CrossMark

Optimization of the monitoring network on the River Tisza (Central Europe, Hungary) using combined cluster and discriminant analysis, taking seasonality into account

Péter Tanos · József Kovács · Solt Kovács ·
Angéla Anda · István Gábor Hatvani

Abstract The most essential requirement for water management is efficient and informative monitoring. Operating water quality monitoring networks is a challenge from both the scientific and economic points of view, especially in the case of river sections ranging over hundreds of kilometers. Therefore, spatio-temporal optimization is vital. In the present study, the optimization of the monitoring system of the River Tisza, the second largest river in Central Europe, is presented using a generally applicable and novel method, combined cluster and discriminant analysis (CCDA). This area for the study was chosen because, spatial inhomogeneity of a river's monitoring network can more easily be studied in a mostly natural watershed - as in the case of the River Tisza - since the effects of man-made obstacles: e.g water barrage systems, hydroelectric power plants, artificial lakes, etc. are more pronounced. Furthermore, since the temporal sampling frequency was bi-weekly, the opportunity of optimizing the monitoring system on a temporal (monthly) scale arose. In the research, 15 water quality parameters measured at 14 sampling sites in the Hungarian section of the River Tisza were assessed for the time period 1975–2005. First, four within-year sections ("hydrochemical seasons") were determined, characterized with unequal lengths, namely 2, 4, 2, and 4 months long starting with spring. Homogeneous groups of sampling sites were determined in space for every season, with the main separating factors being the tributaries and man-made obstacles. Similarly, an overall pattern of homogeneity was determined. As an overall result, the 14 sampling sites could be grouped into 11 homogeneous groups leading to the possibility of reducing the number of sampling locations and thus making the monitoring system more cost-efficient.

Solt Kovács is currently an M.Sc. student in Mathematics at ETH Zurich e-mail: kovacssolt@gmail.com

P. Tanos · A. Anda
Department of Meteorology and Water Management, Georgikon Faculty, University of Pannonia, Keszthely H-8360, Hungary

P. Tanos
e-mail: tanospeter@gmail.com

A. Anda
e-mail: anda-a@georgikon.hu

J. Kovács
Department of Physical and Applied Geology, Eötvös Loránd University, Budapest H-1117, Hungary
e-mail: kevesolt@geology.elte.hu

I. G. Hatvani (✉)
Institute for Geological and Geochemical Research, Research Center for Astronomy and Earth Sciences, MTA, Budaörsi út 45., Budapest H-1112, Hungary
e-mail: hatvaniig@gmail.com

⚫ Springer

Introduction

The aims of the Clean Water Act (1948) and EC Water Framework Directive 2000/60/EC cannot be met without a proper water quality monitoring system. The most important criteria of monitoring systems are that they should be as representative in time and space and as cost-efficient as possible (Chilundo et al. 2008). The optimization of sampling networks—generally based on professional and economic considerations—may be a solution in achieving representativeness as the basic goal and requirement of monitoring. However, the optimization of a monitoring system can only be executed when the whole system (with all the sampling sites) is analyzed at once, because only then can the differences or similarities between the sampling locations (SL) be revealed.

In the case of monitoring system optimization, the most frequent approaches are either deterministic or the stochastic. Additionally, geographic information system (GIS)-based tools exist as well, which can be combined with deterministic or stochastic ones. To give a view of the diversity of methods used in monitoring network optimization, in the following paragraphs, examples will be shown of (i) deterministic methods both in the planning of monitoring networks and the optimization of already existing ones and (ii) their combination with GIS-based methods. This will be followed by (iii) a selection of studies of the use of stochastic modeling, and finally (iv) the combination of stochastic and GIS-based methods will be discussed.

The first step when one wants to set up a monitoring system is precise planning. Telci et al. (2009) used deterministic methods in their study. First, flow dynamics were determined using numerical modeling, and then an optimization model was created, which was in fact a genetic algorithm and Pareto optimal analysis. The methods of Sharp (1971) and Sanders (Sanders and Adrian 1978; Sanders 1980; Sanders et al. 1983) are also frequently used in the planning of monitoring networks (Strobl and Robillard 2008). These approaches can be successfully applied in the case of larger riverine systems as well. Using Sharp's method, the sampling locations are set out based on the locations of the tributaries' mouths, while Sanders and his coauthors suggest a more general consideration of point sources. If it is supposed that the tributaries function as point sources, the two methods are very similar. However, if there are multiple branches which differ from each other to a great

extent, these approaches will not provide a satisfactory result. Another problem which arises with the use of these approaches is that the sampling site locations are not determined precisely enough. Do et al. (2012) found a solution to this problem by combining the deterministic approach with data obtained from GIS. In this way, they were able to study the effects of the non-point sources (e.g. pollutant of anthropogenic origin arriving to surface waters, mainly due to agricultural land use) as well. Continuing along this line, Do et al. (2011) in their previous study dealt with the question of using non-point sources to design the sampling network. They combine a nutrient export coefficient model and Sharp's method in designing a monitoring network. Another example of combining deterministic (genetic algorithm) and GIS methods is presented by Park et al. (2006), who managed to determine the effectiveness of a riverine monitoring network in the Nakdong River system, Korea, while, e.g. Preziosi et al. (2013) used solely GIS methods successfully in the design of a ground water monitoring system in a pilot area in central Italy.

Besides planning new networks, there are many studies which deal with the optimization of already functioning monitoring networks, such as that of Chen et al. (2012). Here, a one-dimensional flow model, a water quality model, and a matter element analysis (MEA) were used to find similar sections of the 1890-km-long upper and middle reaches of the Heilongjiang River in northeast China. They state in their paper that "the definition of the reach of homogeneous water quality is not absolute and depends on the methodology used." The detailed description of MEA can be found in the work of Wang (2001). Additionally, entropy theory is also a frequently applied approach, e.g. (i) Masoumi and Kerachian (2010) successfully applied entropy theory and the transinformation–distance curves to determine the efficiency of sampling sites and the optimal sampling frequency of ground water quality monitoring system, while (ii) Mahjouri and Kerachian (2011) used a micro-genetic algorithm-based optimization method to optimize the spatio-temporal functioning of the monitoring network of the Jajrood River (Iran). Icaga (2005) also used a genetic algorithm to optimize the monitoring network of the River Gediz's watershed (Turkey). Lee et al. (2014) determined the optimal water quality monitoring locations in the Logan and Albert River network (USA) applying cost functions coupled with a genetic algorithm. Staying with the deterministic approaches, Kao et al. (2012) suggested the application of two

different deterministic linear optimization models instead of the previously used simulated annealing method to optimize the monitoring network of the Derchi Reservoir watershed.

Naturally, besides deterministic, stochastic approaches are used as well when it comes to optimization of monitoring networks. Many studies use trend analysis to explore the redundancy of a certain system (Naddeo et al. 2007, 2013; Scannapieco et al. 2012). Other frequently used methods, such as cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA), and factor analysis (FA), have been combined in the studies of Alberto et al. (2001) on the watershed of the Suqia River (Argentina), Varol et al. (2012) on the catchment of the Tiger River (Turkey), and Wang et al. (2012) in their investigation of the Xiangxi River (China) catchment. Combining only CA and PCA, Fan et al. (2010) and Razmkhah et al. (2010) sought the spatio-temporal variability between riverine sampling sites. Hatvani et al. (2011; 2014a) also used CA to group similar sampling sites in the Kis-Balaton Water Protection System (the mitigation wetland of Lake Balaton, Hungary) and the shallow groundwater system of the Seewinkel (Hatvani et al. 2014c), while Kovács et al. (2012a) used a special coded CA to determine the water quality areas of Lake Balaton.

Multivariate statistical methods are frequently combined with GIS-based methods. For example, Shen and Wu (2013) used kriging during the optimization of the monitoring in the Yangtze estuary. Naturally, kriging can be used alone as well (Karamouz et al. 2009). Júnez-Ferreira and Herrera (2013) also combined two multivariate statistical methods (static Kalman filter combined with a sequential optimization method) to optimize spatio-temporarily the Valle de Querétaro aquifer monitoring network (Mexico).

The aim of this study is to provide an alternative to the previously discussed methods when it comes to the optimization of monitoring networks in time and space. This alternative is called combined cluster and discriminant analysis (CCDA, Kovács et al. 2014). CCDA is capable of handling spatio-temporal data obtained from an entire system over the course of decades. It facilitates the frequently difficult decision of whether an obtained grouping should be further divided or not to find homogeneous groups. This capability distinguishes it from the previously mentioned methods, be they deterministic, stochastic, or even combined with GIS.

The detailed aims of this study are (i) to point out within-year similarities (hydrochemical seasons[1]) and (ii) to find spatially homogeneous sampling sites on the River Tisza, with special attention paid to anthropogenic activity. These homogeneous groups can later on serve as the bases for the optimization of a monitoring system.

Materials and methods

Description of the study area

The River Tisza collects the waters of the Carpathian Basin's eastern region; it is therefore a highly important ecological corridor (Zsuga et al. 2004). It stretches from its source in the Eastern Carpathians in Ukraine to its confluence with the Danube at Titel in Serbia. According to Lászlóffy (1982), the area of its watershed is 157,186 km^2, almost one third of which is located in Hungary (approx. 47,000 km^2). The average amount of water brought by the Tisza into the Danube is 25.4 billion m^3 year^{-1} (Pécsi 1969). The main branch (966 km; Sakan et al. 2007) stretches through five countries (Ukraine, Romania, Hungary (594.5 km), Slovakia, and Serbia). Heading downstream in the river's Hungarian section, its tributaries are the following: the Szamos, Bodrog, Sajó, Zagyva, Körös, and Maros, all of which—except for the Zagyva—come from abroad (Fig. 1). It becomes clear from the runoff values that the affluent having the strongest effect on the main flow is the Szamos (at its mouth, its average runoff exceeds half of the average runoff of the Tisza) and a considerable "changing effect" is expected from the Bodrog, Sajó, Zagyva, Körös, and Maros Rivers (Table 1).

It has been documented that, besides the tributaries, other, mostly anthropogenic, factors such as water barrage systems (WBS; e.g. Tiszalök WBS; Fig. 1) or lakes (e.g. Lake Tisza; Fig. 1) can also have an effect on the water quality of the analyzed river section (Kentel and Alp 2013; Moreira and

---

[1] Hydrochemical seasons (e.g. Kolander and Tylkowski 2008) are the seasonal characteristics reflected in the chemical parameters analyzed in the waters of the river. These will be referred to as "seasons" from now on and should not be confused with the general concept of the climatic seasons, even if the former are highly affected by the latter.
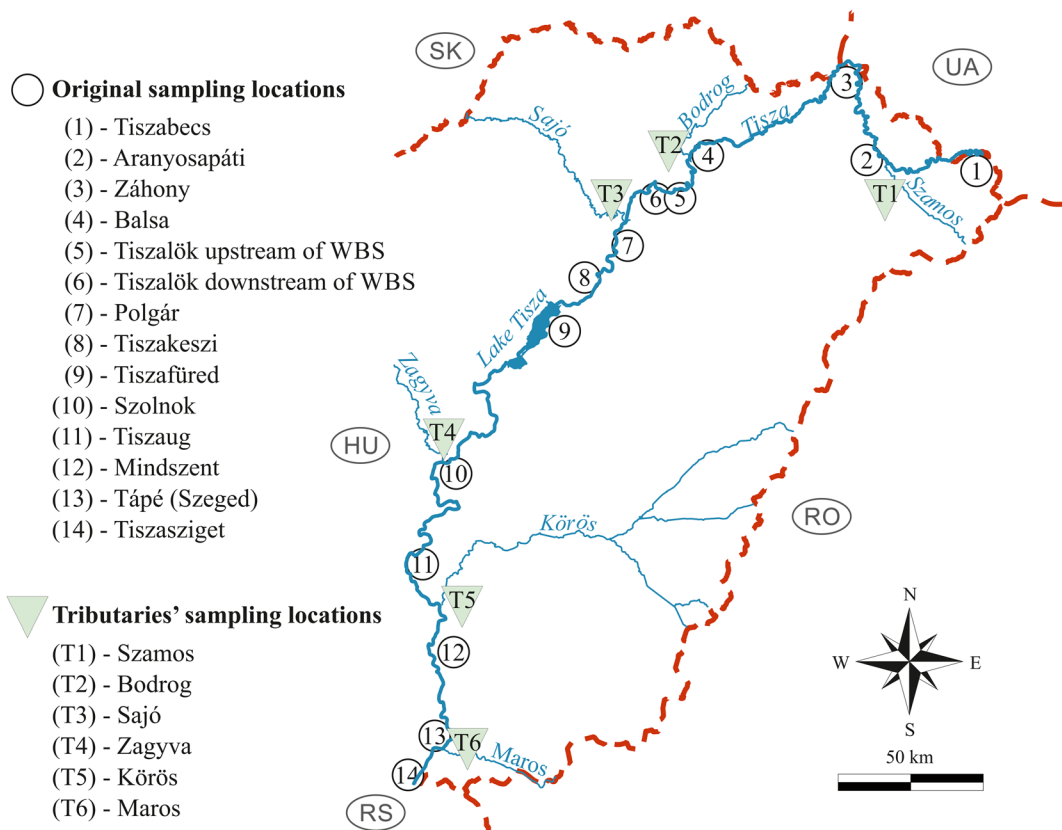
Original sampling locations

(1) - Tiszabecs
(2) - Aranyosapáti
(3) - Záhony
(4) - Balsa
(5) - Tiszalök upstream of WBS
(6) - Tiszalök downstream of WBS
(7) - Polgár
(8) - Tiszakeszi
(9) - Tiszafüred
(10) - Szolnok
(11) - Tiszaug
(12) - Mindszent
(13) - Tápé (Szeged)
(14) - Tiszasziget

Tributaries' sampling locations

(T1) - Szamos
(T2) - Bodrog
(T3) - Sajó
(T4) - Zagyva
(T5) - Körös
(T6) - Maros

Fig. 1 Hungarian catchment of the River Tisza

Poole 1993). Even the river ice regime may change due to the installation of WBSs (Takács et al. 2013; Takács and Kern 2015). Lake Tisza is an artificial reservoir on the river, built in 1973. It was planned to function as a part of a future WBS. Nowadays, it is a frequented recreation zone and nature reserve. The lake's length is 27 km, its mean depth is 1.3 m, and it has a total area of 127 km^2. Moreover, non-point source nutrient loads arriving from agricultural areas have to be accounted for as well (Mandera and Forsberg 2000). Regarding large cities, there are several along the river (e.g. Szolnok at SL-10 and Szeged at SL-13; Fig. 1) which could also have an environmental impact on the river's water quality.

Table 1 Characteristics of the Hungarian section of the River Tisza

		Tisza	Szamos	Bodrog	Sajó	Zagyva	Körös	Maros
Watershed size (km^2)		157,186	15,015	13,579	12,708	5677	27,537	27,049
Average runoff (m^3 s^{-1})	SL1-Tiszabecs	226						
	SL9-Tiszafüred	503	118	123	32	4	116	188
	SL14-Tiszasziget	805						
Average ratio (in %) of runoff between the waters leaving the tributaries and the runoff in the main branch		100	52	36	7	1	23	30
Full length (rkm)		962	411	65	223	179	741	683
Length of Hungarian section (rkm)		594.5	52	51	125	179	180	29

Dataset used

Regarding systematic measurements, the Hungarian section of the Tisza (594.5 rkm) fell under the jurisdiction of many different water authorities over the studied decades (1975–2005). These different authorities did not harmonize the timing of the sampling; hence, samples were taken on different days (~2250 sampling dates) at different SLs. Fortunately, they did take the trouble to intercalibrate the sampling methodologies (details can be found in Hungarian Standard No. MSZ 12749:1993).

In the course of the analyses, the time series of the period between 1975 and 2005, the following parameters from 14 sampling sites (Fig. 1) were examined: runoff ($m^3 s^{-1}$), pH, dissolved O_2, BOD-5, Ca^{2+}, Mg^{2+}, Na^{2+}, K^+, Cl^-, SO_4^{2-}, HCO_3^- (mg l^{-1}), NH_4-N, NO_2-N, NO_3-N, and PO_4-P ($\mu g\ l^{-1}$). After 2005, the sampling frequency was rarefied and the set of parameters changed. The total number of data analyzed added up to ~33,500.

Data preparation was performed so that the dataset would meet some basic requirements of the applied method: (i) no missing data were allowed; all time points were discarded where there was missing data in any of the parameters; (ii) no mistyped extreme values were allowed. These were sought manually, because there were occasions when an "act of God" (e.g. flood) caused a certain parameter to behave differently from the general tendencies, although its measurements were probably accurate.

Data from the tributaries (rivers—Szamos, Bodrog, Sajó, Zagyva, Körös, and Maros) were acquired as well with the same set of variables as the ones used for the analysis of the Tisza. Regarding the tributaries, the data measured at the closest sampling location to the confluences on the tributary were used.

Methodology

A new method, combined cluster and discriminant analysis (CCDA; Kovács et al. 2014), formed the backbone of the research. The main aim of CCDA is to find homogeneous groups based on data with known origins, i.e. sampling locations in this case. It consists of three main steps (Fig. 2):

(I)   a basic grouping procedure, e.g. using hierarchical cluster analysis (HCA), to determine possible groupings;

(II) a core cycle where the goodness of the groupings from step I and the goodness of random classifications are determined using linear discriminant analysis (LDA); these are then compared in the form of a "difference value";

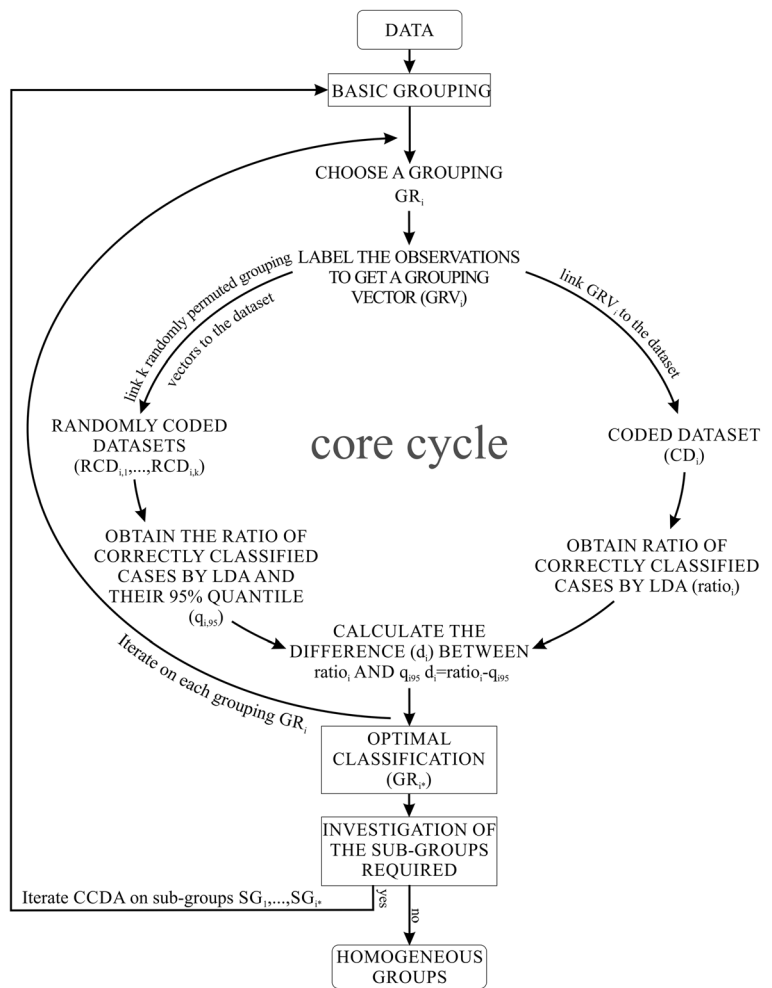(III) a final evaluation step, where a decision about iterative further investigation of sub-groups is taken.

The main idea here is that once the ratio of correctly classified cases for a grouping is higher than at least 95 % of the ratios for the random classifications (i.e. the difference value is positive), then at the level of $\alpha = 0.05$, the given classification is considered as inhomogeneous. For a detailed description of the method, see Kovács et al. (2014) and the corresponding R package ("ccda", http://cran.r-project.org/web/packages/ccda/) used for the computations. It has to be underlined that CCDA is generally applicable for many types of data, be they water chemistry data originating from surface (lake, wetland, river), subsurface water systems (the watershed of a steppe lake; Kovács et al. 2015).

HCA and CCDA were applied to the data in order to (Fig. ESM 1)

- determine the time interval of interest for investigations (HCA based on annual averages),
- determine the seasons (HCA for monthly averages),
- determine the similarities between the consecutive months pairwise using CCDA (note that in order to ensure the comparability of pairwise difference values, an equal number of samples was taken for every month by resampling),
- find homogeneous groups of sampling sites in space in the chosen time interval using CCDA to provide the basis for monitoring network recalibration ("overall picture"), as well as to
- assess the monitoring network in space within the determined seasons (also using CCDA).

In every case where a cluster group is formed, the role of each parameter in determining the formation of the previously obtained cluster pattern can be analyzed. Using Wilks' λ statistics (1932), a Wilks' λ quotient is

Fig. 2 Flowchart of CCDA (after Kovács et al. 2014)

Flowchart elements:

DATA

BASIC GROUPING

CHOOSE A GROUPING GR$_i$

LABEL THE OBSERVATIONS TO GET A GROUPING VECTOR (GRV$_i$)

link k randomly permuted grouping vectors to the dataset

link GRV$_i$ to the dataset

core cycle

RANDOMLY CODED DATASETS (RCD$_{i,1}$,...,RCD$_{i,k}$)

CODED DATASET (CD$_i$)

OBTAIN THE RATIO OF CORRECTLY CLASSIFIED CASES BY LDA AND THEIR 95% QUANTILE (q$_{i,95}$)

OBTAIN RATIO OF CORRECTLY CLASSIFIED CASES BY LDA (ratio$_i$)

CALCULATE THE DIFFERENCE (d$_i$) BETWEEN ratio$_i$ AND q$_{i95}$ d$_i$=ratio$_i$-q$_{i95}$

Iterate on each grouping GR$_i$

OPTIMAL CLASSIFICATION (GR$_{i*}$)

INVESTIGATION OF THE SUB-GROUPS REQUIRED

Iterate CCDA on sub-groups SG$_1$,...,SG$_{i*}$ yes / no

HOMOGENEOUS GROUPS

---

assigned to every parameter, where the quotient is

$$\lambda = \frac{\sum_i \sum_j \left(x_{ij} - \overline{x}_i \right)^2}{\sum_i \sum_j \left(x_{ij} - \overline{x} \right)^2}$$

where $\overline{x}_{ij}$ is the jth element of the ith group, $\overline{x}_i$ the ith group's mean, and $\overline{x}$ is the total mean.

The value of λ is the ratio of the within-group sum of squares to the total sum of squares. It is a number between 0 and 1. If $\lambda = 1$, then the mean of the discriminant scores is the same in all groups and there is no inter-group variability, so, in this case, the parameter did not affect the formation of the cluster groups (Afifi et al. 2004). On the contrary, if $\lambda = 0$, then that particular parameter affected the formation of the cluster groups the most. The smaller the quotient is, the more it

determines the formation of the cluster groups (Kovács et al. 2012b).

Results

The first step in the analysis was to find the most appropriate time period for the detailed analyses. Based on the time series of the water quality variables, a breakpoint was suspected (e.g. Fig. ESM 2). Therefore, HCA was applied to all the annual averages of the variables together (1975–2005). It pointed out that the dataset consisted of two main time intervals: 1975–1992 and 1993–2005 (Fig. ESM 3). In further investigations, the time interval 1993–2005 was examined for the following reasons: (i) data from an additional sampling site were available in the latter period and
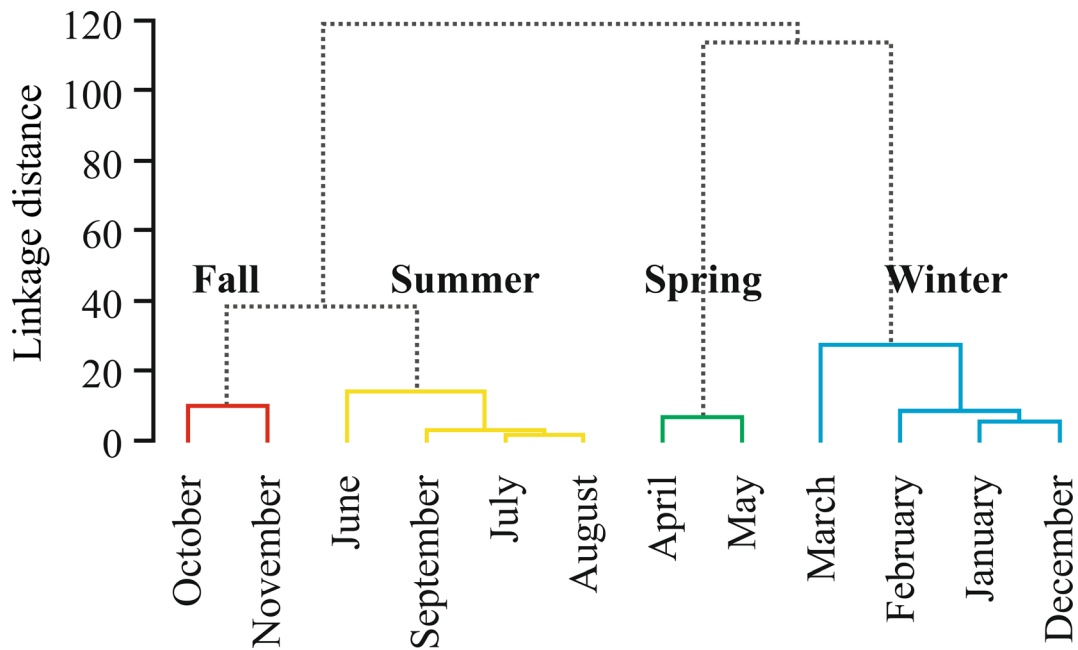
Fig. 3 Results of HCA conducted on monthly averages to find seasons

(ii) it is to be expected that the second period could provide the more accurate picture of the current status of the monitoring system.

As a next step, the intra-annual similarities were sought on a monthly scale. HCA pointed out that there are basically four seasons (Fig. 3).
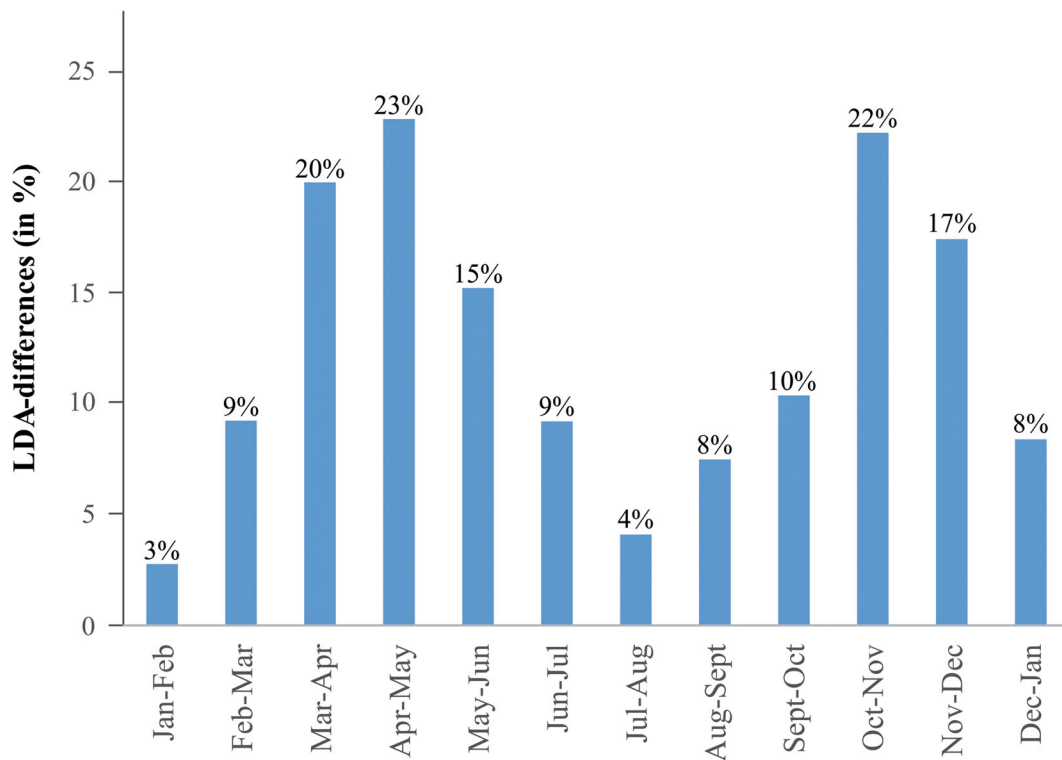

Fig. 4 Pairwise differences between the months pointed out by CCDA

CCDA also provided the opportunity to explore homogeneity within the seasons. It showed that even though the previously determined seasons consist of similarly behaving months, these seasons cannot be considered as homogeneous. Moreover, the months would even form 12 separate homogeneous groups, each month forming a "group" of its own. To point out the significant differences, CCDA was applied to the pairwise comparison of consecutive months (Fig. 4). To ensure the comparability of the resulting difference values, an equal number of samples for each month was selected using resampling.

The next step was to find homogeneous spatial patterns, to be more specific those SLs which contain redundant information. For this, CCDA was used on all the available data for the time interval 1993–2005, resulting in an "overall picture" where 11 homogeneous groups were formed from the 14 SLs considered. There were eight groups consisting of one SL in each. Although some of these are quite close to each other, such as SL-5 and SL-6 at 2 rkm distance from each

other, the information obtained from them makes them highly informative, even alone. In the meanwhile, three groups consisting of two sampling sites were obtained. These pairs of sampling sites were always located next to each other (Fig. 5). However, these were sometimes quite a long distance, ~70 km, from each other (SL-3 and SL-4).

The other approach to determining the spatial pattern in the sampling sites is to analyze their similarity, not on a long but on a much shorter temporal scale, to be more precise seasonally. This was considered to be an important step since the hypothesis is that the sampling sites' relationship to each other changes with the seasons. Therefore, the question was raised, to what extent does the "overall picture" coincide with the seasonal one?

As a first step, let us explore all the winter months for the time period between 1993 and 2005. At this time of the year, the 14 SLs can be divided into 10 homogeneous groups (Fig. 6a) using CCDA. Six SLs form groups alone while there are four groups consisting of two sampling sites each.
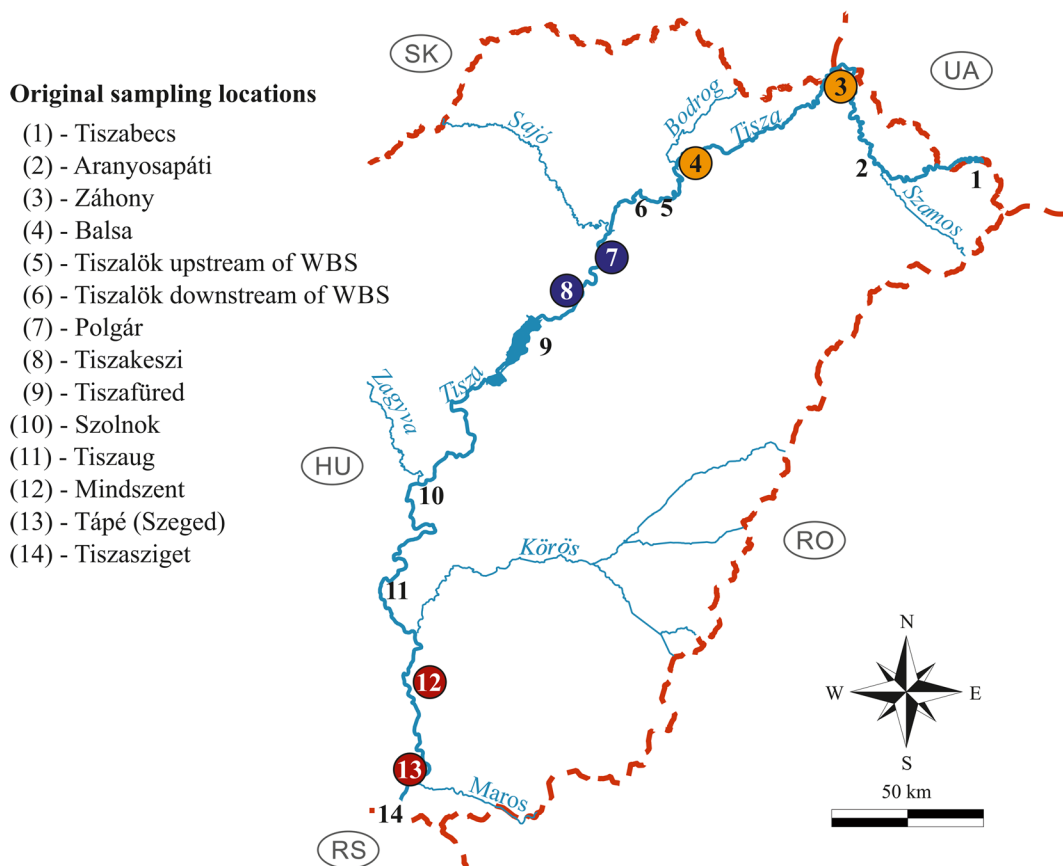
Original sampling locations

(1) - Tiszabecs
(2) - Aranyosapáti
(3) - Záhony
(4) - Balsa
(5) - Tiszalök upstream of WBS
(6) - Tiszalök downstream of WBS
(7) - Polgár
(8) - Tiszakeszi
(9) - Tiszafüred
(10) - Szolnok
(11) - Tiszaug
(12) - Mindszent
(13) - Tápé (Szeged)
(14) - Tiszasziget

Fig. 5 Overall picture with colored sampling locations forming homogeneous groups
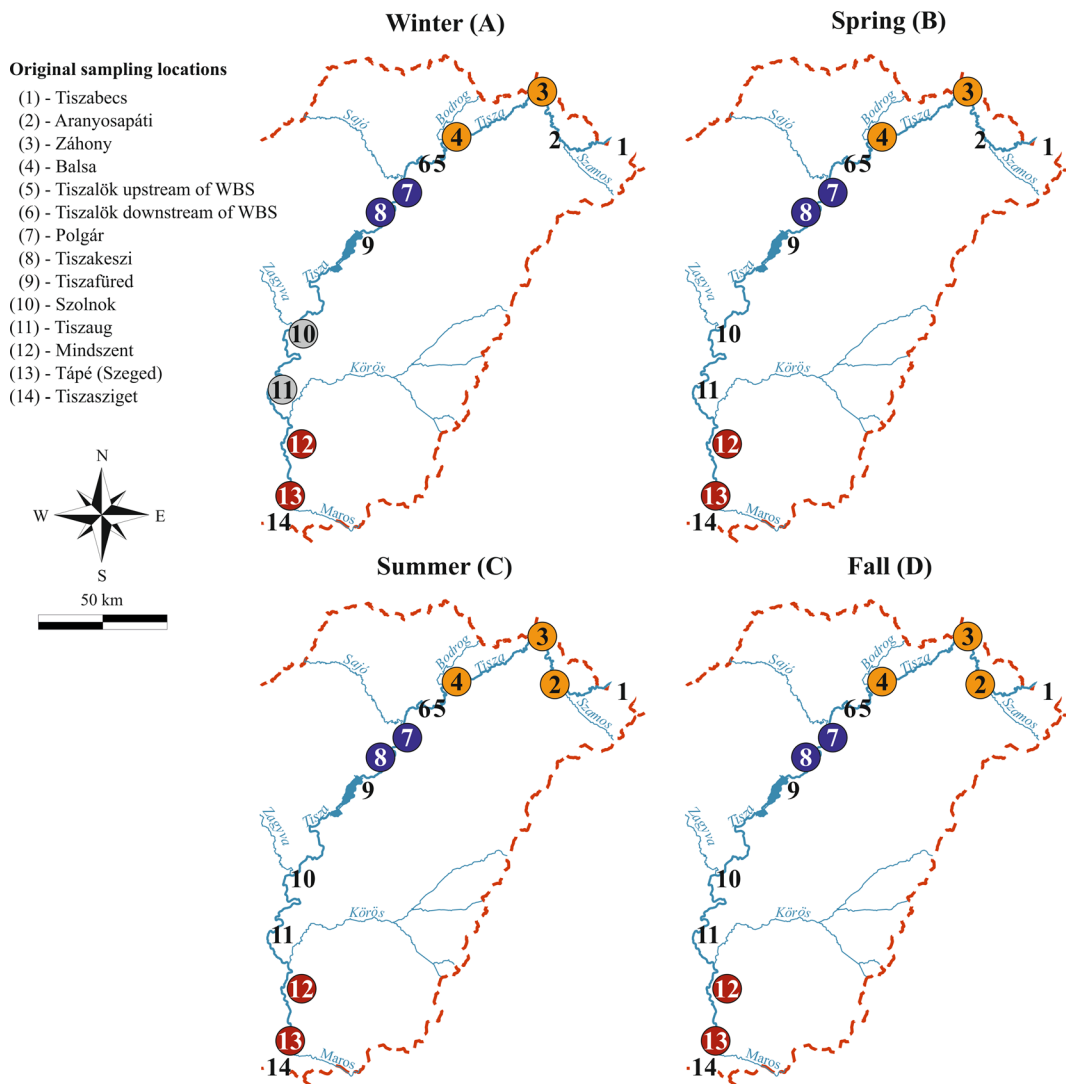
Fig. 6 Homogeneous groups of sampling sites (*colored ones*) on the Hungarian section of the River Tisza for each of the four seasons

In spring, there are only three SL pairs, because SL-10 and SL-11 separated into two distinct groups. Apart from this change, all the other SLs remained in the same group as observed in winter. Therefore, there were 11 homogeneous groups in the spring compared to the 10 in winter (Fig. 6b). This pattern is in accordance with the overall picture obtained from the data for the whole year.

Moving onto the two last seasons, a similar pattern was obtained. In summer and fall, SL-2 connected to the group containing SL-3 and SL-4, unlike in winter or spring. This was the only group containing three SLs (Fig. 6c, d). As a result, the number of homogeneous groups decreased by one from 11 to 10 in summer and in fall.

To summarize, concerning the similarities and differences of the SLs in space (downstream North to South) with regard to the "overall" and the seasonal picture, the following can be said:

i.   SL-1 formed one group alone in the case of every approach;
ii. SL-2 in summer and fall connects to the group formed by SL-3 and SL-4: the latter two were together in all cases;
iii. SL-5 and SL-6 always remained separated;
iv. SL-7 and SL-8 always formed one group together;
v.   SL-9 formed one group alone in the case of every approach;

vi.   SL-10 and SL-11 in winter form one group;

vii. SL-12 and SL-13 always formed one group; and

viii. SL-14 formed one group alone in the case of every approach.

Discussion

Separation of the time period

As presented, the dataset was split at 1992/1993 into two similarly behaving time intervals. The reason for this separation was the change in the concentration of the parameters measuring the trophic conditions and saprobity at the turn of the decade (1990). For example, PO_4-P, BOD-5, and NO_3-N started decreasing at the beginning of the 1990s, while DO started to increase (Fig. ESM 2). The presence of these phenomena concurs with the findings of Mandera and Forsberg (2000), who pointed out that anthropogenic activity was responsible for the elevated concentrations of nutrients in the waters of the Tisza. According to the new political thinking at the end of the 1980s and the beginning of the 1990s, high state subsidies were withdrawn from artificial fertilizers, while high fertilizer prices and the unpredictable future of state farms and cooperatives resulted in a drastic reduction in fertilizer usage. Hungarian agriculture and industry were severely restructured after the collapse of the Soviet Union in 1991. There was a tenfold increase in fertilizer prices in the 1980s. The fall in fertilizer use was much more dramatic (between 10 and 20 times) than in the Western European countries (Csathó et al. 2007; Hatvani et al. 2014b).

These two facts played a great role in the decrease of PO_4-P concentrations in inland waters located in agricultural areas in last years of the twentieth century (Grimvall et al. 2000).

Separation of the seasons

Since Hungary is located in the continental climate zone, from a meteorological perspective four seasons exist. Huschke (1959) in his work entitled "The Glossary of Meteorology" defines the seasons as follows: The warmest period in the year is summer everywhere in the world, except for a couple of tropic regions, while the coldest is winter. According to Trenberth (1983), between these two, spring and fall serve as transitional periods. Here, Trenberth suggested that the seasons should be determined based on annual mean temperatures.

The seasonal pattern obtained did not concur with the general meteorological aspect of the four seasons with equal length in the continental climate zone. Summer and winter were both 4 months long, while spring and fall were each 2 months long (Fig. 3). This observation is supported by the difference values obtained for consecutive months using CCDA (Fig. 4). Since the difference value between February and March (9 %) is much smaller than the one for March–April (20 %), it is reasonable that March rather joins the winter months. Similarly, September is closer to August than October (differences of 8 and 10 % respectively). The other bordering months between the seasons, December and June show a similar behavior as well, justifying that they belong to the winter and the summer seasons, respectively. This finding that the seasons are of unequal length concurs with the research of Alpert et al. (2004) conducted on meteorological data. It is suspected that the reason behind this phenomenon lies in the fast transition of the characteristic processes of winter into summer in spring and the other way around in fall.

To answer the question of which parameters were responsible for the separation of the consecutive months, their Wilks' lambda statistics were determined (Table 2). Parameters closely related to water

Table 2 Wilks' lambda statistics of the parameters. The smaller the coefficient is, the more the parameter was responsible for the separation of the months

Parameter	Wilks' lambda
NO_3-N	0.640
Dissolved O_2	0.667
Runoff	0.744
NH_4-N	0.756
Na^{2+}	0.876
Cl^-	0.882
K^+	0.885
HCO_3^-	0.916
Ca^{2+}	0.924
pH	0.927
SO_4^{2-}	0.955
NO_2-N	0.961
BOD_5	0.963
Mg^{2+}	0.966
PO_4-P	0.984

temperature (e.g. dissolved oxygen) and/or seasonal effects such as floods, composition, and decomposition processes (e.g., ammonium or nitrate; Table 2; Fig. ESM 4A) influenced the separation of the months the most and are likely to vary between the different parts of the year. Regarding the less influential variables, their distribution was less variable across the months (Table 2; Fig. ESM 4B). Nitrate and dissolved oxygen are both periodic. In the meanwhile, runoff reaches its maxima during green floods, as a consequence becoming an important grouping factor. However, not all of the changes in the parameters' concentrations were connected to the transitional seasons or even the points of transition. Ammonium reaches its maximum in winter because of the dominating decomposition processes (Hatvani et al. 2011), standing out from the other seasons and, therefore, becoming another considerable factor, while other (less determining) parameters are more or less stable during the year and, thus, do not play such an important role in the grouping procedure.
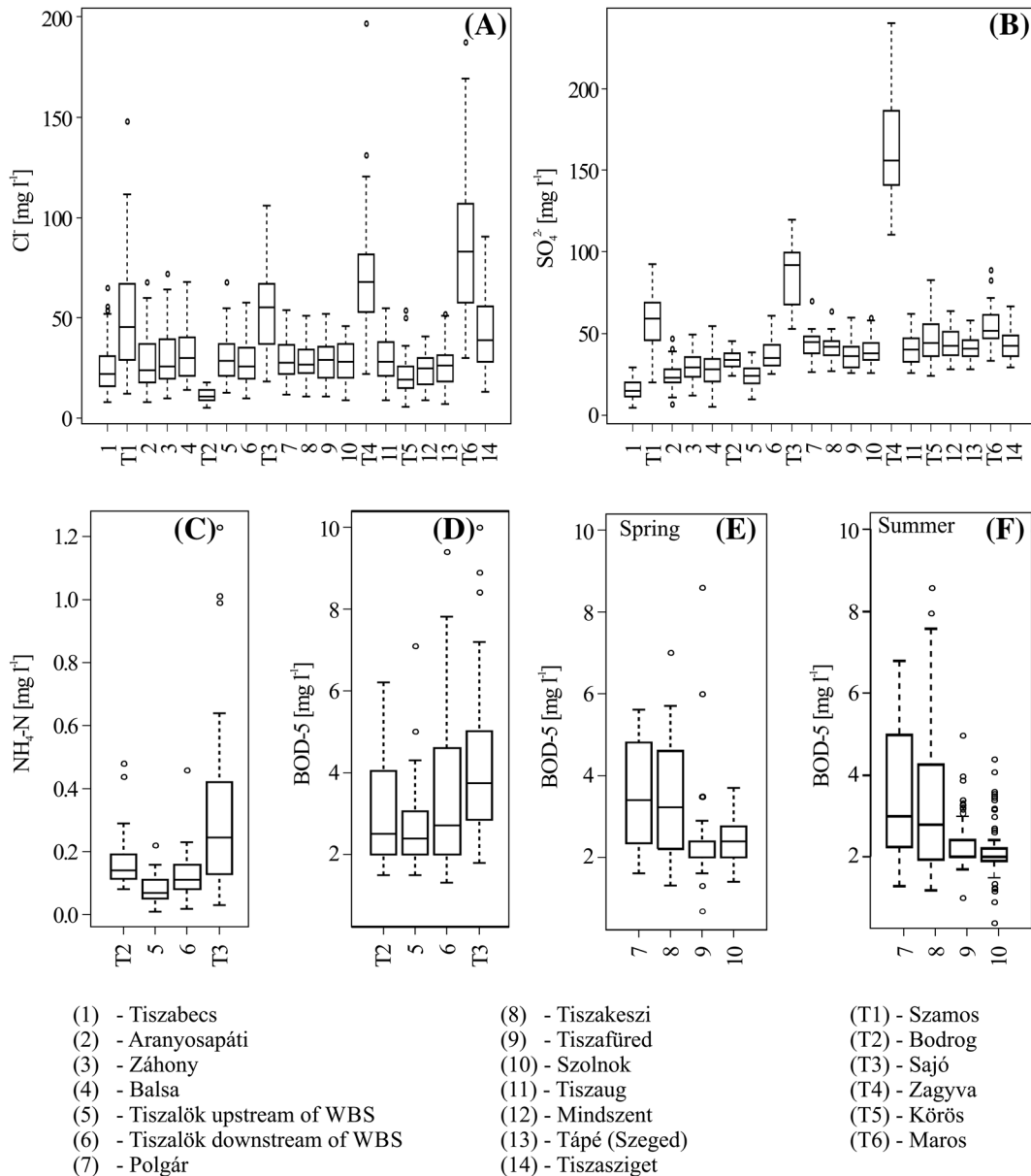

(1) - Tiszabecs
(2) - Aranyosapáti
(3) - Záhony
(4) - Balsa
(5) - Tiszalök upstream of WBS
(6) - Tiszalök downstream of WBS
(7) - Polgár
(8) - Tiszakeszi
(9) - Tiszafüred
(10) - Szolnok
(11) - Tiszaug
(12) - Mindszent
(13) - Tápé (Szeged)
(14) - Tiszasziget
(T1) - Szamos
(T2) - Bodrog
(T3) - Sajó
(T4) - Zagyva
(T5) - Körös
(T6) - Maros

Fig. 7 *Box-and-whiskers plots* showing separating effects

Separation of the sampling sites in space

The separation of the seasons was related to the intra-year fluctuations of the parameter values, while the spatial patterns were primarily caused by the tributaries and anthropogenic obstacles. Irrespective of the seasons, the tributaries separated the SLs on every occasion. This finding concurred with those of Sharp (1971) and Sanders (1980). The effect of the tributaries consisted of two factors, (i) the geochemical composition of the waters brought from their watershed, and (ii) their load of anthropogenic origin. Latter mainly consisted of nutrients. After the mouths of the tributaries the concentrations of the parameters changed (increased or decreased). These phenomena were enough to separate two SLs into two distinct groups. For example SL-13 and SL-14 were separated by the Maros River (T6) (Fig. 7a, b).

Unlike in the studies of Sharp (1971) and Sanders (1980), in which only the tributaries were identified as being responsible for the separation of the sections of the river, here, other separating factors were identified as well. Besides the tributaries, man-made obstacles also played a role in the separation of the SLs into different groups. This finding concurs with the new wave of perception when it comes to recalibrating monitoring networks, i.e. taking anthropogenic effects into account (Do et al. 2011). SL-5 and SL-6 (up- and downstream of WBS) separated, supposedly because of the WBS. Although the riverbed morphology changed in the area resulting in slower water velocity, the concentration of the parameters related to halobity did not change, unlike the concentration of parameters related to biological activity. Their values changed significantly up- and downstream of the WBS; DO decreased between the two, while BOD-5 and NH_4-N increased (Fig. 7c, d). It is suspected that the effect of the WBS could be decreased by fine-tuning its operation through integrated basin management e.g. using the water quantity and quality optimization model of Zhang et al. (2011).

A further separation between the SLs was caused by Lake Tisza. It buffers the nutrient loads arriving from upstream and stabilizes their concentrations (e.g. BOD-5; Fig. 7e, f). Therefore it separated the group consisting of SL-7 and SL-8 and the group formed by SL-9 alone into two groups, although they are relatively close to each other.

Conclusions and outlook

In this paper, homogeneous groups of sampling sites were sought. First, it was shown that there are four temporally similar sections for the period 1993–2005. The months within hydrochemical seasons, however, can only be regarded as similar and not as homogeneous. The most explicit changes could be observed in April–May and October–November. Hence, a higher temporal sampling frequency is suggested in spring and fall, compared to summer and winter months, to obtain a more precise picture of the underlying processes. Nevertheless, for similar cases of unequal season lengths, it should be kept in mind that many statistical methods require equidistant sampling, for which the harmonization of the temporal sampling frequency is essential. Thus, the temporal sampling frequency should be aligned according to the seasons with the highest difference values, i.e., spring and fall; this will comply with the needs of the whole monitoring as well.

In space, the borders between the homogeneous spatial groups tend to be located at the mouths of the tributaries, as shown by Sharp (1971), but not exclusively. Human-made obstacles such as the WBS can also cause such disturbances, leading to the separation of close neighboring sampling sites. As a result, out of the 14 SLs, only 11 are necessary.

The study shows a good example of the spatio-temporal optimization of a monitoring network using the generally applicable combined cluster and discriminant analysis. The authors are convinced that similar situations can be found all over the world; therefore, the conclusions, the method, and the message can be generalized.

References

Afifi, A., Clark, V. A., May, S., Raton, B., et al. (2004). *Computer-aided multivariate analysis (4th ed.)* (p. 489). USA:Chapman & Hall/CRC.

Alberto, W. D., del Pilar, D. M., Valeria, A. M., Fabiana, P. S., Cecilia, H. A., de los Ángeles BM, (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía river basin (Córdoba–Argentina). Water Research, 35 (12), 2881-2894. Doi 10.1016/S0043-1354(00)00592-3

Alpert, P., Osetinsky, I., Ziv, B., Shafir, H. (2004). A new seasons definition based on classified daily synoptic systems: an example for the eastern Mediterranean. International Journal of Climatology. 24, 1013–1021. Doi 10.1002/joc.1037.

Chen, Q., Wu, W., Blanckaert, K., Maa, J., Huang, G. (2012). Optimization of water quality monitoring network in a large river by combining measurements, a numerical model and matter-element analyses. Journal of Environmental Management, 110, 116-124. Doi 10.1016/j.jenvman.2012.05.024

Chilundo, M., Kelderman, P., & O'keeffe, J. (2008). Design of a water quality monitoring network for the Limpopo river basin in Mozambique. *Physics and Chemistry of the Earth, Parts A/B/C*, *33*(8-13), 655–665.

Csathó, P., Sisák, I., Radimszky, L., Lushaj, S., Spiegel, H., Nikolova, M. T., Nikolov, N., Čermák, P., Klir, J., Astover, A., Karklins, A., Lazauskas, S., Kopinski, J., Hera, C., Dumitru, E., Manojlović, M., Bogdanović, D., Torma, S., Leškošek, M., & Khristenko, A. (2007). Agriculture as a source of phosphorus causing eutrophication in Central and Eastern Europe. *Soil Use and Management*, *23*(Suppl. 1), 36–56.

Do, H. T., Lo, S.-L., Chiueh, P-T., Thi, L. A. P. (2012). Design of sampling locations for mountainous river monitoring. Environmental Modelling & Software, 27-28, 62-70, doi: 10.1016/j.envsoft.2011.09.007.

Do, H. T., Lo, S.-L., Chiueh, P-T., Thi, L. A. P., Shang, W-T. (2011). Optimal design of river nutrient monitoring points based on an export coefficient model. Journal of Hydrology, 406, 129-135.

European Commission. 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. Off J Eur Communities 2000; L 327:1. [22.12.2000].

Fan, X., Cui, B., Zhao, H., Zhang, Z., Zhang, H. (2010). Assessment of river water quality in Pearl River Delta using multivariate statistical techniques. Procedia Environmental Sciences, 2, 1220-1234. Doi 10.1016/j.proenv.2010.10.133

Federal Water Pollution Control Act (Clean Water Act) (33 U.S.C. 1251–1376; Chapter 758; P.L. 845, June 30,1948; 62 Stat. 1155

Grimvall, A., Stålnacke, P., & Tonderski, A. (2000). Time scales of nutrient losses from land to sea—a European perspective. *Ecological Engineering*, *14*, 363–371.

Hatvani, I. G., Kovács, J., Székely Kovács, I., Jakusch, P., & Korponai, J. (2011). Analysis of long-term water quality changes in the Kis-Balaton Water Protection System with

time series-, cluster analysis and Wilks' lambda distribution. *Ecological Engineering*, *37*, 629–635.

Hatvani, I. G., Clement, A., Kovács J., Székely Kovács, I., Korponai, J. (2014a). Assessing water-quality data: the relationship between the water quality amelioration of Lake Balaton and the construction of its mitigation wetland. Journal of Great Lakes Research. 40 (1), 115-125. Doi 10.1016/j.jglr.2013.12.010

Hatvani, I. G., Kovács, J., Márkus, L., Clement, A., Hoffmann, R., Korponai, J. (2014b). Determining background factors in the time series of an agricultural watershed using dynamic factor analysis (River Zala, W-Hungary). *Journal of Hydrology, 521*, 460–469. doi:10.1016/j.jhydrol.2014.11.078.

Hatvani, I. G., Magyar, N., Zessner, M., Kovács, J., & Blaschke, A.P. (2014c). The water framework directive: can more information be extracted from groundwater data? A case study of Seewinkel, Burgenland, eastern Austria. *Hydrogeology Journal, 22*(4), 779–794. doi:10.1007/s10040-013-1093-x.

Huschke, R. E. (1959). *Glossary of meteorology*. Boston: American Meteorological Society.

Icaga, Y. (2005). Genetic algorithm usage in water quality monitoring networks optimization in Gediz (Turkey) River basin. *Environmental Monitoring and Assessment*, *108*, 261–277.

Júnez-Ferreira, H. E., & Herrera, G. S. (2013). A geostatistical methodology for the optimal design of space–time hydraulic head monitoring networks and its application to the Valle de Querétaro aquifer. *Environmental Monitoring and Assessment, 185*, 3527–3549.

Kao, J-J., Li, P-H., Hu, W-S. (2012). Optimization models for siting water quality monitoring stations in a catchment. Environmental Monitoring and Assessment, 184, 43–52.

Karamouz, M., Kerachian, R., Akhbari, M., & Hafez, B. (2009). Design of river water quality monitoring networks: a case study. *Environmental Modeling and Assessment*, *14*, 705–714.

Kentel, E., & Alp, E. (2013). Hydropower in Turkey: economical, social and environmental aspects and legal challenges. *Environmental Science & Policy*, *31*, 34–43.

Kolander, R., & Tolkowsky, J. (2008). Hydrochemical seasons in the Lake Gardno catchment on Wolin Island (north-western Poland). *Limnological Review*, *8*(1-2), 27–34.

Kovács, J., Nagy, M., Czauner, B., Székely Kovács, I., Borsodi, A. K., & Hatvani, I. G. (2012a). Delimiting sub-areas in water bodies using multivariate data analysis on the example of Lake Balaton (W Hungary). *Journal of Environmental Management*, *110*, 151–158.

Kovács, J., Tanos, P., Korponai, J., Székely Kovács, I., Godár, K., Godár-Sőregi, K., Hatvani, I. G. (2012b). Analysis of water quality data for scientists, in: Voudouris, K., Voutsa, D. (Eds.), Water quality monitoring and assessment. InTech, 65-94. ISBN 978-953-51-0486-5

Kovács, J., Kovács, S., Magyar, N., Tanos, P., Hatvani, I. G., Anda, A. (2014). Classification into homogeneous groups using combined cluster and discriminant analysis. *Environmental Modelling and Software, 57*, 52–59. doi:10.1016/j.envsoft.2014.01.010

Kovács, J., Kovács, S., Hatvani, I. G., Magyar, N., Tanos, P., Korponai, J., Blaschke A. P. (2015). Spatial optimization of monitoring networks on the examples of a river, a lake-

wetland system and a sub-surface water system. *Water Resources Management*. doi:10.1007/s11269-015-1117-5.

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)*. Budapest: Akadémiai Kiadó.

Lee, C., Paik, K., Yoo, D. G., & Kim, J. H. (2014). Efficient method for optimal placing of water quality monitoring stations for an ungauged basin. *Journal of Environmental Management*, *132*, 24–31.

Mahjouri, N., & Kerachian, R. (2011). Revising river water quality monitoring networks using discrete entropy theory: the Jajrood River experience. *Environ Monitoring Assessment*, *175*, 291–302.

Mandera, Ü., & Forsberg, C. (2000). Nonpoint pollution in agricultural watersheds of endangered coastal seas. *Ecological Engineering*, *14*, 317–324.

Masoumi, F., & Kerachian, R. (2010). Optimal redesign of groundwater quality monitoring networks: a case study. *Environmental Monitoring and Assessment*, *161*, 247–257.

Moreira, J. R., & Poole, A. D. (1993). *Hydropower and its constraints*. United States.

Naddeo, V., Zarra, T., Belgiorno, V. (2007). Optimization of sampling frequency for river water quality assessment according to Italian implementation of the EU Water Framework Directive. Environmental science & Policy, 10, 243–249. Doi 10.1016/j.envsci.2006.12.003

Naddeo, V., Scannapieco, D., Zarra, T., Belgiorno, V. (2013). River water quality assessment: implementation of non-parametric tests for sampling frequency optimization. Land Use Policy, 30, 197-205. Doi 10.1016/j.landusepol.2012.03.013

Park, S.-Y., Choi, J. H., Wang, S., & Park, S. S. (2006). Design of a water quality monitoring network in a large river system using the genetic algorithm. *Ecological Modelling*, *199*(3), 289–297.

Pécsi, M. (1969). *A tiszai Alföld*. Budapest:Akadémiai Kiadó.

Preziosi, E., Petrangeli, A. B., & Giuliano, G. (2013). Tailoring groundwater quality monitoring to vulnerability: a GIS procedure for network design. *Environmental Monitoring and Assessment*, *185*, 3759–3781.

Razmkhah, H., Abrishamchi, A., & Torkian, A. (2010). Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: a case study on Jajrood River (Tehran, Iran). *Journal of Environmental Management*, *91*, 852–860. doi:10.1016/j.jenvman.2009.11.001.

Sakan, S., Gržetić, I., & Đorđević, D. (2007). Distribution and fractionation of heavy metals in the Tisa (Tisza) river sediments. *Environmental Science and Pollution Research*, *14*(4), 229–236.

Sanders, T. G., & Adrian, D. D. (1978). Sampling frequency for river quality monitoring. *Water Resources Research*, *14*(4), 569–576.

Sanders, T. G. (1980). *Principles of network design for water quality monitoring*. Colorado:Colorado State University.

Sanders, T. G., Ward, R., Loftis, J., Steele, T., Adrian, D., & Yevjevich, V. (1983). *Design of networks for monitoring water quality*. Colorado:Water Resources Publications.

Scannapieco, D., Naddeo, V., Zarra, T., & Belgiorno, V. (2012). River water quality assessment: a comparison of binary- and fuzzy logic-based approaches. *Ecological Engineering*, *47*, 132–140. doi:10.1016/j.ecoleng.2012.06.015.

Sharp, W. E. (1971). A topographical optimum water sampling plan for rivers and streams. *Water Resources Research*, *7*, 1641–1646.

Shen, Y., & Wu, Y. (2013). Optimization of marine environmental monitoring sites in the Yangtze River estuary and its adjacent sea, China. Ocean & Coastal Management, 73, 92-100. Doi 10.1016/j.ocecoaman.2012.11.012

Strobl, R. O., & Robillard, P. D. (2008). Network design for water quality monitoring of surface freshwaters: a review. *Journal of Environmental Management*, *87*(4), 639–664.

Takács, K., Kern, Z., & Nagy, B. (2013). Impacts of anthropogenic effects on river ice regime: examples from Eastern Central Europe. *Quaternary International*, *293*, 275–282.

Takács, K., & Kern, Z. (2015). Multidecadal changes in the river ice regime of the lower course of the River Drava since AD 1875. *Journal of Hydrology*. doi:10.1016/j.jhydrol.2015.01.040.

Telci, I. T., Nam, K., Guan, J., & Aral, M. M. (2009). Optimal water quality monitoring network design for river systems. *Journal of Environmental Management*, *90*, 2987–2998. doi:10.1016/j.jenvman.2009.04.011.

Trenberth, K. E. (1983). What are the seasons? Bulletin American Meteorological Society 64 (11).

Varol, M., Gökot, B., Bekleyen, A., & Şen, B. (2012). Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. *Catena*, *92*, 11–21. doi:10.1016/j.catena.2011.11.013.

Wang, J. (2001). Extension set theory, extension engineering method and extension system control. Academic Open Internet Journal 5. http://www.acadjournal.com/2001/v5/part5/p1

Wang, X., Cai, Q., Ye, L., Qu, X. (2012). Evaluation of spatial and temporal variation in stream water quality by multivariate statistical techniques: a case study of the Xiangxi River basin, China. Quaternary International, 282, 137-144. Doi 10.1016/j.quaint.2012.05.015

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, *24*, 471–494.

Zhang, Y., Xia, J., Chen, J., & Zhang, M. (2011). Water quantity and quality optimization modeling of dams operation based on SWAT in Wenyu River Catchment, China. *Environmental Monitoring and Assessment*, *173*, 409–430.

Zsuga, K., Tóth, A., Pekli, J., & Udvari, Z. (2004). A Tisza vízgyűjtő zooplanktonjának alakulása az 1950-es évektől napjainkig. *Hidrológiai Közlöny*, *84*(5-6), 175–178.