Central European Geology, Vol. 54/4, pp. 345–356 (2011)
DOI: 10.1556/CEuGeol.54.2011.4.3

Exploratory data analysis on the Upper-Tisza section using single and multi-variate data analysis methods

Péter Tanos*, József Kovács
Department of Physical and Applied Geology
Eötvös Loránd University, Budapest

Ilona Kovácsné Székely
Institute of Methodology
Budapest Business School, Budapest

István Gábor Hatvani
Department of Physical and Applied Geology
Eötvös Loránd University, Budapest

The River Tisza is one of Central Europe's most important rivers. In the last one and a half century numerous anthropogenic activities have influenced its watershed. As a result measures need to be taken to protect its water quality, necessitating a comprehensive picture of the spatial and temporal variability of its processes, which this study aims to extend further. In this study five sampling locations were analyzed in the upper section of the Tisza over the time interval 1974–2005, dealing with 24 parameters using multi-variate data analysis methods. Employing time series analysis and taking the river's tributaries into account, the strong influence of the River Szamos was pointed out, while stochastic connections indicated the influence of the Tiszalök Water Barrage System on the spatial variation of the Tisza's processes. Finally, by using principal component analysis (PCA), the different background factors were revealed in space and time (seasonal separation) as well. During summer the processes tended to be nitrogen-related, while during winter inorganic compounds play a greater role. Most importantly, spatial variety was observable in the factors.

Key words: data analysis, principal component analysis, River Szamos, River Tisza, stochastic connections, Tiszalök Water Barrage System (WBS), water quality

Introduction

The River Tisza collects the waters of the Carpathian Basin's eastern region. According to Lászlóffy (1982), the area of its watershed is 157,186 km^2. Less than one third of it is located in Hungary (Fig. 1).

From its source in the range of the Maramorosszkiy Massiv (in Hungarian: Máramarosi-havasok) to its confluence with the Danube, it stretches for 966 km

Addresses: *Corresponding author, address: H-1117 Budapest, Pázmány P. stny. 1/c, Hungary
e-mail: tanospeter@gmail.com
Received: August 28, 2010; accepted: December 22, 2011

1788-2281/$ 20.00 © 2011 *Akadémiai Kiadó, Budapest*

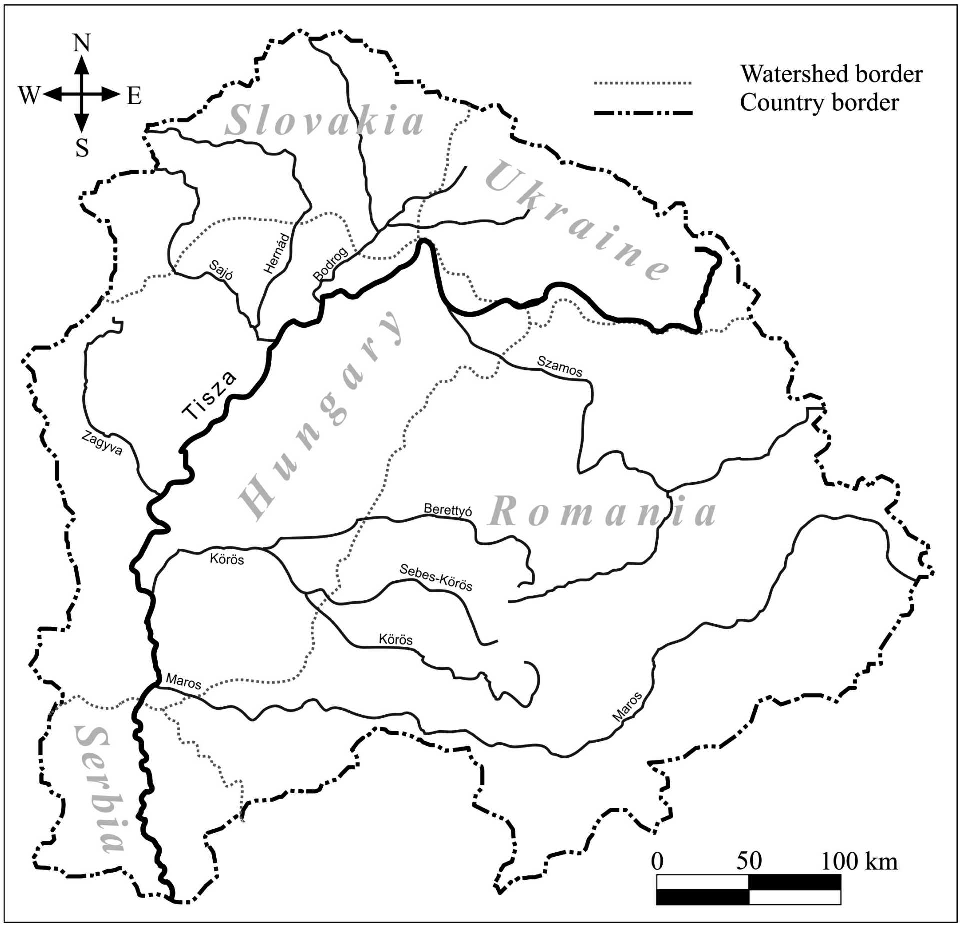


Fig. 1
Watershed of the River Tisza (Based on Istvánovics et al. 2010)

across the Ukraine, Romania, Slovakia, Hungary and Serbia (Sakan et al. 2007). The Hungarian section of the Tisza from border to border is 594.5 km. Its average runoff is 25.4 billion m^3 per annum (Pécsi 1969).

Despite the fact that in the last one and a half century numerous anthropogenic activities have influenced this area, in comparison to Europe's other large rivers it is still considered to have one of the most natural river valleys in Europe (Zsuga and Szabó 2005). Within Hungary alone, approximately 400 settlements and 1.5 million inhabitants' lives depend on its runoff and water quality.

Keeping these facts in mind our main aim was to give an overview, from a different perspective, of the processes of temporal and spatial evolution in the

Hungarian upper section of the Tisza, thus providing further information for protective measures.

Materials and methods

For many decades both systematic and occasional sampling of the River Tisza has been carried out in Hungary. As a result, a dataset is available which is sufficient for researchers to conduct comprehensive studies regarding the river's processes and water quality. Up to now there have hardly been any other methods apart from single and multi-variate analysis applied to these datasets. It is important to mention a few of these studies:

– Csépes et al. (2000) underlined the importance of the much higher conductivity of the first wave during flooding.

– Oláh et al. (2000) stressed the importance of decreasing the use of fertilizers, and emphasized the nitrogen consumption of plants in the river bank buffer zone.

– Szabó et al. (2004a, b) paired the runoff, chlorophyll-a and conductivity parameters, and analyzed their movement.

The application of multivariate methods is to be found only in the work of Lajter et al. (2009). However, this study only deals with ten "ecologically important" parameters, leaving out, for example, the anions and cations.

Sampling locations

Many surface waters are monitored as part of the National Sampling Network. In the case of the Tisza, data from the first five Hungarian sampling locations were analyzed (along a length of 258.7 river km) (Fig. 2).

The River Tisza reaches the Hungarian border at Tiszabecs. The next sampling location (SL) is at Záhony. There are two fairly large tributaries between these two locations, the Szamos and the Kraszna. The next SL is at Balsa, just upstream of the Tokaj and the Bodrog Rivers. The Tiszalök SL is located just below the Eastern Trunk Sewer and the Tiszalök Water Barrage System. The last analyzed SL is at Polgár below the mouth of the River Sajó.

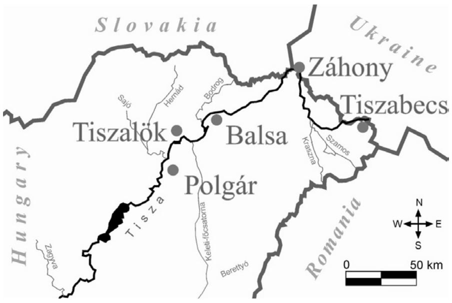


Fig. 2
The first five sampling locations on the River Tisza in Hungary

Acquired dataset

During the research a dataset covering 31 years (1974–2005) was analyzed, consisting of 300 000 data. From 1970 only one sample was taken a week at Tiszabecs and Polgár, according to Comecon specifications (Nagy et al. 2004). At Balsa only one sample was taken per month. At Záhony 26 samples were taken every year. In 1994 the Hungarian Standard No. MSZ 12749:1993 came into force. Since then 26 samples per annum have been taken uniformly at every SL.

The parameters used were as follows: Runoff ($m^3 s^{-1}$), pH, conductivity ($\mu S cm^{-1}$), M-alkalinity ($mval l^{-1}$), oxygen saturation (%), dissolved oxygen, BOD-5, CODC, Ca^{2+}, Mg^{2+}, Na^+, K^+, total hardness, carbonate hardness, Cl^-, SO_4^{2-}, HCO_3^-, NH_4-nitrogen, NO_2-nitrogen, NO_3-nitrogen, mineral nitrogen ($mg l^{-1}$), PO_4-phosphorous, chlorophyll-a (chl-a) ($\mu g l^{-1}$).

Applied methods

After dealing with the outlying and extreme values, the basic statistics of the dataset were prepared, such as the median, mean, standard deviation, lower and higher quartiles, as well as minimum and maximum values. These were used to obtain a better understanding of the temporal and spatial differences in the river, so that the datasets could be separated in the most appropriate fashion.

Thereafter correlation analysis was applied. This determines the linear connection between certain parameters. It is one of the basic methods in the analysis of stochastic processes. Its value varies between –1 and +1. The closer the coefficient is to ±1, the stronger the connection is. If the correlation coefficient is 0, then there is no linear connection.

Since the input of principal component analysis (PCA) is the correlation matrix, we subsequently analyzed the Tisza's background processes with this method. During multi-variate data analysis the assumption that the parameters are uncorrelated cannot be realized in most of the cases. Using PCA there is a possibility of linearly transforming the original variables to form hypothetic ones, called factors. With these factors the entire original system can be described without information loss.

Every hypothetic variable can be called a principal component if:

– It is a linear combination of the observed variables,

– It is uncorrelated, and for every component the sum of the coefficients squared is 1,

– The standard deviation of the components continuously decreases (the highest standard deviation belongs to the first component).

PCA was developed by Hotelling in 1933. The method basically corresponds to the positive semi-definite matrix's eigenvalue eigenvector problem" (Füstös et al. 1986).

Results

Basic statistics

Runoff is a river's most basic parameter, while the chl-a represents its biological activity. To begin with this basic statistic is presented.

Runoff (Fig. 3) between Tiszabecs and Záhony increased from 187 m^3 s^{-1} to 431 m^3 s^{-1}, because of the confluence of the Rivers Szamos and the Tisza. The runoff of the two rivers is almost equal. However, the watershed of the Szamos is much larger (15 882 km^2) than that of the Tisza in the corresponding section (13 173 km^2) (Somogyi 2003). Between Záhony and Polgár runoff again increased though the excess is only 77 m^3 over the course of 138.5 river km.

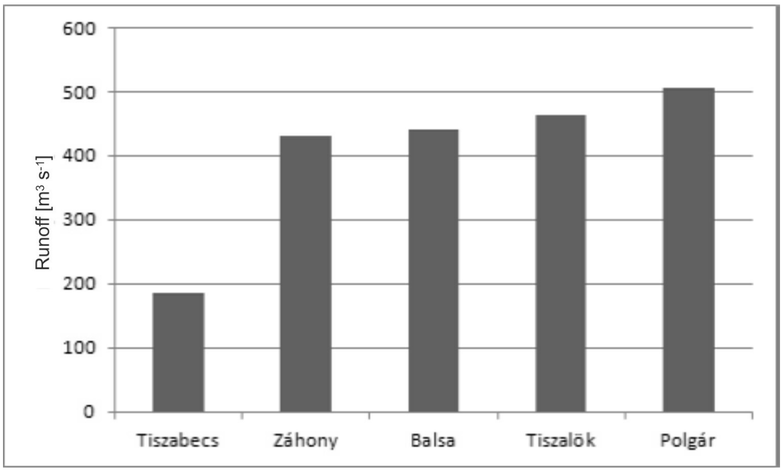


Fig. 3
Average runoff (1974–2005) at each sampling location

At the border, the Tisza can be described as having low primary production. This statement was backed up by a low chl-a content of 2.86 μg l^{-1}. However at Záhony (after the mouth of the Szamos), the chl-a content was almost eight times higher (22.85 μg l^{-1}). Afterwards, a gradual decrease was observed downstream (Fig. 4).

Because processes in a river are highly dependent both on temperature conditions and low and high-water stages, the first step in analyzing the temporal and spatial patterns was to separate the dataset into winter and summer data. As can be seen in Figure 5, winter runoff exceeded that of summer by 40–60%. This was one of the facts (along with temperature difference) that prompted us to separate summer and winter data during the PCA, so that the low and high-water stages and the different temperature conditions could be

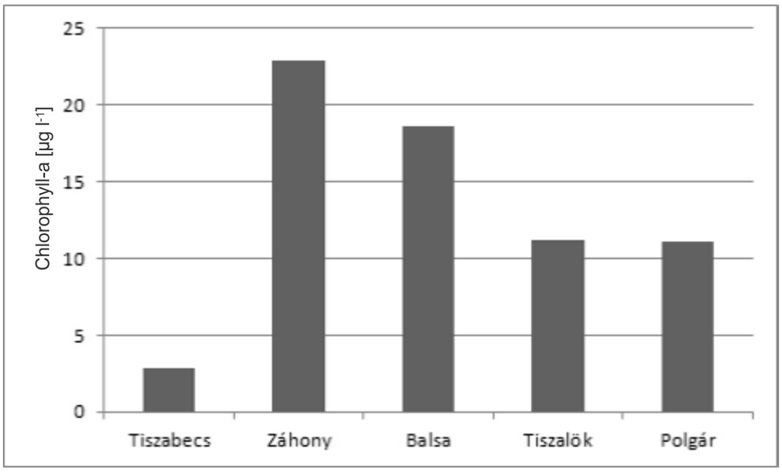


Fig. 4
Chlorophyll-a average (1974–2005) at every sampling location

approached and handled in a single step. This does not mean, of course, that the low and high-water stage separation is unnecessary in subsequent research.

The winter analyses were conducted on the data from November to February, and the summer ones went June to September.

With regard to the most important anions and cations, only small differences could be noticed between the winter and summer data. For example, in the case of the K^+, Ca^{2+} and Mg^{2+}, only a minimal increase could be observed during winter (Fig. 6).

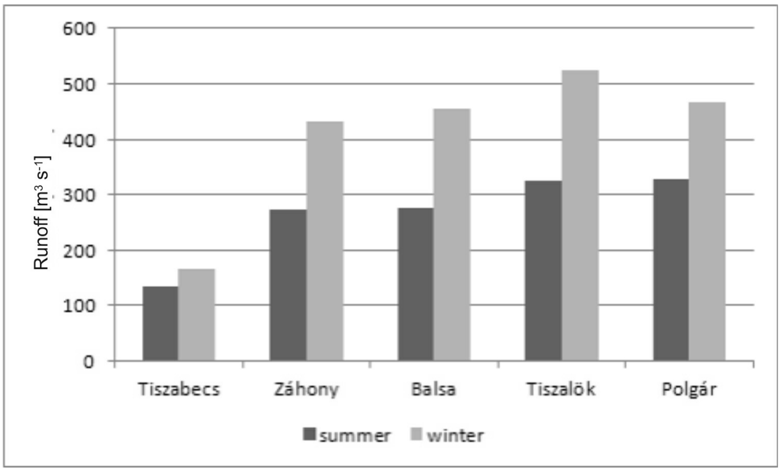


Fig. 5
Summer and winter average runoff at the sampling locations

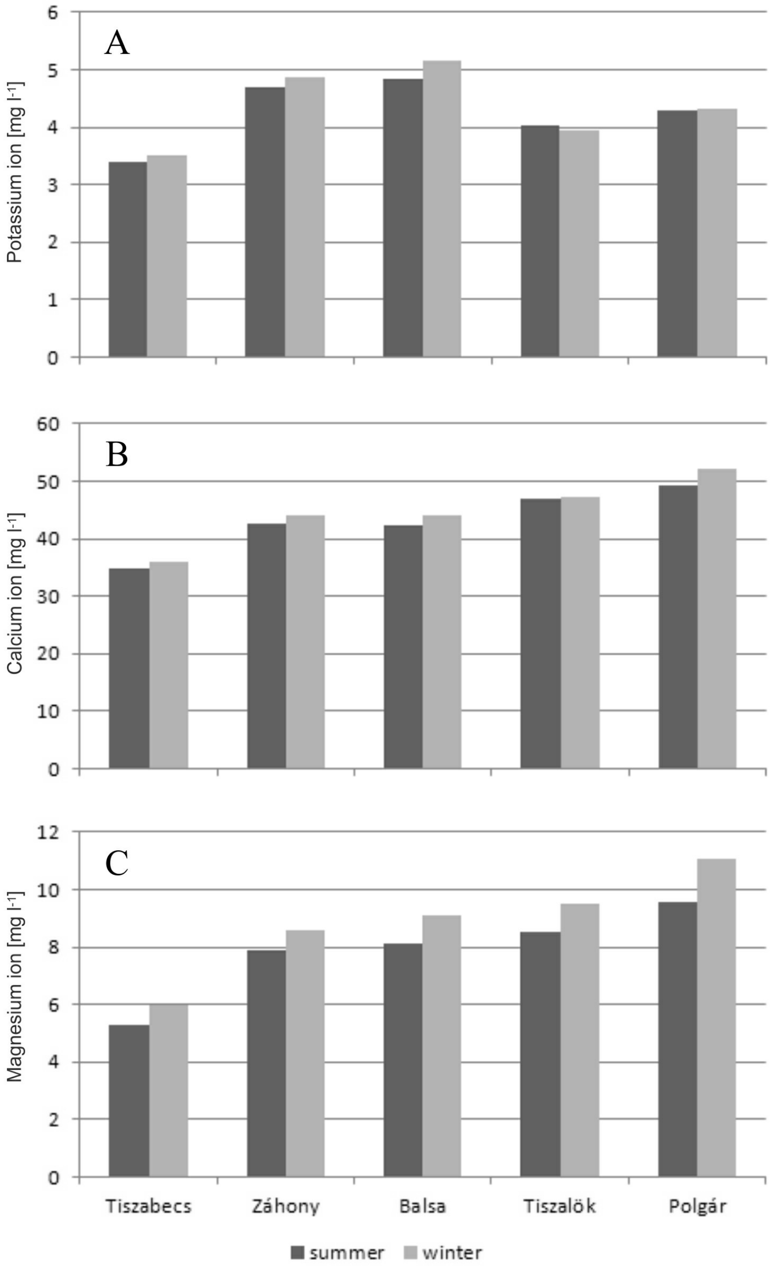


Fig. 6
K^{+} (A), Ca^{2+} (B) and Mg^{2+} (C) summer and winter concentration averages at different sampling locations

As had been expected, seasonal differences in the chl-a content were unambiguously noticeable (Fig. 7). It was surprising, however, that at the first sampling location (Tiszabecs) the difference between the winter and summer chl-a content was smaller. There the summer concentration was only 1.63 times the winter one, in contrast to any other location, where the summer value exceeded the winter one by 5–19 times, respectively.

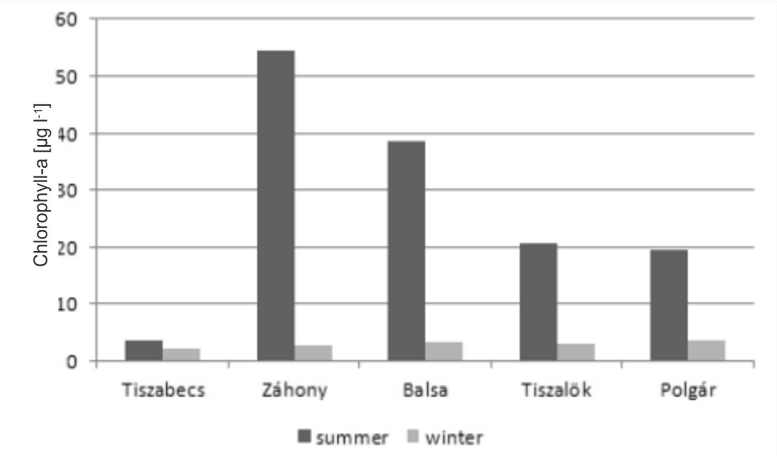


Fig. 7
Chlorophyll-a concentration in winter and summer distribution

Stochastic connections

"Global" correlations

After the descriptive statistics, the connection between the parameter pairs was analyzed using correlation analysis. Only those linear connections were considered as strong where the absolute value of the correlation coefficient was higher than 0.7. The connection between the parameters was analyzed using different approaches: first the entire year, then the different seasons (winter, summer) were taken into account.

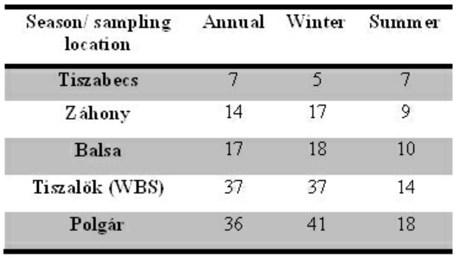
During any time period (whole year, winter/summer) the number of strong correlations increased downstream. At Tiszabecs, the number of strong correlations was only 7 (Table 1), at Tiszalök it reached 36.

During the summer months there are less strong linear connections than in the winter ones. If the results from the entire year are compared to the summer and winter ones, it can be stated that the correlation matrix obtained from the winter data resembles the annual correlation matrix more closely than the summer data. According to Table 1, in the Tiszalök area the number of correlating parameters rises suddenly both in winter and summer.

PCA results

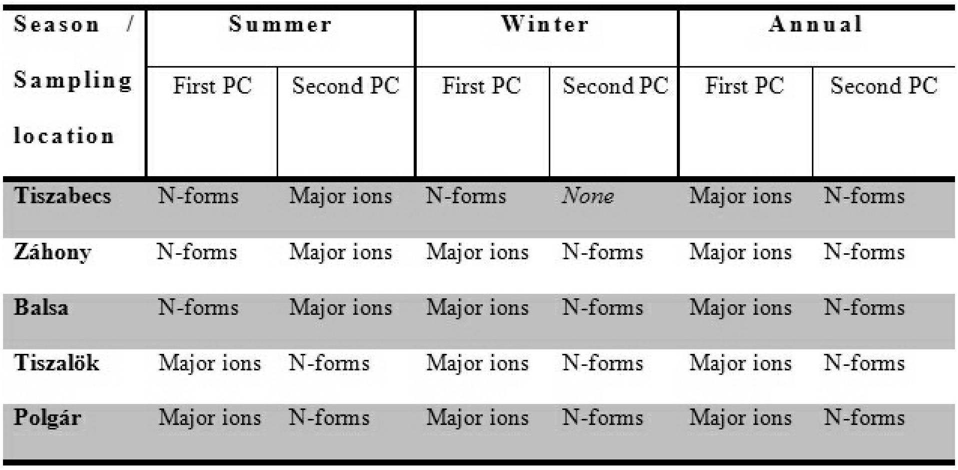
In order to determine the background processes in the Tisza's water, PCA was applied to the summer, winter and whole year data. Before PCA was conducted the number of parameters had to be decreased, either because the parameter was not sampled during certain time periods or because it was unsystematically sampled over the entire investigated time period. In other cases even the parameter itself contained information concerning other parameters (e.g. specific conductance). Only parameters with a factor score (in absolute value) higher than 0.7 were taken into account in the first and second principal components. The summarized results of the PCA can be found in Table 2, where the first two components explain approximately 50% of the data's total variance, independently of their spatial and temporal distribution.

Table 1
The number of strong linear connections (|R| ≥ 0.7) at the sampling locations and with regard to temporal distribution



Season/ sampling location	Annual	Winter	Summer
Tiszabecs	7	5	7
Záhony	14	17	9
Balsa	17	18	10
Tiszalök (WBS)	37	37	14
Polgár	36	41	18

Table 2
Summarized results of the PCA, 'None' indicates that there were no significant factor scores



Season / Sampling location	Summer		Winter		Annual	
	First PC	Second PC	First PC	Second PC	First PC	Second PC
Tiszabecs	N-forms	Major ions	N-forms	*None*	Major ions	N-forms
Záhony	N-forms	Major ions	Major ions	N-forms	Major ions	N-forms
Balsa	N-forms	Major ions	Major ions	N-forms	Major ions	N-forms
Tiszalök	Major ions	N-forms	Major ions	N-forms	Major ions	N-forms
Polgár	Major ions	N-forms	Major ions	N-forms	Major ions	N-forms

In the summer results at Tiszabecs, Záhony and Balsa, mostly the N-forms and the ions responsible for halobity (Mg^{2+}, Na^{+}, K^{+}, Cl^{-}) occurred in the first and second PCs, respectively. Between Balsa and Tiszalök the background processes showed a peculiar change: the scale shifted from the organic components toward the inorganic ones. According to the first PC the major ions (e.g. Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}) were determinant at Tiszalök in the summer. The Polgár sampling location showed the same pattern.

From the perspective of the winter results the first PC's explanatory power varied only between 20 and 40%. Except for Tiszabecs, the ions determined the background processes at all sampling locations. In the second factor, the N-forms were dominant with a factor score higher than 0.7.

Regarding the whole year's PCA results it can be stated that, as in the case of the correlation results, the annual conditions resemble the winter ones to a high degree. In the first PC the ions take on the determining role, while in the second the N-forms are dominant.

Conclusions

According to the runoff and chl-a results it can be said that over the investigated river section, the Szamos determines the Tisza's water quality to the greatest degree. The sudden production excess witnessed at Záhony is caused by the Szamos and its organic nutrient load arriving from across the border. The origin of the nutrients is assumed to be outside of Hungary (Istvánovics et al. 2010), since merely 4% of the Szamos' watershed is located in Hungary (Konecsny et al. 2010). All the other tributaries have a local effect only. The Tisza's water quality is influenced by other tributaries (the Bodrog and Sajó) as well. Since their runoff measured at their mouths is much smaller in comparison to the Tisza's at the same location, their effect is of less importance. Although they increase the mineral-N content of the water, this does not affect the chl-a content to such a degree as it did upper stream. Besides the Szamos, every other tributary only has a local effect. After analyzing the annual statistics the data was separated into winter and summer data. With regard to the former, a minimal excess (Fig. 4) was observed in the case of the K^{+}, Ca^{2+} and Mg^{2+} ions, which may be the result of precipitation (which causes a higher runoff as well) (Csépes et al. 2000) and a more intense flow. This flow transports greater amounts of substances from the soil into the river. Much larger seasonal differences were observed in the case of chl-a than in the case of the ions. The reason may be that primary production is much higher in summer than in winter. The explanation for the more equal distribution of primary production at Tiszabecs may be that at the border the Tisza's waters contain only a small amount of mineral-N and phosphorus. Hence, the processes there show smaller seasonal differences.

The peculiarly high number of correlations at Tiszalök can be explained by the flow conditions. In the area of the water barrage system the water-flow slows down and (according to the spiral model) so does the physical transport. Suspended solids are deposited, the water becomes more transparent and light limitation decreases. This provides an opportunity for organisms to absorb nutrients into their systems more rapidly and more accurately (according to

Zsuga and Szabó 2005). Summarizing the correlation analysis, it can be noted that the number of correlations increases downstream. The annual, winter and summer results are different regarding the number of correlations and the parameters which correlate as well. During summer there are less linear connections, but those few are between the parameters related to organic processes. As mentioned earlier the correlation matrix obtained from the winter data resembles the annual correlation matrix more than the summer one. Thus, if only the whole year had been analyzed, vital differences between summer and winter would have been lost.

The fact that at Tiszabecs, Záhony and Balsa the N-forms were dominant in the first PC lead us to the assumption that biological processes such as saprobity and trophic conditions are responsible for the background processes. In contrast, the results from Tiszalök and Polgár revealed a shift in the determining processes. After Tiszalök, the inorganic processes (e.g. aggregation, dissolution) substituted the N-forms in the first PC. Regarding temporal distribution, inorganic processes determined the Tisza's water quality during winter. Both in correlation analysis, and PC, results which are not temporally divided are unsatisfactory. However, this can only be revealed if the summer and winter data are analyzed separately. Again, the results obtained from the winter data represent the annual conditions much more than the summer ones do. During winter, inorganic processes are dominant except for Tiszabecs, where biological processes play a determining role throughout the year.

Summary

Twenty-four parameters from five sampling locations on the Upper Tisza were analyzed for the time interval 1974–2005. The explorative data analysis methods highlighted the temporal and spatial variability of its processes.

The various basic statistics obtained from the parameters' annual datasets illustrated the tributaries' impact on the Tisza's water quality. Due to its high amount of transported organic and inorganic substances throughout the year, the River Szamos had the greatest influence of all tributaries. However, descriptive statistics revealed the differences in the seasonal water qualities dramatically. The same pattern emerged from the results of the stochastic connections. They indicated an increase in the number of correlations downstream, independent of seasonal distribution. The strongest linear connection noticed was at the river barrage system at Tiszalök. Here, the longer water retention time formed ideal conditions for parameter to develop significant stochastic connections.

By ascertaining background processes affecting water quality, serious differences were brought to light: in winter biological processes were dominant (spatially) as far as Záhony, in summer as far as Tiszalök. It is important to emphasize that the winter results resembled the annual ones much more than the result obtained from the summer data. With regard to such seasonal differences, the dataset of a whole year cannot be handled as a single unit.

References

Csépes, E., M. T. Nagy, I. Bancsi, P. Végvári, P. Kovács, E. K. Szilágyi 2000: A vízminőségi jellemzők alakulása az évszázad egyik legnagyobb árvizének tükrében (Water quality changes in the wake of the centuries biggest flooding). – Hidrológiai Közlöny, 80/5–6, pp. 285–287.

Füstös, L., Gy. Meszléna, N. Simonné Mosolygó 1986: A sokváltozós adatelemzés statisztikai módszerei (Statistical methods of multi-variate data analysis). – Akadémiai Kiadó, Budapest, 525 p.

Istvánovics, V., M. Honti, L. Vörös 2010: Phytoplankton dynamics in relation to connectivity, flow dynamics and resource availability in the case of a large, lowland river, the Hungarian Tisza. – Hidrobiologica, 637, pp. 121–141.

Konecsny, K., G. Bálint 2010: Main hydrological statistical characteristics of low water on the Somes/Szamos river. – http://riscurisicatastrofe.reviste.ubbcluj.ro Volume/XI_Nr_2_2010/PDF/ Konecsny.pdf

Lajter, I., C. Schnitchen, Gy. Dévai, S. Nagy 2009: Hosszú távú adatsorok felhasználási lehetőségei vízfolyások ökológiai minősítésében a Tisza vízrendszerének példáján (The use of long-term data series in the ecological evaluation of streams on the example of the River Tisza). – Hidrológiai Közlöny, 89/6, pp. 141–144.

Lászlóffy, W. 1982: A Tisza, vízi munkálatok és vízgazdálkodás a tiszai vízrendszerben (Works on the River Tisza and water management on the Tisza's water system). – Akadémiai Kiadó, Budapest, 609 p.

Nagy, T.M., E. Csépes, A. Aranyné Rózsavári, I. Bancsi, P. Kovács, P. Végvári, K. Zsuga, K. 2004: A hosszú-távú adatsorok értékelésének korlátai (The barriers of long-term data analysis). – Hidrológiai Közlöny, 84/5–6, pp. 162–165.

Oláh, M., J.A. Tóth, J. Oláh, T. Bodea 2000: Parti pufferzóna a folyóvölgyi nitrogén anyagcserében a Tisza mentén (Puffer zone on the Tisza River's banks regarding the nitrogen cycle). – Hidrológiai Közlöny 80/5–6., pp. 339–341.

Padisák, J. 2005: Általános limnológia (General limnology). – ELTE Eötvös Kiadó, Budapest, 310 p.

Pécsi, M. 1969: A tiszai Alföld (The Tisza's plain). – Akadémiai Kiadó, Budapest, 381 p.

Sakan, S., Grzetic, I., Đordevic, D. 2007: Distribution and Fractionation of Heavy Metals in the Tisa (Tisza) River Sediments. – Environmental Science and Pollution Research, 14/4, pp. 229–236.

Somogyi, S. 2003: A Tisza vízgyűjtőjének földrajzi helyzete (Geographical setting of the Tisza's watershed). – In: Teplán, I. (Ed): A Tisza és vízrendszere (The Tisza and its water system) – MTA Társadalomkutató Központ, Budapest, pp. 17–27.

Szabó, A., Gy. Dévai, K. Zsuga 2004a: Javaslat a vízjárás és a vízminőségi mutatók összefüggésének egy lehetséges hosszú távú elemzési módszerére a Tisza példáján (Suggestion on the analysis method of the Tisza's long-term data series regarding the runoff's and water quality parameters' connection). – Hidrológiai Közlöny, 84/5–6, pp. 139–142.

Szabó, A., Gy. Dévai, K. Zsuga, K. Kaposvári 2004b: A vízjárás, az elektromos vezetőképesség és a kémiai oxigénigény összefüggésének elemzése a szolnoki Tisza-szakasz napi adatsorai alapján (An analysis of runoff, conductance and chemical oxygen demand's connection on the daily data series from Tisza's section by Szolnok). – Hidrológiai Közlöny, 84/5–6, pp. 143–146.

Zsuga, K., A. Szabó 2005: A Tisza hazai vízgyűjtőterületének ökológiai állapota, környezetvédelmi problémái (The ecological state and environmental problems of the Tisza's Hungarian watershed). – Hidrológiai Közlöny, 85/6, pp. 168–170.