Gas- u. Wasserfach **96**: 604.

RÁKÓCZI, L. (1989): Vízlépcsők hatása a hordalék- és mederviszonyokra. [Impact of the impoundments on the sediment-pattern and the river channel morphology.] – Vízügyi Közl. **71** (1): 5–24. [in Hungarian with English summary]

REVENGA, C, BRUNNER, J, HENNINGER, N, KASSEM, K & PAYNE, R. (2000): Pilot analysis of global ecosystems: freshwater ecosystems. – 83 pp., World Resources Inst., Washington, DC. In: http://www.unep.org/vitalwater/23.htm

REYNOLDS, C.S. (2000): Hydroecology of River plankton: the role of variability in channel flow. – Hydrol. Processes **14**: 3119–3132.

REYNOLDS, C.S., CARLING, P.A. & BEVEN, K.J. (1991): Flow in River Channels: New Insights into Hydraulic Retention. – Arch. Hydrobiol. **121** (2): 171–179.

REYNOLDS, C.S., DESCY, J.P. & PADISÁK, J. (1994): Are phytoplankton dynamics in rivers so different from those in shallow lakes? – In: DESCY, J.P., REYNOLDS, C.S. & PADISÁK, J. (eds.): Phytoplankton in turbid environments: Rivers and shallow lakes. Hydrobiologia **289**: 1–7. Kluwer Acad. Publ.

REYNOLDS, C.S. & GLAISTER, M.S. (1993): Spatial and temporal changes in phytoplankton abundance in the upper and middle riches of the River Severn. – Arch. Hydrobiol. Suppl. **101**: 1–22.

REYNOLDS, C.S., HUSZAR, V., KRUK, C., NASELLI-FLORES, L. & MELO, S. (2002): Towards functional classification of the freshwater phytoplankton. – J. Plankton Research **24**: 417–428.

ROTT, E., HOFMANN, G., PALL, K., PFISTER, P. & PIPP, E. (1997): Indikationslisten für Aufwuchsalgen in österreichischen Fließgewässern. Teil 1: Saprobielle Indikation. – Wasserwirtschaftskataster. Bundesminsterium f. Land- u. Forstwirtschaft, Wien.

SCHIEFELE, S. & SCHREINER, C. (1991): Use of diatoms for monitoring nutrient enrichment, acidification and impact of salt in rivers in Germany and Austria. – In: B.A. WHITTON, E. ROTT & G. FRIEDRICH (eds.): Use of algae for monitoring rivers: 103–110. Inst. f. Botanik, Univ. Innsbruck, ISBN 3-9500090-0-0.

SCHMIDT, A. (1994): Main characteristics of the phytoplankton of the Southern Hungarian section of the River Danube. – Hydrobiologia **289**: 97–108.

SKIDMORE, R.E., MABERLY, S.C. & WHITTON, B.A. (1998): Patterns of spatial and temporal variation in phytoplankton chlorophyll-*a* in the River Trent and its tributaries. – The Science of the Total Environment **210/211**: 357–365.

STOYNEVA, M.P. (1994): Shallows of the lower Danube as additional sources of potamoplankton. – Hydrobiologia **289**: 171–178.

THORNTON, J., STEEL, A. & RAST, W. (1996): Reservoirs. – In: CHAPMAN, D. (ed.): Water Quality Assessments. UNESCO/WHO/UNEP Second ed., Cambridge Univ. Press, 626 pp.

UHERKOVICH, G. (1971): A Tisza lebegő paránynövényei. (A Tisza fitoszesztonja.) [Phytoplankton of the River Tisza.] – Szolnok Megy. Múz. Adattár **20–22**: 1–282 I-CXLIII. [in Hungarian]

VÍGH, GY. (2002): Vizeink minősége. [Annual report on the quality of the Hungarian surface waters.] – Környezetgazdálkodási Intézet Környezetvédelmi Igazgatóság. Környezetvédelmi és Vízügyi Minisztérium Budapest, 2003: 155 pp. [in Hungarian]

WATANABE, T., ASAI, K. & HOUKI, A. (1986): Numerical estimation to organic pollution of flowing water by using the epilithic diatom assemblage – Diatom Assemblage Index (DAIpo). – The Science of the Total Environment **55**: 209–218.

WFD (2000): Directive 2000/60/ec of the European Parliament and of the Council 22.12.2000. – Off. J. Europ. Communities L327/1-72.

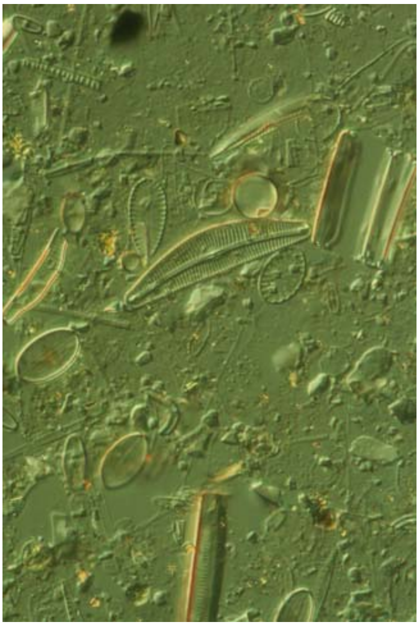
Accepted: 18 April 2007

Paper 5.

Implementation of the European Water Framework Directive: Development of a system for water quality assessment of Hungarian running waters with diatoms

Van Dam, H., Stenger-Kovács, Cs., Ács, É., Borics, G., Buczkó, K.,
Hajnal, É., Soróczki-Pintér, É., **Várbíró, G.**, Tóthmérész, B.,
Padisák, J.

Large Rivers Vol. 17, No. 3-4
Arch. Hydrobiol. Suppl. 16113-4, p. 339-364, Oktober 2007



Large Rivers Vol. 17, No. 3–4
Arch. Hydrobiol. Suppl. 161/3–4, p. 339–364, Oktober 2007

Implementation of the European Water Framework Directive: Development of a system for water quality assessment of Hungarian running waters with diatoms

Herman van Dam[1], **Csilla Stenger-Kovács**[2], **Éva Ács**[3], **Gábor Borics**[4], **Krisztina Buczkó**[5], **Éva Hajnal**[2], **Éva Soróczki-Pintér**[2], **Gábor Várbíró**[4], **Béla Tóthmérész**[6] and **Judit Padisák**[2]

With 4 figures and 8 tables in the text

Abstract: The purpose of the paper is (1) to make an inventory of the diatom assemblages of the Hungarian running water types; (2) to assess relationships between the species composition and environmental variables; (3) to adapt existing metrics to the Hungarian situation; and (4) to present a map of the of diatom inferred quality of the country's running waters. In spring 2005, diatom samples were taken in 339 streams. Samples for analysis of water chemistry were taken at most of the stations. The diatom species composition was investigated using standard methods and 496 taxa were found. Ordination by detrended correspondence analysis (DCA) revealed that current velocity, altitude, shading, oxygen and alkalinity are master variables for the species composition of diatoms. Nutrients are of less importance in determining the diatom distribution over the whole country, but regionally they can have a considerable impact. Twelve diatom river types (IndVal method) were distinguished. The existing physiographic river typology is too detailed for diatom assemblages. Human impact has caused shifts in the assemblages. For the estimation of water quality in the running waters, the IPS (Index of Pollution Sensitivity) was calculated and the class boundary limits were adapted to the Hungarian geomorphology. Different class boundaries are set for high-,

---

Authors' addresses: [1] Grontmij | AquaSense, P.O. Box 95125, NL-1090 HC Amsterdam (present address: Consultancy for Water and Nature, P.O. Box 37777, NL-1030 BJ Amsterdam), The Netherlands; E-mail: herman.vandam@waternatuur.nl

[2] Department of Limnology, Pannon University, P.O. Box 158, H-8200 Veszprém, Hungary.

[3] Institute of Ecology and Botany, Hungarian Danube Research Station, Hungarian Academy of Sciences, Jávorka S. u. 14, H-2131 Göd, Hungary.

[4] Environmental Protection, Nature Conservation and Water Authority, Trans-Tiszanian Region, Hatvan u. 16, H-4025 Debrecen, Hungary.

[5] Botanical Department of Hungarian Natural History Museum, P.O. Box 222, H-1476 Budapest, Hungary.

[6] Department of Ecology, Debrecen University, P.O. Box 14, H-1410, Debrecen, Hungary.

0945-3784/07/0161-0339 $ 6.50

mid- and low-altitude streams. Using these boundaries, 81 % of the streams have a high or good quality. Poor or bad conditions of the streams are often due to domestic sewage load, but industrial pollution may also be important. Provisionally, reference sites are selected.

Key words: Water Framework Directive (WFD), phytobenthos, diatoms, ordination, cluster analysis, master variables, IPS, water quality assessment.

Introduction

An important biological quality element (BQE) for the estimation of water quality in the European Water Framework Directive (WFD) is the phytobenthos (European Union 2000). In each of the member states metrics (indices) for the calculation of water quality (Ecological Quality Ratio, or EQR) have to be developed. As diatoms represent the most important component of the phytobenthos, the taxonomic composition of attached diatom assemblages is used for the calculation of the EQR in many countries, as is shown by several papers in this volume.

The WFD recommends that assessment of water quality should be simple, straightforward and robust, i.e. the information that is given by the diatom-inferred water quality should be related to the most important stressors for the water bodies in question. The construction of yardsticks or metrics to calculate Ecological Quality Ratio's thus needs a strong insight in the environmental variable(s) that are responsible for the taxonomic composition of diatom assemblages of the relevant water types.

Although Hungary has a long tradition in research on taxonomy and ecology of attached diatoms (e.g. PANTOCSEK 1901; CHOLNOKY 1929; SZEMES 1956; PADISÁK 1982; ÁCS & KISS 1991) recent evaluations are scarce (BUCZKÓ & ÁCS 1994; ÁCS et al. 2004; KOVÁCS et al. 2006). A nation-wide survey of diatom assemblages covering all important types of running waters with concurrent and relevant environmental variables was absent until now, as it is for most of the countries in the world. Hungary is a rather flat country, where over 80 % of the surface area is used for agriculture. So, diatom assemblages in Hungarian streams should be rather similar to those in most western and central European countries.

Therefore, we started an inventory of diatom assemblages in 339 Hungarian streams, in order to assess the relationships between the species composition of the diatom assemblages and measured environmental (chemical, physical) variables by means of multivariate analyses. These relationships provide evidence for the master variables regulating the species composition of diatoms. These relations will be used to test whether the existing diatom indices (e.g. PRYGIEL et al. 1999) can be applied in Hungary. Subsequently, boundaries for ranges of metrics describing different classes of water quality will be set and used to map the quality of diatoms in Hungarian surface waters. Moreover, cluster analysis will be used in order to describe the taxonomic composition of the diatom assemblages of the water types in a compact way and to explore the correspondence between diatom-based river types and physiographic river types.

The diatom results will be compared with surveys of fish and macro-invertebrates, which were performed at the same river locations in the so-called ECO-SURV-project (ARCADIS 2005). All technical details and full results of this paper are to be found in VAN DAM et al. (2005).

Materials and methods

Selection of sampling stations

For the purposes of the implementation of the WFD 25, physiographic natural types of running waters are distinguished in Hungary on basis of their altitude, geology, substrate and dimensions. An additional type for artificial water bodies (channels) is added (Table 1).

Table 1. Description of Hungarian physiographic river types and number of sites included in this study.

Type	Altitude	Geology	Substrate	Dimension	Sites
1	high	silicious	rough	small	18
2	high	calcareous	rough	small	19
3	high	calcareous	rough	medium	6
4	mid	calcareous	rough	small	22
5	mid	calcareous	rough	medium	21
6	mid	calcareous	rough	large	10
7	mid	calcareous	rough	very large	3
8	mid	calcareous	medium	small	34
9	mid	calcareous	medium	medium	18
10	mid	calcareous	medium	large	7
11	low	calcareous	rough	small	5
12	low	calcareous	rough	medium	10
13	low	calcareous	rough	large	12
14	low	calcareous	rough	very large	5
15	low	calcareous	medium	small	20
16	low	calcareous	medium	very small (very flat area)	19
17	low	calcareous	medium	very small (flat area)	10
18	low	calcareous	medium	medium	27
19	low	calcareous	medium	large	19
20	low	calcareous	medium	very large	15
21	low	organic	fine		4
22	low	organic	medium		3
23	low	calcareous	rough	very large	12
25	low	calcareous	medium	very large	2
26	artificial channel				14
	unknown type				4
	total				339

Sampling sites were selected proportionally among the WFD types, as evenly over the country as possible, and across the whole gradient from nearly pristine to highly impacted streams. Large streams were sampled at more than one site. In total 339 sites were sampled all over the country. Hungary has an extensive network of national and regional water quality monitoring stations (PADISÁK et al. 1991), which covers the main rivers (Danube, Tisza, Dráva) and their tributaries. These stations were used as much as possible. However, in smaller streams, particularly in the more remote areas, no permanent monitoring stations exist; thus additional stations were established. Heavily shaded stream sections and sections directly below point sources and villages were avoided as much as possible for sampling.

Sampling and laboratory analysis

Sampling was performed from March to July 2005. The largest number (163 samples) was taken in April, while 67, 80 and 26 samples were taken in March, May and June, respectively. In July three supplementary samples were taken at locations where few diatoms were present in the previous samples. Permanency, altitude and the presence of diffuse pollution (arable land) in the watershed above the sampling stations was estimated from topographical and land-use maps. Current velocity, width, and depth were measured or estimated visually in the field. Data on the presence of point sources of pollution and sewage inflows were obtained from regional water managers. Valley form was recorded in the field on a six-point scale (V-shaped [1] to flat [6]). Shading of shoreline and recent removal of aquatic vegetation were recorded in the field.

Water temperature, pH, conductivity and oxygen were measured at the site with calibrated equipment. Samples for routine laboratory determination of total dissolved solids (TDS), total suspended solids (TDS), loss on ignition (LOI), chemical oxygen demand (COD), of alkalinity (bicarbonate), calcium, chloride, Kjeldahl-nitrogen, ammonium, nitrite, nitrate, total phosphorus and silicate were taken at 225 out of the 339 sites and transported to a laboratory for analysis with standard Hungarian methods.

Diatom sampling was carried out according to recommendations by KELLY et al. (1998). Natural stones (but sometimes also concrete or bricks) were the preferred substrate (168 samples), but particularly in lowland stretches other substrates had to be sampled in the absence of stones: from *Phragmites* stems (79), other emergent macrophytes (54), tree stems (24) and other substrates (15).

Diatom samples were treated with hot hydrogen-peroxide method and then were mounted in Zrax or Hyrax. They were identified at least to species level according to KRAMMER & LANGE-BERTALOT (1986–1991) and LANGE-BERTALOT (1993) by four different analysts. We retained the 'old' names as much as possible as the flood of 'new' names is often not informative for applied limnologists (LANGE-BERTALOT 1998). Comparability of their identifications was ensured by preliminary intercalibration. A minimum of 400 valves were counted on each slide.

Data processing

Prior to all calculations, the abundance values of diatom taxa in the samples were transformed into percentage abundance values by dividing the number of valves of a taxon in a sample by the total number of valves counted in the sample and multiplying this quotient by 100.

Diatom-based indications of water quality values were calculated using the formula for the IPS or Index de Polluosensitivité Specifique (COSTE in CEMAGREF 1981) with the Omnidia-program (LECOINTE et al. 1993). Updated indicator values for nearly all taxa were obtained form M. COSTE (pers. comm.). Moreover, weighted average indicator values for R (acidity), H (salinity), N (organic nitrogen), O (oxygen), S (saprobity), T (trophic state) and M (moisture) were calculated according to VAN DAM et al. (1994).

As data on chemical variables were not available for all sites no direct ordination (canonical correspondence analysis or CCA) could be performed on the whole dataset. To include all diatom data and all chemistry data, an indirect ordination method, detrended correspondence analysis (DCA) was applied, using the computer programme Canoco 4.5 (TER BRAAK & ŠMILAUER 2002) and then environmental parameters were related to the DCA axes. The ordination was carried out with 113 taxa (all taxa with a mean relative abundance >0.05 %), which comprised over 97 % of the total abundance of the taxa. In each sample, at least 95 % of the total abundance was used for the calculations. The default settings of the program were used, but detrending was by 2nd degree polynoms (not by segments), percentage abundance values were log(x+1) transformed and the importance of species which were abundant at one or a few stations and did not occur in the dataset was reduced.

After completing the DCA, Pearson correlation coefficients were calculated between the ordination axes and environmental data, including chemistry and hydromorphological variables to identify master environmental master variables regulating the species composition of diatoms. Prior to calculating correlation coefficients, environmental variables were tested for their skewness by testing if the quotient of their mean and median was greater than 1.1. If so, these variables were log-transformed.

As a clustering program IndVal analysis (DUFRÊNE & LEGENDRE 1997) was used. Indicator species* were defined as the most characteristic species of each group, which are present in the majority of the given group's sites and were found

---

* Note that the expression 'indicator species' in this classification method has a different meaning form that in the IPS-method.

chiefly in this single group. The program calculated an indicator index for each species and this index, for a given species, is independent of the other species' relative abundances. Euclidean distance was used as a similarity measure, as it stresses the importance of the most abundant species. The calculation of the indicator value followed the equation: $\mathrm{IndVal}_{ij} = \mathrm{A}_{ij} * \mathrm{B}_{ij} * 100$, where A_{ij} is the mean abundance of species *i* in the sites of group *j* compared to all groups, and B_{ij} is the relative frequency of occurrence of species *i* in the sites of group *j*. The program compared the observed distribution of the species to the random simulation. When this deviation was significant the species was said to be a characteristic species for the group of samples where the deviation is detected. The significance level of the character species was 0.01, if character species was not found on this level, the significance level was changed to 0.05.

The significance of differences of environmental variables and diatom indicator values according to VAN DAM et al. (1994) and number of species between clusters was tested with simple one-way analysis of variance, using the data analysis package of the Excel spreadsheet program.

Results

Taxonomic composition

In the 339 samples a total of 496 taxa (species and varieties) were identified. Many of these taxa were very rare and occurred only in one or in a few samples. The commonest genera are *Navicula* (104 taxa), followed by *Nitzschia* (80), *Fragilaria* (43), *Achnanthes* (41), *Gomphonema* (30) and *Cymbella* (26). The most common taxa are listed in Table 2. The most common species, *Achnanthes minutissima*, is also the most common freshwater diatom species in the world. The majority of the other common taxa are characteristic of eutrophic, alkaline waters. Some of these are tolerant to organic pollution (e. g. *Gomphonema parvulum, Nitzschia paleacea*), others require cleaner waters (e. g. *Achnanthes biasolettiana, Gomphonema micropus*). *Meridion circulare* is one of the very few diatom species which are characteristic for running waters (VAN DAM et al. 1994). Species from acidic or oligotrophic waters are rare in the dataset: the most common one is *Eunotia bilunaris*.

Ordination

The first DCA axis accounts for 8.5 % of the total variance; the cumulative variance for second and third axes is 13.3 and 17.3 %, respectively (Table 3). These values are not extremely high, indicating that the species composition is relatively diverse. Such low values are characteristic of really large data sets.

The first axis (Fig. 1) is highly positively correlated with the water type number (size), valley form (higher numbers indicating flatter valleys), and negatively with

Table 2. Acronyms, ecology, mean percentage abundance and percentages of occurrence of the most common taxa. For each species the sensitivity number (IPSs) and indication value number (IPSv) is given (M. Coste pers. comm., adapted from CEMAGREF 1982). Taxa with high IPSs-values (5) are 'clean water species', taxa with high IPSv-values (3) are taxa with a narrow ecological range.

Acronym	IPSs	IPSv	Taxon	author	Mean abundance	Perc. of samples
ACHNBIAS	5.0	2	*Achnanthes biasolettiana*	Grunow in Cleve & Grunow	1.6	17
ACHNDAON	5.0	2	*Achnanthes daonensis*	Lange-Bertalot	0.2	2
ACHNLAfr	3.4	1	*Achnanthes lanceolata* ssp. *frequentissima*	Lange-Bertalot	1.4	53
ACHNLAla	4.6	1	*Achnanthes lanceolata* ssp. *lanceolata*	(Brébisson) Grunow	0.9	34
ACHNLANC	4.6	1	*Achnanthes lanceolata*	(Brébisson) Grunow	2.5	32
ACHNLATE	5.0	3	*Achnanthes laterostrata*	Hustedt	0.1	2
ACHNMINU	5.0	1	*Achnanthes minutissima*	Kützing	15.4	87
ACHNPLOE	5.0	2	*Achnanthes ploenensis*	Hustedt	0.2	11
ACHNSPEC			*Achnanthes* sp.		0.2	13
AMRAINAR	5.0	1	*Amphora inariensis*	Krammer	0.4	18
AMRALIBY	4.0	2	*Amphora libyca*	Ehrenberg	0.2	23
AMRAPEDI	4.0	1	*Amphora pediculus*	(Kützing) Grunow	6.8	74
CANEBACI	4.0	2	*Caloneis bacillum*	(Grunow) Cleve	0.2	24
CCNEPEDI	4.0	2	*Cocconeis pediculus*	Ehrenberg	0.1	13
CCNEPLAC	4.0	1	*Cocconeis placentula*	Ehrenberg	3.5	70
CYCLCOMT	5.0	1	*Cyclotella comta*	(Ehrenberg) Kützing	0.1	8
CYLACIST	4.0	3	*Cymbella cistula*	(Ehrenberg) Kirchner	0.1	10
CYLAMINU	4.8	2	*Cymbella minuta*	Hilse ex Rabenhorst	0.3	32
CYLASILE	5.0	2	*Cymbella silesiaca*	Bleisch ex Rabenhorst	0.1	7
CYLASINU	4.8	1	*Cymbella sinuata*	Gregory	0.4	24
CYPHDUBI	3.0	2	*Cyclostephanos dubius*	(Fricke) Round	0.1	12
CYTEATOM	2.0	1	*Cyclotella atomus*	Hustedt	0.4	9
CYTEMENE	2.0	1	*Cyclotella meneghiniana*	Kützing	0.6	35
CYTEPSST	4.0	1	*Cyclotella pseudostelligera*	Hustedt	0.1	9
DIATMONI	4.0	2	*Diatoma moniliformis*	Kützing	0.6	19
DIATTENU	3.0	1	*Diatoma tenuis*	Agardh	0.4	14
DIATVULG	4.0	1	*Diatoma vulgaris*	Bory	0.1	14
EPITADNA	4.0	3	*Epithemia adnata*	(Kützing) Brébisson	0.1	1
EPITTURG	5.0	2	*Epithemia turgida*	(Ehrenberg) Kützing	0.1	1
EUTIBILU	5.0	2	*Eunotia bilunaris*	(Ehrenberg) Mills	0.2	11
FRLACAgr	1.0	3	*Fragilaria capucina* var. *gracilis*	(Oestrup) Hustedt	0.2	8
FRLACAPU	3.0	1	*Fragilaria capucina*	Desmazières	1.7	40
FRLACAru	4.0	1	*Fragilaria capucina* var. *rumpens*	(Kützing) Lange-Bertalot	0.1	6
FRLACAva	3.4	1	*Fragilaria capucina* var. *vaucheriae*	(Kützing) Lange-Bertalot	0.8	30
FRLADELI	4.0	1	*Fragilaria delicatissima*	(W. Smith) Lange-Bertalot	0.1	4
FRLAFASC	2.0	3	*Fragilaria fasciculata*	(Agardh) Lange-Bertalot	0.2	14

Table 2 (continued).

Acronym	IPSs	IPSv	Taxon	author	Mean abundance	Perc. of samples
FRLAPULC	3.0	3	*Fragilaria pulchella*	Ralfs	0.1	3
FRLASPEC			*Fragilaria* sp.		0.1	6
FRLAULac	4.0	1	*Fragilaria ulna* var. *acus*	(Kützing) Lange-Bertalot	0.2	19
FRLAULNA	3.0	1	*Fragilaria ulna*	(Nitzsch) Lange-Bertalot	1.9	56
GOMPHANGU	5.0	3	*Gomphonema clevei*	Fricke	0.7	16
GONEANGU	5.0	1	*Gomphonema angustum*	Agardh	0.3	13
GONECLAV	5.0	2	*Gomphonema clavatum*	Ehrenberg	0.2	16
GONEMINU	4.0	1	*Gomphonema minutum*	(Agardh) Agardh	0.2	18
GONEMIPU	3.0	1	*Gomphonema micropus*	Kützing	1.7	33
GONEOLUM	4.6	1	*Gomphonema olivaceum*	(Hornemann) Brébisson	4.7	71
GONEPARV	2.0	1	*Gomphonema parvulum*	(Kützing) Kützing	2.5	74
GONEPUMI	5.0	1	*Gomphonema pumilum*	(Grunow) E. Reichardt & Lange-Bertot	0.6	17
GONESPEC			*Gomphonema* sp.		0.4	23
GONETRUN	4.0	1	*Gomphonema truncatum*	Ehrenberg	0.1	9
GYSIACUM	4.0	3	*Gyrosigma acuminatum*	(Kützing) Rabenhorst	0.1	15
MEDICIRC	5.0	2	*Meridion circulare*	(Greville) Agardh	1.4	31
MELOVARI	4.0	1	*Melosira varians*	Agardh	0.9	33
NAVIACCO	1.0	3	*Navicula accomoda*	Hustedt	0.1	10
NAVIATOM	2.2	1	*Navicula atomus*	(Kützing) Grunow	0.3	19
NAVIATpe	2.3	1	*Navicula atomus* var. *permitis*	(Hustedt) Lange-Bertalot	2.1	40
NAVICATO	3.0	2	*Navicula capitatoradiata*	Germain	0.1	14
NAVICINC	3.0	1	*Navicula cincta*	(Ehrenberg) Ralfs	0.1	13
NAVICONT	4.0	1	*Navicula contenta*	Grunow in Van Heurck	0.2	6
NAVICRCE	3.5	2	*Navicula cryptocephala*	Kützing	0.4	20
NAVICRTE	4.0	1	*Navicula cryptotenella*	Lange-Bertalot	1.3	63
NAVIEXIL	4.8	2	*Navicula exilis*	Kützing	0.1	8
NAVIGOEP	2.0	2	*Navicula goeppertiana*	(Bleisch) H. L. Smith	0.1	4
NAVIGREG	3.4	1	*Navicula gregaria*	Donkin	1.3	50
NAVIHALA	2.0	3	*Navicula halophila*	(Grunow ex Van Heurck) P.T. Cleve	0.5	14
NAVILANC	3.8	1	*Navicula lanceolata*	(C. Agardh) Ehrenberg	9.0	67
NAVIMARG	2.0	3	*Navicula margalithii*	Lange-Bertalot	0.4	23
NAVIMELU	4.0	1	*Navicula menisculus*	Schum.	0.4	40
NAVIMILA	3.0	1	*Navicula minuscula*	Grunow	0.4	15
NAVIMImu	2.0	1	*Navicula minuscula* var. *muralis*	(Grunow) Lange-Bertalot	0.1	8
NAVIMINI	2.2	1	*Navicula minima*	Grunow	0.7	28
NAVIMUTI	2.0	2	*Navicula mutica*	Kützing	0.2	9
NAVIPHYL	2.6	3	*Navicula phyllepta*	Kützing	0.1	5
NAVIRADI	5.0	2	*Navicula radiosa*	Kützing	0.2	13
NAVIREIC	3.6	1	*Navicula reichardtiana*	Lange-Bertalot	0.1	18

Table 2 (continued).

Acronym	IPSs	IPSv	Taxon	author	Mean abundance	Perc. of samples
NAVISANA	2.6	2	*Navicula salinarum*	(Grunow) Hustedt	0.1	7
NAVISLES	3.0	3	*Navicula slesvicensis*	Grunow	0.2	20
NAVISPEC			*Navicula* sp.		0.2	17
NAVISUMI	2.0	1	*Navicula subminuscula*	Manguin	1.9	46
NAVITRIP	4.4	2	*Navicula tripunctata*	(O.F.Müll.) Bory	0.7	47
NAVITRIV	2.0	3	*Navicula trivialis*	Lange-Bertalot	0.1	10
NAVIVENE	1.0	2	*Navicula veneta*	Kützing	0.6	42
NITZACIC	2.0	2	*Nitzschia acicularis*	(Kützing) W.Smith	0.2	24
NITZAMPH	2.0	2	*Nitzschia amphibia*	Grunow	0.4	28
NITZCAPI	1.0	3	*Nitzschia capitellata*	Hustedt	0.4	22
NITZCONS	2.4	2	*Nitzschia constricta*	(Kützing) Ralfs	0.1	15
NITZCOTA	2.0	3	*Nitzschia commutata*	Grunow	0.1	6
NITZDISS	4.5	3	*Nitzschia dissipata*	(Kützing) Grunow	2.6	69
NITZFONT	3.5	1	*Nitzschia fonticola*	Grunow in Van Heurck	0.6	29
NITZFRUS	2.0	1	*Nitzschia frustulum*	(Kützing) Grunow	0.6	27
NITZGRLI	3.0	2	*Nitzschia gracilis*	Hantz. ex Rabenhorst	0.2	11
NITZHEUF	4.0	1	*Nitzschia heufleriana*	Grunow	0.2	21
NITZINSP	2.8	1	*Nitzschia inconspicua*	Grunow	2.4	54
NITZLINE	3.0	2	*Nitzschia linearis*	W. Smith	0.2	25
NITZPACE	2.5	1	*Nitzschia paleacea*	(Grunow) Grunow in Van Heurck	2.1	50
NITZPALE	1.0	3	*Nitzschia palea*	(Kützing) W. Smith	1.4	50
NITZRECT	3.0	2	*Nitzschia recta*	Hantzsch in Rabenhorst	0.2	22
NITZSIMO	3.0	1	*Nitzschia rosenstockii*	Lange-Bertalot	0.1	15
NITZSOBI	3.0	3	*Nitzschia sociabilis*	Hustedt	0.4	20
NITZSPEC			*Nitzschia* sp.		0.1	8
NITZTUBI	2.8	2	*Nitzschia tubicola*	Grun. in Cleve & Grun	0.1	6
NITZVERM	1.5	2	*Nitzschia supralitorea*	Lange-Bertalot	0.2	17
PSAMGRIS	5.0	2	*Psammothidium grischunum*	Wuthrich	0.1	0
RHSPABBR	4.0	1	*Rhoicosphenia abbreviata*	(Agardh) Lange-Bertalot	2.3	62
STDIHANT	1.8	1	*Stephanodiscus hantzschii*	Grunow	0.7	23
STDIMINU	4.0	1	*Stephanodiscus minutulus*	(Kützing) Cleve & Möller	0.5	27
STDIPARV	3.0	1	*Stephanodiscus parvus*	Stoermer & Hakansson	0.2	9
STEPINVI	2.6	1	*Stephanodiscus invisitatus*	Hohn & Hellermann	0.2	12
STEPTENU	2.8	1	*Stephanodiscus tenuis*	Hustedt	0.1	7
SURIANGU	4.0	1	*Surirella angusta*	Kützing	0.1	19
SURIBREB	3.0	2	*Surirella brebissonii*	Krammer & Lange-Bertalot	2.0	63
SURIOVAL	2.0	2	*Surirella ovalis*	Kützing	0.2	6
THSIPSNA	2.0	2	*Thalassiosira pseudonana*	Hasle & Heindal	0.1	9

current velocity, altitude, shading, oxygen and nutrients (nitrate, but also total phosphorus and Kjeldahl-nitrogen) (Table 3). Thus, there is a major gradient from smaller, fast-running shaded upland streams to more nutrient-rich, larger lowland rivers on the first axis. The second axis is correlated highly negatively with average width, indicating that the smallest streams are in the left top part of Fig. 1. The negative correlation with nutrients indicates that these sites are the most nutrient-poor ones in the data set. Consequently, diatoms from nutrient rich waters would be on the right-hand side of Fig. 1 and species from cleaner waters on the left-hand side. The most common species are in the centre of the diagram and the scarce ones at the margins.

It is apparent from the correlation table that nutrient status, including organic materials, is not among the master variables in determining the taxonomic composition of diatom assemblages in Hungarian running waters. This is corroborated by the fact that none of the diatom indications (IPS or Van Dam indices) has very high correlations with the first axis. Only acidity has a fairly high (-0.37) correlation with the first axis. Diatom indications show that nutrient status and organic matter contents are becoming more important on the third and even the fourth axis of the ordination.

Cluster analysis

The whole set of 339 samples was split into 12 clusters (Fig. 2). In the process of cluster analysis the final number of clusters has to be decided by the analysts. We tried to identify as many clusters as was meaningful, using our knowledge about the ecological and other conditions of the water bodies. In Table 4, a selection of species is listed with the clusters placed in their logical order, as appears from Fig. 2. In the ordination diagram of Fig. 3 the samples are marked by their cluster numbers from Fig. 2 and Table 4.

A whole suite of more or less common species is present throughout all clusters, although no one of these species is a significant indicator for all groups. Twelve

Table 3. Performance of ordination and Pearson correlation coefficients of ordination axes with timing of sampling, environmental variables, human impact and diatom indicator values. Average values and number of observations for each parameter are given also. Variables in *italics* were transformed logarithmically. Average values for these parameters are geometric means for the untransformed values (all other averages are arithmetic means). *** = $p \leq = 0.001$, ** = $p \leq 0.01$, * = $p < 0.05$. Correlations < -0.4 or > 0.4 are underlined. Fraction as a unit means the fraction of sites that obeys the mentioned physical factor or human impact, e.g. 3 % of the sites has point source pollution in the watershed. IPS = Index de Polluosensitivité Specifique (Coste in CEMAGREF 1981), R, H, N, O, S, T and M are weighted averages of indicator values according to Van Dam et al. (1994).

Variable	Average	Unit	Axis 1	Axis 2	Axis 3	Axis 4	n
Performance of ordination							
Eigenvalue			0.32	0.18	0.15	0.13	339
Percentage of variance explained			8.5	4.8	4.0	3.5	339
Timing							
Sampling date	20-4-2005		0.20 ***	-0.25 ***	-0.01	0.03	339
Physical factors							
Calceareous substrate	0.94	fraction	0.23 ***	-0.01	-0.34 ***	0.14 *	316
Altitude	137	m	-0.48 ***	0.23 ***	0.12 *	-0.11 *	339
Valley form (scale 1-6)	4.58		0.44 ***	-0.23 ***	-0.02	0.05	339
Predefined water type	12.0		0.53 ***	-0.21 ***	-0.07	0.01	335
Shading	24	%	-0.36 ***	0.23 ***	0.08	-0.11 *	339
Current velocity (max)	0.61	m·s^{-1}	-0.53 ***	-0.20 ***	0.04	-0.02	336
Current velocity (mean)	0.39	m·s^{-1}	-0.51 ***	-0.28 ***	0.03	0.02	333
Secchi-depth	0.61	cm	0.14 *	0.07	0.08	-0.21 ***	258
Temperature	12.4	°C	0.48 ***	-0.27 ***	-0.04	0.08	336
Width	*8.0*	m	0.11 *	-0.51 ***	-0.16 **	-0.02	339
Mean depth	1.0	m	0.17 **	-0.17 **	-0.04	0.00	339
Maximum depth	1.4	m	0.16 **	-0.27 ***	-0.11 *	0.01	339
Permanency	0.94	fraction	-0.06	-0.23 ***	-0.08	0.03	339
Maior ions							
pH	8.0		-0.26 **	0.08	-0.18	0.06	116
Conductivity	794	µS·cm^{-1}	0.29 ***	0.16	-0.04	0.26 ***	334
Loss on ignition	336	mg·l^{-1}	0.26 ***	0.15 *	-0.16 *	0.23 ***	224
Total dissolved solids	512	mg·l^{-1}	0.28 ***	0.17 **	-0.18 **	0.22 ***	225
Total suspended solids	*31*	mmol·l^{-1}	-0.10	-0.09	0.09	-0.06	225
Total alkalinity	5.0	mg·l^{-1}	0.46 ***	0.07	-0.16 *	0.10	225
Calcium	87	mg·l^{-1}	0.07	0.28 ***	-0.14 *	0.12	225
Chloride	38	mg·l^{-1}	0.32 ***	-0.08	-0.02	0.27 ***	225
Sulphate	*72*	mg·l^{-1}	0.12	0.24 ***	-0.11	0.25 ***	225
Oxygen economy							
Oxygen	10.3	mg·l^{-1}	-0.43 ***	0.10	-0.21	-0.11 *	336
Oxygen saturation	97	%	-0.28 ***	0.03	-0.22	-0.09	336
Chem.oxygen demand	53	mg·l^{-1}	0.13	-0.03	-0.02	0.06	221
Nutrients							
Kjeldahl nitrogen	*0.97*	mg N·l^{-1}	0.31 ***	-0.09	0.01	0.31 ***	225
Ammonium	*0.11*	mg N·l^{-1}	-0.08	-0.03	0.04	0.30 ***	225
Nitrite	*0.05*	mg N·l^{-1}	0.06	0.01	-0.18 **	0.28 ***	225
Nitrate	*5.66*	mg N·l^{-1}	-0.45 ***	0.11 *	-0.08	0.07	225
Total phosphate	*0.18*	mg P·l^{-1}	0.20 **	-0.21 ***	-0.01	0.32 ***	225
Silicium	*3.9*	mg·l^{-1}	-0.27 ***	0.02	0.26 ***	0.02	225
Human impact							
Hydrologically modified	0.33	fraction	0.38 ***	-0.17 **	0.05	0.02	323
Point source pollution	0.03	fraction	0.00	0.04	0.00	0.07	339
Diffuse pollution	0.87	fraction	0.08	-0.10	-0.13 *	0.08	339
Sewage overflow	0.01	fraction	-0.05	-0.05	-0.01	-0.03	339
Removal of vegetation	0.91	fraction	0.08	0.05	-0.02	-0.02	339
Diatom indicator values							
IPS (Ind. Poll. Sensit.)	13.89		-0.18 ***	0.24 ***	0.01	-0.67 ***	339
R (acidity)	3.81		-0.37 ***	-0.22 ***	-0.42 ***	0.19 ***	339
H (salinity)	2.22		-0.24 ***	-0.13 *	-0.47 ***	0.25 ***	339
N (organic nitrogen)	2.24		0.16 **	-0.15 **	0.24 ***	0.72 ***	339
O (oxygen)	2.43		0.00	-0.29 ***	-0.12 *	0.68 ***	339
S (saprobity)	2.61		0.12 *	-0.15 **	0.01	0.74 ***	339
T (trophic state)	4.75		-0.20 ***	-0.01	-0.06	0.18 ***	339
M (moisture)	2.57		-0.20 ***	-0.11 *	0.39 ***	0.09	339

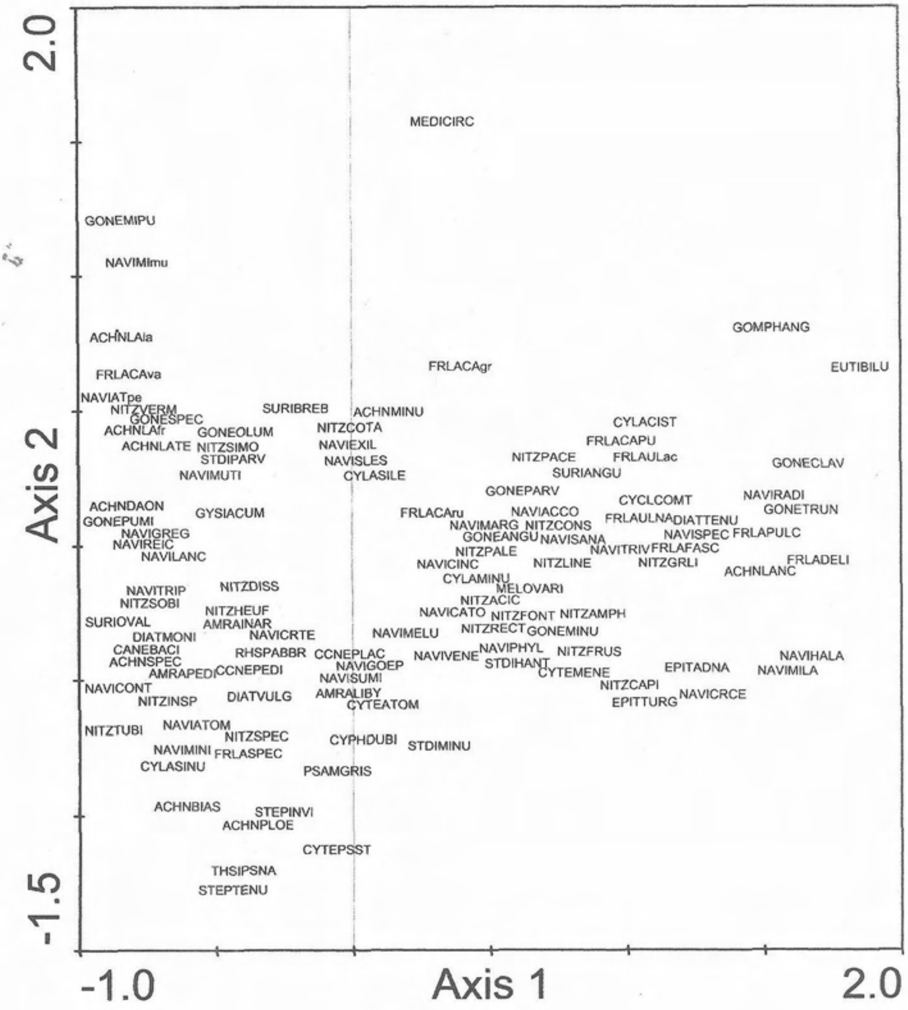


Fig. 1. First two axes of species scores in the ordination.

species are indicative for the closely related five clusters 2, 7 and 10–12, but most indicator species are significant for three or less clusters only.

The results of the analysis of variance shows that most of the measured environmental variables have significant differences between clusters. The same is true for the diatom indicator values. According to the IPS values, the ecological conditions of most clusters are not very different from each other, although the indicator species for cluster 2 have low IPS values (polluted water) (Table 5).

Cluster C10 (13 samples) is differentiated from all the other groups by the relatively large presence of the indicator species *Diatoma moniliformis, Surirella*

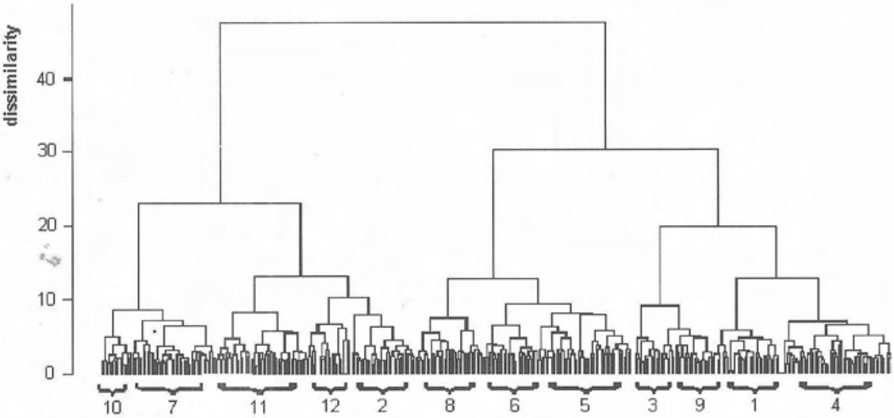


Fig. 2. Dendrogram showing the relationship between the 12 clusters.

ovalis and *Fragilaria capucina* var. *rumpens.* The sites of this cluster are in the upper left of the ordination diagram in Fig. 3. It has relatively high concentrations of oxygen, low temperatures, phosphorus, Kjeldahl-nitrogen and chemical oxygen demand. These are mainly (often shaded) high- or mid-altitude streamlets with a high sinuosity. The IPS values are high (clean water), the diatom indicator values for oxygen and saprobity are intermediate: the acidity, salinity and trophic indications are even high.

Cluster C7 (36 samples) has only one fairly typical species: *Navicula salinarum*, which is only present at 5 of the 36 sites. The sites are in the left part of the ordination diagram (Fig. 3). It shares its other indicator species with the related clusters 2 and 11–12. Current velocity is not particularly high, but a relatively large fraction of these streams has torrents. Therefore, oxygen concentrations are rather high. However, it has elevated values of calcium, sulphate, nitrogen species, nonpoint pollution and sewage inflow. This agrees with the high values of the diatom indices for oxygen, saprobity and trophic state. The IPS values are not particularly high. Most of these samples are from high- or mid-altitude ranges.

Cluster C11 (39 samples) has the aerophilous diatom *Navicula contenta* as a unique indicator species. In a large number of samples, *Nitzschia inconspicua* is present, which is also known to thrive under such conditions. Its centre of occurrence in the ordination diagram is down-left. In comparison with the other clusters, it has relative low values for pH (although absolutely high with 7.8), alkalinity, Kjeldahl-nitrogen and nitrite, but ammonium and nitrate are high (nitrogen is mineralized quite well). Silicate is also high here, probably because this cluster has the largest fraction of siliceous geology. Most of the sites are in high- and mid-altitude regions. The diatom indices for salinity and moisture are high, indicating temporary desiccation with increasing concentrations of solutes. The IPS is intermediate.

Table 4. Cumulative abundance of selected cluster indicator species. Significant ($p < 0.05$) and non-significant cluster indicator species are bordered by solid and non-solid borderlines, respectively. For each species the sensitivity number (IPSs) and indication value number (IPSv) is given. Taxa with high IPSs-values (5) are 'clean water species', taxa with high IPSv-values (3) are taxa with a narrow ecological range. 0 means <1, empty cells are real zeros.

	Ecology		Cluster											
Species	IPSs	IPSv	C10	C7	C11	C12	C2	C8	C6	C5	C3	C9	C1	C4
Number of sites			13	36	39	19	28	29	23	41	12	23	28	48
Gomphonema pumilum	5.0	1	5	1	8	2	12	0	0	2	25	1	139	2
Achnanthes laterostrata	5.0	3			2							0	7	17
Gomphonema angustatum	3.0	1	0	2		0		24	36	9			7	169
Epithemia adnata	4.0	3								2				34
Navicula mutica	2.0	2	2	11	1	4	1	0	4	3	47	0	1	1
Navicula accomoda	1.0	3		7	0	0		2	2	9	7	1	0	0
N. atomus var. *permitis*	2.3	1	3	27	40	11	16	30		5	430	101	29	11
N. minuscula var. *muralis*	2.0	1	0	2	1	0	0	1			2	9	8	1
Meridion circulare	5.0	2	3	13	7		2	3	53	43	1	195	13	126
Stephanodiscus parvus	3.0	1	4	14	1	0	3	0		1	3	32	2	2
Cyclotella comta	5.0	1		0	1	0		6	1	3		0	0	17
Nitzschia capitellata	1.0	3	3	7	3	5	6	4	5	89	5	1	0	5
Eunotia bilunaris	5.0	2		0			0	3	1	67	0			10
Cymbella cistula	4.0	3		2			0	1	20	9		2	0	10
Fragilaria capucina	3.0	1	5	18	76	4	12	61	126	142	1	19	16	80
Navicula cryptocephala	3.5	2	1	3	0	9	2	54	16	45			1	6
Cyclotella meneghiniana	2.0	1		23	12	6	14	44	15	56	1	5	4	17
Navicula halophila	2.0	3		1		3	8	35	9	110	0	0		3
Nitzschia amphibia	2.0	2		4	13	1	18	37	24	13		1	3	11
Navicula trivialis	2.0	3		2	0	2	1	2	6	5			0	2
Gomphonema truncatum	4.0	1		0			1	1	4	9		0		5
Achnanthes minutissima	5.0	1	39	51	220	124	69	123	482	255	74	397	876	2527
Amphora pediculus	4.0	1	11	44	533	138	658	121	68	15	36	130	398	158
Gomphonema parvulum	2.0	1	13	35	19	39	20	117	59	296	63	16	98	73
Gomphonema olivaceum	4.6	1	492	150	213	56	52	60	143	30	41	257	64	24
Cocconeis placentula	4.0	1	4	22	126	35	210	519	23	25	3	8	136	86
Nitzschia dissipata	4.5	3	35	69	115	229	65	37	51	78	11	92	63	22
Nitzschia palea	1.0	3	8	56	4	37	20	40	40	183	31	7	10	25
Nitzschia paleacea	2.5	1	5	51	7	31	9	13	74	370	59	79	6	14
Amphora inariensis	5.0	1	3	4	25	2	9	9	8	6		16	22	18
Navicula minima	2.2	1	2	5	50	6	128	19		12	3	6	2	7
Navicula atomus	2.2	1	3	6	10	6	39	34	0	1	2	2	3	4
Stephanodiscus invisitatus	2.6	1	1	12	1	31	18	0		0			3	4
Cyclotella atomus	2.0	1	1	12		74	2	3	0	22				17
Stephanodiscus minutulus	4.0	1	1	3	4	33	22	13	9	68	1	2	0	6
Nitzschia inconspicua	2.8	1	1	22	376	25	185	26	9	44	53	16	24	17
Caloneis bacillum	4.0	2	0	5	17	14	7	2	1	2	0	6	17	3
Achnanthes ploenensis	5.0	2		0	30	2	16	11	1	5	1		3	1
Navicula contenta	4.0	1	1	3	54	0		0		0	4		2	0
Navicula lanceolata	3.8	1	252	1573	743	81	54	45	112	35	62	18	55	18
Navicula gregaria	3.4	1	6	173	90	22	27	39	14	3	22	16	23	11
Navicula salinarum	2.6	2		23	0	1	2	4	2	4	0			0
Diatoma moniliformis	4.0	2	63	38	6	10	11	0	1	8	1	4	1	44

Cluster C12 (19 samples) has planktonic diatoms such as *Stephanodiscus invisitatus, S. minutulus, Cyclotella atomus* and five other, pennate species, including *Navicula goeppertiana* and *N. phyllepta* as indicator species. The samples are more concentrated down-left in the ordination diagram than the in the previous cluster. Values for pH, conductivity and major ions are typically low, but total suspended solids (turbidity) are high. Nutrient concentrations are low. According to the morphological parameters, large rivers, like Duna, Dráva and Tisza are in this cluster.

Table 5. Averages of environmental data and diatom indicator values. Variables in *italics* were transformed logarithmically. Average values for these parameters are geometric means for the untransformed values (all other averages are arithmetic means). IPS = Index de Polluosensitivité Specifique (Coste in CEMAGREF 1981), R, H, N, O, S, T and M are indicator values according to Van Dam et al. (1994). * = variables showing significant differences ($p \leq 0.05$) between clusters.

Variable	Units	C10	C7	C11	C12	C2	C8	C6	C5	C3	C9	C1	C4	Mean
Number of samples		13	36	39	19	28	29	23	41	12	23	28	48	28.3
Number of samples for lab. chemistry		10	24	31	6	13	17	14	21	8	23	22	36	18.8
Timing														
Date of sampling*	d-m	1-4	14-4	2-4	23-5	4-5	9-5	25-4	6-5	18-4	3-4	8-4	8-4	20-4
Physical factors														
Altitude*	m	159	140	168	103	135	105	114	99	155	171	183	132	137
Current velocity (mean)*	m·s^{-1}	0.57	0.41	0.55	0.49	0.65	0.30	0.19	0.24	0.61	0.26	0.50	0.22	0.39
Shading*	%	38	25	36	10	13	12	17	9	38	41	51	18	24
Temperature*	°C	8.4	11.6	8.0	16.9	12.7	15.3	13.8	16.7	11.3	10.0	9.4	12.5	12.4
Secchi depth	m	0.57	0.60	0.58	0.56	0.57	0.68	0.66	0.53	0.46	0.71	0.59	0.72	0.61
*Width**	m	6.2	7.4	4.0	57.4	34.0	9.2	8.3	10.2	4.5	2.5	4.0	7.2	8.0
Mean depth	m	0.7	0.8	0.7	1.6	1.4	1.1	1.2	1.1	0.7	0.8	0.6	1.3	1.0
Permanency	fraction	1.00	0.97	1.00	1.00	1.00	0.97	1.00	0.93	0.92	0.78	0.89	0.85	0.94
Maior ions														
pH*		8.2	8.1	7.8	7.8	8.2	7.9	8.0	7.8	8.0	8.3	8.0	7.9	8.0
Conductivity*	µS·cm^{-1}	646	929	641	440	546	1084	940	1051	896	1031	570	681	794
Total alkalinity*	mmol·l^{-1}	3.9	5.3	3.5	3.6	3.9	6.9	7.4	7.4	4.1	5.6	3.4	4.9	5.0
Total suspended solids	mg·l^{-1}	22	24	36	51	39	35	31	18	36	29	39	32	31
Ca*	mg·l^{-1}	76	104	75	59	70	94	88	79	91	124	65	89	87
Cl*	mg·l^{-1}	27	39	27	17	29	68	51	74	43	33	21	30	38
*Sulphate**	mg·l^{-1}	64	106	67	24	46	111	75	88	85	126	44	55	72
Oxygen and nutrients														
Oxygen*	mg·l^{-1}	11.9	12.2	11.6	10.5	10.6	8.4	9.9	6.9	10.2	11.9	11.8	10.1	10.3
Chemical oxygen demand	mg·l^{-1}	39	42	62	58	52	40	91	82	33	44	37	45	53
*Kjeldahl nitrogen**	mg N·l^{-1}	0.76	1.02	0.76	0.89	1.13	1.55	1.27	1.59	0.89	0.81	0.67	0.94	0.97
Ammonium	mg N·l^{-1}	0.12	0.15	0.14	0.05	0.10	0.16	0.11	0.12	0.10	0.10	0.08	0.09	0.11
*Nitrite**	mg N·l^{-1}	0.05	0.09	0.03	0.04	0.06	0.08	0.09	0.04	0.03	0.05	0.02	0.04	0.05
*Nitrate**	mg N·l^{-1}	7.51	9.78	8.65	4.56	5.57	3.35	5.56	2.28	11.09	7.61	5.61	4.15	5.66
*Total phospate**	mg P·l^{-1}	0.10	0.19	0.17	0.13	0.23	0.56	0.21	0.36	0.21	0.16	0.13	0.11	0.18
*Silicium**	mg·l^{-1}	4.0	4.1	7.1	1.9	3.1	5.4	3.4	2.3	5.8	3.0	6.4	2.9	3.9
Human impact														
Hydrologically modified*	fraction	0.08	0.19	0.18	0.25	0.15	0.52	0.41	0.54	0.33	0.04	0.08	0.40	0.29
Point source pollution*	fraction	0.00	0.00	0.08	0.00	0.00	0.03	0.13	0.00	0.17	0.09	0.00	0.00	0.03
Nonpoint pollution*	fraction	0.92	0.94	0.82	0.89	0.89	0.93	0.96	0.93	0.83	0.91	0.64	0.79	0.87
Sewage overflow	fraction	0.00	0.03	0.03	0.00	0.00	0.07	0.00	0.00	0.00	0.04	0.00	0.00	0.01
Diatom indicator values														
IPS (Index of Pollution Sensitivity)*		15.4	13.2	14.0	13.3	14.4	12.5	14.0	10.4	10.0	14.9	15.8	17.0	13.9
R (acidity)*		4.4	4.0	3.9	3.9	4.0	3.9	3.7	3.8	3.9	3.9	3.6	3.4	3.8
H (salinity)*		2.4	2.6	2.4	2.2	2.2	2.1	2.1	2.2	2.2	2.1	2.1	2.0	2.2
N (organic nitrogen)*		2.0	2.2	2.2	2.3	2.3	2.3	2.2	2.6	2.8	2.2	2.1	2.0	2.2
O (oxygen)*		2.3	2.9	2.5	2.4	2.5	2.8	2.3	2.9	3.2	2.2	2.0	1.6	2.4
S (saprobity)*		2.4	2.9	2.5	2.5	2.5	2.7	2.6	3.1	3.4	2.4	2.3	2.2	2.6
T (trophic state)*		4.9	5.0	4.8	4.8	4.6	4.9	4.6	4.7	5.0	4.9	4.8	4.5	4.8
M (moisture)*		2.0	2.7	2.7	2.4	2.6	2.4	2.4	2.5	2.8	2.4	2.7	2.7	2.6
Number of species in count *		29	32	33	45	38	33	32	29	23	27	29	24	31

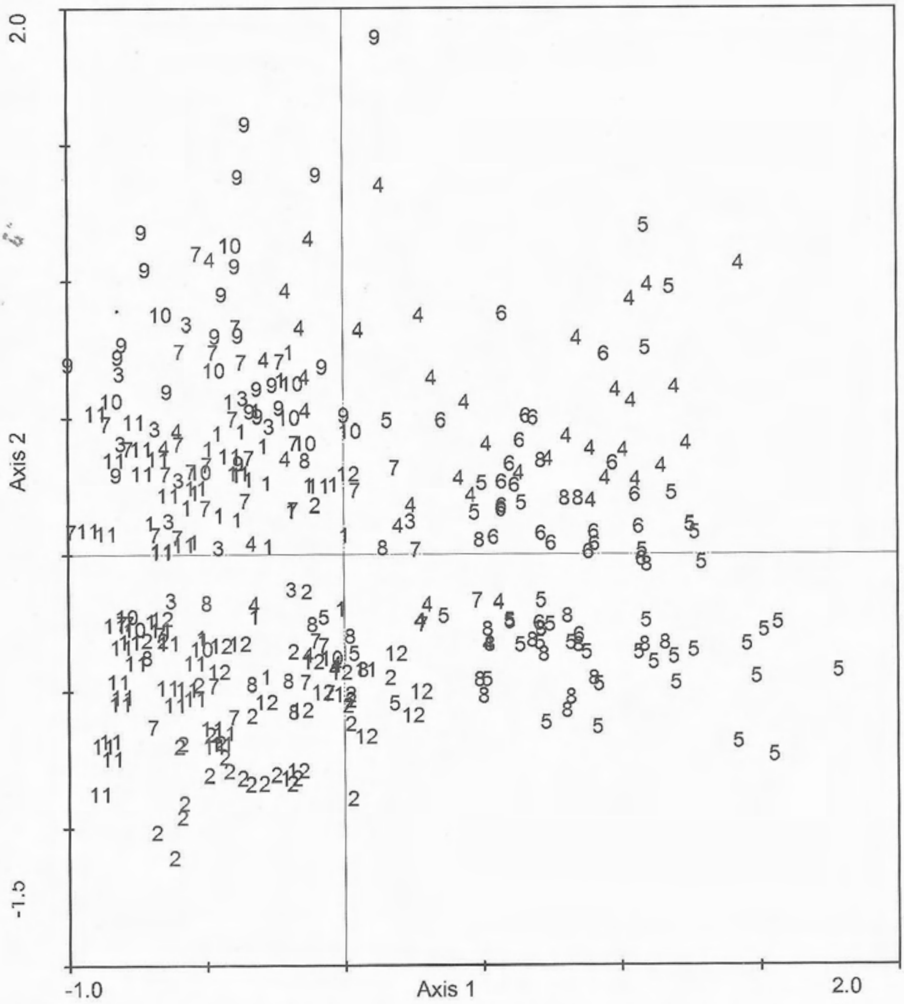


Fig. 3. Sample scores on the first two axes of the ordination. Samples are marked by cluster numbers of Fig. 2 and Table 4.

The long residence time in these rivers permits the development of planktonic diatoms, which partly settle in the phytobenthos. Therefore, the number of species in the counts is high. Due to the presence of planktonic diatoms the moisture indication is low.

Cluster C2 (28 samples) has *Achnanthes biasolettiana* and *Navicula minima* as indicator species. It shares *N. atomus* and *Cyclotella pseudostelligera* with the previous cluster. Due to many misidentifications by earlier authors, the ecology of *Achnanthes biasolettiana* is not well known, but the species is considered to be indicative for meso- to eutrophic water bodies, which is in conflict with the ecology of the other species characteristic of this group. The samples are more ore less in the same range of the ordination diagram as those from the previous cluster. It has relatively high values for pH and total suspended solids, but low to intermediate values for major ions. Values for nutrients indicate that this cluster is from more eutrophic waters than the previous one, with high concentrations (0.23 mg l^{-1}) of total phosphorus. The current velocity is high, as is the average width. The cluster includes sites from the river Duna, but also from smaller streams. Many of these are artificial. According to the diatom indicator values, the pH is high indeed, and the trophic state would be low. This cluster is difficult to interpret unambiguously.

Cluster C8 (29 samples) is the first cluster that belongs in the right-hand group of samples in the first main split in the dendrogram in Fig. 2, so it does not share indicator species with the previous groups, apart from those species that occur typically in all groups. The right-hand clusters (1, 3, 4–6, 8, 9) have the diatom *Cyclotella comta* in common as an indicator, although this diatom from oligotrophic waters is present in relatively few samples. Cluster C8 has no unique indicator species, but shares 18 species with the next two clusters, of which *Achnanthes lanceolata* and *Fragilaria ulna* are the most common. The samples are in the down-side in Fig. 3. It has high values for major ions and most nutrients in typically small to very large lowland or mid-altitude streams, with modified channels with non-point pollution and a relatively high proportion of sewage overflows. The number of species in the count is high, but the IPS is low (pollution). The diatom indictor values for organic nitrogen, oxygen, saprobity and trophic state are also fairly high, indicating pollution.

Cluster C6 (23 samples) has unidentified *Navicula*-taxa and *Cymbella cistula* as unique indicator species. It shares *Nitzschia gracilis* with the next cluster. Samples are in the upper right quarter of the ordination diagram (Fig. 3). Values for alkalinity (bicarbonate) and chloride are high. Low concentrations of oxygen and high values for chemical oxygen demand, Kjeldahl-nitrogen and nitrate indicate harsh oxygen conditions. Current velocities are low, depths are high and morphological parameters indicate modified and artificial medium-size channel forms as well as non-point pollution at many of the sites. The number of species is, nevertheless, fairly high. The IPS values are not particularly high. The indicator values for acidity and trophic state point to **relatively** acid (pH = 8.0!) and low trophic state.

Cluster C5 (41 samples) has five species with very different ecological properties as indicator taxa: species from mesotropic waters, such as *Fragilaria delicatissima* and *Eunotia bilunaris,* to species from more or less brackish waters, like *Fragilaria pulchella* and *Nitzschia capitellata.* Samples from this cluster are present in the right part of the ordination diagram in Fig. 3. In contrast to the previous cluster, the water is turbid (low Secchi depth), but alkalinity and chloride are high and the oxygen budget seems to be bad, too. Current velocities are low, due to the high depths and heavily modified hydromorphological conditions. There are mainly lowland small to very large rivers with a medium to fine substrate in these clusters, although some mid-altitude stations are also included. The number of species is fairly high and the dominance is low. The IPS and diatom indicator values indicate a high saprobity and bad oxygen budget.

Cluster C3 (12 samples only) has *Navicula mutica* and *N. accomoda* as unique indicator species. These taxa are well adapted to waters with a high organic load and fluctuating water levels (shallow waters), as are some of the indicator taxa shared with the next group, e.g. *N. atomus* var. *permitis* and *N. minuscula* var. *muralis.* The sites are concentrated in the lower left part of of Fig. 3. The water bodies are shallow (which may be the reason that the reported Secchi depth is low), rich in nutrients and low in oxygen, although the current velocity is relatively high. The fraction of sites affected by point pollution is relatively high (0.17). Most of the sites are small- or medium-sized high- and mid-altitude stations. The number of species in the count is low and the IPS and diatom indicator values indicate to severe organic pollution and low depth (temporary desiccation).

Cluster C9 (23 samples) shares three indicator species with the previous cluster, of which *N. atomus* var. *permitis* and *N. minuscula* var. *muralis* are adapted to organic load and temporary desiccation. Three indicator species are unique for this cluster, of which *Meridion circulare,* one of the very few typical running water diatoms, is most common. pH and major ions are high here, as is Secchi depth. Oxygen conditions are much better than in the previous cluster. Total nitrogen is high, due to high nitrate concentrations. These stations occur in high- and mid-altitudes (small streams with rough or medium-fine substrate) with intensive shading. The channel form is often more or less natural, but point pollution may be a problem at some of the sites. The number of species is relatively low, but the IPS is fairly high, which concurs with low values for the diatom indices for oxygen and organic nitrogen.

Cluster C1 (28 samples) is characterized by the unique indicator species *Gomphonema pumilum* and two species which are shared with the next cluster: *Achnanthes laterostrata* and *A. daonensis.* These species are characteristic of oligo-mesotrophic water bodies and that seems also to be the case for *G. pumilum.* Samples are mainly in the upper left quarter of Fig. 3. Major ions and nutrient concentrations are in accordance with the autecological observations. Concentrations of silicate are high here, as a relatively large part of the sites have a siliceous bedrock, although there are also calcareous sites. They are often shaded and shallow streams with no hydromorphological changes. The IPS is high (good water quality) and diatom indicator values for acidity, salinity, organic nitrogen, oxygen and saprobity are low.

Cluster C4 (48 samples) shares *Achnanthes laterostrata* and *A. daonensis* with the pervious cluster. They are characteristic for oligo- and mesotrophic water bodies. *Gomphonema angustatum* is the only significant unique indicator species in this cluster. The ecology of this species is not well known. Samples from this cluster are mainly spread out over the upper half of the ordination diagram. Concentrations of major ions are intermediate and those of nutrients are low. Both Secchi depth and mean depth of these sites are high. Nevertheless, there seem to be many non-permanent sites in this cluster. Many of them are (straightened) channels and current velocity is low. Most of the samples derive from (very) small streams with a medium-fine substrate, but some are from very large rivers. The number of species in the counts is low and the dominance percentage is high (low diversity). The average IPS has its maximum in this cluster (good water quality). The good water quality is also reflected by low values of most of the diatom indicator values.

Comparison between physiographic river types and diatom river types

Generally, there was some correspondence between diatom and physiographic river types, but the original diatom type is replaced by deviating types, according to human impact (Table 6). As an example: in high altitude streams (types 4 and 5) the original diatom cluster C10 is replaced by other clusters, including C7, due to eutrophication. At the other hand: the clusters C5, C6 and C8 are nearly confined to the lowlands and are only rarely found in mid-altitude streams.

Table 6. Correspondence between Hungarian physiographic river types (see Table 1 for definitions) and river diatom clusters according to Table 4. Entries are numbers of samples. U = type unknown.

	Hungarian river type																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	u	sum
C1	6	2	1	2	2			7			1	1	2		1			1									2	28
C2	1				1	6	1			2			6						2	2			1	4		1	1	28
C3	2		1	1					2	2	1	1			1									1				12
C4	3	1	1	3	1			6	1	1		4	1		7	5		3			1	3		2		5		48
C5				1	1			1	1						6	10	4	6	4	6						1		41
C6					1			1	2						3	2	1	3	1	1	2					6		23
C7		2		2	6		2	6	5	1		2		2	1			5	1	1								36
C8				1				4	1			1				2	4	7	7	1						1		29
C9		9	1	4	1			4	2						1			1										23
C10			2	3	1	2					1		1	1						1	1							13
C11	6	5		5	7	2		3	4		2	1	1					1	2									39
C12								2		1			1	2			1		2	3			2	2	2		1	19
sum	18	19	6	22	21	10	3	34	18	7	5	10	12	5	20	19	10	27	19	15	4	3	3	9	2	14	4	339

Class boundaries and the present quality of Hungarian rivers

For Hungarian running waters, the IPS seems to be a suitable index to use, as the major variation in diatom assemblages is due to hydromorphology and concentrations of major ions. Naturally, as stream order increases, water courses get rich in nutrients, and become more and more turbid. Such conditions favour shade-adapted species with higher nutrient demand. Presence of these species must not be understood as deterioration of the ecological status, without taking into account

Table 7. IPS-class boundaries for high-, mid- and low-altitude streams. Symbols correspond to those in Fig. 4.

Quality	High altitude	Mid altitude	Low altitude	Symbol
high	$17 \leq IPS \leq 20$	$16 \leq IPS \leq 20$	$15.5 \leq IPS \leq 20$	○
good	$13 \leq IPS < 17$	$12 \leq IPS < 16$	$11.5 \leq IPS < 15.5$	□
moderate	$9 \leq IPS < 13$	$8.5 \leq IPS < 12$	$8 \leq IPS < 11.5$	△
poor	$5 \leq IPS < 9$	$5 \leq IPS < 8.5$	$5 \leq IPS < 5$	▽
bad	$IPS < 5$	$IPS < 5$	$IPS < 5$	✝

the natural variation in expected conditions as stream size increases. Therefore there is a decrease in average values of IPS in high attitude streams (14.96), via 13.98 in mid-altitude streams to 13.52 in low-altitude streams. The IPS has its strongest correlation with total phosphorus ($r = -0.40$), Kjeldahl-nitrogen (-0.39) and chloride (-0.38). These variables are also highly intercorrelated.

Class boundaries for the IPS were not set by the original authors, but generally the classification of the related Indice Biologique des Diatomées (IBD) by PRYGIEL & COSTE (2000), as developed in north-western France, is used for distinction of water quality classes. Starting from these class boundaries set (as kept for Hungarian high-altitude streams) the adapted class boundaries for these different stream types are presented in Table 7.

Nearly one quarter (23 %) of the sites has a very good (high) quality and another 58 % has a good quality, so 81 % fulfil the criteria for phytobenthos for the Water Framework Directive. About 15, 4 and 0.6 % of the sites have a moderate, poor or bad quality, respectively and measures have to be taken at these sites for improvement of water quality (Fig. 4).

Discussion

Data-set

The present study is fairly unique, because it is the first time that a nation-wide survey of benthic river diatoms has been carried out in any European country. The dataset is robust as the primary sources of uncertainty in diatom training sets (choice of site and substrate for sampling, inter-operator differences in diatom taxonomy and counting techniques; BESSE-LOTOTSKAYA et al. 2006) were eliminated as much as possible. Moreover, parallel with the diatoms, investigations were made on other biological quality elements, including macro-invertebrates and fish.

In total, nearly 500 taxa were found in the 400 valve counts from 339 sites. This certainly does not include all diatom taxa in the inland waters of a country like Hungary, which probably number between 1000 and 2000 'conservative' taxa.

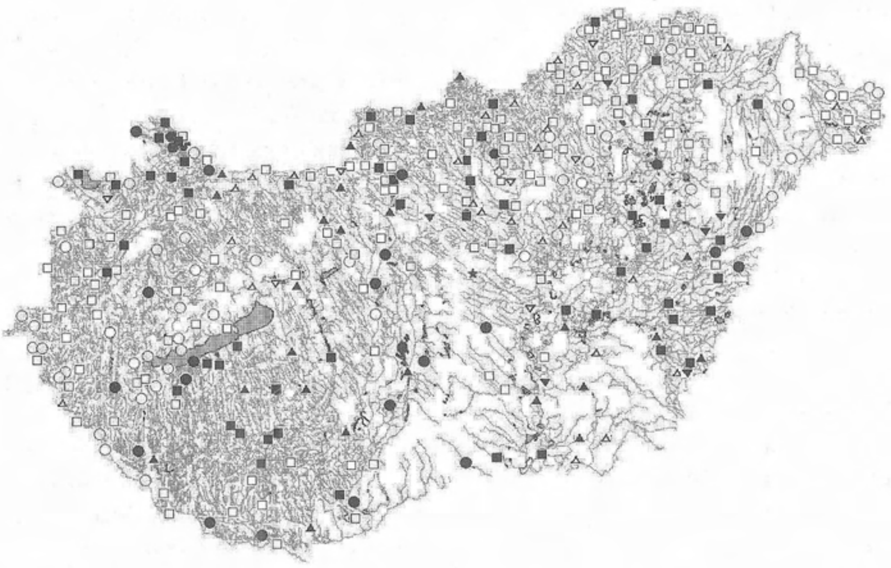


Fig. 4. Diatom-inferred quality of Hungarian running waters. Symbol-types correspond to those in Table 7. Heavily modified water bodies are marked with black symbols.

Many of the taxa which were not recorded in the present survey (probably some hundreds) are naturally rare at the sampling stations, e. g. the larger ones, while others are present in habitats that have not been included in this report, like lakes, fens, bogs and springs.

The commonest genus is *Navicula* (sensu lato), which is present in many different kinds of water. The second important genus is *Nitzschia,* indicating that many of the sites have a load of organic, biodegradable material. The genera *Achnanthes* and *Cymbella* are not rare in the dataset and they have many species which prefer clean water that is rich in oxygen.

Reference conditions and river typology

For a proper development of a system of ecological status classes within the implementation process of the Water Framework Directive at first reference conditions are to be set for each of the biological water quality elements, than relevant stressors (human impact) are to be identified and finally metrics have to be constructed.

Reference conditions do not necessarily equate to totally undisturbed, pristine conditions. They include very minor impacts which means that human pressures are allowed as long as they are of no or only very minor ecological effects (CIS Working Group 2.3 – REFCOND 2003). In this study, it has been assumed that all sites with IPS-values above the threshold values for high water quality in Table 7 are reference sites. Formally all sites in the dataset should be screened against the criteria for the presence of diffuse and point source pollution, morphological alterations, flow regulation, etc. as listed by the CIS Working Group.

During a pilot survey describing reference conditions, the main conclusion (supported by more than only one indicator group of organisms) was that the number of physiographic river types existing at that time was too high: some (especially some lowland types, but also others) have to be pooled, otherwise it is impossible to find more or less distinct assemblages of biota that characterize types (Szilágyi 2004). Unfortunately, the number of physiographic river types did not change in the ECOSURV project and therefore we faced the same dilemma. However, results of this study provide a much more solid basis for developing practically the same conclusion.

Our ordinations have clearly shown that typological elements (stream order [size of catchments], depth [temporary desiccation], altitude, current velocity, shading) are master variables determining distribution and abundance of diatom taxa and they rather clearly separate along such axes. Nutrients and organic content becomes important only on the third or fourth axes. Locally they may affect the composition of the diatom assemblages seriously, e. g. at the sites in cluster C3. Acidification is not important in these streams: with alkalinity values above 3 mmol l^{-1} they are well buffered (Table 5).This result is promising because it allows the conclusion that description of reference conditions for well defined types is possible.

Cluster analysis of the samples resulted in twelve distinct groups. These groups do not accord to the physiographic typology, especially since excellent and poor conditions exist in each of them. For example, in cluster-10 samples from high/mid altitudes with conditions close to reference are pooled, while cluster 7 contains samples from similar types but far from reference conditions. Similar grouping can be traced in the big cluster 3-9-1-4. Therefore, diatom assemblages allow differentiation clearly only between some physiographic 5–7 types that cannot be matched with the existing 26 types of rivers. Reduction in types needs consideration of results from other indicator groups, especially macro-invertebrates.

Suitability of metrics

Recently an intercalibration exercise was performed on the phytobenthos scales in running waters of twelve central and Baltic European countries (Kelly et al. 2007). In these countries as many as ten different metrics are used. Many of them are based on the IPS (Coste in CEMAGREF 1982) or on the Trophic Index (Rott et al. 2003). The latter one (TI) has been developed in the predominantly oligotrophic Austrian rivers.

TI and IPS have their largest discriminating power in the oligotrophic-mesotrophic and mesotrophic-eutrophic range, respectively. The IPS seems to develop more or less as a standard in the more eutrophic range, due to the inclusion of nearly all common European taxa and to its geographic flexibility; i.e. the possibility to set different class limits in different regions. Therefore the IPS is becoming a popular tool in many European countries (e. g. PRYGIEL et al. 1999; HLUBIKOVA et al. 2006; KAHLERT et al. 2006; PICIŃSKA-FAŁTYNOWICZ et al. 2006; VILBASTE et al. 2006; WOJTAL 2006).

Comparison with other biological quality elements

The results from macro-invertebrates and fish from the same sites are entered in Table 8. For fish only about half of the sites was sampled and for these sites the total score was calculated, using the minimum scores of those for diatoms, macro-invertebrates and fish, according to the 'one out all out' principle of the Water Framework Directive.

Table 8. Percentage of sites in each water quality class, using different biological quality elements (data from ARCADIS 2005).

Biological Quality Element	Number of sites	Quality				
		High	Good	Moderate	Poor	Bad
Phytobenthos	339	23	58	15	4	0
Macroinvertebrates (organic load)	340	1	66	31	1	1
Macroinvertebrates (general degradation)	340	14	18	24	30	14
Fish	184	1	22	34	27	16
All (minimum)	183	0	14	30	31	25

It is clear that the phytobenthos gives the most positive impression of water quality, as it is much less sensitive to hydromorphological changes than the macro-invertebrates (general degradation) and fish. The phytobenthos results are more or less comparable to those for the macro-invertebrates (organic load).

Concluding remarks

The present study provided a relatively homogeneous data set for a whole country, and compared to diatoms and diatom studies in most other European countries. The study shows that despite the intensive human use of over 90 % of the area of the country, natural variation is still the predominant factor for the composition of river diatom assemblages. The present survey is a snapshot of this variation at one moment in time. A better understanding of the relationships between diatom assemblages and environmental factors (including stressors) will be gained by a more time-intensive monitoring of selected stations in the present network.

Acknowledgements

We gratefully acknowledge our colleagues Dr. B. Kiss, Dr. P. Juhász and Dr. Z. Müller (BioAquaPro, Debrecen) for performing the field work, Mr. R. Nieuwenhuis (Arcadis, Arnhem [presently: Aqua Integra Consult, Utrecht], Mr. J. Dennisøe (Dansk Hydrological Institute, Copenhagen) and Dr. C. Dieperink (Waterframe, Ry, Denmark) for valuable discussions and technical assistance. Dr. M. Coste (CEMAGREF, Bordeaux) provided an updated list of IPS indicator values. The project was funded by the European Union and the Hungarian Ministry of Environment and Water (project EuropeAid/114951/D/SV/2002-000-180-04-01-02-02). The paper improved by the comments of two anonymous referees.

References

Ács, É. & Kiss, K.T. (1991): Investigation of periphytic algae in the Danube at Göd (1669 river km, Hungary). – Arch. Hydrobiol. Suppl. **89**: 47–67.

Ács, É., Szabó, K., Tóth, B. & Kiss, K.T. (2004): Investigation of benthic algal communities, especially diatoms of some Hungarian streams in connection with reference conditions of the Water Framework Directives. – Acta Botan. Hung. **46**: 255–277.

ARCADIS (2005): Ecosurv: final technical report, part A. EuropeAid 114951/D/SV/2002-000-180-04-01-02-02. – 48pp. Ministry of Environment and Water, Budapest.

Besse-Lototskaya, A., Verdonschot, P.F.M. & Sinkeldam, J.A. (2006): Uncertainty in diatom assessment: sampling, identification and counting variation. – Hydrobiologia **566**: 247–260.

Buczkó, K. & Ács, É. (1994): Algological studies on the periphyton in the branch-system of the Danube at Cikolasziget (Hungary). – Verh. Intern. Verein. theor. angew. Limnol. **25**: 1680–1683.

CEMAGREF (1982): Étude des méthodes biologiques d'appreciation quantitative de la qualité des eaux. Rapport Q.E. Lyon. – Agence de l'Eau Rhône-Mediterrane Corse, 218 pp.

Cholnoky, B.J. (1929): Epiphyten-Untersuchung im Balatonsee. – Int. Rev. ges. Hydrobiol. **22**: 313–345.

CIS Working Group 2.3 – REFCOND (2003): Rivers and lakes – Typology, reference conditions and classification systems. – Guidance document **10**: 87 pp. Office for Official Publications of the Europ. Comm., Luxembourg.

Dufrène, M. & Legendre, P. (1997): Species assemblages and indicator species: the need for a flexible asymmetrical approach. – Ecol. Monogr. **67**: 345–366.

European Union (2000): 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. – PE-CONS 3639/1/00, REV 1, ENV 221, CODEC 513: 1–152. Brussels.

Hlubikova, D., Miettinen, J. & Hindakova, A. (2006): Testing of diatom water quality indices in monitoring and reference sites od Slovakian rivers (Central Europe). – In: Ács, É., Kiss, K.T., Padisák, J. & Szabó, K.E. (eds.): 6th International Symposium on Use of Algae for Monitoring Rivers: 62–67. Magyar Algológiai Társaság, Göd.

Kahlert, M., Andrén, C. & Jarlman, A. (2006): Diatoms as biomonitoring tools in

Sweden. – In: Ács, É., Kiss, K.T., Padisák, J. & Szabó, K.E. (eds.): 6th International Symposium on Use of Algae for Monitoring Rivers: 71–75. Magyar Algológiai Társaság, Göd.

Kelly, M., Bennett, C., Coste, M., Delmas, F., Denys, L., Ector, L., Fauville, C., Ferreol, M., Golub, M., Jarlman, A., Kahlert, M., Lucey, J., Ni Chathain, B., Pardo, I., Pfister, P., Picinska-Faltynowicz, J., Schranz, C., Schaumburg, J., Tison, J., Van Dam, H. & Vilbaste, S. (2007): Central/Baltic GIG Phytobenthos Intercalibration Exercise. – 59 pp. Bowburn Consultancy, Durham.

Kelly, M.G., Cazaubon, A., Coring, E., Dell'Uomo, A., Ector, L., Goldsmith, B., Guasch, H., Hürlimann, J., Jarlmann, A., Kawecka, B., Kwadrans, J., Laugasta, R., Lindstrøm, E.-A., Leitão, M., Marvan, P., Padisák, J., Pipp, E., Prygiel, J., Rott, E., Sabater, S., Van Dam, H. & Vizinet, J. (1998): Recommendations for the routine sampling of diatoms for water quality assessments in Europe. – J. Appl. Phycol. **10**: 215–224.

Kovács, Cs., Kahlert, M. & Padisák, J. (2006): Benthic diatom communities along pH and TP gradients in Hungarian and Swedish streams. – J. Appl. Phycol. **18**: 105–117.

Krammer, K. (2002): Diatoms of Europe. Diatoms of the European Inland Waters and Comparable Habitats. Vol. **1–4**. – Gantner, Ruggel.

Krammer, K. & Lange-Bertalot, H. (1986–1991): Bacillariophyceae 1.–4. Teil: Naviculaceae. – In: Pascher, A. (ed.): Süßwasserflora von Mitteleuropa. Band **2**/1–4. G. Fischer, Heidelberg/Berlin.

Lange-Bertalot, H. (1993): 85 neue Taxa und über 100 weitere neu definierte Taxa ergänzend zur Süßwasserflora von Mitteleuropa Vol. **2**/1–4. – Bibl. Diatomol. **27**: 454 pp.

– (1998): As a practical limnologist, how does one deal with the flood of new diatom names? – Limnologica **28**: 153–156.

LeCointe, C., Coste, M. & Prygiel, J. (1993): Omnidia: a software for taxonomy, calculation of diatom indices and inventories management. – Hydrobiologia **269/270**: 509–514.

Padisák, J. (1982): The periphyton of Lake Fertõ: species composition and chlorophyll-a content. – BFB-Bericht **43**: 95–115.

Padisák, J., Ács, É., Rajczy, M. & Kiss, K.T. (1991): Use of algae for monitoring rivers in Hungary. – In: Whitton, B.A., Rott, E. & Friedrich, G. (eds.): Use of algae for monitoring rivers: 123–128. Inst. f. Botanik, Univ. Innsbruck, Innsbruck.

Pantocsek, J. (1901): Die Kieselalgen oder Bacillarien des Balaton. – Resultate der wissenschaftlichen Erforschung des Balatonsees **II**: 112 pp., Pls. 1–17 (Figs. 1–377). Budapest.

Picińska-Fałtynowicz, J. (2006): Epilithic diatom communities and their indicative value of streams of Sudety Mts. (SW Poland). – In: Ács, É., Kiss, K.T., Padisák, J. & Szabó, K.E. (eds.): 6th International Symposium on Use of Algae for Monitoring Rivers: 109–113. Magyar Algológiai Társaság, Göd, ISBN 963-06-0497-3.

Prygiel, J. & Coste, M. (eds.) (2000): Guide méthodologique pour la mise en oevre de l'indice biologogique diatomées. – NF T90-354: 134 pp. CEMAGREF, Gazinet.

Prygiel, J., Coste, M. & Bukowska, J. (1999): Review of the major diatom-based techniques for the quality assessment of rivers – State of the art in Europe. – In: Prygiel, J., Whitton, B.A. & Bukowska, J. (eds.): Use of algae for monitoring rivers. Proc.

Internat. Symposium held at the Agence de l'Eau Artois-Picardie, Douai, France, 29 Sept. – 1 Oct. 1997: 224–238. Agence de l'Eau Artois-Picardie, Douai.

ROTT, E., PIPP, E. & PFISTER, P. (2003): Diatom methods developed for river quality assessment in Austria and a cross-check agianst numerical trophic indication methods used in Europe. – Algol. Stud. **110**: 91–115.

SZEMES, G. (1959): Systematische und floristisch-ökologische Bearbeitung von Bacillariophyceen des Szelider Sees. – In: DONÁSZY, E. (Hrsgb.): Das Leben des Szelider Sees: 301–360. Akadémiai Kiadó, Budapest.

SZILÁGYI, F.(ed.) (2004): Referencia helyek jellemzése, passzportok véglegesítése. 2. Melléklet: a folyópasszportok véglegesítése. – 93 pp. Budapest Univ. of Technology and Eçonomics, Dept. of Sanitation and Environmental Engineering, Budapest.

TER BRAAK, C.J.F. & ŠMILAUER, P. (2002): CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5). – 500 pp. Biometris, Wageningen and České Budějovice.

VAN DAM, H., MERTENS, A. & SINKELDAM, J. (1994): A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands. – Netherl. J. Aquat. Ecol. **28**: 117–131.

VAN DAM, H., PADISÁK, J. & KOVÁCS, C. (2005): Ecosurv BQE report phytobenthos. Under the authority of the Ministry of Environment and Water, Hungary. – EuropeAid 114951/D/SV/2002-000-180-04-01-02-02: 54 pp. BioAquaPro, Debrecen/MottMac Donald, Croydon/DHI, Copenhagen/Arcadis, Arnhem/Grontmij | AquaSense, Amsterdam (downloadable from www.eu-wfd.info).

VILBASTE, S., LEISK, Ü. & IITAL, A. (2006): Diatom indices and species composition in relation to environmental parameters in Estonian streams. – In: ÁCS, É., KISS, K.T., PADISÁK, J. & SZABÓ, K.E. (eds.): 6th International Symposium on Use of Algae for Monitoring Rivers: 173–177. Magyar Algológiai Társaság, Göd.

WOJTAL, A. (2006): Use of diatoms for monitoring springs in Southern Poland. – In: ÁCS, É., KISS, K.T., PADISÁK, J. & SZABÓ, K.E. (eds.): 6th International Symposium on Use of Algae for Monitoring Rivers: 179–183. Magyar Algológiai Társaság, Göd.

Accepted: 11 April 2007