Application of state-of-the-art geomathematical methods in water protection

- on the example of the data series of the Kis-Balaton Water Protection System -

by

István Gábor Hatvani
M.Sc., Environmental Sciences, Eötvös Loránd University, Budapest 2010
Eötvös Loránd University, Faculty of Science, Department of Physical and Applied Geology

Dissertation submitted to the Ph.D. program for Environmental Earth Sciences at the Ph.D.
School of Environmental Sciences, Eötvös Loránd University, Budapest

Chair of the doctoral school & the Environmental Earth Sciences Doctoral Program: Prof. András Galácz (Department of Paleontology)

Supervisor: József Kovács Ph.D. (Department of Physical and Applied Geology, Budapest)
External advisor: János Korponai Ph.D. (West Transdanubian Water Authority, Department Kis-Balaton, Keszthely & University of West Hungary, Department of Chemistry and Environmental Sciences, Szombathely)

Budapest, 2014

Table of contents

Kányavári Bridge at Kányavári Island at Kis-Balaton. Photo taken by János Korponai.

List of figures

List of tables

List of abbreviations

AA	Autocorrelation analysis
aNL	Annual nutrient load
ANOVA	Analysis of variance
ARD	Annual relative difference
BOD_5	Five day biological oxygen demand
CA	Cluster analysis
Chl-a	Chlorophyll-a
COD_{ps}	Chemical oxygen demand….
CV	Coefficient of variation
DA	Discriminant analysis
DF	Dynamic factor
DFA	Dynamic factor analysis
KBWPS	Kis-Balaton Water Protection System
Kszt	Keszthely Bay sampling site
LOESS	Locally weighted scatterplot smoothing
MANOVA	Multivariate analysis of variance
N	Nitrogen
OECD	Organization for Economic Co-operation and Development
Org. N	Organic nitrogen
P	Phosphorus
PCA	Principal component analysis
POT	Peaks over Threshold
PP	Particulate phosphorus
Q	Daily runoff
SDOM	Standard deviation of the mean
SRP	Soluble reactive phosphorus
SS	Sampling site
TN	Total nitrogen
TP	Total phosphorus
TSS	Total suspended solids
VA	Variogram analysis
WSA	Wavelet spectrum analysis
WW-P	Waste water phosphorus; effluent total P concentration of the Zalaegerszeg waste water treatment plant
Wλ	Wilks' λ statistics

Acknowledgements

First of all I would like to express my gratitude to my supervisor József Kovács Ph.D. and my external supervisor, János Korponai Ph.D.

I would like to express my sincere gratitude to my closest family, my Grandmother Márta Chiovini who supported me all the way…

I am thankful for the professional help of Dr. Adrienne Clement, Dr. László Márkus, Prof. Lajos Pap, Prof. Antal Füst and Prof. Alfred P. Blaschke and for the stimulating discussions I had with Prof. Ádám Kiss, Prof. Ákos Török, Dr. István Tátrai.

*I owe thanks to György Süle for providing me with the datasets, Dr. Nóra Szilágyi for providing the Matlab *m program, Péter Tanos for his help with the artwork, Paul Thatcher for the English review, Péter Völgyesi for the months-and-months spent together studying, and each and every colleague with whom I worked.*

This work was financially supported by:

- *the Ph.D. scholarship of Hungary to I.G. Hatvani 2010-2013,*
- *the TÁMOP 4.2.4. A/1-11-1-2012-0001 „National Excellence Program,*
- *the "Habilitas" scholarship of the Hungarian Development Bank Plc. (MFB Zrt.),*
- *the Ernst Mach Scholarship of the Austro-Hungarian Action Fund.*

Dedicated to the memory of my parents, Mária Aszódi & Prof. Dr. István László Hatvani…

1. Foreword

Water is perhaps our most precious common resource, and with urbanization and the spread of agriculture and industry it has become, and is becoming more and more exposed to contamination and misuse. Therefore sufficient monitoring of its quality and quantity is a vital issue. With problems related to water quality becoming more frequent, more sampling sites are being included in national and international monitoring networks, and their sampling frequency is showing a tendency to increase as well. This trend has as its result an ever-growing amount of data, to a point where this amount is greater than is usual within the scope of "simple" statistical analyses. Thus, in the last few decades stochastic modeling, with the use of time series analysis and multivariate statistical techniques, has increased dramatically in surface and groundwater research. The reason is the increase in the amount and time span of the available data. As a consequence, it has become possible to investigate the relationship between various natural parameters (i.e. random variables in the statistical modeling) and the temporal evolution of the natural processes, observed in discrete time periods as time series.

The often-used deterministic models suppose - implicitly - the invariability of the environment where the described phenomenon evolves. However the environment of hydrological processes is so complex and diverse that the assumption of their invariability allows only for description of average conditions, or on larger scales. The perturbances, fluctuations and extreme behavior, as well as the reliability of the obtained results are tractable primarily through the techniques of probability theory and statistics. In addition the characteristics of hydrological processes, driven by precipitation, are in general considered to be stochastic. These thoughts appear in the international literature as well, stating that besides deterministic, fuzzy and combinations of stochastic and deterministic models the future may be stochastic modeling (Kovács & Szanyi, 2005; Wilkinson, 2006), which is one of the most straightforward tools when it comes to geomathematics. Geomathematics is the discipline which deals with the qualitative and quantitative problems of the Earth's systems using the tools of mathematics/statistics combined with the knowledge of earth sciences (Agterberg, 1974; Freeden et al., 2010).

2. Introduction

- General discussion of applicable methods on water quality datasets -

Water quality datasets obtained from surface water monitoring can be described by a four dimensional graph, as in the work of Kovács et al., 2012a (Fig. 2-1). Before geomathematical tools and their detailed application and description can be presented on the example of the Kis-Balaton Water Protection System's (KBWPS) time series, the general properties of environmental (surface water quality) datasets and their handling should be discussed. This is an important step since the tools of multivariate data analysis are diverse; thus, the datasets could be approached from many perspectives. It is therefore it is possible that in the course of such a complex investigation, aspects are either left out or duplicated. This is the reason why the Introduction aims to provide a clear indication of the paths along which my analyses were conducted.

In most aspects of water research the events analyzed are described by different parameter samples obtained from more than one spatial sampling site.

Here I would like to stress the fact that the terminus technicus "parameters" is used differently in the two disciplines concerned i.e. environmental sciences and mathematical statistics. Often what is called a parameter in environmental sciences is considered as a variable in mathematics. I feel that the environmental approach is closer to the perspective of the present study; therefore chemical substances will be addressed as "parameters", although they are the variables of the statistical analyses.

The location of a planar sampling site is determined by two spatial coordinates (x_i & y_i) giving two dimensions, while the parameter types (e.g. runoff, biological oxygen demand, potassium, soil phosphorus surplus etc.) are located in a third dimension. A limnological process is often described by status parameters sampled only once. If, however, samples are taken over time, one is dealing with a time series, and these three dimensions are extended to four, with time as the fourth axis (Fig. 2-1).

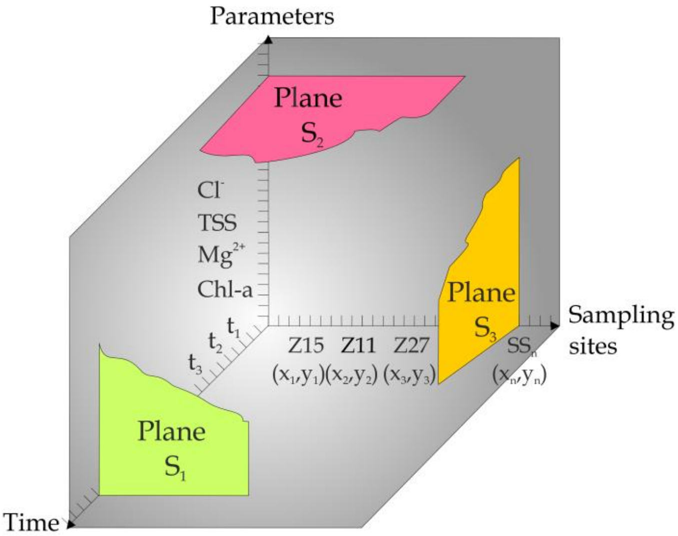


Figure 2-1 Data in four dimensions (Sampling sites [x_i; y_i], parameters, time) based on Kovács et al. (2012a)

2.1. Plane S1

As an example let us imagine that in a certain area many parameters are sampled from more than one water quality sampling site (SS)[1] at the same time (plane S1). If the data obtained is then recorded on worksheets, it is usual practice that each row corresponds to a SS and each column to a parameter. This data matrix is considered to be static. Besides the common univariate statistical methods, multivariate ones can also be used, such as cluster-, discriminant-, principal component-, and factor analysis or even multidimensional scaling, which is able to map an *N* dimensional space onto two or three dimensions. Its use in environmental sciences is becoming more and more popular, as one may see in the book of Borg & Groenen (2007).

Cluster analysis and multi-dimensional scaling can be used at SSs in cases where it is necessary to reveal similarities. Another aim can be the selection of natural parameters (variables) that explain most of the variance of the original dataset. This can be achieved using principal component analysis or factor analysis. Usually in the course of cluster analysis the rows of the data matrix, and in the case of principal component analysis or factor analysis its columns, are the object of the analyses. Naturally, if there are multiple sampling sites sampled at sufficient density, the effective range can be determined using spatial sampling density analysis, by means of the semivariogram, which is the basic tool of geostatistics (Matheron, 1965). Continuing in this vein, from the actual concentration values of the measured parameters one can plot isoline maps. The most frequently used interpolation (linear estimation) method is kriging, which uses the parameters of the semivariogram to characterize the association between data points (Füst 2004, Molnár 1987, Stein, 1999). This is highly informative in cases where anomalies are sought (Hatvani et al., 2014a). However, maps can only be created if the monitoring system is sufficiently densely spread in space. In the case of a linear "1D" system, such as the KBWPS, such a result would not make sense.

2.2. Planes S2 and S3

In most cases data depend also on time, spanning the fourth dimension of the scheme (Fig. 2-1). In such cases the data matrix is not static. Staying with the previous example, if more than one SS is sampled in time, one is dealing with problems described in plane S2, while if multiple parameters are sampled multiple times at one SS, one is dealing with problems

---

[1] Further on SS will be used for abbreviating „sampling site", as opposed to general statistics where it is commonly used to abbreviate „sum of squares".

described in plane S3. The approach in most cases depends only on whether the analyses are to be conducted on the time series of one natural parameter at different sampling sites and then compared with the results for the other parameters, or the time series of multiple parameters are assessed at only one sampling site, and then the sampling sites are compared.

In both planes time series methods can be employed, as e.g. in the monograph of Hans (2005); in particular, this implies the identification of the parameters' trend and periodicity. In many cases determining these is a key question in the study, since, if extracted, both can be used for forecasting, but only if one is certain that the driving processes still exist and will exist in the distant future as well. However, further expansion on this topic is not an aim of this work. Alongside those previously mentioned, other methods can be applied as well, e.g. estimation of the autocorrelation function or variogram function, etc.

Regarding the planes, the only question is, whether does one first checks the time series of multiple parameters at one SS and then compares them - if more than one SS exists -, or the other way around.

Besides the questions that can be answered with the aid of "elementary" time series methods, a more complex one, concerning the background processes[2] which drive the sampled parameters' temporal fluctuation, can also be addressed. Because consecutive temporal samples are not independent of each other, dynamic factor analysis can provide a solution to explore the driving background factors of processes while taking into account the lagged correlation structure (Kovács et al., 2004; Kovács, 2007; Márkus et al., 1999). When it comes to the application of dynamic factor analysis there are two known cases: (1) in plane S2 only one parameter was sampled at many SSs (see e.g. the studies of Kovács et al., 2004; Kovács, 2007; Márkus et al., 1999), or the other way around, when (2) the time series of multiple parameters are assessed together at only one sampling site (Hatvani et al., 2014b). In the latter case it can be described as an analysis conducted in S3. Naturally if one is fortunate enough to have multiple sampling sites and parameter time series, the only question is in which plane to explore first (acting both ways see Muñoz-Carpena et al., 2005; Ritter & Muñoz-Carpena, 2006).

However if a parameter's simultaneous spatial and temporal changes are of interest, a more complex approach is needed (e.g. Dryden et al., 2005). Methods such as artificial neural networks (Gallant, 1995; Golden, 1996; Haykin, 1999) can simultaneously work with data from all the three planes. These are most frequently used for forecasting; however artificial neural

---

[2] Background processes are directly unobservable processes, which are present as a weighted additive part in every observable process of the given phenomenon. These can be inferred through a mathematical model, in this case PCA and DFA using directly measured/observed variables.

networks were not used in the case of the KBWPS study, because a part of the system (the as-yet unfinished wetland area, Phase II of the KBWPS) is due to be inundated and completed in 2013-2014, which will change the flow conditions, thereby causing difficulties in the prediction of the processes.

In the case of S1, one can explore the sampling frequency using semivariograms; however, in S2, not only can one explore in space but in time as well for one parameter, while in the case of S3, we are limited to time.

Based on the previous approach, the methodology described in the 4D graph can be actualized and projected onto 2D as a flowchart. This will set the framework of the research conducted on the Kis-Balaton Water Protection System, the study area of the present work.

2.3. Aims of the study

The present work has three main aims:

(1) to demonstrate the applicability of the proposed methods in water quality evaluation, in particular on the example of the long-term data series of the Kis-Balaton Water Protection System (KBWPS);

(2) to describe the processes evolving in the KBWPS over three decades of water quality samples using explorative data analysis methods, and provide excess information to support a better understanding of the system;

(3) to suggest a means of recalibrating the (monitoring) system, meaning that it could be fine-tuned to follow the evolving processes.

Furthermore, the study aims to deal with a segment of the paradox: “In general we are rich in data, but poor in information” (Clement, 2013). This is particularly true for Hungary, where hundreds of thousands of data “lie around” without being assessed in sufficient depth. In the case of the study area the more than 250,000 data assessed represented the fruit of approx. 30 years of hard work and monitoring. This fact urged me to try to extract as much excess information as possible with the aid of stochastic modeling. For a short discussion on the novelty of the methodology used please see Section 4.3. This study (and more generally the approach used in it) could hopefully help scientists to gain a broader perspective on processes evolving in such water systems as the KBWPS; and with regard to the study area, when it comes to finishing the second phase of the reservoir system (2013-2014), more knowledge will be available on what can be expected regarding the quality of the water entering Lake Balaton, and the conservation of the nature preserve wetland area.

The thesis synthesizes a number of topics which have formed the subject of separate studies (in temporal order: Kovács et al., 2010; Hatvani et al., 2011; Kovács et al., 2012b, 2012c; Hatvani et al., 2014a, 2014b, 2014c). The specific aims of these separate studies will be presented from a broader perspective in a coherent context in Section four (Fig. 4.2-1) in conjunction with the corresponding methodology. In my opinion, the study evaluates the results of an explorative research in which many questions were suggested by the personnel working at the West Transdanubian Water Authority, Department Kis-Balaton.

3. Description of the study area

Lake Balaton is the largest shallow freshwater lake in Central Europe (Padisák & Reynolds, 2003). It is located in W Hungary (Fig. 3-1) in the S-SE foreground of the Transdanubian Central Range. It is a nationally-important tourist attraction and recreation area. The lake's surface area is 596 km^2, its average water depth is 3.2 m, and its morphologically diverse watershed covers approximately 5181 km^2, in which 51 inflows are located. 20 of these are constant (Láng, 1986; Pomogyi, 1991). The mean depth and surface area of the lake's geographical basins increases from west to east, while the area of the corresponding sub-watersheds decreases. The largest tributary, the River Zala, supplies almost 50% of the lake's total water input, and - before a number of measures for nutrient control were taken, such as phosphorus (P) precipitation at the Zalaegerszeg waste water treatment plant, a dramatic decrease in fertilizer use, etc. – the River Zala accounted for 35-40% of the lake's nutrient input (Istvánovics et al., 2007; Lotz, 1988). The Zala enters the lake at its westernmost and smallest basin, Keszthely Basin (Kszt in Fig. 3-1), whereas the only outflow is the Sió Canal at the easternmost end of the lake. The canal was built in the late 19th century to maintain a steady water level in Lake Balaton, so that the Budapest-Fiume (today Rijeka, Croatia) railway would be protected from floods.

According to paleolimnological research, about 5000 cal. yr BP the water level of the area began to decrease. Since then, a wetland with a fluctuating water surface and dense macrophyte vegetation has formed in the delta of the River Zala; this is the Kis-Balaton Wetland (Korponai et al., 2010). Prior to the water level regulations of the 19th - 20th Century, the region of the Kis-Balaton lying within the area of the lower Zala Valley was in close connection with Lake Balaton. The recently regulated water level in the lake has been fixed at 2-3 meters below the mean level of the original one. Because of this nineteenth century artificial modification of water level, and the fact that the bed of the Zala River was regulated from 1832 (Virág, 1998) and indeed completely separated from the former wetland areas of Kis-Balaton Wetland and the Lower Zala Valley, the Kis-Balaton Wetland partially dried up and decreased in effectiveness as a filtering area for the River Zala’s waters (Lotz, 1988). It should be mentioned that there are other approaches, based on historical maps, which argue that the Kis-Balaton Wetland separated from Lake Balaton long before the Sió Canal was put into operation (Zlinszky & Timár, 2013; Zlinszky, 2011).

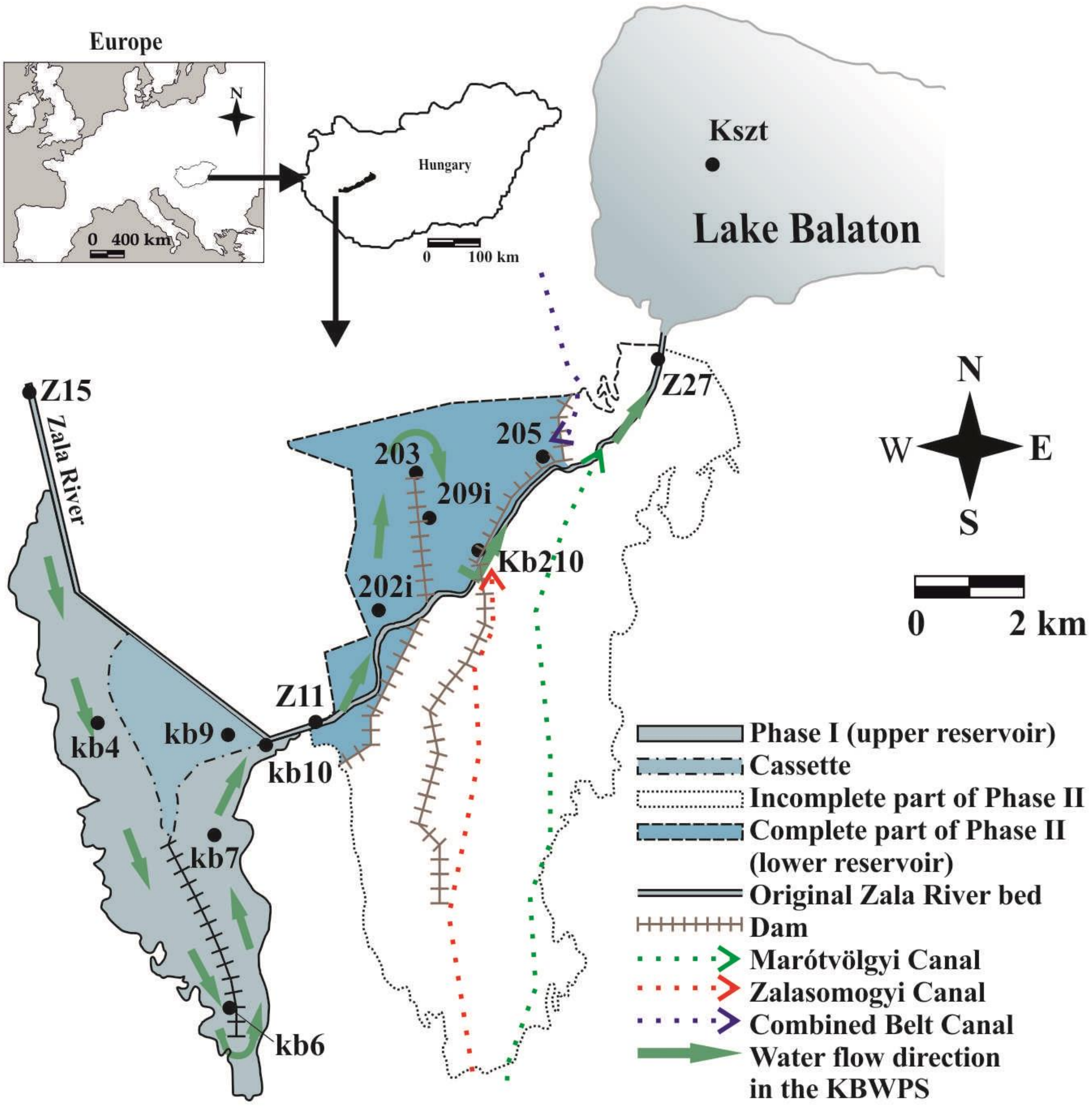


Figure 3-1 The KBWPS and Keszthely Bay, with their sampling sites

In the absence of the Kis-Balaton Wetland functioning as a filter, the waters of the Zala entered Lake Balaton without having been naturally filtered. This, along with accelerated anthropogenic activity in the catchment (waste water production, use of fertilizers etc.), resulted in a significant increase in external nutrient loads and a deterioration of Lake Balaton's water quality, leading to occasional considerable economic losses in the tourism sector (Istvánovics et al., 2007). The first disturbing sign of the deterioration was the blooming of the N_2-fixing *Aphanizomenon flos-aque* cyanobacteria in Keszthely Bay in 1965, while since the eighties the *Cylindrospermopsis raciborskii* has been the primary source of water blooms (Istvánovics et al., 2007; Padisák, 1997; Parpala et al., 2003; Sprőber et al., 2003). The presence of this cyanobacteria indicates high phosphorous loads and hypertrophy (Padisák, 1997). These phenomena were in turn due to the increased nutrient loads reaching the lake. The first signs of heavy eutrophication were observed during the 1970s, at first only in Keszthely Bay, which was then followed by the other basins, producing a significant trophic state gradient from West to East (Somlyódy and Van Straten, 1986). The specific morphometry of the lake gave rise to environmental gradients along its longitudinal axis, due to the increasing water retention time and decreasing area-specific nutrient loads (Fig. 3-2; Istvánovics et al., 2007). Recorded algae blooms peaked in 1982 in Keszthely Bay, exceeding 60 mg m^{-3} Chl-a as an annual mean, and with a maximum of 200 mg m^{-3}, respectively (Herodek, 1984). The worst trophic conditions along the longitudinal section of the lake, however, were observed in 1994. In the following years algal biomass was driven by the decreasing external loads and the hydrometeorological conditions. The decrease of nutrient loads shifted the composition of the phytoplankton towards the “former” mezoeutrophic communities in the Lake (Istvánovics et al., 2002).

One of the problems with blue-green algae growth in relation to the public, is algae toxins. These can cause acute skin irritant effects (Pilotto et al., 2004), and in the case of ingestion during recreational water contact stomach cramps, vomiting, nausea, diarrhea, fevers, sore throat, headaches, muscle and joint pains, blisters in the mouth and liver damage (Bartram & Chorus, 1999). They can even cause the water to have an extremely unpleasant smell. Naturally the above-mentioned symptoms can only occur if the water in question in consumed in great quantities. In the algae-dominated years the periods of blue-green algae blooms lasted for as long as two or three weeks. However, in subsequent years this decreased to only a couple of days. Nevertheless, after 2000, although external loads decreased, Chl-a content remained somewhat elevated, due to the high internal loads (Fig. 3-2b).

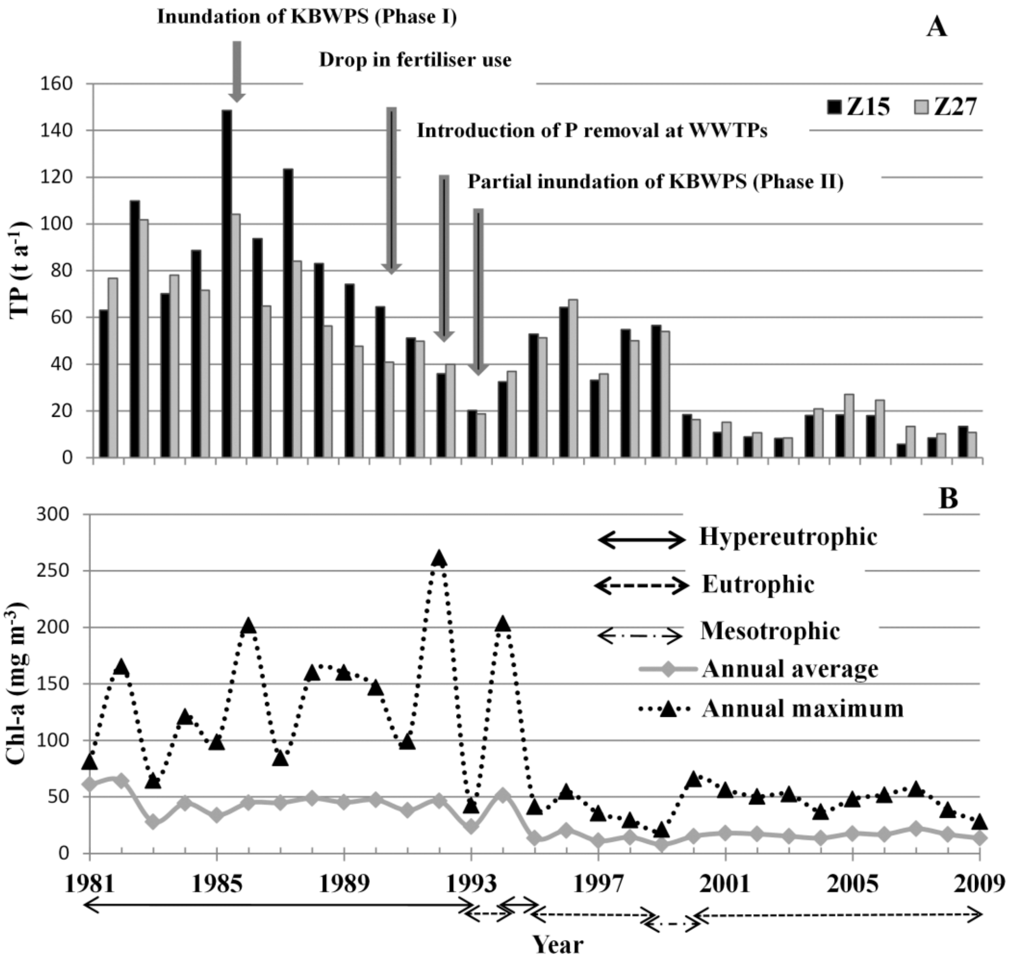


Figure 3-2 TP loads arriving to the KBWPS through the inlet (Z15) and the outlet (Z27) of the Zala (based on Istvánovics et al, 2007) A) and Chl-a concentrations measured in Keszthely Bay with its trophic status marked in 1981-2009 B) (reproduced from Hatvani et al., 2014c)

By the end of the 1980s the P load carried by the River Zala had doubled (Fig. 3-2a) as compared to the beginning of the 1970s. For this reason a regional nutrient load control strategy was worked out for Lake Balaton in the early 1980s, and approved by the government in 1983 (Somlyódy & Van Straten, 1986). The most important management measures included (i) sewage diversion from the eastern and southern shoreline settlements, (ii) construction of a new waste water treatment plant with tertiary treatment and the introduction of P removal (chemical P precipitation) at the waste water treatment plants in the western part of watershed (Zalaegerszeg and Keszthely), (iii) the downsizing of several large livestock farms, and (iv) the construction of the KBWPS, consisting of two reservoirs in sequence located at the mouth of the River Zala to protect Lake Balaton from high nutrient loads which would otherwise result in water quality deterioration.

Review of the literature on the water quality problems of Lake Balaton and the historical background and structure of the Kis-Balaton Water Protection System

Since the 19th century agriculture and urbanization have increased the nutrient load of many surface waters, causing eutrophication and other ecological side-effects which lead to the deterioration of water quality (Cleneghan, 2003; Yates & Prasher, 2009). The primary pollutants in question are phosphorus and nitrogen forms. Both are limiting factors in such waters. With regard to phosphorus, 70% of the source is considered to be diffuse (Stämpfli, 2006). The contribution of mitigation wetlands to the filtering and retention of these diffuse nutrient loads is significant (Baker, 1992; Chambers et al., 1993; Hammer, 1992; Hattermann et al., 2006; Li et al., 2003; Mitch, 1992). They are employed used around the world to clean all kinds of wastewater (Scholz, 2006) and improve the water quality of catchment areas (Mustafa et al., 2009; Verhoeven et al., 2006) and rivers as well (Wu et al., 2011).

The major goal in establishing the KBWPS was to reduce the diffuse and remaining point source pollution originating in the catchment area of the River Zala and arriving at Lake Balaton. The original 1977 plan was to remove nutrients - primarily P - using macrophytes before the water entered Lake Balaton (Kárpáti ed., 1980; Tátrai et al., 2000), in order to prevent further deterioration in water quality, and to achieve the trophic conditions typical of the early 1960s by 2005–2010 (Istvánovics et al., 2007; Láng, 1986).

The construction of the KBWPS was planned in two phases. Phase I (the 18 km^2 Upper Reservoir; Fig. 3-1) was inundated and began operating in 1985. However, instead of becoming a wetland populated by macrophytes (Kárpáti ed., 1980), it became an open lake dominated by algae. The system was designed along the same lines as wetlands for treating sewage waters, and after inundation efforts were made to plant the area. However, wind direction (which is in parallel to the reservoir's main axis), water depth, the suddenness of the inundation, and the stirring up of sediment meant that this effort did not meet with success. Until 1991 the Upper Reservoir removed about 50 % of the external P loads annually. However, retention efficiency decreased to 20-30% after the reduction of the external load upon the commencement of P removal at the Zalaegerszeg wastewater treatment plant, the largest town in the catchment of the River Zala (Somlyódy, 1998). This result can be explained by the increased contribution of internal loading, followed by external load reduction (Clement et al., 1998; Istvánovics & Somlyódy, 1999; Sas, 1990), which in fact again resulted in decreased retention. It is a well-known fact that if the input decreases the retention does not decrease proportionately, but to a much greater degree.

Phase II (the Lower Reservoir, total area 51 km^2, Fig. 3-1) is still under construction. A part of it (16 km^2), which is covered by reed-dominated macrophytes, was flooded at the end of 1992, and the rest is planned to be completed in 2013-2014. Processes in this area are dominated by decomposition. Observations reveal that P retention capacity has been considerably lower here in comparison to Phase I (Somlyódy, 1998). These observations indicate that the operation of both reservoirs differs from the original assumptions, and raises the question of what effect the KBWPS actually has on the water quality of Lake Balaton, and primarily on that of Keszthely Bay. As was mentioned before, in natural systems retention is influenced by many factors, and the uncertainties with regard to the KBWPS are therefore not surprising. In the case of P assimilation we must consider short-term (uptake and release by vegetation, microorganisms, periphyton and detritus) and long-term (deposition in soil and accretion of organic matter) storage (Reddy et al., 1999). Given that abiotic processes controlling P retention are dominated by sediment sorption and chemical precipitation with metal ions; if external P loads are low, wetlands tend to export P instead of retaining it. This is significantly different from wetlands which are constructed as buffers or for the treatment of domestic or industrial waste waters, since in such systems the concentration ranges for nutrients and organic matter are higher by several orders of magnitude (Kovacic et al., 2006; Sheng-Bing et al., 2007). As a result they do not lend themselves readily to comparison with those treating surface waters. Considering the fact that the KBWPS's retention capacity is strongly influenced by inflow concentrations, any changes in the Zala catchment will therefore affect its efficiency (Clement et al., 1998).

For these reasons it is also important to note that, in the years following the inundation of Phase I, Hungarian agriculture and heavy industry were thoroughly restructured following the political and economic changes of 1989/90. In line with the new thinking, high state subsidies were withdrawn from mineral fertilizers, while high fertilizer prices and the unpredictable future of state farms and cooperatives resulted in a drastic reduction in fertilizer usage. Because of the ten-fold increase in fertilizer prices since the beginning of the 1980s, there has been a resulting significant drop in commercial fertilizer use (between 1974-1990 ~10,170 t of P were sold annually in Zala County in form of fertilizer, while between 1991-2000 this value dropped to ~920 t yr^{-1}; HCSO, 2012). As a consequence, the previously applied P quantity (approx. 26 kg/ha/yr) was replaced with extracted P originating from soil P reserves (Sisák, 1993). These changes contributed to the significant reduction in diffuse pollution resulting from agricultural activity. Continuing the load reduction measures, in the late 1990s additional nutrient removal was implemented at all waste water treatment plants in the Balaton region. All of these measurements, however, concern point loads and not diffuse ones - which is not surprising - since the estimation of non-point sources is much more difficult (Brigault & Ruban, 2000).

Monitoring shows that the total load carried by the River Zala has changed dramatically in the past 30 years as a result of the load reduction measures implemented in the watershed. There was an unambiguous increase between the 1960s and 1980s of loads from both diffuse and point sources. The initiation of the Upper Kis-Balaton reservoir (Phase I), the improvement of waste water treatment and the restructuring of the agricultural sector resulted in a load reduction of more than 50% in TP compared to the 1980s. Between 1981 & 1990 an annual average of ~92 t of TP load arrived into the KBWPS, while this value was only ~42 t in the 90s.

After an initial delay, the formerly hypertrophic Basin I of Lake Balaton (Keszthely Bay) showed a fast recovery. The reason behind the delay was that the internal phosphorus loads originating from the sediment were enough to maintain a high algal biomass, and the regeneration of the sediment took several years (Istvánovics & Somlyódy, 2001; Istvánovics et al., 2004). Istvánovics et al. (2007) also found that the long-term behavior of the highly calcareous sediment of Lake Balaton determines the potential maximum of internal P load. In the years when the internal load approached its maximum, there was strong correlation between the biomass of phytoplankton and the estimated concentration of mobile P, which is influenced by the carbonate content of the sediment (Istvánovics et al., 1989). In other years, physical constraints dependent on hydrometeorological conditions might have played a role in keeping the biomass of phytoplankton below the highest possible level.

Consequently, it became evident that – because of this shifted response in lakes (Sas, 1989) –it is sometimes hard to find a direct correlation between external load reduction and water quality improvement, particularly over a short time period. However, it is clear that in the case of Lake Balaton this relation does exist, and load reduction measures have led to changes in the trophic state and the restructuring of the lake’s ecosystem (Istvánovics et al., 2007). The construction of the KBWPS was an important measure - albeit, not the only one taken - to facilitate these oligotrophication processes. However, in addition to its major role in diminishing loads carried by the Zala River by the natural filtration of its area, the KBWPS can also entirely modify river load conditions in its role as a pre-reservoir system. Among others, the formation of particulate P resulting from intensive algae growth in the Upper Reservoir (Phase I) and the reformation of dissolved P in the reed belt (Phase II), contribute to transformation of different nutrient forms and P/N ratios (Istvánovics & Somlyódy, 1999). Furthermore, it should be mentioned that this unique habitat, rich in flora and fauna, is a highly protected nature conservation area, and is under the de-jure protection of the Ramsar Convention (1971). It is also considered a wetland of European importance as well.

As such, the water level management of the system is of great importance. A brief overview of this will now be given. This is a key question because this knowledge will be of great importance in the discussion section. The water level management of the system is also a key question because there are two different habitats with different ecological needs. However, the Water Authority’s policy since the project’s initiation has been to keep the water at a constant level. To achieve this, in Phase I (eutrophic pond) the water’s hydrological retention time was kept at approximately 30 days, while in Phase II (wetland) it was planned to be 60 days. Even in the case of floods or droughts, the water level is artificially modified to keep it as constant as possible.

4. Materials and methods

This section first presents the acquired datasets, and in fine, the detailed description of the applied methodology based on the previously mentioned flowchart actualized for the present research of the KBWPS.

4.1. Acquired dataset and software used

As a first step the most basic characteristics of the dataset should be discussed, meaning, which state of development the analyzed monitoring system is in and what is the intensity of its processes. After the work of Füst & Geiger (2010), three different intensities can be separated: quasi stable-, changing-, and unknown (Fig. 4.1-1a). As for the KBWPS the processes of its input (Z15) can be described as quasi stable (1993-2009; Fig. 4.1-1b) while its sampling is “equidistant” (weekly and daily).

However, over a longer time scale (1985-2009; Fig. 4.1-1c) the sampling site Z15 of the KBWPS could be described as having a “classical” changing intensity, where all three sections are present. It is affected mainly by external influences (discussed in Section 3). In the meanwhile, the output (Z27) has multiple (two) starting-, intensive and consolidated sections, since, being the output, it was affected by the inundation of Phase I (1985) and the inundation of Phase II (1992), as well as the inflow arriving from Z15 and from other canals entering the system in Phase II. Here the starting and intensive sections overlap, because of the rapidity of the changes caused by the measures taken (inundation), so the consolidation section is reached in shorter time (Fig. 4.1-1). After both events (1985 & 1992), the characteristics of the processes suddenly changed.

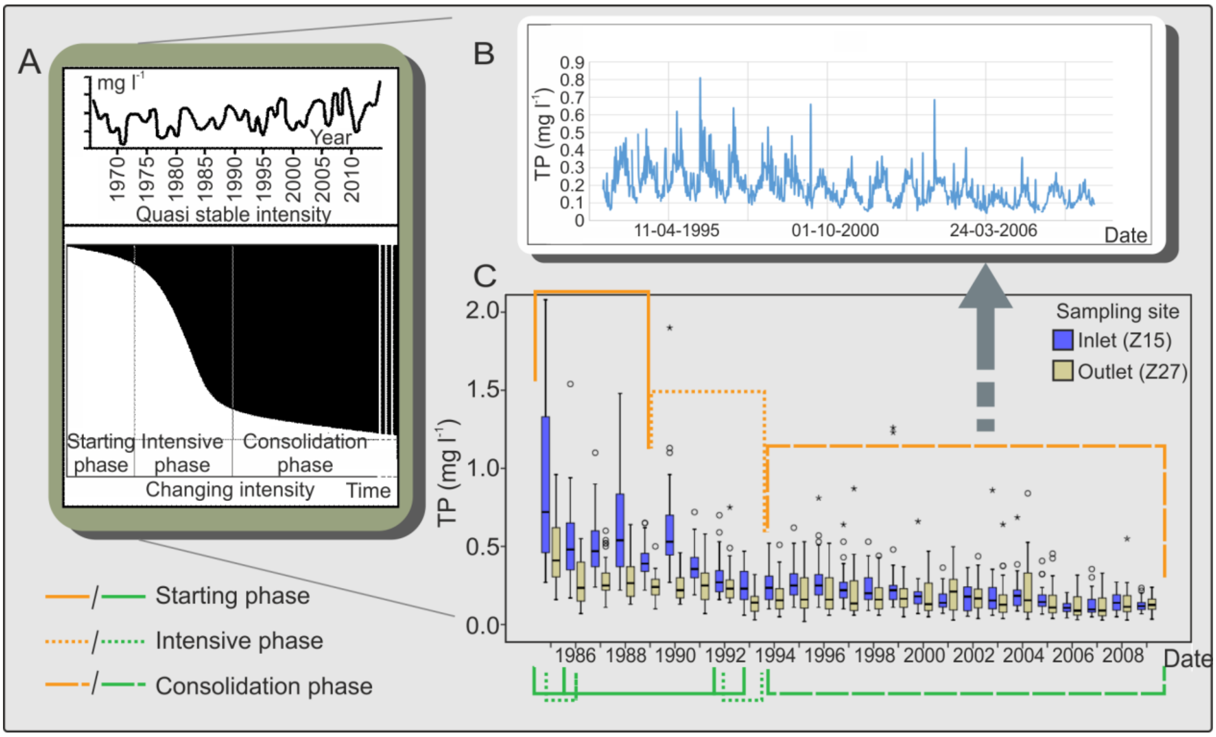


Figure 4.1-1 Characterization of the KBWPS' monitoring network where the system's characteristic TP time series is presented based on the analogy of Füst & Geiger (2010) A) at sampling site Z15 (1993-2009) B) &at sampling sites Z15 & Z27 (1985-2009) C)

In fine it can be stated that I was fortunate. On the one hand, I was able to investigate the time series of the KBWPS as a quasi-stable system, on the other, as one with a changing intensity, giving me the opportunity to investigate the changes caused in the system.

In the course of the research, the time series of a total of 24 weekly- and 3 daily- sampled parameters were used (Table 4.1-1). The weekly ones were sampled at 14 sampling sites in the system, while the daily ones at four sites (Z15, Z11, Kb210 & Z27; Fig. 3-1). Since the sampling began at different time points at the different sites, and furthermore the time interval of the analyses varied between 1977 and 2009.For a better overview the exact values are given in Table 4.1-1.

In addition to the "general" set of parameters, when latent or background effects governing the variability of the response parameters are sought, and in order to identify these mathematically obtained theoretical latent effects with real life phenomena, annual averages of explanatory parameters are necessary when dynamic factor analysis is used. These were: soil P surplus calculated for the watershed of the River Zala; effluent total P concentration of the

Zalaegerszeg waste water treatment plant; water temperature at Z15; annual averages of daily precipitation; relative humidity; and daily runoff (Q, Table 4.1-1).

In order to calculate gross nutrient balances for agricultural fields, manure and fertilizer displacement, crop yield and mineralization effects were taken into account based on county-level statistical data, while WW-P concentrations were obtained from the operator of the Zalaegerszeg waste water treatment plant. Since exact effluent values were not available before the year 2002, the predetermined threshold limits were used, since the source (the operator at the Zalaegerszeg waste water treatment plant) stated that the plant always aimed to keep the P effluent under and/or around the predetermined threshold limit. The explanatory parameters describing the watershed were all valid for the Zala's catchment. Water discharge, representing total runoff, was measured at the same location as the response parameters, whereas the meteorological ones were sampled at the meteorological station of the town of Keszthely, in the vicinity of the river's sampling site.

Of the 14 sampling sites one was located in the River Zala, six in the KBWPS Phase I, six in the KBWPS Phase II, and one at Keszthely Bay (Kszt; Fig. 3-1). Five of the 14 SSs are described in more detail because the system can be viewed as a cascade of "black boxes", with SS Z15, Z11, Z27 & Kszt representing the system inlets and outlets, while Kb210 is of particular interest regarding the periodicity and autocorrelation analyses.

- "Z15", located at the mouth of the River Zala (at the settlement Zalaapáti) in Phase I of the KBWPS, which characterizes the River Zala and its upstream catchment (1528 km^2),
- "Z11", located at the interface between Phase I and Phase II of the KBWPS, which typifies the water generated by Phase I, the eutrophic pond,
- "Kb210", the sampling site that typifies Phase II solely as a wetland without the extra water inputs, which after Kb210 bring an approximately 25-30% increase in nutrient loads through canals arriving from Somogy County, south of Kb210,
- "Z27", located at the outlet of KBWPS (at the settlement Fenékpuszta), which typifies the water output of Phase II, the macrophyte-dominated area of the system, including the extra inputs mentioned above. It represents the entire Zala catchment (2622 km^2),
- "Kszt", represents the SS of Keszthely Bay, where the waters of the River Zala run into and deposit their loads in Lake Balaton.

In the case of Kszt, the period examined was 1981-2009, of which the years 1993-2009 were used for statistical evaluation, and 1981-1992 were analyzed for verification with the same set of parameters (Table 4.1-1).

The majority of the measurements on the KBWPS were conducted weekly (except daily observations of water flow, phosphorus- and nitrogen forms), by the West Transdanubian Water Authority's Kis-Balaton Department. Measurements of macrophyte biomass were not conducted, and it would not have even provided significant information about the total production of the system. All the graphs representing the nutrient loads arriving and leaving the system (input-output analysis) were derived from the daily sampled values (TP, TN, Q). In the meanwhile, at Keszthely Bay the Central Transdanubian Water and Environmental Inspectorate performed this task bi-weekly as a part of the National Water Quality Monitoring System. The acquired dataset contained more than 250,000 records.

The software used for data preparation, and visualization in the current work were MS Excel 2003-2013, Daniel's XL Toolbox and CorelDRAW Graphics Siute X6, while the ones used for the analyses can be found in the last row of Table 4.1-1.

Table 4.1-1 Details of the dataset used with the time periods and parameters used with each method along with the software. The parameters measured in the KBWPS & Keszthely Bay can be found between the dotted lines, while the methods appear beneath the line at the bottom of the table. The parameters under the dotted line were used as explanatory parameters during DFA. ***Because my research began in 2007 and with each year the dataset had to be extended with new measurement data, when a series of results was published (I reached a milestone) I did not therefore run the analyses again, or rather only in cases when it was unavoidable for interpretation. This is the main reason why different time intervals exist in the results section. In 1993 (*) the obtained correlation matrix was singular, therefore it could not have been used for PCA.***

Parameters / methods	Measurement unit	Basic statistics	CA, DA & Wλ	PCA *	PCA for validation	WSA	AA	DFA	Time series analysis - Box & whikers plots	VA	Input-output analysis
Conducted on SSs		Z15; Z11; Z27 & Kszt	All SSs, including Kszt & excluding Z15	All SSs, including Kszt & excluding Z15	Kszt	Z15; Z11; Kb210 & Z27	Z15; Z11; Kb210 & Z27	Z15 and samples of explanatory parameters from the Zala catchment	Z15; Z11; Z27	Z15; Z11; Kb210 & Z27	Z15 & Z27
COD_{ps}	mg l^{-1}	1993-2009	1993-2009 on annual averages	1994-2009	1981-1992	1993-2006	1993-2006	1978-2006 on annual averages			
BOI_5											
Cl^-											
SO_4^{2-}											
HCO_3^-											
CO_3^-											
Mg^{2+}											
Ca^{2+}											
Na^+											
K^+											
NH_4-N											
NO_2-N											
NO_3-N											
TN									1993-2009		
TP											
SRP								1978-2006 on annual averages			
TSS											
Fe^{2+}											
Mn^{2+}											
Org. N											
Dissolved P						1993-2006	1993-2006				
daily TN										1993-2006	
dailiy TP											1977-2009
PP											
Chl-a	mg m^{-3}	1993-2009	1993-2009 on annual averages	1994-2009	1981-1984 & 1985-1992	1993-2006	1993-2006		1983-2009 & 1981-2009 at Kszt		
pH	-										
daily Q	$m^3 s^{-1}$							1978-2006 on annual averages	1993-2009	1993-2006	1977-2009
Soil P surplus	kg $ha^{-1} y^{-1}$										
WW-P	mg l^{-1}										
Water temperature	°C										
Realtive humidity	%										
daily precipitation	mm							1979-2006 on annual averages			
Software used		MS Excel 2013	IBM SPSS Statistics 20	STATISTICA 8		Matlab 2003b	MS Excel 2013	Brodgar 2.7.2, E-DFA	IBM SPSS Statistics 20	Surfer 11	MS Excel 2013

Water quality is a characteristic defined by the physical, chemical and biological content of the water, taken together. In the following paragraphs, I would like to give a brief discussion on the water quality parameters which played an important role in this study concerning the discussed processes of the Kis-Balaton Water Protection System. Of the available parameters (Table 4.1-1), the nutrient forms (phosphorus and nitrogen), chlorophyll-a (Chl-a), total suspended solids, and biological & chemical oxygen demand should be discussed in more detail.

Chl-a and TP are used by the Organization for Economic Co-operation and Development (OECD) to classify water bodies into trophic categories (Vollenweider & Kerekes, 1982). In the course of this practice transparency is considered as well. However, since in Hungary standing waters are mainly shallow and eutrophic it does not provide excess information unlike in deep ("real") lakes. Besides Chl-a and total phosphorus, other forms of nutrients are frequently used for defining the trophic state of the water as well: total nitrogen, ammonium, nitrate, particulate phosphorus, organic reactive phosphorus etc.

In detail, chlorophyll-a provides information about all the primary producers. More specifically, as in the case of the present study, it corresponded to the algal biomass in the water, thus providing a basis for comprehensive comparisons between the different habitats and ecological states in the KBWPS.

The nutrient forms in general (phosphorus and nitrogen) are limiting factors, since they are either present in small quantities, or require a vast amount energy to retrieve. If their quantities increase – in forms which are accessible – the biomass of algae increase and the amount of nutrients decrease (Wetzel, 2001).

Total suspended solids include all organic (with algae as well) and inorganic particles suspended in the water. From an analytical perspective it equals the dry weight of sediment from a known volume of water filtered through a 0.45 μm filter. It mainly provides information about the level of algae growth, and therefore about the level of eutrophication in standing waters, whilst at the same time shedding light on turbidity as well. If turbidity increases mainly due to eutrophication, diversity in the water body decreases. Besides these phenomena, dissolved oxygen decreases along with the level of photosynthesis. Naturally, as less oxygen is produced by plants and algae, there is a further drop in dissolved oxygen levels.

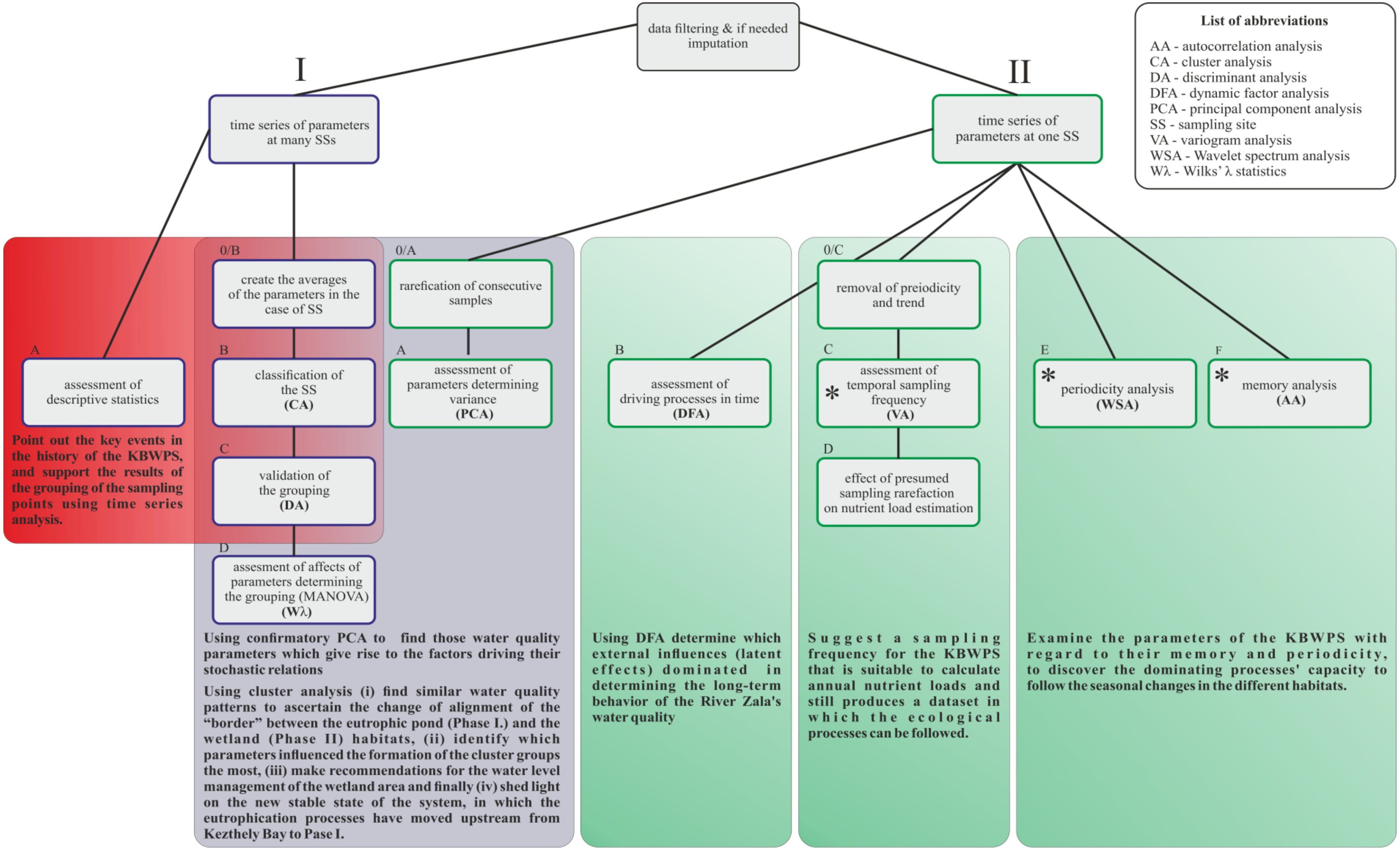
In fine oxygen demand, be it biological or chemical, indirectly indicates the amount of oxygen needed to decompose all organic matter by microorganisms during a certain period of time, and the amount of oxygen consumed by a certain quantity of chemical solution needed for oxidation respectively. The higher the ratio of biological and chemical oxygen demand is, the more biologically decomposable matter is present in the water.

4.2. Framework of the study, applied methods and detailed aims

The previously listed methods cover a wide range of general applicability; however, there are cases when there is either no need to apply them, or only certain ones were applicable to the datasets of the KBWPS. In the following, only those methods will be discussed which were used during the thesis work.

This chapter is intended to introduce a framework, a proposed methodology of those exploratory data analysis techniques which were applied to the study of the spatio-temporal relationship between the sampling sites and parameters measured on the KBWPS based on a modified flowchart after the work of Hatvani et al. (2014a; Fig. 4.2-1). In that particular paper the original framework, which included a number of other techniques/approaches (some of which were not applicable in the present case), was the subject of a survey, in which scientists working with water quality data at governmental institutes or research facilities were asked whether the proposed methodology would be useful or not in the course of their work. Based on the 57 replies received from all over Europe the answer was: “51% said that it would be certainly useful to “provide employees dealing with water quality data with a flowchart which determines step-by-step which methods should be used in sequence” and 42% said that “yes, it would probably be useful”, while only 3.5% & 3.5% thought that it would be either “probably not useful” or “not useful at all” (Hatvani et al., 2014a).

All of these exploratory data analysis methods are useful and important tools for obtaining an overview of systems which can be described by many different parameters, for determining the latent and explicit connections between the parameters, and for sorting and grouping the obtained data on a mathematical basis.



Point out the key events in the history of the KBWPS, and support the results of the grouping of the sampling points using time series analysis.

Using confirmatory PCA to find those water quality parameters which give rise to the factors driving their stochastic relations

Using cluster analysis (i) find similar water quality patterns to ascertain the change of alignment of the "border" between the eutrophic pond (Phase I.) and the wetland (Phase II) habitats, (ii) identify which parameters influenced the formation of the cluster groups the most, (iii) make recommendations for the water level management of the wetland area and finally (iv) shed light on the new stable state of the system, in which the eutrophication processes have moved upstream from Kezthely Bay to Pase I.

Using DFA determine which external influences (latent effects) dominated in determining the long-term behavior of the River Zala's water quality

Suggest a sampling frequency for the KBWPS that is suitable to calculate annual nutrient loads and still produces a dataset in which the ecological processes can be followed.

Examine the parameters of the KBWPS with regard to their memory and periodicity, to discover the dominating processes' capacity to follow the seasonal changes in the different habitats.

Figure 4.2-1 Flowchart of the applied methodology and the detailed aims of the research. Here the methods were placed in trees I or II based on the characteristics of their input data. The same color rectangles indicate that the methods within them were used together for the same purpose. The results of those methods which are marked with an asterisk (*) were later on compared at the different SSs this way covering all the four dimensions. The tree codes will be used in the text to help the reader to better follow the methodology (based on Hatvani et al., 2014a)

4.2.1. Data filtering and general preparation

Accurate results can only be expected from multivariate methods if the datasets used contain the desired information that describes the investigated processes precisely, and the amount of data obtained is sufficient. Determining the amount of information is the duty and responsibility of the given discipline. As an important requirement the number of observations must be greater than the number of analyzed parameters [*n*>>*p*] (Füstös et al., 1986; Stodden, 2006). Environmental data often fails to meet this criterion, so having it present in the dataset takes us closer to statistical stability.

The first step during my research was data preparation, where all the mistyped, outlying and extreme values were handled. In the case of ions the procedure could have been done based on the charge balance error; this however, this is not suitable for non-ionic parameters. Therefore in the case of both ionic and non-ionic ones some sort of filtering based on statistics (e.g. discarding all the data which falls outside the average ±3 times the interval of the standard deviation; box-and-whiskers plots etc.) or manual data filtering based on professional experience could be used. Because of the characteristics of the system: (1) changing external and internal loads, (2) new areas being inundated, it was not suggested that any type of statistics be used, because in this case non-outlying values may be considered as outlying ones. *Therefore in the case of all the used data series filtering was performed manually.* I have checked each and every datum and verified whether it is simply outlying - because of a flood or "act of God" -, mistyped, or simply wrongly measured. *During this procedure the variability of a certain parameter and the fact that it may have been the linear combination of one or more others was of great help. In each case when I had doubts as to whether it was a true value or not, my external advisor provided me with the historical booklets where the data was stored on paper, helping to verify typos.*

The above-mentioned steps actions were general ones; all the specific transformations and data preparation steps will be discussed at the beginning of each corresponding section.

4.2.2. Imputation

After the data was filtered the second step was to ensure that there should be no missing data. Two of the most common solutions for dealing with missing data is linear or multiple regression. These can be applied if there is a strong linear relationship between the data of different parameters or sampling sites as well. If they are applied it is important to verify the accuracy /quality of the estimation by comparison with real data.

However, in my case - besides the need for a dataset with no missing data - equidistant sampling was a requirement as well. The problem was that the available data were supposed to be sampled weekly, but if the sampling day fell on a holiday, then 8 or 6 days passed between the two samples instead of 7. Another problem occurred when a sample was skipped. These sampling dates were either left out of the analyses or interpolated, because certain methods - such as periodicity (*tree II.D in Fig. 4.2-1*)[3] - would not work without an equidistant and continuous time series.

The solution was to reconstruct the original data, with a cubic spline interpolation (Fig. 4.2.2-1) and resample on every 7th day. It can be seen clearly that the interpolation's margin of error is slight. The data from 15.12.2003 - 05.01.2004 are missing, and only the interpolated data can be seen in the gap (Fig. 4.2.2-1).

As a result an equidistant database complying with the conditions of periodicity analysis was obtained. If none of the previously described methods had worked, there are many in the literature (Cook & Rubin, 2007; Schneider, 2001; Yuan, 2010); however, a thorough discussion of these is not the aim of the study.

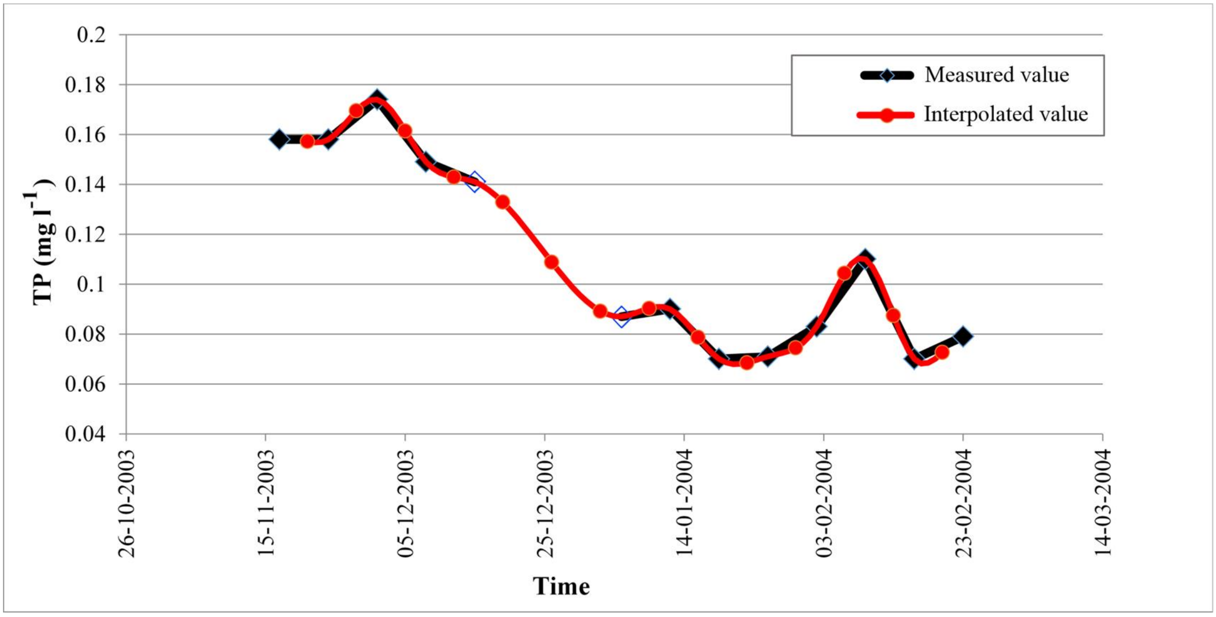


Figure 4.2.2-1. Spline interpolation on the TP time series, with data missing between 15.12.2003 - 05.01.2004 (empty rectangles mark the beginning and the end of the missing segment)

---

[3] Roman numbers combined with letters in parenthesis are used to help the reader to better follow which "stage" is the study at, with respect to whole methodology (Fig. 4.2-1).

Every criterion can only be met if the data matrix is checked for these and other kinds of errors before analysis. This was the most annoying and time-consuming part of the research. However, skipping this step would have inevitably lead to incorrect results and conclusions.

4.2.3. Methods used on time series of parameters measured at many SSs (*I*)

In the case of tree No. I (Fig. 4.2-1) the data at hand – as discussed previously (Section 2) - can be described in four dimensions. Therefore compression of the information (in most of the cases, time) into statistics was needed (*tree I.0/B in Fig. 4.2-1*), because it is complicated to assess time and space simultaneously. However, before this step was taken the descriptive statistics (*tree I.A in Fig. 4.2-1*) were analyzed. If descriptive statistics are obtained, it is highly useful - especially during discussion of the results - to visualize them on box-and-whiskers plots, in which multiple statistics can be studied simultaneously. Here the boxes show the interquartile range and the black line in the box is the median. Two upright lines represent the data within the 1.5 interquartile range. The data between 1.5 and 3 times the interquartile range are indicated with circles (outliers), and those with values higher than 3 times the interquartile range are considered to be extreme values indicated with an asterisk (Norušis, 1993).

Cluster analysis – tree I.0/B & I.B in Fig. 4.2-1

After the analysis of the descriptive statistics the question of similarity of sampling sites was raised. To obtain an answer, annual averages were formed first (*tree I.0/B in Fig. 4.2-1*) to compress the temporal dimension, thus making the dataset compatible with CA (*tree I.B in Fig. 4.2-1*).

Clustering is a kind of coding, in which a certain sampling site - originally described with many parameters (Chl-a, COD_{ps} etc.) - is described with only one value, its group code (cluster number). It is important to note that during clustering not the parameters but the sampling sites are contracted by classifying a set of observations into two or more mutually exclusive “unknown” groups by combining interval variables. The purpose of cluster analysis is to find a system of organizing observations, usually sampling sites, into groups, where members of the groups share properties in common (Stockburger, 2001). The main aim is to settle the similar sampling sites into groups based on similarity, which is measured by assigning a distance (metrics) to each sampling site placed in an *N* dimensional space. On the one hand, if the distance between two sampling sites is small, then they are highly similar to each other. If the distance is zero they are perfectly similar. From this, it should be clear that choosing the right distance between sampling sites is a key question. On the other hand, the groups have to differ to the highest extent possible. To ensure this, another distance is introduced into sets of sampling sites. One has to find the grouping consisting of sets of the largest possible distances.

There are basically five types of clustering methods: partitioning (e.g. K-means), hierarchical, density-based, grid-based and model-based methods (e.g. Self-Organizing Maps; Han et al., 2012). In this study divisive hierarchical cluster analysis was applied to Z-scored (standardized) annual averages using Ward's method, along with squared Euclidian distance. Standardization is vital to minimize the influence of different variances of variables and eliminate the effect of different measurements units (scaling). Furthermore, it renders the data dimensionless (Kazi et al., 2009). The sampling sites were clustered for each year (1993-2009; i.e. 17 dendrograms produced) based on the annual averages of 19 parameters (Table 4.1-1). The input dataset was prepared so that the sampling sites (to be clustered) corresponded to each row and the annual averages of the parameters to each column. This structure was repeated for each and every year. In this particular study each case starts in a separate cluster and joins up to the other clusters as the linkage distance grows, and only one cluster remains (Day & Edelsbrunner, 1984).

Discriminant analysis and Wilks' λ statistics - tree I.C & I.D in Fig. 4.2-1

The existence of the previously obtained groups remains only a hypothesis until these are validated. In order to validate the grouping (*tree I.B in Fig. 4.2-1*), besides the "elementary" univariate hypothesis testing tools - which test whether the parameters of the groups are similar or not - discriminant analysis can also be used (*tree I.C in Fig. 4.2-1*). Discriminant analysis unlike "elementary" hypothesis testing tools takes into account all of the parameters at once. Its results indicate to what extent (%) the planes separating the groups can be distinguished by building a predictive model for group membership. The model is composed of a discriminant function based on linear combinations of the predictor variables (Johnson & Wichern, 1992). The functions are generated from a sample of cases for which the group membership is known; the functions can then be applied to new cases that have measurements for the predictor variables but their group membership is as yet unknown (Landau & Everitt, 2004). The result of discriminant analysis is often visualized on a surface stretched between the first two discriminating functions (function 1 & function 2; Ketskeméty & Izsó, 2005).

After the grouping was validated, the parameters determining the cluster formations were sought for on an annual basis using (*tree I.D in Fig. 4.2-1*) Wilks' lambda statistics (Wilks, 1932). Here a Wilks' λ quotient is assigned to each parameter, where the quotient is:

$$\lambda = \frac{\sum_i \sum_j \left(x_{ij} - \bar{x}_i\right)^2}{\sum_i \sum_j \left(x_{ij} - \bar{x}\right)^2} \qquad (1)$$

Where x_{ij} is the j^{th} element of group i, $\bar{x}_i$ the average of group i, and $\bar{x}$ the total group's average. 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. The smaller the quotient is, the more it determines the formation of the cluster groups (Afifi, 2004).

4.2.4. Methods used on time series of parameters measured at one SSs (*II*)

Principal component analysis – tree II.0/A & II.A in Fig. 4.2-1

After grouping the observations with the same behavior, variables were sought which would explain most of the variance. Despite the fact that ANOVA "is robust enough that it can be used in the case of moderate departures from normality" (Kuzma & Bohnenblust, 2005), the non-normal distribution of the data does rule out the application of certain formal tests. However in the case of PCA these and similar problems do not matter, since normality is not an issue (Rencher, 1995). Therefore confirmatory PCA was applied to the data (Rogerson, 2001). Here it is worth remarking that while CA works with various distances, PCA works with a correlation matrix, so its results reflect the stochastic interdependences. Confirmatory PCA is a technique usually used in the advanced stages of the research process to test a theory concerning latent processes (Tabachnick & Fidell, 1996). The measured chemical and physical parameters - the input of PCA - are correlated, whereas the hypothetical variables - called principal components - are uncorrelated and are obtained as a linear combination of the original parameters. PCA decomposes the original variables into principal components that explain the original total variance of the dataset in a component-wise monotonically decreasing order. The correlation coefficients between the original parameters and the principal components are the factor loadings. They explain the weights of the principal components in the original parameters. They do not, however, give an exact answer as to whether a weight has to be considered as significant or not. Nor do they answer the question of how many principal components are important. In the course of PCA analyses, weekly sampled summer data were explored. Van Straten & Herodek (1982) found that in areas of Lake Balaton with different trophic conditions, at temperatures below 10°C the activity of "cold"-water organisms becomes significant, while above 15°C that of the "warm"-water types. On this basis, data collected when the corresponding water temperature was higher than 15°C were considered as summer data.

In the course of PCA the cases must be independent, in my case, it means temporarily independent observations. For this reason the weekly data had to be rarefied to bi-weekly before they were used in the analysis (*tree II.0/A in Fig. 4.2-1*). At Z15 and Z27, for example, based on semivariogram analysis, the temporal-dependence of nutrients, TP in particular, resulted in a temporal range of more than 5 days (for details see Section 5.2.3, or Kovács et al., 2012b). It should be noted that TP is a parameter with a high degree of variability based on its CV (for details, see Table 5.1.1-1). Therefore, if inorganic parameters - which fluctuate to a lesser degree (again, see Table 5.1.1-1) - are also the subject of analysis, a seven day sampling could result in high autocorrelation in the measured parameters. Furthermore, after two weeks the mean rate of autocorrelation of the parameters drops beneath $r^2 < 35\%$ at one lag (for details see Section 5.2.4/*Periodicity and memory analyses*). In conjunction with the variogram results, this fact reinforced the decision in favor of bi-weekly rarefication.

Summer data, the subject to PCA, were fully able to describe the driving forces of the changes in the growing season. In my research, factor loadings were the characteristics which provided the most useful information on the behavior of the phenomena observed. Only factor loadings outside the ±0.7 interval were considered as influential. Turning to principal components, influence was accepted over 20% explanatory power in the case of the first, and over 15% in case of the second component.

During PCA a correlation matrix was formed from the bi-weekly sampled data of the same set of parameters as in the case of cluster analysis at every sampling site (except Z15; Table 4.1-1). After the correlation matrices were obtained the principal components were computed for each year, except for 1993 where the correlation matrix was singular.

Dynamic factor analysis – tree II.B in Fig. 4.2-1

According to the goals set in Fig. 4.2-1/*Dynamic factor analysis*, the identification of latent effects requires the application of a certain kind of factor analysis modeling. However, in view of the time series' structure - meaning that consecutive annual averages were temporally dependent on each other - factor analysis modeling (PCA for example) is not suitable in its conventional form. Dynamic factor analysis (DFA; *tree II.B in Fig. 4.2-1*) is the proper method to find background factors and in the meanwhile take into account the lagged correlation structure (Geweke, 1977).

DFA modeling is consolidated in the literature into state space representation with common trends. Two computational realizations of DFA models were applied during my thesis work, E-DFA & Brodgar. The difference in the two realizations is the prescribed structure of the common trends, which is - in the present case - a first order autoregressive process in one instance, and a random walk in the other.

The model using autoregressive structure for factors was developed and first published in Ziermann & Michaletzky (1995), and is called E-DFA for short in the present paper. A program written in FORTRAN was used for the actual calculations. The other model, using a random walk structure, is based on Zuur et al. (2003a; 2003b), and a commercial program called Brodgar[4] is available for computing the factor solution. It was used for the comparison of the main results of the study. The concept of a solution in E-DFA is based on minimizing the static and dynamic error terms, which is errors in state estimation and prediction. This is in contrast to Brodgar, in which the likelihood value is maximized via an algorithm. Turning to the details of E-DFA, the basic factor model equation is:

$$Y(t) = \mathbf{A} \cdot F(t) + \varepsilon(t) . \tag{2}$$

In this equation (2) the observations (response variables) and the factors have a time series structure. The factors $F(t) = (F_1(t),\dots,F_M(t))'$, $0 \le t \le T$, are specified to be uncorrelated autoregressive processes of the first order in the present application:

$$F_j(t) = c_{j,0} + c_{j,1} \cdot F_j(t-1) + \delta_j(t) \tag{3}$$

where $\delta(t) = (\delta_1(t),\dots,\delta_M(t))'$ stands for a Gaussian white noise independent from $\varepsilon(t)$. A linear transformation expressed through the non-random and non-time dependent $\mathbf{N} \times \mathbf{M}$ matrix **A** creates the observations $Y(t) = (Y_1(t),\dots,Y_N(t))'$ from the factors *F(t)* and an additive white noise $\varepsilon(t) = (\varepsilon_1(t),\dots,\varepsilon_N(t))'$. Within this framework two types of errors are considered. One of them is the prediction error of the factors themselves according to the autoregressive structure (3),

---

[4] www.brodgar.com

while the other is the state estimation error obtained as the squared sum of the differences between observations and their prediction from the factors. For a detailed description see Márkus et al. (1999). The simultaneous minimization of these errors leads to the factor solution of (3). Obtaining the exact solution leads to a very complicated optimization problem on Stiefeld Manifolds that is theoretically intractable except in very low dimensional cases (Rapcsák, 2002). Therefore algorithmic solutions are needed. A zig-zag algorithm developed in Ziermann & Michaletzky (1995), also described in Márkus et al. (1999), leads to the desired factor solution; thus the dynamic factors are obtained.

As an alternative to this procedure, two DFA methods provided by Brodgar were used for the sake of comparison with the E-DFA in order to evaluate the model's adequacy. Of these, the M common trends + noise applied using an ***H*** non-diagonal matrix gave the better-comparable results. It is important to notice here that Brodgar gives a smoother factor solution, and that might be caused by the random walk specification, which enforces the use of less variable-generating noise. In our view, the autoregressive specification follows the fluctuations more closely.

In practice, the steps of the analysis were the following:

(i) as an initial step, DFA was applied to the time series (annual averages) of the response variables, resulting in dynamic factor time series;

(ii) then the factor loadings of the DF time series were assessed. Those loadings were considered as significant which fell outside the ±0.6 interval;

(iii) and, as a last step, the DF time series were correlated with the explanatory variables' time series to find the governing background processes.

Since the explanatory variables (Table 4.1-1) were only available from the vicinity of sampling site Z15 (representing the River Zala), the possibility of finding the background processes in the system was restricted to this particular site, and that for the time interval of 1978-2006. The model was computed from annual averages of 21 weekly sampled response parameters measured in the River Zala (Z15) and 6 explanatory parameters (for details see Table 4.1-1).

Sampling frequency analysis – tree II.0/C-D in Fig. 4.2-1

Temporal sampling frequency analysis is one of the most important analyses that has to be performed during any kind of scientific research. In my case, the first step in preparing the daily datasets (*tree II.0/C in Fig. 4.2-1*) for temporal sampling frequency estimation was their division into summer and winter data for each year, as discussed previously (*tree II.A in Fig. 4.2-1*; Kovács et al., 2012b). The only amendment to this would be that in between the summer and winter data, where the corresponding temperature was between 10°C and 15°C, data were considered as a period of transition regarding the ecosystem, and were therefore discarded.

According to the additive model (Shumway & Stoffer, 2000), these time series (X_t) are considered to be the sum of trend (T_t), period (P_t) and the irregular component (I_t):

$$X_t = T_t + P_t + I_t \tag{4}$$

This statement implies that to be able to determine the ideal sampling frequency for nutrient load estimation and other analyses, trend (if it exists) has to be subtracted from the time series and periodicity has to be removed as well. By this procedure, the analyzed time series gets closer to stationarity, necessary for sampling frequency estimation. It should be remarked here that other forms of non-stationarity (in variance and distribution) are not significant for the dataset at hand. This subtraction could have been performed by removing an annual average period where the period length is constant; however, this was not satisfactory, because the length could vary or eventually disappear from the process for a certain time period (Kovács et al., 2010; Section 5.2.4/*Periodicity and memory analyses*). Had the latter been the case, I could have created periodic behavior artificially (Kovács, 2007; Slutsky, 1937). Another approach could have been the removal of polynomial trends (Chan et al., 1977), but this was not satisfactory either, because the trend did not have a polynomial character.

The Locally weighted scatterplot smoothing (LOESS) procedure, however, could overcome the aforementioned difficulties, and thus was the one used. It is a local polynomial regression fitting. The method is described in detail in two publications by Cleveland (1979, 1988), and the program used here was developed by Jon Peltier in 2011 at Peltier Technical Services Inc.

I feel it is necessary to explain one of its most important characteristics which allows the smoothing to retain the original N points, unlike e.g. a k point moving average, which would result in a time series k points shorter than the original one.

LOESS, for each x value where a smoothed y value is calculated, a local linear regression is performed on the N points closest to the target x value. These N points can fall before or after the target x value, and are not necessarily evenly distributed, either before or after. So for the first x value, no points are before it, and the rest are on it or after it. The regression is weighted, so that among the N points in the local regression, those points closest to the target x value are given more importance. The smoothed y value is calculated from this regression. Then the analysis moves to the next target x value, using the N closest points. For a while the same points are used (with different weightings) in the moving regression. What can and does happen is that the smoothed y values calculated near the ends of the data set result in "strange" values, because so many of the N points are on one side of the calculations. This produces straighter curves near the ends than might be expected. Many researchers have tried to adjust for this by using a second order moving regression, but with the weighting, this becomes complicated (Peltier, 2013).

An example of trend removal can be seen in Fig. 4.2.4-1. The residual obtained was suitable for the requirements of the sampling frequency estimation method, the semivariogram.

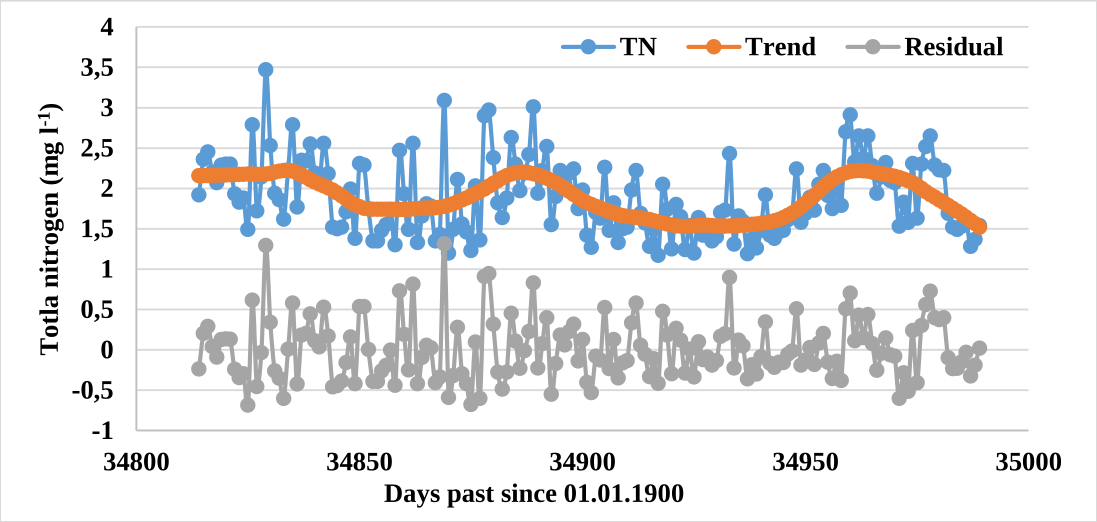


Figure 4.2.4-1 Example of the use of LOESS in the removal of the trend from TN time series (summer of 1995) at Z27, revealing the residual

Returning to periodicity, because the annual time series were split due to the separately analyzed winter and summer data only half-length or even shorter periods (if they existed) were the subject of variogram analysis. There are many functions which can be applied to describe spatial or temporal dependence. Of these the variogram - which is most commonly applied in geostatistics - was used. The variogram (*tree II.C in Fig. 4.2-1*) can be described mathematically as follows (Füst, 2004; Molnár et al., 2010; Molnár & Füst, 2002): Let $Z(x)$ and $Z(x+h)$ be the values of a parameter sampled at distance $|h|$ from each other. The length of h is measured in time or space for temporal or spatial processes respectively. The variance of the difference of $Z(x)$ and $Z(x+h)$ is

$$D^2[Z(x)-Z(x+h)] = D^2[Z(x)]+D^2[Z(x+h)]-2COV[Z(x),Z(x+h)] \quad (5)$$

In the case of samples taken from the same population (stationarity) we could assume that

$$D^2[Z(x)]=D^2[Z(x+h)] \quad (6)$$

so that

$$D^2[Z(x)-Z(x+h)] = 2D^2[Z(x)]-2COV[Z(x),Z(x+h)] = 2\gamma(h). \quad (7)$$

The function $2\gamma(h)$ is called the parameter’s variogram, while γ(h) is its semivariogram. If we introduce the simplified notation $D^2[Z(x)] = D^2(x)$, and $COV[Z(x),Z(x+h)] = g(h)$, then $\gamma(h) = D^2(x)-g(h)$. The semivariogram could be calculated by the Matheron algorithm (Matheron, 1965):

$$\gamma(h)=\frac{1}{2N(h)}\sum_{i=1}^{N(h)}\left[Z(x_i)-Z(x_{i+h})\right]^2 \quad (8)$$

where N(h) is the number of pairs within the lag interval h.
In practice,

$$Z(x_i) \geq 0 \ (i=1,2,\ldots,n) \text{ and } D^2[Z(x)] \geq g(h) \geq 0 \quad (9)$$

so that theoretically the semivariogram can only take values from the

$$0 \leq \gamma(h) \leq D^2[Z(x)] \quad (10)$$

range.

The most important properties of the function are as follows (Fig. 4.2.4-2):

(i) When the function does not start from the origin of the coordinates ($C_0>0$), this is called the “nugget effect”. The value C_0 of the semivariogram at the origin withholds information regarding the error of the sampling.

(ii) If the semivariogram does not have a rising part, the empirical semivariogram’s points will align above the abscissa parallel to it. A semivariogram like this is called of a nugget effect type. In this case no range can be estimated, i.e. the sampling frequency is insufficient.

(iii) The level at which the variogram stabilizes is the sill ($C+C_0$); it is equal to the variance for stationary processes. The value C itself is the reduced sill.

(iv) Range is the distance within which the samples still have an influence on each other (Webster and Oliver, 2007).

If $\gamma(h)$ is an increasing function (if $h \to \infty$ then $\gamma(h) \to \infty$), the parameter is non-stationary.

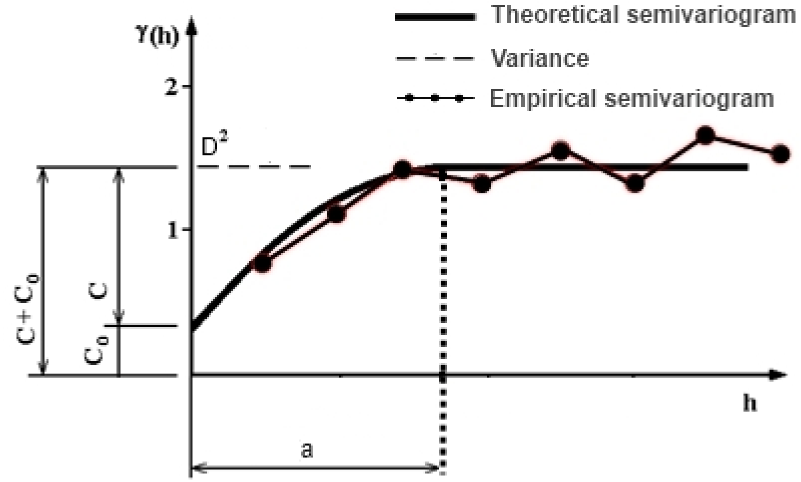


Figure 4.2.4-2 Properties of the semivariogram, where "a" stands for range, "C" for the reduced sill and "C_0" for the nugget effect, "h" for lag distance and "D^2" for the variance (based on Füst & Geiger, 2010)

Empirical semivariograms can be approximated by many theoretical functions (Füst, 2004; Molnár et al., 2010; Molnár & Füst, 2002; Webster & Oliver, 2007). However, a discussion of these is not an aim of the present study. The estimation of sampling frequency using variograms is based on the fact that in space samples outside the variogram range (be it temporal or spatial) are uncorrelated (Chilès & Delfiner, 2012, Chapter 2.2.7/Range and sill, p. 49). This idea gave the basis for sampling frequency estimation. The hypothesis is that if the sample is taken outside the range, inferences can only be made on quantities depending on the marginal distribution of the process, the interdependence structure remains undetectable. Therefore, in order to be able to describe the processes, samples should be taken inside the spatial, or - as in my case - temporal range. The rarefaction within the range should be chosen so as to preserve a substantial part of linear interdependencies in the sample. The choice may depend on external circumstances as well (source of funds, personnel, need for preserving interdependence etc.).

In practice the steps of the analysis conducted on daily measured total phosphorus and total nitrogen at the four cardinal sampling sites (Z15, Z11, Kb210 and Z27) in the KBWPS for 1993-2006 were the following:

(i) the time series of the nutrients were split into summer and winter data,

(ii) trend was subtracted from the summer and winter time series using LOESS (e.g. Fig. 4.2.4-1),

(iii) empirical semivariorams were computed for each summer and each winter separately,

(iv) the average temporal range was determined for all the sampling sites and parameters in question.

Since one of the primary purposes of the KBWPS is nutrient retention, it was a logical step to use annual nutrient load values for validation (*tree II.D in Fig. 4.2-1*). The initial idea was to consider the already-available daily sampling to be the most accurate measurement possible. The question was therefore raised of how a difference alters when an annual nutrient load calculated from the original daily sampling and from a rarefied sample is filtered to every 3-7th, 10th and 14th day.

The annual nutrient load of nutrients such as phosphorus or nitrogen (aNL; in tons) is calculated as follows:

$$\text{aNL} = \sum_{i=1}^{n} \frac{C_i * Q_i * 60 * 60 * 24}{10^6} \qquad (11)^5$$

Where n is the number of samples (=365 or 366) of the water quality parameter in question (in this study daily measured phosphorus and nitrogen), C_i in mg l^{-1} is the unit of concentration (again for the water quality parameter), and Q_i is the runoff in m^3 s^{-1}. Definition 11 is based on general practice and Stenback et al. (2011).

Using the equation above the annual nutrient loads (aNLs) were obtained for both TP and TN parameters from each sampling site for the daily and all the rarefied samples for each year. The nutrient loads calculated from the daily and the rarefied samples were then compared for each year as follows:

$$ARD = \left|\overline{\frac{aNL_r}{aNL} - 1}\right| * 100\ \% \qquad (12)$$

---

[5] The same nutrient load estimation was used throughout the whole study.

Where *r* is the degree of rarefication of the original daily sampling, aNL is the value of the nutrient loads. Therefore aNL_r represents the annual nutrient load values rarefied to a certain degree. For example aNL_3 means that the annual nutrient loads were calculated from a dataset which has been rarefied from daily to a three day sampling frequency. It thus represents an annual nutrient load as it originated from a three day temporal sampling frequency. These values were called annual relative differences (ARD). These provide a picture of how a presumed rarefaction of the sampling frequency would affect the accuracy of nutrient load estimation in percentages at a certain sampling site in a certain year.

The main goal here was to obtain information on the reliability of the nutrient load estimations, if the sampling frequency is to be changed, in particular, the occurrence of large fluctuations as a result of the rarefaction. A usual approach is the peak over threshold (POT) method (i.e. Leadbetter (1991) in hydrology), when values exceeding a certain level are statistically analyzed. The problem in the present case is that a sufficiently large sample is necessary for a proper estimation of the generalized Pareto distribution characterizing the peak over threshold model, but unfortunately the current sample size is insufficient for that. However, continuing along this line, it was possible to compare (i) the number years when the annual relative differences at a certain degree of rarefication exceeded the 5% level and (ii) the averages of those annual relative differences (ARDs) at the given levels of rarefication.

Because of the limited number of samples, a low threshold limit of 5% (roughly above the 0.8 - 0.5 quantile, depending on the sampling site and degree of rarefaction) was chosen, in accordance with Raisin et al. (1997).

Periodicity and memory analyses – tree II.E & II.F in Fig. 4.2-1

To achieve the aims determined in Fig. 4.2.-1/*tree II.E&F*, two methods, namely wavelet spectrum analysis and autocorrelation analysis (AA) were used to detect the periodicity and the memory (autocorrelation) of the processes evolving in the KBWPS.

During the analysis of natural processes it frequently occurs that the data were not registered equidistantly (Section 4.2.2). Therefore “classical” spectrum analysis methods cannot be used. In such cases the application of Lomb-Scargle type of periodogram (Lomb, 1976) may be a solution. Still, even using this method the question remains of whether the period sought for is significantly present in the whole time period, and if not, when it is present, and when it is not (as an example, see Figure A1 in the Appendices Section). On the contrary the Morlet wavelet transformation, unlike the standard Fourier transformation, permits of time-frequency mapping. The wavelet spectrum analysis (WSA) is an increasingly discussed and very accurate method for determining periodicity in a database (Dianmo & Zhongwei, 2000). The Fourier transformation decouples the examined data to trigonometric (sine, cosine) curves, thus rendering it inadequate for the estimation of a power spectral density (PSD). However, the Wavelet transformation overcomes the non-stationary problem with its adaptability, by fragmenting the data into short “wavelets” instead of long sine waves. Wavelet transformation W_n(s) could be defined as the convolution of the data and the wavelet function (13):

$$W_n(s) = \sum_{n'=0}^{n} X_{n'} \psi^{*}\left[\frac{(n'-n)\delta t}{s}\right] \tag{13}$$

Whereby the asterisk (*) represents the complex conjugate, ‘x_n’ the original data stream, ‘s’ the scale, ‘ψ’ the wavelet function, and ‘δ’ the degree of the resolution.

Its adaptability lies in the scaling method. The mother wavelet provides a source function to generate daughter wavelets, by scaling and transforming it. The purpose of the wavelet transformation is multiple dissociation; by decomposing the data in the scaling space. In this way it is possible to reveal its self-similarity structure, and as a result the data’s periodicity can be examined.

To estimate a power spectral density (PSD) graph, a special wavelet was used. In this particular case the “Morlet” (Fig. 4.2.4.-3) has been applied to examine the weekly measured data (Benedetto & Frazer (Eds.), 1994; Vidakovic, 1999).

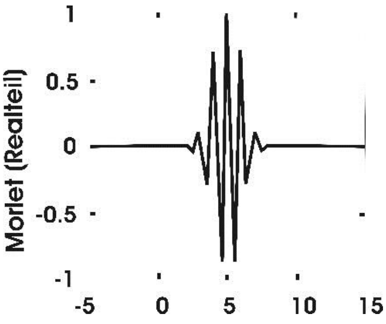


Figure 4.2.4-3 Graph of the mother Morlet wavelet

One could have argued that, as weekly data was at hand, on what basis was I using it for periodicity analysis? As a matter of fact, as the annual period was sought, and therefore, based on the theory of Shannon (1949), only four observations per year would have been enough to obtain a representative result; in my case I had approx. 50-51. With regard to the question of equidistance, as discussed in Section 4.2.2, the weekly sampled data were resampled for every 14 days using a cubic spline interpolation.

In practice a power spectral density graph was created for each and every weekly sampled parameter (see detailed list in Table 4.1-1) for the time interval 1993-2006 for all four cardinal sampling sites (Z15, Z11, Kb210 and Z27). It resulted in 88 power spectral density graphs (e.g. Fig. 4.2.4-4), 22 at each sampling site, one for each parameter covering the whole time period investigated. After these were obtained, their results were transformed in order to be able to make a visual evaluation (as an example, see Fig. 5.2.4-1. In those particular graphs the lines were the ones indicating the absence of, and the gaps the presence of the periods from the examined years. The length of the lines was converted into time, so I was able to summarize the non-periodic time segments of a parameter, and calculate this as a percentage of the total time at a certain sampling site for the time interval 1993-2006.

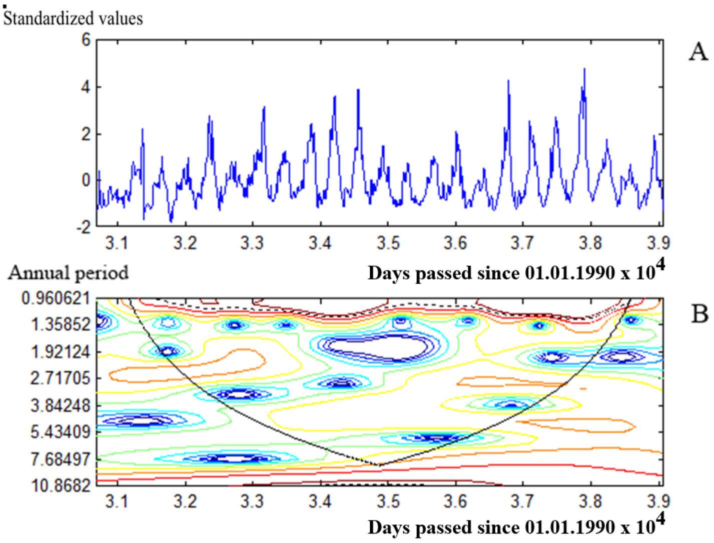


Figure 4.2.4-4 Output of the WSA: normalized original A) and PSD B) of COD$_{ps}$ at Z11 (1993-2006), where it is clear that the annual period is present within the confidence interval

Autocorrelation (*tree II.F in Fig. 4.2-1*) is a frequently used mathematical tool in environmental and other sciences for determining a process' memory. Basically, it is the cross-correlation of a signal with itself (Boersen, 2006). It shows the linear dependence of the time series on themselves from a certain point in time, and it indicates the memory of the variable. The further a point is on the offset from the one before it, the greater the change in the process' autocorrelation. Memory here is represented by the autocorrelation function, long memory (Beran, 1994), however, is an asymptotic characteristic, which requires the hyperbolic temporal decay of the autocorrelation function. In the study the change of the autocorrelation coefficient at a one week lag was considered as the most representative feature which suited our needs. During the research an autocorrelation function was plotted for each of the 22 weekly sampled parameters at the four cardinal sampling sites. This resulted in again 88 graphs, where the maximum lag of the function was determined in 52 weeks. Afterwards, regardless of their origin (which sampling site the data came from), these were Z-score transformed and clustered using Ward's method and squared Euclidian distance. It means that four of each parameter - marked with the name of the sampling site where it was measured (e.g. "Z15_TP"; "Kb210_TN"; "Z27_TP") - formed the input data of the clustering (88 rows), where the autocorrelation coefficients at the different lags regarding each parameter formed the 52 columns. As a last step the obtained autocorrelation groups were analyzed to see the similar autocorrelation patterns.

In the study both periodicity analysis and autocorrelation analysis were used on the time series of 22 parameters sampled weekly at the same four cardinal sampling sites, as in the previous case (for a detailed list of parameters, please see Table 4.1-1).

4.3. Discussion of the novelty of the applied methodology

One can view methodology of the present study in two main ways. One, by looking at the methods by themselves, and the other by taking the methodology as a whole.

The methods by themselves are widely used in natural sciences (except for dynamic factor analysis). However, there are some, which have not gained significant ground in the field of limnology as yet. On the one hand let us first consider variogram analysis. To my knowledge it has not been used before to estimate a temporal sampling frequency of a limnological system, although it is widely used in geosciences. The same statement is true for dynamic factor analysis as well, but with restrictions. On the other hand dynamic factor analysis is becoming more-and-more popular. There have been studies in the field of hydrology and hydrogeology where it was used to deal with the background processes of time series, e.g. Kovács et al. (2004), Kovács (2007), Márkus et al. (1999) or Muñoz-Carpena et al. (2005), Ritter & Muñoz-Carpena (2006). However, none of these deal with a riverine system where the set up involves only one sampling site for the multiple response parameters.

During my research in the case of periodicity analysis, I have found only a few similar case studies where wavelet spectrum analysis was applied to a water system. In oceanology it has been used to analyze significant wave height in the East China Sea (Yanyou et al., 2006), while Zhang et al. used it to identify the climate changes and flood-drought risks in the Yangtze delta (Zhang et al., 2006; 2007; 2008). The case where its use was most similar to ours was in a study by Lafreniére and Sharp (2003) who analyzed the inter-annual variability in the runoff regimes of glacial and nival stream catchments.

As for the other methods one should look at their application as a whole. Cluster and principal component analyses are frequently used in water sciences in a broad variety of combinations. E.g. the cluster groups are formed for instance in space, and then their background processes compared (Kazi et al., 2009; Zhao et al., 2008); the sampling sites are grouped and afterwards the stochastic relationship of the parameters is assessed using PCA for every year (Magyar et al., 2013), the dimensions of the parameters are reduced and then clustered (Xu et al., 2009) or the parameters monitored in the water column and the sediment are compared using both methods (Hanrahan et al., 2008) and the list goes on. One of the most outstanding differences between these works and the present one, is the lack of using Wilks' lambda statistics to determine which variables influence the formation of the cluster groups the most, and the fact that in most of the cases the cluster results are not validated. In contrast I used discriminant analysis to make good this lack.

In the recent past the database in question has been examined mostly using deterministic methods and not in such a complex way as in the present study. In these deterministic models the changes in different shorter time segments were assessed using basic descriptive statistics and - in certain cases - approximated with descriptive models (Clement, 2013).

In summary, the novelty of my approach lies in the application of various stochastic methodologies. Cluster-, principal component- and discriminant analyses have not been used together to analyze such an ecosystem like the River Zala and the two reservoirs of the Kis-Balaton Water Protection System. Foremost the complex result, such as the ones discussed, have not been obtained so far to my best knowledge. Furthermore variogram, dynamic factor-, and wavelet spectrum analyses have not been applied in general to assess limnological systems, in particular mitigation wetlands. This complex approach permits a better, new and wider overview on the relationship/nexus of the biotic and abiotic processes in the system.

5. Results

The results of my research will be presented according to the aims described in the flowchart (Fig. 4.2-1). Since the structure of the study is complex, the variety of the applied methods is diverse, and the set of parameters and time periods differed during the different analyses, I will give a short reminder at the beginning of the major results sections of the data and methodology used, based on the overview table describing these in detail (Table 4.1-1).

5.1 Results obtained from time series of parameters at many sampling sites (I)

5.1.1 Descriptive statistics (*I.A*)

To obtain a concise overall picture of the dataset its descriptive statistics should be assessed (Table 5.1.1-1). These indicate the previously discussed differences in the system. In the case of Chl-a, its content is highest in Phase I, lowest in The River Zala, then in Phase II, and increases again in Keszthely Bay. This is no surprise, because Phase I is hypertrophic, Phase II is covered by macrophytes, and Kszt is phytoplankton-dominated. Heading downstream, the CV values of Chl-a increase continuously, indicating a growth in the parameters' variability. A somewhat similar pattern is true for TP, except it is much more stable regarding CV, and it continuously decreases downstream starting with Z11. An average CV for all the parameters in the KBWPS remained between 60-80%, while in Kszt it was ~120% and without nitrate ~ 53%.

.

Table 5.1.1-1 Statistics of analyzed data (1993-2009) at sampling sites Z15; Z11; Z27 & Kszt (reproduced from Hatvani et al., 2014c)

Statistic / Parameter	COD_{ps}	BOI_5	pH	Cl^-	Ca^{2+}	Na^+	K^+	SO_4^{2-}	NH_4-N	Chl-a	Mg^{2+}	HCO_3^-	CO_3^-	TSS	NO_2-N	NO_3-N	SRP	TN	TP
Z15																			
M	24.0	3.1	8.1	33.3	95.5	34.3	5.3	63.6	0.1	5.3	34.3	400.4	4.0	32.7	0.0	2.0	0.094	3.3	0.2
MED	22.0	2.8	8.1	33.0	96.0	33.0	5.0	59.0	0.1	4.4	35.0	414.8	0.0	20.0	0.0	1.9	0.080	3.1	0.2
SD	8.6	1.7	0.3	5.1	13.5	10.9	1.4	19.5	0.1	3.9	5.6	59.4	11.9	62.8	0.0	0.7	0.055	1.0	0.1
CV%	35.8	54.1	3.1	15.3	14.1	31.8	26.2	30.7	81.8	72.5	16.3	14.8	301.1	191.9	84.0	34.9	58.554	30.1	58.7
MIN	9.0	0.0	7.0	15.0	41.0	4.1	2.0	24.0	0.0	1.0	10.7	97.6	0.0	1.0	0.0	0.2	0.000	1.2	0.0
MAX	69.0	14.0	8.9	53.0	134.0	74.0	11.4	150.0	0.8	43.4	55.6	523.0	126.0	910.0	0.3	7.1	0.500	10.7	1.3
Z11																			
M	52.2	6.1	8.3	39.4	68.9	36.6	5.0	66.6	0.1	95.2	35.9	319.3	9.6	27.1	0.0	0.4	0.023	3.0	0.2
MED	46.0	5.6	8.3	38.0	70.0	33.0	4.9	62.0	0.1	67.0	36.0	329.4	0.0	25.0	0.0	0.1	0.010	2.6	0.1
SD	25.9	3.2	0.3	8.8	19.4	12.6	1.0	27.2	0.3	86.3	5.9	69.7	14.4	17.5	0.0	0.6	0.036	1.6	0.1
CV%	49.6	51.7	3.7	22.3	28.2	34.4	19.4	40.9	171.5	90.7	16.4	21.8	150.7	64.7	92.5	153.7	157.044	52.3	68.3
MIN	14.0	0.1	7.0	18.0	24.0	9.0	2.1	20.0	0.0	2.0	18.4	90.0	0.0	0.0	0.0	0.0	0.000	0.7	0.0
MAX	174.0	27.0	9.2	82.0	129.0	94.0	9.4	196.0	2.7	578.0	70.0	500.2	73.0	160.0	0.2	3.8	0.425	12.4	0.8
Z27																			
M	39.6	2.3	7.9	37.8	84.6	33.4	4.6	66.8	0.1	6.3	37.1	389.7	1.4	6.0	0.0	0.3	0.103	1.8	0.2
MED	39.0	1.8	8.0	37.0	83.0	33.0	4.6	63.0	0.1	4.4	37.0	390.0	0.0	4.0	0.0	0.1	0.077	1.7	0.1
SD	11.5	1.7	0.3	6.4	12.8	8.5	1.2	25.5	0.2	6.3	5.7	47.3	6.6	6.3	0.0	0.3	0.086	0.6	0.1
CV%	29.0	73.3	3.9	17.0	15.1	25.5	26.7	38.2	137.1	101.5	15.5	12.1	454.4	105.7	129.0	132.0	84.114	35.0	66.3
MIN	13.0	0.0	6.6	15.0	28.0	14.0	1.6	21.0	0.0	0.0	16.0	159.0	0.0	0.0	0.0	0.0	0.000	0.6	0.0
MAX	110.0	18.4	8.8	63.0	133.0	60.0	9.8	174.0	1.1	81.4	67.7	550.0	78.0	45.0	0.3	2.6	0.500	5.8	0.9
Kszt																			
M	26.6	2.6	8.5	32.6	47.8	32.7	6.6	124.5	0.0	19.2	49.7	250.8	16.3	22.8	0.0	0.2	0.017	1.2	0.1
MED	26.0	2.2	8.5	31.9	46.8	31.9	6.5	128.2	0.0	11.8	48.9	252.6	14.4	16.0	0.0	0.1	0.013	1.1	0.1
SD	5.8	1.2	0.2	4.9	11.0	5.7	1.1	20.4	0.0	23.3	7.3	33.2	9.1	23.8	0.3	0.3	0.021	0.5	0.1
CV%	21.7	47.4	2.2	14.9	23.0	17.6	16.5	16.4	90.8	121.2	14.6	13.2	56.0	104.5	1314.8	170.4	117.790	41.6	63.4
MIN	16.0	0.0	7.6	21.3	26.1	19.1	3.3	48.0	0.0	0.0	24.9	151.3	0.6	2.0	0.0	0.0	0.000	0.0	0.0
MAX	49.0	7.8	9.1	44.0	72.9	47.0	10.0	177.0	0.5	203.4	64.0	335.6	48.0	176.0	4.9	1.8	0.221	3.8	0.3

Besides descriptive statistics, it is useful to see the nutrient (TP) loadings' fluctuation over the years (1993-2009), after the partial completion of the KBWPS (Fig. 5.1.1-1), to get a glimpse of the system's behavior and characteristics. It is clear that with decreasing runoff, loadings decrease as well, and the distinct difference between the River Zala and the two reservoirs seen in the case of high runoff years diminishes. In parallel it is quite obvious that the biggest difference with regard to runoff is between Phase I (Z11) and Phase II (Z27), where the runoff of Phase II is almost 40% higher than that of Phase I. It seems that the Z27/Z15 discharge ratio is higher during the wet periods than during the dry periods due to the different impacts that evapotranspiration has on the water balance in the two conditions. The most significant differences are observable mostly in the case of the P forms. Regarding TP (Fig. 5.1.1-1b), the KBWPS' output is generally slightly higher than the input "overall picture". However, looking at the processes operating behind this, the SRP (Fig. 5.1.1-1c), it becomes clear that in Phase I SRP is taken up by phytoplankton, thus decreasing as it reaches Z11. However at Z27 SRP again exceeds the system's input levels because of the decomposition processes in the reed area (substance transformation). Although these might be simple transformations, if the reasons behind this phenomenon and the retention capacity of the system are taken into account, the fact cannot be ignored that Phase II's own watershed - which is by itself approx. 1100 km^2, i.e. around 40% of the watershed of the Zala River - brings excess loads to the system (Hatvani et al., 2011, 2014c). Furthermore the release of phosphorus from the sediment cannot be ignored either.

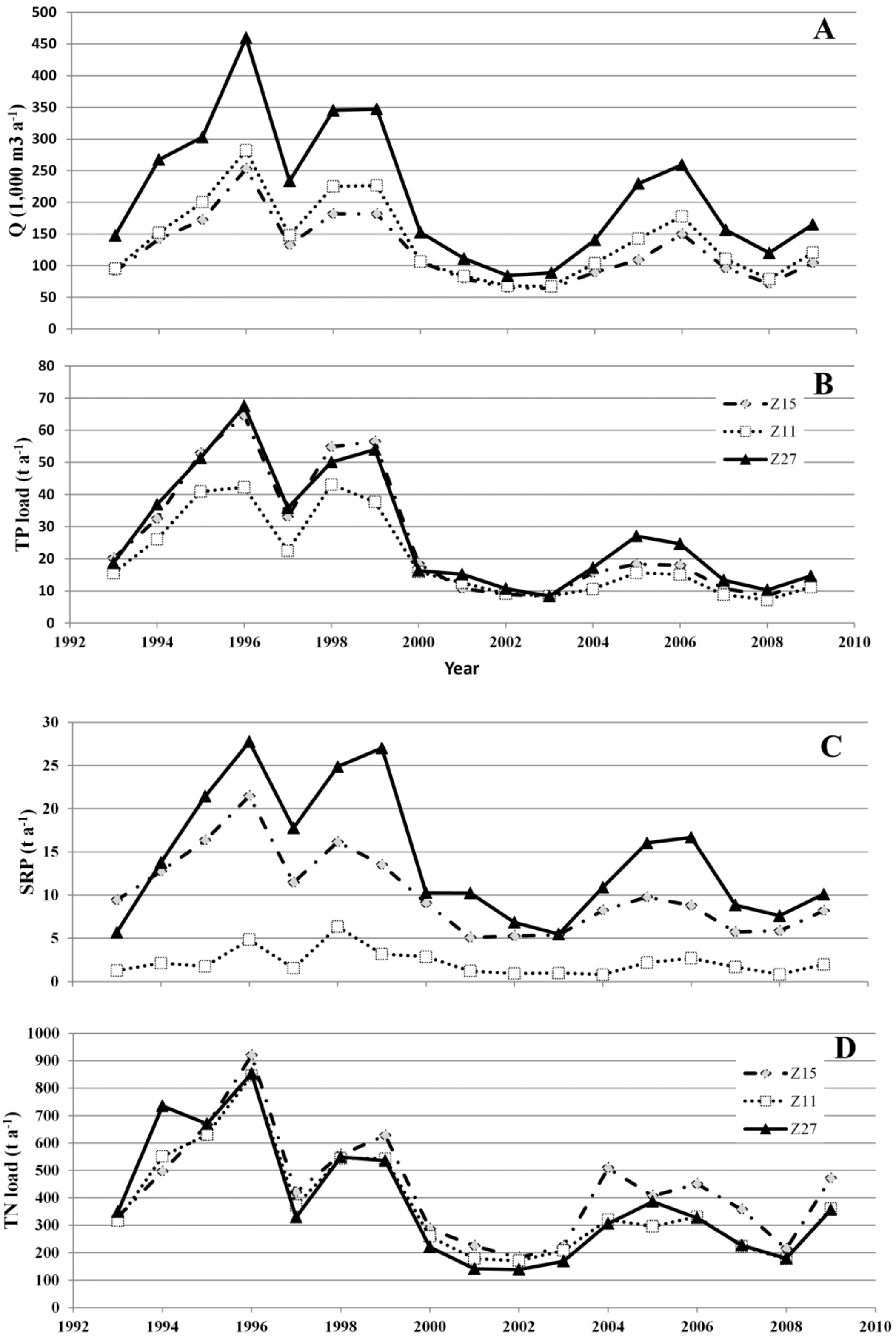


Figure 5.1.1-1 Runoff and nutrient annual loads at three sampling sites in the system, representing the input to KBWPS (Z15), the interface between the two reservoirs (Z11) and the output to Keszthely Bay (Z27). A) Runoff, B) TP, C) SRP and D) TN (reproduced from Hatvani et al., 2014c)

If the statistics for annual phytoplankton biomass - discussed previously - are visualized on box-and-whiskers plots (Fig. 5.1.1-2) it becomes even clearer that, from the point when Phase I became fully operational, Chl-a started to rise at both Z11 and Z27, while over the whole investigated time period the River Zala (Z15) was bringing algae-poor water into the system. The rise in the Chl-a values after 1985 at Z11 and Z27 can be explained by the fact that by putting Phase I into operation an open water body was created which gives ideal conditions for algae to reproduce. At sampling site Z27 only after 1992, when the 16 km^2 of Phase II was inundated, did the Chl-a values decrease as a result of the partially finished state of Phase II, because it is covered by dense macrophytes. The shade provided by the dense vegetation prohibits the growth of algae (Kovács et al., 2010). This resulted in Chl-a values lower than any experienced before 1992.

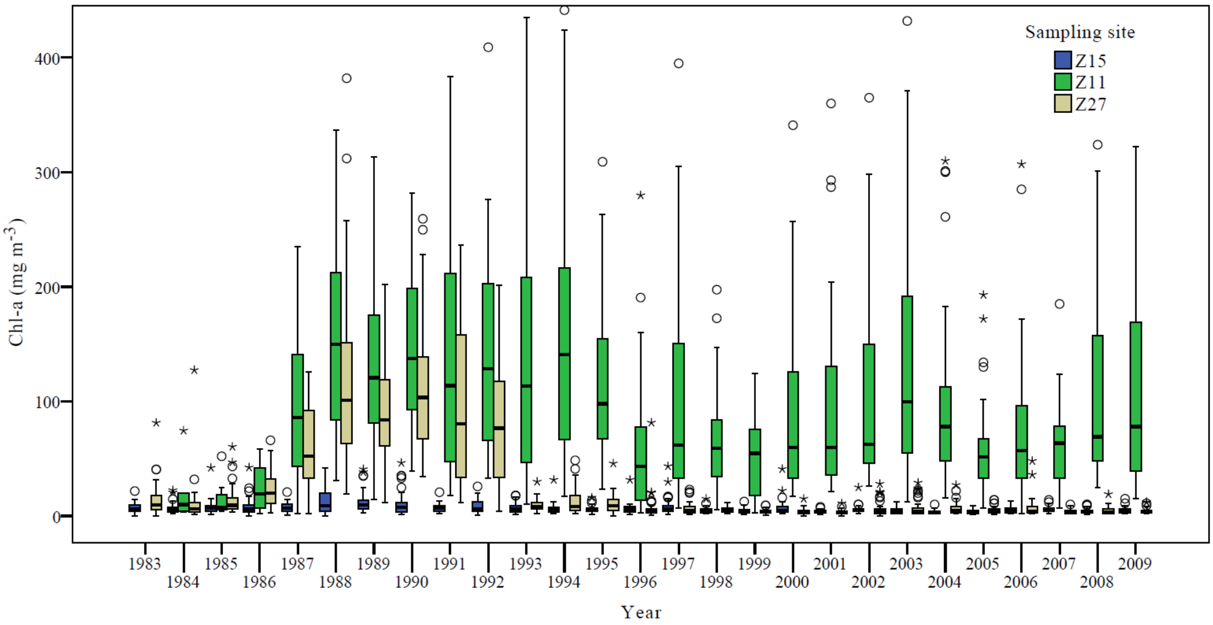


Figure 5.1.1-2 Annual trends of Chl-a at Z15 (inlet), Z11 (interface), Z27 (outlet) from 1983-2009

To explore the amount of nutrients entering the system through the Zala and leaving it towards Lake Balaton, an input-output analysis was conducted on the sampling sites Z15 & Z27 for the period 1977-2009, based on Somlyódy et al. (2003). It became clear that before the inundation of Phase I (1977-1983), the outlet of the Zala, at that time entering directly into Lake Balaton (at Z27 containing the all the inflows of the KBWPS), was higher than the input of the KBWPS, the River Zala, at Z15 alone (Fig. 3-2). In the years 1984-1985, when the inundation of Phase I was started, the output started to decrease, and between 1985-1990 Phase I was removing a significant amount of TP in the form of algae. This is reflected in the arrangement of the points, which were way under the 1:1 line. After Phase II was set into operation - at the beginning of the nineties - the TP input of the system significantly decreased. In the meanwhile the output became comparable to the input (1991-2009; Fig. 5.1.1-3a), which is reflected in the arrangement of the points on the 1:1 line close to the arbitrary point. Nevertheless it can be stated that the amount of phosphorus entering Balaton from the watershed of the Zala throughout the KBWPS decreased by two thirds (Fig. 5.1.1-1) if a comparison is made between the 80s and the end of the first decade of the new century. Regarding runoff, (5.1.1-1a) it is obvious that in years of high precipitation it increases downstream (Z15-Z11-Z27), because of the extra inputs of Phase II; furthermore, it seems as if there are two distinct time intervals, 1993-1999 and 2000-2009. It is a well-known fact that if input loads decrease, the effectiveness of nutrient retention decreases as well; the decrease cannot compensate for the loads arriving from the watershed of Phase II, for example in 2004-2005 (Fig. A2 in the Appendix). However, when the concentrations are considered in addition to loads, the picture changes, and the retention of the system becomes much more explicit (Fig. 5.1.1-3b). Here the groups identified are similar to the groups of the years seen in Fig. 5.1.1-3a. Until 1991 the KBWPS was able to maintain an almost constant TP concentration at the outflow (Z27) independently from the inflows. This concentration was significantly higher before the inundation. When the impact of the socio-economic changes and P removal began to be felt in 1991, a further decrease in inflowing TP concentrations of the River Zala forced the outflow concentrations to decrease even more, resulting a linear relationship.

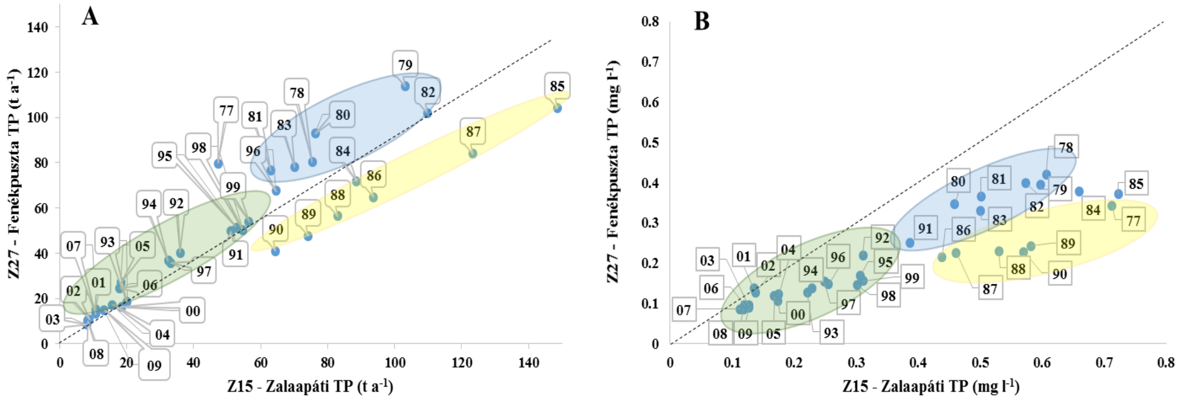


Figure 5.1.1-3 Input-output analysis on average annual phosphorus loads A) and average annual concentrations B) at the inlet (Z15) and outlet (Z27) of the KBWPS (1977-2009; reproduced from Hatvani et al., 2014c)

5.1.2 Cluster, discriminant and Wilks' lambda analysis results (*I.0/B - I.D*)

The variability of the parameters and the SSs became quite obvious in the light of the descriptive statistics, and it even gave a brief insight into the characteristics of the different habitats. Nevertheless, the exploration of the SSs on the basis of their similarity was plausible. As a reminder (Table 4.1-1. & Section 4.2.3/Cluster analysis), divisive hierarchical cluster analysis was applied to Z-scored annual averages formed from weekly sampled data (for the list of the 19 parameters see Table 4.1-1) using Ward's method, along with squared Euclidian distance for the time interval 1993-2009.

Cluster analysis subdivided the twelve SS of the KBWPS (excluding Z15, the Zala) and Kszt (altogether 14 sampling sites) into three groups for the time interval 1993-2009, with the dendrograms intersected at 80% of the maximum linkage distance. One group was formed mainly of the SSs in Phase I, a second of the SSs in Phase II. Therefore, the sampling sites fell into groups which in most of the cases cover the two habitats (two constructional phases, Fig. 5.1.2-1), and the third group included Kszt alone. According to discriminant analysis, these three groups proved to be valid in 97.7% of the cases.

The sampling sites in Phases I & II almost always remained separate. There was only one occasion when a sampling site, SS *202i*, disconnected from the cluster group covering the macrophyte covered wetland (Phase II), and connected to the group related to the eutrophic pond (Phase I), with its open water body (Fig. 5.1.2-1; Hatvani et al., 2011). This occurred between 1997 and 1998. This is important because it is the only point that changes alignment between the two phases and keeps its connection over the whole research period. There are two further SSs that are worthy of mention. In the study of Hatvani et al. (2011) SSs Kb9 & 205 occasionally formed separate groups alone. The reason why these were not separated at 80% of the maximum linkage distance in this study is that the overall distances in the N dimensional space changed. This happened because in this particular study (i) Z15 was excluded, (ii) Kszt included and (iii) the parameters used in the analysis also changed to a minor degree. Nevertheless these SSs did separate, but - because of the changed distances - only at a smaller rescaled distance.

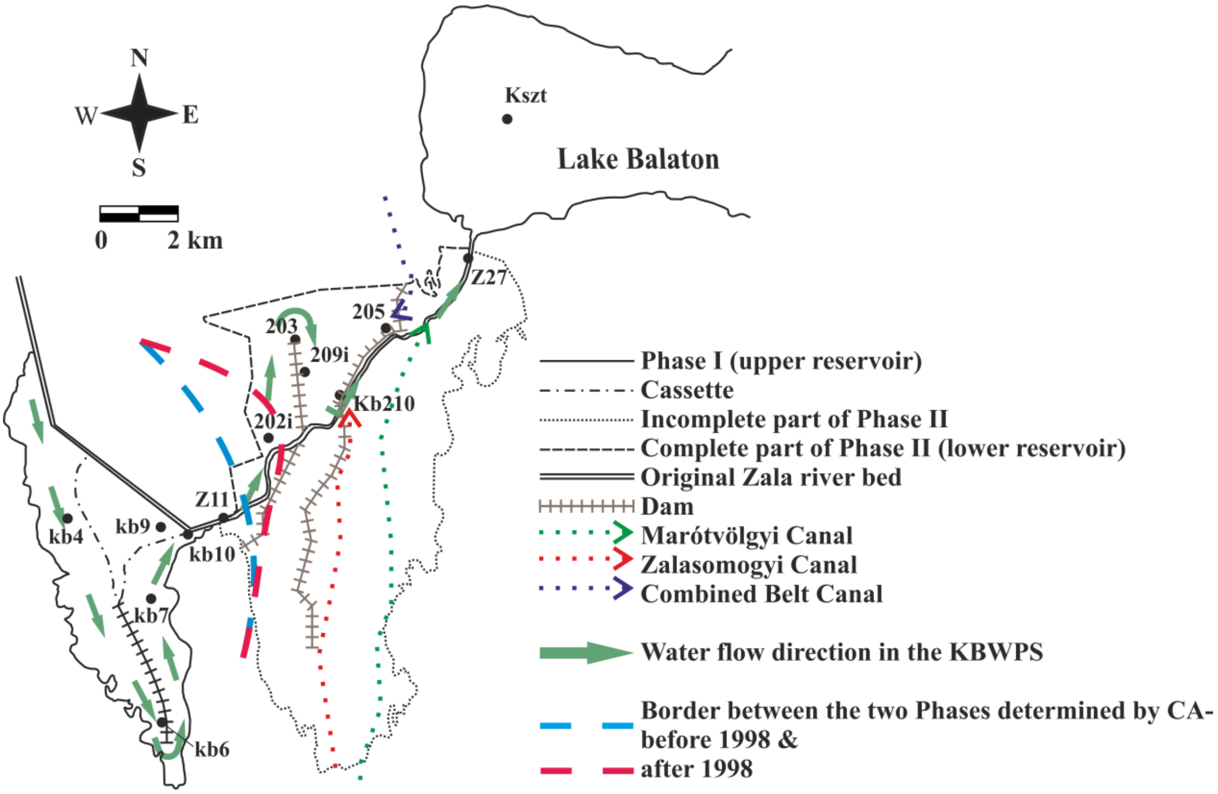


Figure 5.1.2-1 The changing border between the cluster groups covering the two constructional phases of the KBWPS. Changes occurring between 1997 & 1998

Parallel to the changes witnessed in the alignment of SSs in the KBWPS, Kszt stood alone in 1993 and in 2000-2009, while in 1994-1999 it connected to the cluster group covering Phase II, and the number of groups was thus reduced to two. The pattern seen in the cluster dendrograms concurred with the two separated time periods defined by the runoff of the River Zala (Fig. 5.1.2-2). The question needed to be answered of whether the two time periods defined by the runoff exist in the data series of the three spatial cluster groups or not. In order to obtain an answer, the annual averages of the sampling sites for each spatial group were artificially separated into two time periods. These periods were the ones previously defined by runoff (1994-1999; 2000-2009). Then the presence of these periods was tested using discriminant analysis. It was shown that these time periods are valid in 99.2% of the cases in Phase I, 90.0% in Phase II and 100.0% in Kszt when it stood alone.

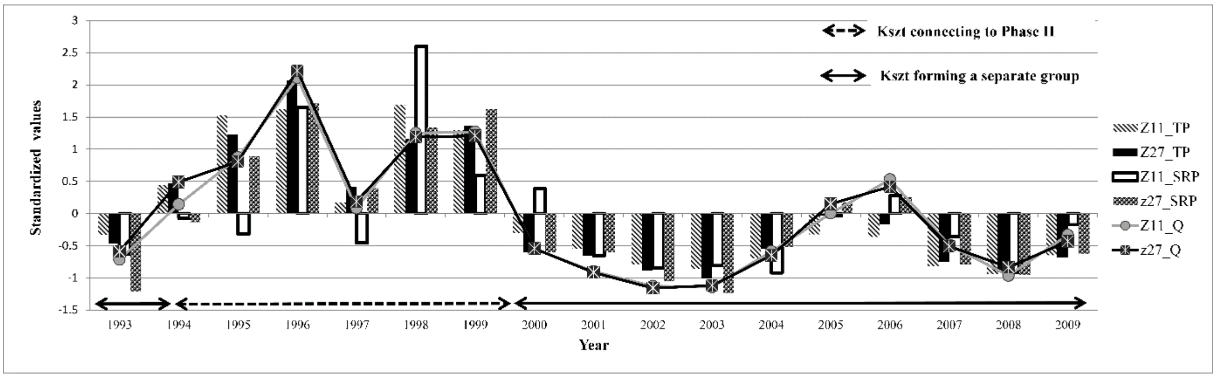


Figure 5.1.2-2 Standardized TP & SRP loadings and runoff with the behavior (connectance) of SS Kszt towards the two constructional phases of the KBWPS. If the values of the bars (Z-score transformed loadings) exceed the values of the line (runoff), there were excess P loads in the system compared to the average P loads at the particular SS. If, however, the bars are under the line, it means that in that year the loads were significantly below the general average, and P removal was taking place in the system (reproduced from Hatvani et al., 2014c)

In the years characterized by low runoff, when Kszt without a doubt stood alone, phosphorus forms had the greatest effect on the formation of the cluster groups. It is considered a low runoff year when the output of Z27 is less than three times the volume of Keszthely Bay (87 km^3) in a year; while if it is more than three times the volume, it is considered a high runoff year. The phosphorus forms' Wilks' λ quotients - when Kszt formed one group alone (2000-2009; Fig. 5.1.2-3) - were practically zero and their concentrations at Kszt - in the case of SRP ~25%, and in the case of TP ~32% - were lower than in 1994-1999. In 1993 SRP was much lower than in the following years. This may have happened because in the first years after Phase II was inundated (1992), internal P loads originating from its sediment increased Keszthely Bay's P loads. This was reflected in the input-output analysis as well. This phenomenon coincided with the high runoff of the following years, resulting in the observed SRP values (Istvánovics et al., 2004). In 1994-1999 there were only a few parameters which could be considered as having a significant influence on the formation of the cluster groups (Table 5.1.2-1), the most important of which were Chl-a, BOD_5 and TSS, all with an average Wilks' lambda quotient of under 0.25. In 1993 & 2000-2009 besides the explicit domination of phosphorus forms in the formation of clusters, other chemical parameters - mostly the ones present in the previous years - had influence as well.

It is another question, however, to track how the similarities of the SSs change if the outlet of the KBWPS (Z27) and Kszt are not analyzed from the side of the KBWPS, but from that of Lake Balaton instead. This means that the set of parameters was slightly different from the one generally used in this study, and the set of sampling sites contained 10 sampling sites, two of which were Z27 and Kszt. In the study of Kovács et al. (2012c) a special coded cluster analysis was used to determine the water-bodies in Lake Balaton. It turned out that instead of the generally accepted four geologically determined areas, in general five water quality areas exist (1985-2004), two of which are the outlet of the KBWPS, Z27 and Keszthely Bay, both forming groups alone in this case after 1998. This again proves that changing the set of parameters, SSs, and circumstances could result in somewhat different alignments.

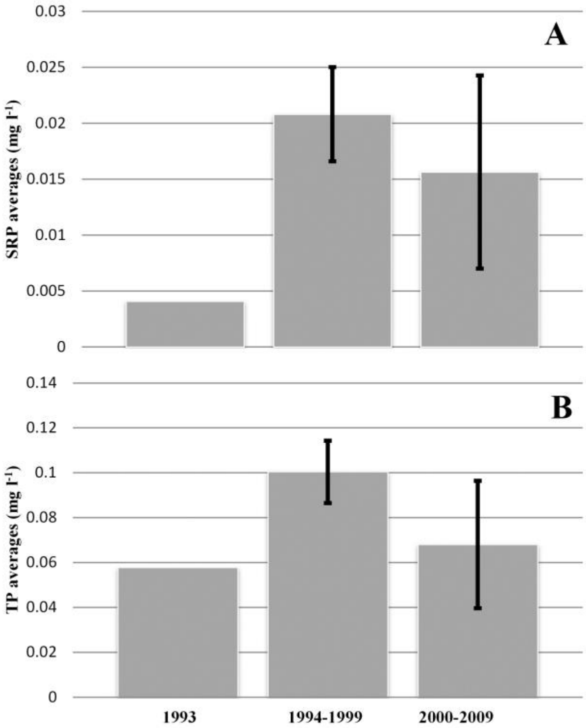


Figure 5.1.2-3 Total annual average SRP A) & TP B) concentrations in 1993, 1994-1999, and 2000-2009 at SS Kszt, with the standard deviation marked on the graph (reproduced from Hatvani et al., 2014c)

Table 5.1.2-1 Wilks' λ quotients for each sampling site and year, with the high runoff years underlined (reproduced from Hatvani et al, 2014c)

Year/ Parameter	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
COD_{ps}	0.444	0.562	0.542	0.884	0.821	0.816	0.751	0.516	0.380	0.401	0.339	0.441	0.570	0.452	0.674	0.509	0.411
BOI_5	0.132	0.090	0.188	0.415	0.162	0.183	0.159	0.312	0.237	0.254	0.180	0.159	0.228	0.149	0.074	0.146	0.160
pH	0.085	0.302	0.320	0.485	0.593	0.521	0.613	0.132	0.158	0.189	0.238	0.158	0.168	0.028	0.061	0.062	0.053
Cl^-	0.375	0.968	0.909	0.942	0.774	0.887	0.942	0.733	0.761	0.849	0.765	0.985	0.998	0.595	0.814	0.874	0.810
Ca^{2+}	0.166	0.730	0.902	0.984	0.920	0.955	0.932	0.494	0.310	0.300	0.251	0.428	0.629	0.539	0.432	0.411	0.425
Na^+	0.688	0.809	0.635	0.768	0.732	0.820	0.809	0.793	0.697	0.765	0.734	0.848	0.896	0.767	0.997	0.924	0.952
K^+	0.745	0.997	0.971	0.841	0.748	0.721	0.726	0.829	0.800	0.815	0.713	0.779	0.862	0.855	0.900	0.901	0.597
SO_4^{2-}	0.082	0.822	0.921	0.747	0.703	0.646	0.708	0.524	0.221	0.151	0.059	0.271	0.549	0.297	0.213	0.133	0.357
NH_4-N	0.965	0.974	0.976	0.956	0.922	0.984	0.989	0.961	0.943	0.921	0.890	0.562	0.672	0.663	0.912	0.904	0.664
Chl-a	0.201	0.040	0.135	0.593	0.205	0.273	0.220	0.238	0.314	0.171	0.194	0.132	0.284	0.159	0.254	0.116	0.212
Mg^{2+}	0.167	0.615	0.853	0.583	0.585	0.620	0.654	0.505	0.246	0.293	0.231	0.161	0.271	0.221	0.100	0.081	0.118
HCO_3^-	0.103	0.469	0.313	0.628	0.734	0.654	0.729	0.286	0.190	0.222	0.277	0.220	0.418	0.294	0.186	0.309	0.313
CO_3^-	0.076	0.168	0.105	0.470	0.673	0.481	0.550	0.134	0.495	0.627	0.287	0.345	0.338	0.171	0.139	0.196	0.259
TSS	0.217	0.150	0.107	0.386	0.590	0.345	0.280	0.434	0.214	0.252	0.165	0.255	0.426	0.106	0.297	0.279	0.303
NO_2-N	0.585	0.937	0.515	0.993	0.794	0.878	0.834	0.841	0.878	0.818	0.645	0.449	0.832	0.000	0.323	0.152	0.353
NO_3-N	0.542	0.863	0.736	0.673	0.851	0.919	0.952	0.841	0.743	0.867	0.847	0.847	0.797	0.837	0.835	0.823	0.857
SRP	0.000	0.924	0.925	0.899	0.898	0.858	0.862	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TN	0.113	0.351	0.319	0.460	0.544	0.728	0.676	0.371	0.549	0.867	0.348	0.637	0.624	0.249	0.522	0.436	0.254
TP	0.000	0.929	0.929	0.902	0.903	0.867	0.867	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

5.2 Results obtained from time series of parameters at one sampling site (*II*)

5.2.1 Principal component analysis (*II.0/A & A*)

To see the background parameters' behavior in the time periods separated by runoff, principal component analysis was conducted on the same set of 19 parameters (Table 4.1-1) as in the case of cluster analysis. The set of parameters in the first principal component (as with the pattern obtained from CA) appears to change corresponding to the system's runoff (Table 5.2.1-1), with a clear change mainly in the case of Phase II. After the year 2000, the parameters describing eutrophication and the nutrient cycle take on a determining role in the processes of Phase II. Despite this, no significant change was observed in either the processes of Kszt or of Phase I. The processes of the former were still determined by the inorganic compounds, and those of the latter mostly by nutrients (Table 5.2.1-1b). It is clear that with reference to stochastic relations, in the years with low runoff, Kszt was not the SS where the set of parameters in the first principal component underwent a change, but it was the dominating parameters of Phase II which became co-dominated by the nutrients (Table 5.2.1-1). In this way Phase II and Kszt began to differ from each other. The absence of the nutrients from Phase II's most determining parameters in the years with high runoff is remarkable, and concurs with the cluster, and more specifically the Wilks' λ results. In both sets of results (before and after 2000) the first factor explained approx. 30% of the original data's variance (as an example for a scree plot on Z11 see Fig. A3 in the Appendix). This is an average value formed from the percentages at the different SSs.

Table 5.2.1-1 Influential factor loadings at all SSs of the KBWPS & Kszt: A) 1994-1999, and B) 2000-2009 (reproduced from Hatvani et al., 2014c)

A

SS / Parameter	kb4	kb6	kb7	kb9	kb10	z11	202i	203	209	Kb210	205	Z27	Kszt
COD_{ps}	X	X	X	-	X	X	-	X	X	X	-	-	-
BOI_5	-	X	-	-	-	-	-	-	-	-	-	-	-
pH	-	-	-	-	X	-	X	-	-	-	-	-	-
Cl^-	-	X	X	-	X	X	-	X	-	-	-	-	X
Ca^{2+}	X	X	X	-	X	X	-	-	-	-	X	-	X
Na^+	X	X	X	-	X	X	X	-	X	-	-	X	X
K^+	-	-	-	-	-	-	-	-	-	-	-	-	X
SO_4^{2-}	-	-	-	-	-	-	-	-	-	-	X	-	-
NH_4-N	-	-	-	-	-	-	-	-	-	-	-	-	-
Chl-a	X	X	X	X	X	X	-	-	-	-	-	-	-
Mg^{2+}	-	-	-	-	-	-	-	-	-	X	-	-	-
HCO_3^-	-	-	-	-	-	X	-	-	X	-	-	-	X
CO_3^-	-	-	-	-	-	-	-	-	-	-	-	-	-
TSS	-	X	X	X	X	X	-	-	-	-	-	-	-
NO_2-N	-	-	-	-	-	-	-	-	-	-	-	-	-
NO_3-N	-	-	-	-	-	-	-	-	-	-	-	-	-
SRP	-	-	-	-	-	-	-	-	-	-	-	-	-
TN	-	X	X	-	X	X	-	-	-	-	-	-	-
TP	-	X	-	-	-	X	-	-	-	-	-	-	-

B

SS / Parameter	kb4	kb6	kb7	kb9	kb10	z11	202i	203	209	Kb210	205	Z27	Kszt
COD_{ps}	X	X	X	X	X	X	X	X	-	-	X	-	-
BOI_5	-	-	-	-	-	-	-	-	-	-	-	-	-
pH	-	-	-	-	-	-	-	-	-	-	-	-	-
Cl^-	-	-	X	-	X	X	X	X	-	-	X	-	X
Ca^{2+}	-	-	X	-	X	X	X	-	-	-	X	-	X
Na^+	X	X	X		X	X	X	X	X	-	X	-	X
K^+	-	-	X	X	-	-	-	-	-	X	-	X	X
SO_4^{2-}	-	-	-	-	-	-	-	-	-	-	X	-	X
NH_4-N	-	-	-	-	-	-	-	X	-	X	-	X	-
Chl-a	X		X	X	X	X	X	-	-	-	-	-	-
Mg^{2+}	-	-	-	-	-	-	-	-	-	-	-	-	X
HCO_3^-	-	-	-	-	-	-	X	-	X	-	-	-	X
CO_3^-	-	-	-	-	-	-	-	-	-	-	-	-	-
TSS	-	-	X		X	X	X	-	-	-	-	-	-
NO_2-N	-	-	-	-	-	-	-	-	-	-	-	-	-
NO_3-N	-	-	-	-	-	-	-	-	-	-	-	-	-
SRP	-	-	-	-	-	-	-	X	X	-	X	X	-
TN	X	X	X	X	X	X	X	X		X	X	X	-
TP	X	X	X	X	X	X	X	X	X	-	X	X	-

Taking an overview of the stochastic and spatial patterns in the Kis-Balaton and Kszt system for 1993-2009, an additional question was raised: Did the putting into operation of the KBWPS trigger any change in the parameters determining the variance at Kszt, or not? By applying PCA on the Kszt data for the time period 1981-1984 the results unambiguously showed that prior to the construction of the KBWPS the processes of Kszt were co-influenced by parameters related to eutrophication, resembling the eutrophic Phase I as it happened some years later (Table 5.2.1-2). It is not surprising that in that time period Keszthely Bay was hypertrophic. In the first two principal components (responsible for 52.16% of the total variance), the parameters closely related to eutrophication are present alongside the other chemical parameters, with loadings outside the ±0.7 interval (Table 5.2.1-2a). It is still a question, however, why P forms did not have influential factor loadings. The third PC explained about 11% of the total variance; it was thus discarded. In the meanwhile in the transient period between the initiation of Phases I & II (1985-1992), the parameters related to eutrophication disappear from the group of components which determine the background processes at Kszt. However, Chl-a has a loading value lower than the one in 1981-1984, though this is still relevant (Table 5.2.1-2b). Here, the first two PCs explained 54.5% of the total variance, while the third only 9% (as an example for a scree plot on K see Fig. A4 in the Appendix).

Table 5.2.1-2 Factor loadings of SS Kszt: A) 1981-1984; and B) 1985-1992. The influential instances are underlined (reproduced from Hatvani et al., 2014c)

Principal component (loadings) / Parameter	1981-1984 A)		1985-1992 B)	
	1st PC	2nd PC	1st PC	2nd PC
COD_{ps}	**-0.876**	-0.134	-0.533	-0.672
BOI_5	**-0.731**	-0.024	-0.337	-0.551
pH	-0.112	-0.688	**-0.894**	0.101
Cl^-	-0.075	-0.422	-0.576	0.578
Ca^{2+}	**0.842**	0.071	**0.924**	0.162
Na^+	0.076	**-0.799**	-0.649	**0.615**
K^+	0.434	-0.670	-0.477	0.392
SO_4^{2-}	**0.713**	-0.544	-0.227	0.761
NH_4-N	-0.545	-0.268	0.446	-0.198
Chl-a	**-0.938**	-0.105	**-0.732**	-0.438
Mg^{2+}	0.544	-0.655	-0.468	0.666
HCO_3^-	**0.873**	0.265	**0.867**	0.361
CO_3^-	-0.686	-0.350	**-0.884**	-0.182
TSS	-0.216	-0.406	-0.003	0.012
NO_2-N	**-0.725**	0.000	0.220	-0.563
NO_3-N	0.015	-0.033	0.368	-0.357
SRP	-0.368	0.202	-0.092	-0.555
TN	**-0.845**	-0.027	-0.283	-0.657
TP	-0.182	-0.218	-0.426	-0.261

5.2.2 Dynamic factor analysis (*II.B*)

Assessment of the whole time period (1978-2006) with both Brodgar and E-DFA

After finding the spatial patterns (*I.B-D*) and the natural parameters determining most of the variance (*tree II.A in Fig. 4.2-1*) in the KBWPS, the question needed to be answered: what background processes govern the time series of the River Zala (Z15), the input of the KBWPS. The concept was to explore first the whole time period 1978-2006 with both Brodgar and E-DFA, and then assess shorter time segments if necessary, that is, if the previously described break in the time series (1991) made it necessary. The dataset used concerned 21 response parameters (ions, nutrients etc.) measured weekly in the water of the River Zala (1978-2006) and 6 explanatory ones (e.g. total runoff (Q); waste water phosphorus (WW-P); water temperature, precipitation etc.) measured in the vicinity of the river, or in the case of runoff, in the river itself, mostly for the time interval 1981-2006 (for the detailed parameter list please see Table 4.1-1).

The first factor obtained with Brodgar shows a very similar – although smoother - course than the E-DFA. The correlation between the two different types of first factors was 0.8. It is apparent from Fig. 5.2.2-1a that there is a clear relation between the two factors. The correlations between the first Brodgar factor and the P surplus of the soil and the WW-P were 0.87 and 0.89 respectively, whereas with the first E-DFA factor they were 0.85 and 0.85 respectively. The figures show similarly decreasing and then stabilizing segments separated according to the breakpoint around 1990-1991, when Hungarian industry and agriculture were thoroughly restructured, and the already-mentioned drop in fertilizer usage took place. Furthermore, this was the time when P removal was introduced at waste water treatment plants, resulting in decreased point source loads in relation to surface waters.

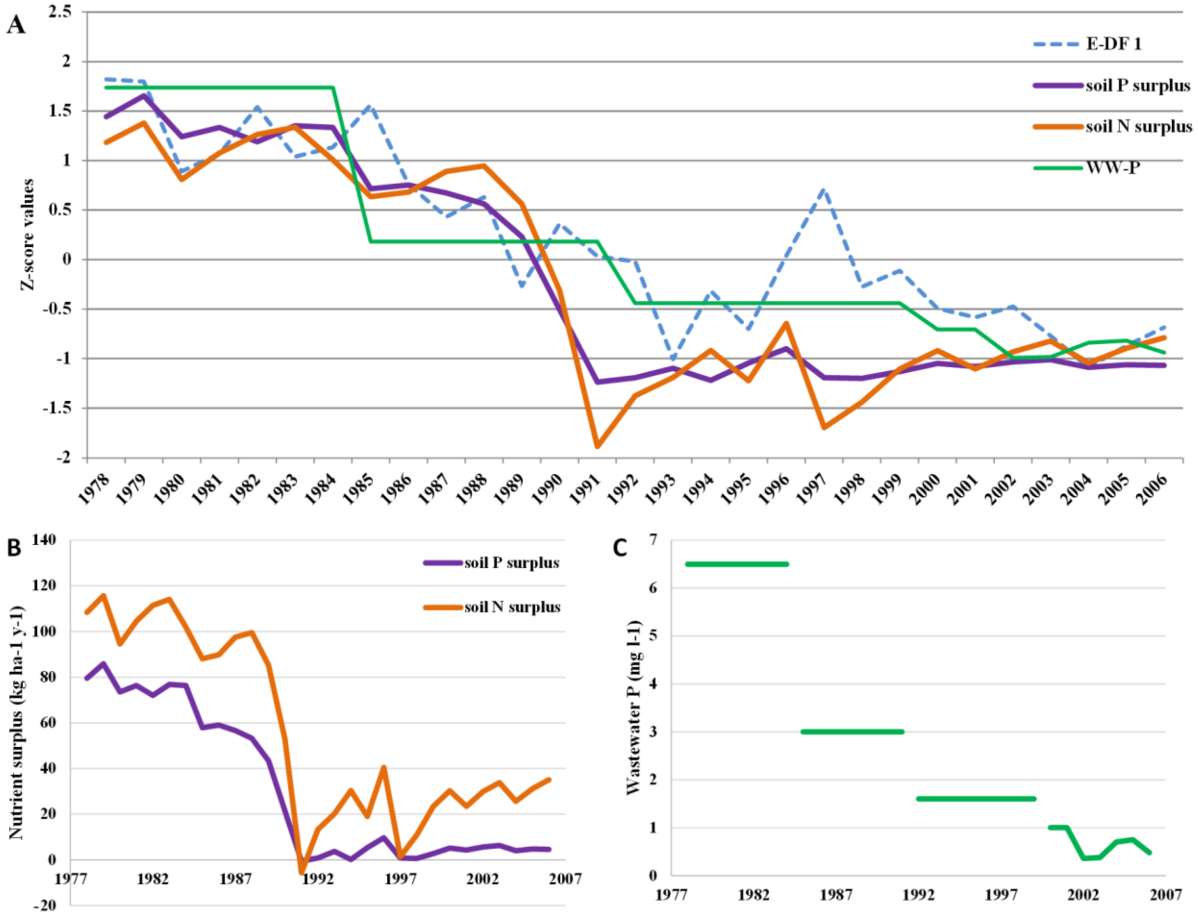


Figure 5.2.2-1 Dynamic factor time series of Brodgar and E-DFA visualized together with Z-score standardized soil P surplus and the P output of the Zalaegerszeg waste water treatment plant A) the actual soil P surplus data B) and the actual WW-P values C) (reproduced from Hatvani et al., 2014b)

It is clear from the results that both programs (Brodgar and E-DFA) reveal essentially the same structure. Although the models are similar, there are differences regarding the principles used to obtain the factor solutions. These differences lie in the temporal structure of the factors, which places different degrees of emphasis on the first factors in replicating the trends of the P surplus (Fig. 5.2.2-1b) and the WW-P (Fig. 5.2.2-1c). While E-DFA closely follows the fluctuations of the explanatory processes, in particular the drop in the P surplus (1986-1990), its weak point lies in its replication of the stabilized fluctuations; Brodgar, on the other hand, is able to follow the main trends, but is less effective in capturing fluctuations in the processes in their entirety. WW-P cannot be considered to be capable of fully representing the detection of any small fluctuations because - as mentioned before - the P data values (for details see Section 3.2) were not exactly measured before the year 2002.

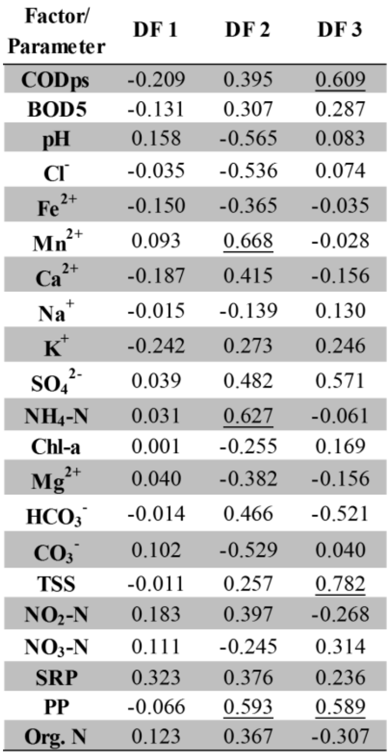
Based on the results obtained from the assessment of the whole time interval the question of whether the background processes change their character over time was raised, as two main trends occur in two different time segments. From 1978-1991 descent dominated, while from 1992 to 2006 a stable fluctuation dominated process (closer to a stationary regime) can be observed (Fig. 5.2.2-1a). These two time periods are presumably separated by the effects of the socio-economic events of 1989/90 and the other water quality amelioration measures discussed earlier. The impact of these might have changed the behavior of the river's processes to such an extent that the separation of the dataset at the breakpoint becomes a necessity. For some the question may arise: Would it make sense to examine the time interval (1978-2006) in two sections, based on runoff, as in the course of PCA. Sampling site Z15 – subject of DFA - was left out of the principal component analyses for a reason: most importantly the characteristics and the processes evolving in the river are different from the ones in the KBWPS. Although it is the primary source for the nutrients, it still has to be examined separately. As can be seen from the dynamic factor time series (Fig. 5.2.2-1), there is no breakpoint around the year 2000, contrary to the case of the KBWPS (Fig. 5.1.2-2). Therefore split at the year 2000 in the case of DFA became groundless.

After comparing the results of the two methods a decision had to be made whether to use both or only one of the software packages. The two DFA methods gave essentially the same trend pattern with differences in the fluctuations. Nevertheless they differ in the prescribed factor correlation structure, as Brodgar works with a random walk and E-DFA with an autoregressive structure. Furthermore Brodgar is known to experience difficulties in handling "smaller" datasets - resulting in the slight perturbation of the result after each run, which is a consequence of the computation and does not represent the reliability of the underlying model, while E-DFA gives one perfectly reproducible path as a result and the reliability of this has to be assessed by theoretical means. Therefore E-DFA was chosen for detailed analyses.

Assessment of the split time series (1978-1991 & 1992-2006) using E-DFA

After splitting the time series in two, the factor loadings of the parameters were analyzed. Between 1978 and 1991 there was no parameter which had an influential factor loading in the first DF time series (DF1). In the second factor (DF2), Mn, NH_4-N and PP, while in the third (DF3), COD_{ps}, TSS and PP had considerable factor loadings (Table 5.2.2-1). Regarding explanatory parameters, the DF1 time series correlated with precipitation (r = 0.77) (Fig. 5.2.2-2a), and DF2 with the P and N surplus of the soil (r = 0.67) and the WW-P (r = 0.59) (Fig. 5.2.2-2b).

Table 5.2.2-1 Factor loadings of the response parameters in the first three factors (1978-1991); significant loadings are underlined (reproduced from Hatvani et al., 2014b)



Factor/ Parameter	DF 1	DF 2	DF 3
CODps	-0.209	0.395	0.609
BOD5	-0.131	0.307	0.287
pH	0.158	-0.565	0.083
Cl^-	-0.035	-0.536	0.074
Fe^{2+}	-0.150	-0.365	-0.035
Mn^{2+}	0.093	0.668	-0.028
Ca^{2+}	-0.187	0.415	-0.156
Na^+	-0.015	-0.139	0.130
K^+	-0.242	0.273	0.246
SO_4^{2-}	0.039	0.482	0.571
NH_4-N	0.031	0.627	-0.061
Chl-a	0.001	-0.255	0.169
Mg^{2+}	0.040	-0.382	-0.156
HCO_3^-	-0.014	0.466	-0.521
CO_3^-	0.102	-0.529	0.040
TSS	-0.011	0.257	0.782
NO_2-N	0.183	0.397	-0.268
NO_3-N	0.111	-0.245	0.314
SRP	0.323	0.376	0.236
PP	-0.066	0.593	0.589
Org. N	0.123	0.367	-0.307

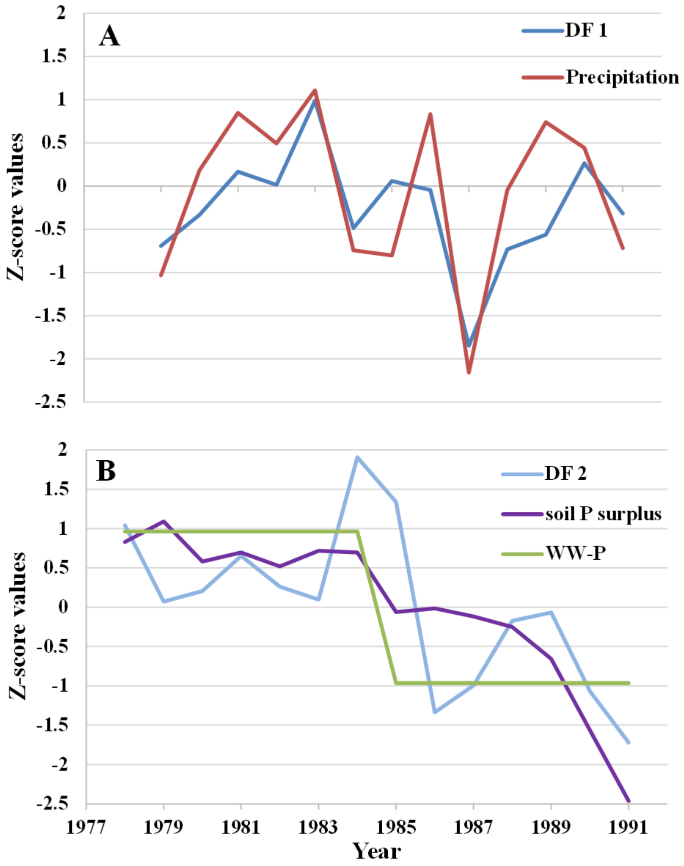
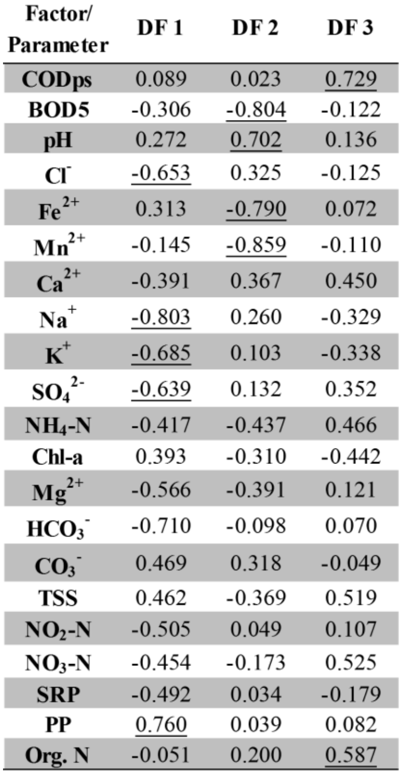


Figure 5.2.2-2 DF time series visualized together with explanatory parameters: DF1 with precipitation for the time interval 1979-1991 A) and DF2 with soil P surplus and WW-P for the time interval 1978-1991 B). The explanatory time series were Z-score standardized before plotting (reproduced from Hatvani et al., 2014b)

In the second time period (1992-2006) in the first DF, ions and PP had significant loadings, while in the second DF parameters were in close relationship to organic matter content and the redox conditions, and in the third they were in relationship rather with Org. N, COD_{ps} and TSS (Table 5.2.2-2).

Table 5.2.2-2 Factor loadings of the response parameters in the first three factors (1992-2006); significant loadings are underlined (reproduced from Hatvani et al., 2014b)



Factor/ Parameter	DF 1	DF 2	DF 3
CODps	0.089	0.023	$\underline{0.729}$
BOD5	-0.306	$\underline{-0.804}$	-0.122
pH	0.272	$\underline{0.702}$	0.136
Cl^-	$\underline{-0.653}$	0.325	-0.125
Fe^{2+}	0.313	$\underline{-0.790}$	0.072
Mn^{2+}	-0.145	$\underline{-0.859}$	-0.110
Ca^{2+}	-0.391	0.367	0.450
Na^+	$\underline{-0.803}$	0.260	-0.329
K^+	$\underline{-0.685}$	0.103	-0.338
SO_4^{2-}	$\underline{-0.639}$	0.132	0.352
NH_4-N	-0.417	-0.437	0.466
Chl-a	0.393	-0.310	-0.442
Mg^{2+}	-0.566	-0.391	0.121
HCO_3^-	-0.710	-0.098	0.070
CO_3^-	0.469	0.318	-0.049
TSS	0.462	-0.369	0.519
NO_2-N	-0.505	0.049	0.107
NO_3-N	-0.454	-0.173	0.525
SRP	-0.492	0.034	-0.179
PP	$\underline{0.760}$	0.039	0.082
Org. N	-0.051	0.200	$\underline{0.587}$

Exploring the factor loadings (response parameters) alone is not enough to obtain sufficient information about the background processes; for this the relationship of the factors and the explanatory parameters also has to be assessed. A more detailed analysis of the factor time series and the explanatory variables reveals that the first factor is closely related to the runoff of the River Zala (r = 0.74), and to the WW-P (r = 0.65; Fig. 5.2.2-3a). Despite the close linear relationship to WW-P, DF1 shows no significant correlation whatsoever with the soil P surplus (r = -0.076). In the meanwhile, the third factor corresponds to water temperature (r = 0.77; Fig. 5.2.2-3b). If there had been an explanatory parameter representing the soil organic matter content and redox conditions, it would presumably have correlated with DF2. This supposition is based on the response parameters’ correlation with the second DF; however, no unambiguous results were obtained.

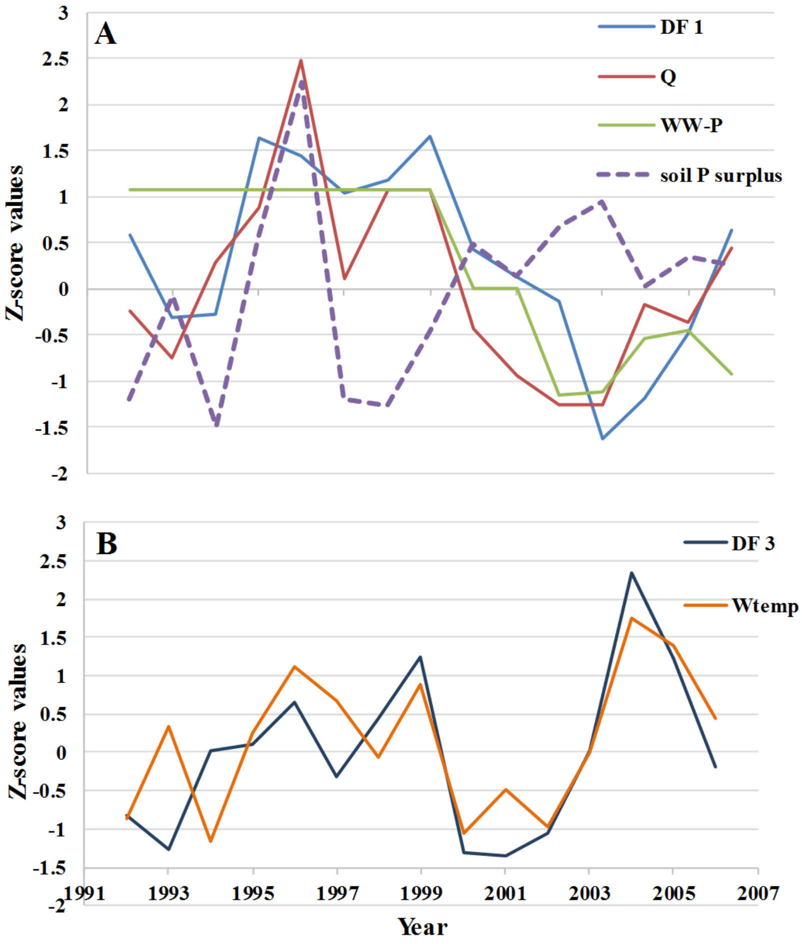


Figure 5.2.2-3 DF time series visualized together with explanatory parameters: DF1 with runoff and WW-P A) and DF3 with water temperature B) for the time interval 1992 – 2006. The explanatory time series were Z-score standardized before plotting (reproduced from Hatvani et al., 2014b)

5.2.3. Temporal sampling frequency analysis (*II.0/C – II.D*)

In the preceding chapters the temporal and spatial patterns were assessed. However, it should always be borne in mind that the primary purpose of the KBWPS is to retain the nutrients (brought by the River Zala) from the westernmost bay (Keszthely Bay) of Lake Balaton. Therefore, in order to track its efficiency - first of all- the N and P forms have to be measured. In this case uncertainty of nutrient load estimation is a key question (Clement, 2001; Somlyódy & Van Straten, 1986), and this in turn mostly depends on temporal sampling frequency. Therefore the aim of the sampling frequency (variogram) analysis was to *suggest a sampling frequency for the KBWPS* that is suitable to calculate annual nutrient loads and still produces a dataset in which the ecological processes can be followed.

As discussed in Section 4.2.4/Sampling frequency analysis the subject of the sampling frequency analysis were the detrended daily summer and winter total phosphorus and total nitrogen time series (1993-2006) at four sampling sites (Table 4.1-1). After removing the trend from the split time series 224 variograms were obtained for summer and winter for all the examined years. An example is presented in Fig. 5.2.3-1. A six day temporal range can be observed. Furthermore, the nugget effect is somewhat low, meaning that the measurement and/or sampling error was low. Of course there were cases when the semivariogram obtained was of the nugget effect type, but this happened in 8.8% of the cases.

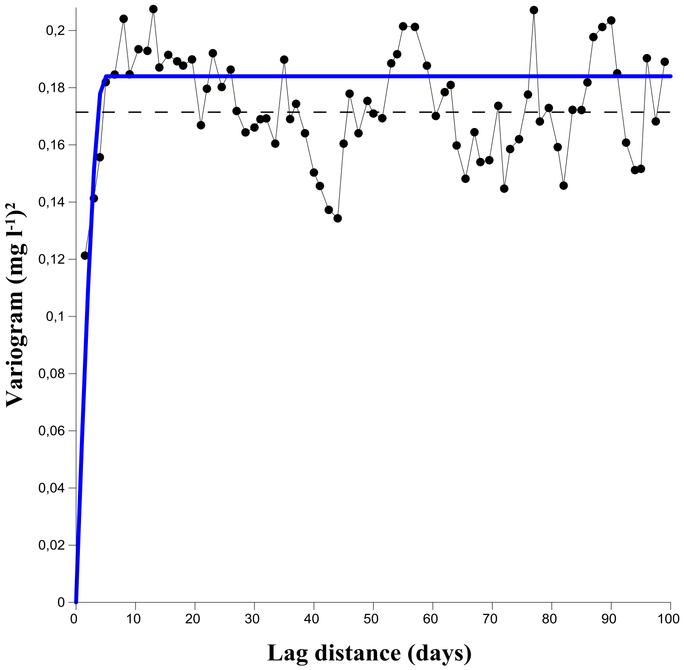


Figure 5.2.3-1 Theoretical semivariogram fitted on the empirical one, which was determined by the total nitrogen residual, sampling site Z27, 1993 winter

The range for every variogram was determined in days. If the variogram was of the nugget effect type then the range was considered to be zero. From these results the average range was calculated (Fig. 5.2.3-2) for TP and TN, 5 and 3.84 days respectively.

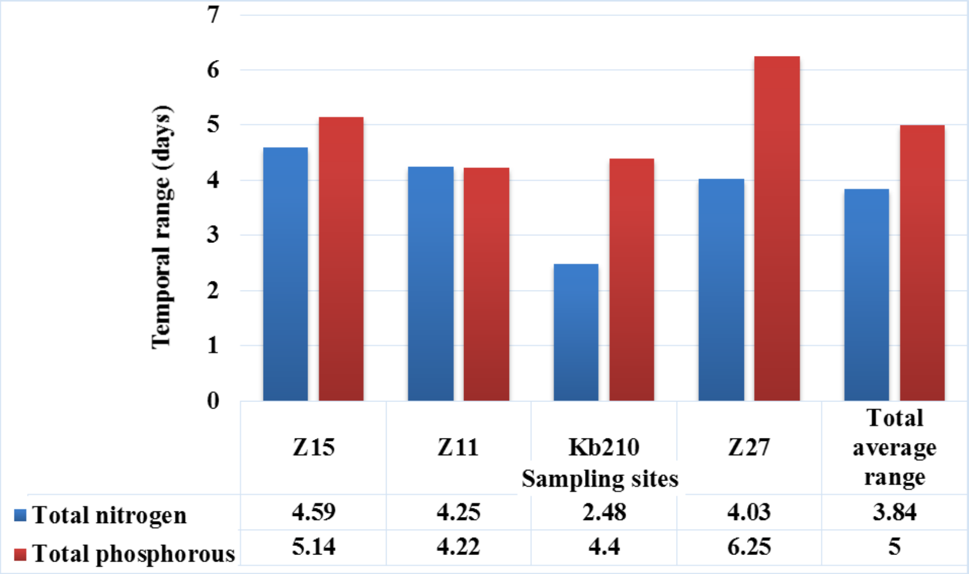


Figure 5.2.3-2 Average of the temporal ranges obtained from the variograms (1993-2006), regarding both TN & TP for each sampling site and the whole system as well (after Kovács et al., 2012b)

Because the sampling frequency of any system should be aligned to the most variable parameter, and from a practical point of view a monitoring system should be operating within a time frame measurable in whole numbers of days, after applying an uncertainty factor -three times the standard deviation of mean 3*SDOM = 0.59 (Taylor, 1997) - 3.84 was rounded downwards to three. Therefore it is suggested that the *temporal sampling frequency for TP and TN taking the system as a whole should be three days***.** It should be noted that the result only stands for the variables taken separately, further research is needed regarding their interrelations.

The results obtained from the summer and winter variograms were compared as well, but using a two-sample t-test for equal means, the difference between the two seasons was proven not to be significant, however, in the case of TP it was +0.53 days, while in the case of TN it was no more than -0.19 days.

Additional perspective on sampling frequency

As discussed in Section 4.2.5/Sampling frequency analysis the annually differenced nutrient loads were calculated in percentages from the daily and from the rarefied samples, whereby the annual nutrient loads of the former were taken as 100%. The percentage values were calculated for each year (1993-2006) then transformed into absolute values, called Annual relative differences (ARD). In fine the number and average of annual relative differences were calculated which are over the 5% threshold limit.

Generally, the number of years when annual relative differences peaked over the 5% threshold increases with the degree of the rarefaction at every sampling site. Naturally, between the areas of the KBWPS the increase in the annual relative difference (ARD) may differ in degree because of the different habitats and inflows.

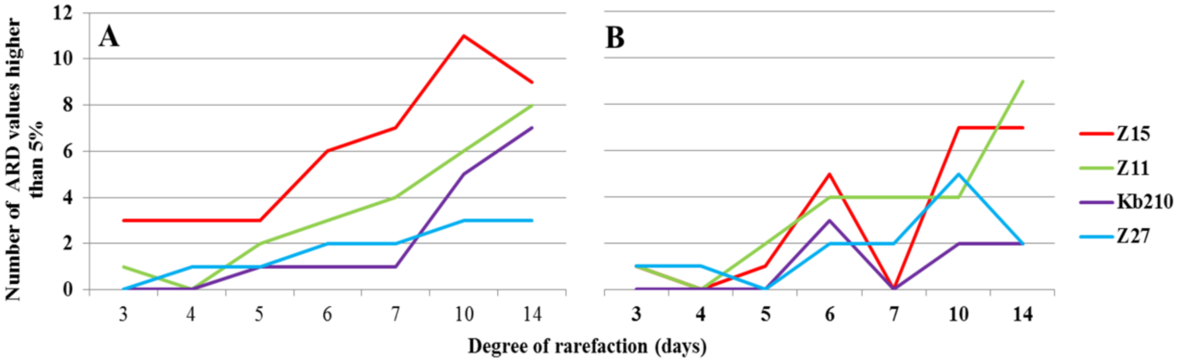


Figure 5.2.3-3 Number of ARDs higher than 5% at every sampling site (1993-2006), TP A); TN B) (after Kovács et al., 2012b)

The final step was to examine the average of annual relative differences (ARDs) exceeding 5% level at each SS. As an example TP is presented at sampling sites Z15 & Z27 (the inlet and outlet respectively; Fig. 5.2.3-4). This example reveals that if the rarefaction is increased from 3 to 4 days the average of annual differences (ADs) higher than 5% almost doubles from 5.4 to 8.2% at Z15. Furthermore, if the same question is put regarding sampling site Z27, which is the outlet to Lake Balaton, the average annual relative differences higher than 5% increases from 0% (indicating that there were no annual relative differences higher than 5% at all) to 6.03% (Fig. 5.2.3-4). The pattern observable in Fig. 5.2.3-4 is true for most of the other sampling sites, and for the TN parameter as well.

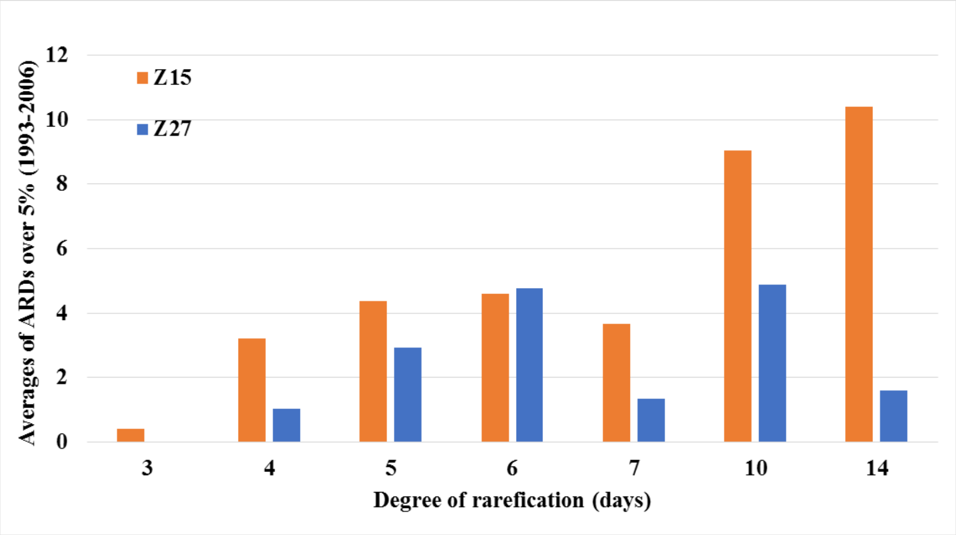


Figure 5.2.3-4 Example of how the average of ARDs (higher than 5%) alter with rarefaction regarding the TP parameter (1993-2006) at sampling sites Z15 and Z27 (after Kovács et al., 2012b)

5.2.4 Periodicity and memory analyses (*II.E & II.F*)

Since Hungary is located in the moderate climate zone, the processes in nature are likely to indicate annual periodicity driven by the seasons. Aquatic ecosystems follow these changes through the temperature and light conditions, so their chemical and biological parameters should show annual periodicity. If these systems are disturbed, for example by overflows, “acts of God”, pollution (be it natural or anthropogenic), or even limited in their connection with the light and temperature conditions, their capability to follow the seasonal changes and to sustain their memory (autocorrelation) is reduced. These were the phenomena explored during the wavelet spectrum (periodicity) and autocorrelation (memory) analysis.

Periodicity analysis (II.E)

During the periodicity analysis conducted on 22 weekly sampled water quality parameters at four sampling sites for 1993-2006 (table 4.1-1), annual periods were sought using the already-discussed wavelet spectrum analysis (WSA). Significant periodicity was expected, because of the shallow characteristics of the KBWPS, which should reflect the seasonal changes in the periodicity of its data. Unlike most methods for indicating periodicity, WSA was able to point out exactly when the annual periods were present in the data. A power spectral density graph (Fig. 4.2.4-4) was created for each parameter at each sampling site. These were then analyzed and the time periods when the annual periods were present were extracted (Figs. 5.2.4-1 & A5 in Appendix) - as discussed in Section 4.2.4/*Periodicity analysis* - and summarized for the whole KBWPS (Table 5.2.4-1).

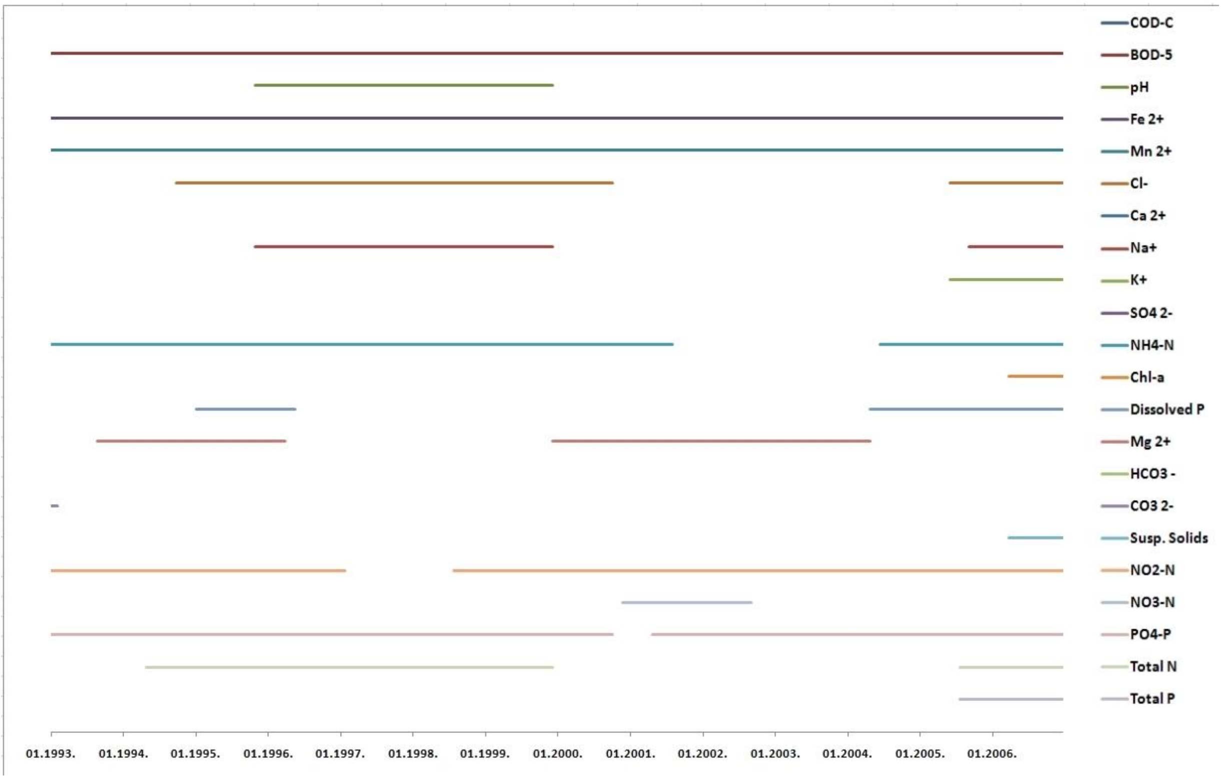


Figure 5.2.4-1 Wavelet spectrum analysis results showing the absence of annual periods at Z11 (interface; reproduced from Kovács et al., 2010)

Table 5.2.4-1 Percentage of annual periods at all four sampling sites (1993-2006)

Sampling site	Z15	Z11	Kb210	Z27
Parameters indicating annual periodicity with respect to the whole time period investigated	36.4%	59.1%	40.9%	31.8%

It can clearly be seen that the River Zala (Z15) is unlikely to indicate annual periodicity. On the contrary the parameters of the eutrophic pond (Z11) show periodic behavior over more than half of the time investigated. Weaker periodicity was expected from the area of Phase II where the extra inflows do not yet reach the wetland (before Kb210), and even weaker from the region after Kb210, where the canals coming from Somogy County reach the system (described by Z27).

Autocorrelation analysis (II.F)

The memory of the parameters at the four cardinal sampling sites was analyzed using autocorrelation analysis. The obtained autocorrelation functions were clustered, and six groups were obtained. As a next step curves were formed for the groups' averages, and these proved to be expressive of their parameters' memory (Fig. 5.2.4-2). Memory here is represented by the autocorrelation function, long memory (Beran, 1994), however is an asymptotic characteristic, which requires the hyperbolic temporal decay of the autocorrelation function. This was therefore not considered here, I only compared different patterns of memories corresponding to different groups.

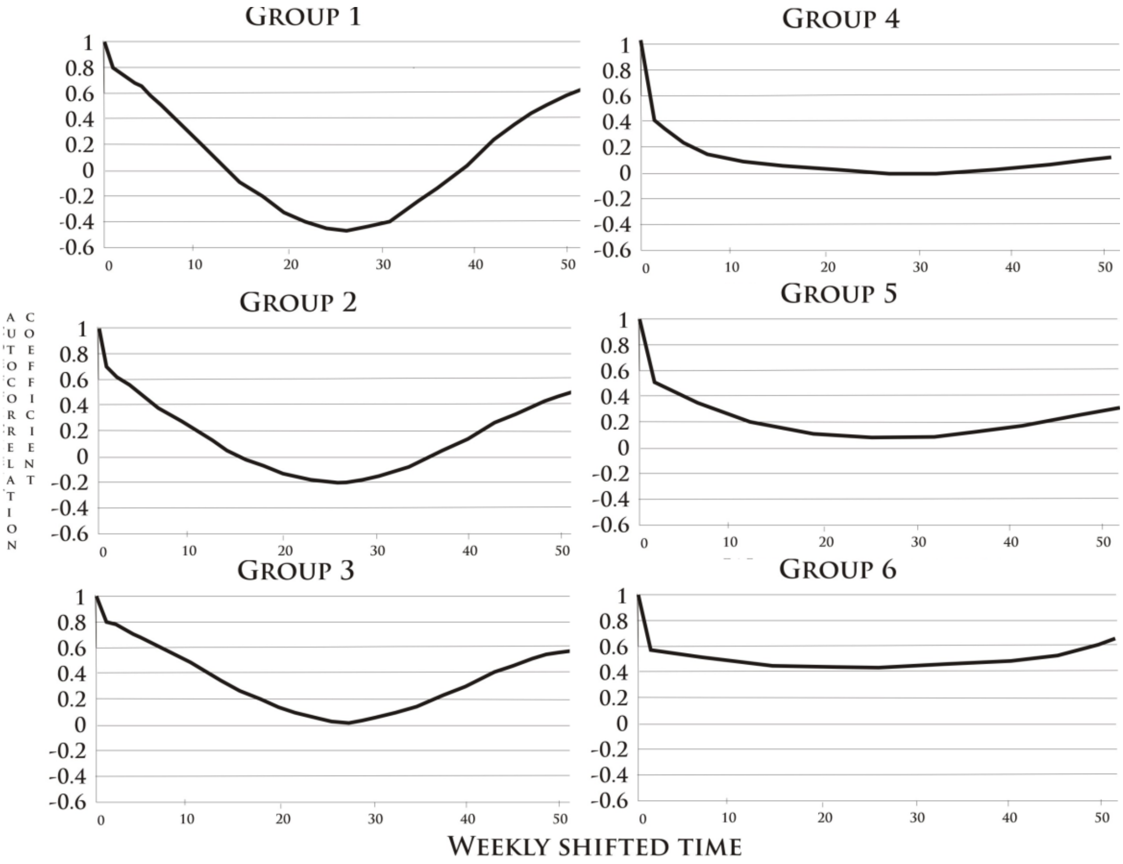


Figure 5.2.4-2 Groups formed from the autocorrelation results regarding the parameters' memory (reproduced after Kovács et al., 2012b)

It can be seen that the value of the autocorrelation function drops to a different extent with every one week lag; this was one of the main features that separated the groups from each other. In this study, it is considered as high autocorrelation where the autocorrelation coefficient ≥ 0.7 at a one week lag. The other feature was the shape of the autocorrelation function over a one year shifted time: the graphs can be classified regarding their shapes into two classes. One contains Groups 1, 2 and 3, which are highly periodic, while Groups 4, 5 & 6 are less or by no means periodic. It was interesting to see the percentage of the parameters in the groups based on their origin. The results of the analysis indicated that the groups with the highest autocorrelation coefficients at a one week lag (Groups 1, 2, 3) predominantly contain parameters from the eutrophic pond (Z11; Table 5.2.4-2).

Table 5.2.4-2 Percentage of parameters in all of the autocorrelation groups regarding their origin

Sampling site/ Group number	Z15	Z11	Kb210	Z27
1	21.4%	35.7%	28.6%	14.3%
2	18.8%	31.3%	25.0%	25.0%
3	28.6%	42.9%	14.3%	14.3%
4	26.5%	17.6%	17.6%	38.2%
5	6.7%	20.0%	26.7%	46.7%
6	0.0%	0.0%	0.0%	100.0%

It does not matter from which perspective the results are viewed from, be it that of the parameters in these three groups identified as 100% (Table 5.2.4-3a), or the number of the parameters on each sampling site (Table 5.2.4-3b), it always turns out that the parameters of the eutrophic pond are the most likely to be located in the groups with the highest autocorrelation coefficients at a one week lag.

Table 5.2.4-3 The distribution of parameters in the groups with the highest autocorrelation coefficients at a one week lag. The parameters in these three groups were identified as 100% A) and where the number of the parameters at each sampling site was identified as 100% B)

	Sampling site/ Group number	Z15	Z11	Kb210	Z27
A	**1, 2 & 3**	21.6%	35.1%	24.3%	18.9%
B	**1, 2 & 3**	36.4%	59.1%	40.9%	38.1%

It can clearly be seen that it does not matter which angle we examine the autocorrelation results from - the groups' or the sampling sites' - the results show the same tendencies. The memory and periodicity of the River Zala (Z15) is weak, while the eutrophic pond (Z11) is more likely to display periodicity, and its processes are more likely to show high autocorrelation, while in the area of the wetland which lies before the extra inflows (Kb210), these values (autocorrelation coeff. at a one week lag) decrease. This tendency also continues into the areas of the wetland after 210 as well, represented by Z27. It must be stated here that these are exactly the same tendencies indicated by the wavelet spectrum analysis.

From the angle of the parameters, the phosphorus forms tend to take place in the groups with the highest autocorrelation coeff. at a one week lag (1, 2 & 3, Table 5.2.4-4), in the entire Zala-KBWP system, . Dissolved- and total phosphorus are always present in the groups 1, 2 & 3 at every sampling site, as are the nitrate, calcium and sulphate ions. The Zala contains the fewest parameters indicating high autocorrelation at a one week lag, while Phase I contains the most. In the eutrophic pond, besides the parameters generally present in the groups with the highest autocorrelation coefficients at a one week lag throughout the whole KBWPS, the appearance of the Chl-a, TN and the TSS is particularly important. In the area of Phase II before Kb210 fundamentally the same parameters showed high autocorrelation at a one week lag and periodicity as they did in the disturbed area.

Table 5.2.4-4 Parameters in autocorrelation groups number 1, 2 & 3 regarding their origin

Sampling site / parameter	Z15	Z11	Kb210	Z27
Dissolved P	X	X	X	X
SRP	X	-	X	X
TP	X	X	X	X
K^+	X	X	X	-
Ca^{2+}	X	X	X	X
Na^+	X	X	X	-
SO_4^{2-}	X	X	X	X
NO_3-N	X	X	X	X
CODps	-	X	X	X
Chl-a	-	X	-	-
TN	-	X	-	-
TSS	-	X	-	-
Cl^-	-	X	-	-
HCO_3^-	-	X	-	-

6 Discussion

6.1 *Cluster-, discriminant, Wilks' λ and principal component analyses - I.0/B - D & II.0/A - A*

The sampling sites in the KBWPS were subdivided using cluster analysis, and basically formed thee groups. Two of these were the two constructional phases in the KBWPS, while the third Keszthely Bay. The border between the two groups covering the two constructional phases prior to 1997 was "located" after Z11, the actual border. However after 1997 the situation changed, and 202i disconnected from the cluster group covering the SSs of Phase II, and connected to those covering Phase I. This event could be explained by the degradation of the reeds in that area of the wetland surrounding 202i. This change was mainly caused by changes in water management. The areas (e.g. 202i) on higher ground used to be under periodical water coverage; however, after the initiation of the KBWPS project, these areas were permanently under water, with water inflow, characteristic of pH 7.76 (in 1995). This constant water coverage and the arrival of water with a higher pH caused the deterioration of the "native" sedge rush-bed vegetation (Pomogyi et al., 1996). As a result 202i turned into an open water body with processes similar to those in the nearby area of the pond. This change could even be followed by the high Chl-a concentration, which reached a summer average of 235 mg m^{-3} (Fig. 6.1-1).

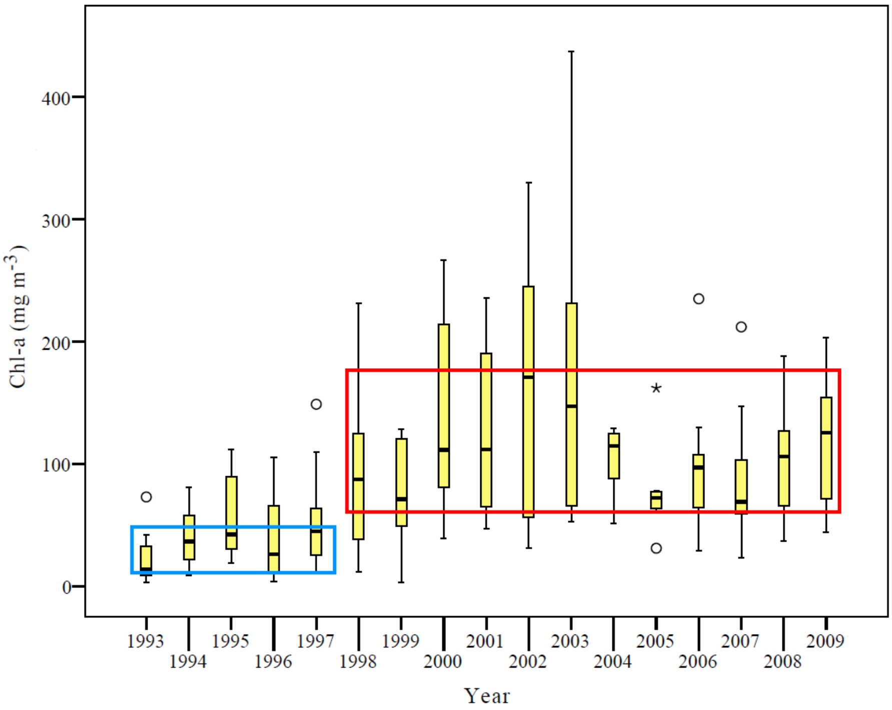


Figure 6.1-1 Chl-a values at 202i, where the blue rectangle marks the minimum and maximum median values when the SS belonged to the cluster group covering Phase II, and the red rectangle after it when it connected to Phase I

The changes witnessed in the alignment of the sampling sites, especially 202i, explicitly shows that the system is continuously changing, and the current border between the habitats stabilized not so long ago. However, if the water management of the system continues to stick to a constant residence time, it will presumably not be able to provide sufficient conditions for the wetland habitat to maintain its present size and vegetation density, although it is known that macrophyte cover and its effects are key questions in the success of mitigating wetland loss (Asaeda et al., 2001; Mitch el al., 2005).

The separation of SS 205 from the cluster group covering the wetland habitat can be explained by (i) the effects of the Combined Belt Canal bringing the outflow of Lake Héviz together with the treated sewage of the city of Keszthely, and (ii) the stagnant water conditions.

In detail, the reasons for the isolated state of SS Kb9 become clear if one is aware of the fact that it is located in a discrete area called the Cassette. This is the most eutrophic area in the whole system (even more eutrophic than the pond (Phase I.); the water is still, with almost no water flow. It operates independently from other parts of the system (Korponai et al., 1997). Bottom-up processes (Pernthaler, 2005) dominate which can be described by the Vollenweider model (Vollenweider & Kerekes, 1982).

These were the explanations which were presumed to lie behind the behavior of the SSs within the KBWPS. However, the discovery of which phenomena were responsible for the connection and disconnection of Kszt was not that straightforward. The connection between runoff and the behavior of Kszt became much clearer in the light of the results, and can be explained as follows: when runoff is low (1993 & 2000-2009) the flow of matter coming from the KBWPS is low as well. The water coming from Phase II (lower reservoir, wetland area) does not have such a great influence on the concentrations of Kszt. These assumptions concur with the high influence factor of the phosphorus forms and with their lower concentrations as well. The adverse effects of the decrease in nutrients were responsible for the separation of Kszt and for its oligotrophic trends. With the significant decrease in nutrient loads, the local environment (phytoplankton communities) gained in influence. At this time external loads decreased to such an extent that internal loads became much more determining (Istvánovics et al., 2004, 2007). Besides the KBWPS, investments in the development of the waste water infrastructure (Istvánovics et al., 2002, 2007; Hajnal & Padisák, 2008; Hatvani et al., 2011), and the significant decrease in the use of fertilizers (Sisák, 1993) in the catchment area of Keszthely Bay played an essential role in the decrease of the external nutrient loads.

In 1994-1999, despite high runoff and the increased external loads, Kszt did not “connect” and become similar to Phase I, where planktonic eutrophication is dominant; instead it displayed a resemblance to Phase II. Only a few parameters were found which could influence the formation of the cluster groups, and the system was considered to be more homogeneous.

However, the influence of Chl-a and BOD_5 could be explained by the low trophic conditions of Kszt enforcing the differences between the hypertrophic Phase I and the rest of the system. Keszthely Bay displayed a greater similarity to the macrophyte-dominated Phase II, where the amount of algae is low (Asaeda et al., 2001; Kovács et al., 2010; Tátrai et al., 2000). It must be stated that this similarity is a distance in an *N* dimensional space. The separated time periods were all distinguishable by reference to the trophic state of Keszthely Bay: whereas in the 1980s and early 1990s it was hypertrophic, in 1993 it was eutrophic, in 1994 hypertrophic again, and only after 1995 was it constantly eutrophic, with a mesotrophic year in 1999 (Vollenweider & Kerekes, 1982; Fig. 3-2b).

The changes seen in the cluster results induced by runoff were noticeable in the case of the PCA results as well, although as far as PCA is concerned, it is not the concentrations, but the stochastic interdependences which provide the bases of the analyses. There have been similar studies which have used the same approach in assessing water quality data; first grouping the observations, then the cases (Filik Iscen et al., 2008; Vega et al., 1998). The only difference is that in these studies the authors took more factors into consideration and used smaller factor loadings as a threshold. Either way, the first PC explained in all cases approximately 35% of the original data’s variance.

Regarding PCA, the background processes of Kszt underwent a change in the years 1981-2009. As time passed and measures were taken (inundation of Phases I & II etc.), the set of parameters closely related to eutrophication withdrew from the group of the most determining variables (Table 5.2.1-2), with the loading of Chl-a decreasing in the 1st PC and then disappearing. These phenomena indicate a change between the first (1981-1984) steady hypertrophic and a transient state of hysteresis (1985-1992) in Keszthely Bay (Scheffer, 1998). The third is an alternative, less eutrophic state (after 1993), with the complete absence of the parameters related to eutrophication from the 1st PC (Table 5.2.1-1). This state seemed to be stable.

Even the change in runoff and loads did not affect its stochastic relations after 1993. Only the initiation of the KBWPS, together with other measures, which had significant impact on the load reduction and changed the nutrient regime of the Zala River, as seen from the input-output analysis and the work of e.g. Somlyódy et al. (2003), was able to cause the ”disappearance” of the parameters corresponding to the nutrient cycle and eutrophication. Despite the fact that the efficiency of the KBWPS decreased (based on the input-output analysis), the results obtained from PCA clearly indicate that its effect cannot be ignored even today in the water quality amelioration processes of Keszthely Bay. The background processes of the eutrophic and mostly open water Phase I remained almost the same during the whole period analyzed. In the meanwhile the background processes of Phase II showed significant changes between years of high and low runoff. The reason behind this phenomenon is the vulnerability of the wetland area of Phase II (Hatvani et al., 2011; Kovács et al., 2010), which can easily be influenced by the fluctuating water regime (Davis et al., 2011; McCosker, 1998; Pomogyi et al., 1996) and other environmental factors (Copeland et al., 2010; Gilvear & McInnes, 1994; Wetzel, 2001). It is suspected that in drier years the dominance of nutrients increases because of the lower water level, warmer water, oxygen depletion in the overlaying water, and more intense internal P loads (Chambers & Odum, 1990). In summary, after reaching the partially finished state of the KBWPS (1992), neither the changes in the background processes of Phase II, nor the change in runoff and loadings was able to influence the solid state of the stochastic interdependences at Kszt, while its trophic processes (observable in continuous decrease 1981-1984 & 1985-1992) moved upstream from Keszthely Bay to Phase I (Tables 5.2.1-1 & 5.2.1-2). This was the basic idea when the KBWPS was planned (Kárpáti ed., 1980; Tátrai et al., 2000). The separation of the different measures' impacts on the external nutrient loads of Lake Balaton would require further analysis, which lies beyond the scope of the current research; however, the results presented herein could provide a good basis for further analysis of the question.

6.2 Dynamic factor analysis - II.B

It should be stated that although significant linear relations were observed between the DF time series and the soil P surplus along with the P output of the Zalaegerszeg waste water treatment plant, it is generally difficult to separate their effect. What does become clear, however, from the results is that there is a transition in the relationship of the DF and both the response and explanatory parameters, with a "border" at the year 1991. In the first time period, the DF time series correspond to both the diffuse and point source nutrient loads of the River Zala, while after 1991 this changes, and the role of the P originating from the waste water treatment plant increases, along with the runoff of the river and other parameters.

In the period 1978-1991 agricultural production was at its peak in almost all the countries of the Soviet Bloc, and Hungary was no exception. The use of fertilizers in agriculture boomed in the beginning of this period thanks to substantial subsidies and the highly industrialized development of fertilizer production. Precipitation generates surface runoff and soil erosion, which are the main transport mechanisms in the catchment (Albek, 2003). In the case of flood events and/or extreme weather conditions P compounds associated with soil particles are transported into the river, causing diffuse nutrient loads. Besides these diffuse loads, there was a significant point source related to the effluent of the Zalaegerszeg waste water treatment plant as well. This may be the explanation behind the fact that precipitation, fertilizer inputs (P surplus in the soil originating explicitly from agriculture), WW-P, and other related parameters (NH_4-N, PP) were represented in the DF time series. PP indicates the relationship of DF1 to agriculture, to WW-P and, presumably, to the stirring up of the sediment while NH_4-N represents only cleaned waste water inputs to the river. In the case of lower river flow velocities – according to the analyzed data in the period 1977 to 1985 by Sisák & Pomogyi (1994) – the TP content of the river was equal to the amount of wastewater origin P. The average SRP content was even lower than that of the wastewater (it is absorbed into solid particles). Above a 5 $m^3 s^{-1}$ runoff, P of surface origin appears in the TP content of the water (Sisák & Pomogyi, 1994).

Comparison of the two time periods

Between the two analyzed periods (1978-1991 & 1992-2006) a major difference was observed in terms of response and explanatory parameters as well. Between 1992 and 2006 none of the explanatory parameters related to agriculture were influential with respect to its loading; however, the relationship between the dynamic factor time series and WW-P remained, and moved up to the first factor. The disappearance of the agriculture-related parameters occurred because of the decrease in the use of fertilizers and other chemical compounds related to water quality deterioration (Herodek et al., 1995).

As for the further changes in the stochastic relations of the river's waters detectable by DFA:

1) Since DFA detects the processes and not the concentrations, it is obvious that the significant relations between runoff, PP and WW-P are due to the increased importance of P originating from the Zalaegerszeg waste water treatment plant, although it decreased in concentration. Measured PP concentrations in the Zala partly originated from effluent wastewater and partly in the sediment stirred up by heavy precipitation. Although these events

lasted for a short time, in terms of the background processes of the river they were quite determining.

A slow microbial using up of the phosphorus, together with runoff by the river water, results in loads of P being found rather in the KBWPS than in the river itself, but phosphorus concentrations will not increase parallel to river flow (Sisák & Pomogyi, 1994). When water discharge was low, the dilution effect became more important in the question of the appearance of parameters associated with waste water treatment plant effluents, e.g. WW-P and the ions. In this way the "constant" - and later on decreased - WW-P load became relatively more important with regard to the stochastic processes. It is clear that the phosphorus forms in rivers originating from agricultural production are mostly bound to soil particles (Tunney, 1997) and their sedimentation in the case of low flow conditions does not increase phosphorus concentrations in the water. Sisák and Pomogyi have also stated that in sub-surface flow (under 5 m^3 s^{-1} and generally as well) TP does not contribute significantly to the concentration of phosphorus forms in river water. On the contrary, flows caused by high runoff periods (floods; $Q > 5$ m^3 s^{-1}) do contribute to the concentration of P in the River Zala, but only for shorter time intervals (24-36 h), thereby causing short-lived diffuse loads. In these periods the variability of SRP and TP is so high that their origin cannot exactly be determined. One option might have been to partition the data based on flow velocity. However, doing this without daily measurements would have led to short intervals, thus providing insufficient information during DFA. Nevertheless, with E-DFA being employed on the annual averages it became quite obvious that it is of waste water, and not diffuse (surface erosion) origin. Thanks to these results, and the fact that the runoff of the Zala is generally under 5 m^3 s^{-1}, a cross can be set against erosion as the origin of P, and again it can be concluded that waste water is the factor behind this phenomena.

2) The phenomena observed in DF2 could be related to the following: soil absorbs nutrients (primarily P) based on its Mn, Fe content, pH, and redox conditions. This means that the external nutrient loads reaching the River Zala river, originating from land use, fertilizer use, waste water, etc. are exchanged with the inner processes of the soil that control its nutrient output.

3) Water temperature - as an influencing process governing the water's organic matter content - has a strict relationship with organic parameters such as Org. N and $COD_{ps,}$ representing a decrease in biological activity, which was the result of the decreased nutrient

loads. Thus, the reduced nutrient load leading to reduced microbial proliferation can explain the correlations found in DF3.

It can be concluded that, with the application of DFA, the consequences of historical events (or any other environmental phenomena) can be followed in time, and their governing background processes can be determined on a temporal scale.

6.3 Sampling frequency analysis - II.0/B – D

The sampling frequency analysis underlined the fact that there remains the possibility of decreasing the daily nutrient measurements to three days, which would still be able to provide comprehensive results in a more economical way; the costs of the sampling could be reduced by approximately 50%.

The result also contains valuable information from a data analysis perspective as well.

1) It provides information about the temporal nexus which goes beyond the recognition of the variance and mean.
2) The results obtained from daily measurements correspond to a longer, weekly time scale.
3) If methods where the temporal dependence of consecutive samples is a key question, such as PCA, are used, knowledge of the parameters' temporal range is vital. For instance, the sampling frequency of TP at Z27 reached 6 days; therefore, if one wanted to conduct PCA on weekly data, caution is required, because if less variable parameters are assessed, consecutive measurements in the dataset may be dependent.

Although there are only a few such studies, Füst & Geiger (2010), Hatvani et al. (2012), Kovács et al. (2005), used exactly the same approach as I did, but without the additional POT method.

6.4 Periodicity and memory analyses - II.E & F

The results of the Wavelet spectrum analysis clearly show that the River Zala (represented by Z15) is unlikely to indicate the annual periodicity. This can be explained by the continuously changing behavior of the riverine ecosystem. Here the biota, characteristic of streams and rivers, is mainly affected by the water level fluctuation. Its stochastic characteristics obscure the seasonal temperature conditions causing less periodic behavior. Furthermore, the river's hydrological regime and sudden diffuse nutrient loads effect annual periodicity in a disruptive way as well. This is the reason why annual periods can only be rarely found here.

On the contrary, the eutrophic pond (represented by Z11) follows seasonal changes. In addition it slows down the water flow and stabilizes the periodicity of the waters coming from the Zala. It has become a eutrophic limnological system. As a result it provides better conditions for algae growth, and these in turn can boost the periodical behavior with their seasonal reproduction and the cyclic planktonic eutrophic processes (Wetzel, 2001).

In the wetland area of Phase II (represented by Kb 210), the periodicity of the processes is reduced by the slower decomposition characteristic of the macrophyte dominated organic matter (Istvánovics et al., 1997), and the shade that the macrophytes provide. The shading results in smaller diurnal- and seasonal- temporal fluctuations, so every process that is light - or temperature - controlled loses something of its periodicity. This is a good example of how macrophyte cover changes the periodic characteristics of a once-eutrophic pond. To my knowledge, measurement of this type has not yet been performed, although macrophyte cover and its effects are key questions in the success of mitigating wetland loss (Asaeda et al., 2001; Mitch et al., 2005).

In fine, the outlet of the KBWPS (Z27) indicates even weaker periodicity than the wetland area represented by Kb210, because of the 25-30% extra incoming nutrient load brought by the Zalasomogyi- and the Marótvölgyi Canals (NYUDUVIZIG, 2012). These canals have the characteristics of a fluvial ecosystem. These have their own natural watersheds, which show much closer resemblance to a river, not to standing water. The unfinished area of Phase II is not capable of buffering these kinds of overflows, and that is the reason why the periodical nature of the processes is further reduced.

With regard to the autocorrelation analysis, regardless of the perspective the results are examined from, i.e. from either the perspective of (i) the groups (with the highest autocorrelation coeff. at one week lag), (ii) or the four sampling point, the tendencies and their causes would be almost exactly the same as those discussed in the explanation of the results of the wavelet spectrum analysis.

The results describe the nature of the different habitats of the KBWPS, but in order to gain a comprehensive picture of the processes, the set of parameters with the highest autocorrelation coeff. at a one week lag were examined. The results in terms of their origin can be explained by the following.

The River Zala and the whole KBWPS are phosphorus-dominated. This is the driving and at the same time the most limiting factor in the eutrophication. The reason why the phosphorus forms show high autocorrelation at a one week lag, and are also periodic in most cases, is because eutrophication itself is highly seasonal, and it is defined by the light and temperature conditions of the water. In the River Zala only a few parameters showed high autocorrelation on account of diffuse loadings and floods. When seasonal floods came, the concentration of these particular parameters rose and afterwards decreased.

In the eutrophic pond, among many others, the Chl-a, TSS, and total-nitrogen instantly took their places in the autocorrelation groups with the highest autocorrelation coeff., which is not surprising. It is a general truth that in a moderate climate biogenic cycles dominate the waters (Padisák, 2005) and compel the processes to display high autocorrelation and periodicity in most cases. In the water, metabolism is light- and temperature limited. In the summer, metabolism is faster, then it slows down in winter, and turns back on in spring. In Phase I planktonic eutrophication dominates with its diurnal and seasonal cycle. These systematic events cause the parameters of the eutrophic pond to show high autocorrelation.

The water then flows into the "undisturbed" area of the wetland (Phase II, Kb210), where the water body is 80-90% shaded by macrophytes. In particular 60-70% of the water surface is continuously covered by reed and sedge, the open water (20-30%) is dominated by *Ceratophyllum* sp. and *Polygonum* sp. and rush clumps can be found as well. The shade provided affects the water's light conditions, and damps the amplitude of the diurnal temperature oscillation. As a result, photoautotrophic processes are forced into the background, and the decomposition processes, along with the microbial loop, take their place. These processes are also defined by the temperature oscillations, but on a much larger scale. This is reflected in the absence of the parameters that describe eutrophication. Only those take place in the "high autocorrelation" groups that were among the ones controlled by the ox-redox conditions and decomposition. In the microbial loop the oxidation and nitrification processes are dominant (for a detailed list see Table 5.2.4-4). In both the undisturbed (Kb210) and disturbed (Z27) areas of Phase II in winter the processes of decomposition slow down, and the water's oxygen content is not used up, so the oxidation processes of sulphate, and in the summer its reduction processes, dominate. The only difference between the "undisturbed", and "disturbed" areas of Phase II, is that after the Zalasomogyi- and Marótvölgyi Canals' inflow reaches and disturbs the wetland, thus the number of components showing high autocorrelation at a one week lag is reduced.

7. Conclusions

After presenting and discussing the results and the applicability of the methods to the time series of the KBWPS, these have then to be considered from the perspective of the system. In the interests of a comprehensive interpretation, conclusions will be drawn from the results in a downstream direction, at first describing the inflow of the KBWPS, the River Zala.

Before 1991 the input of the KBWPS, the River Zala, could be described as having low algal activity and high nutrient loads, which decreased after 1991 as a result of the changes in agriculture and the processes governing diffuse and point source nutrient loads arriving to the River Zala, such as soil erosion, contamination and water treatment. These changes could only have been pointed out by DFA, since they are governed by processes which emerge through a set transcripts of such complexity that they cannot be directly determined. This is the reason why, although WW-P decreased in concentration, the importance of point source loads was nevertheless revealed without the significant presence of agricultural diffuse loads "covering" its effect (after 1991). The non-point loads and the hydrological regime – characteristic of a river – caused the River Zala to follow the seasonal changes in only one third of the time investigated.

These nutrient-rich (prior to 1991) and less nutrient-rich (after 1991) waters reach Phase I, the eutrophic pond of the KBWPS, at Z15 and leave it through Z11, the interface with Phase II, the wetland area. Since Phase I can be described as an open eutrophic water body, its most conspicuous characteristic is phosphorus retention. This retention is largely the work of algae. Although Phase I is highly eutrophic at all of its SSs, Kb9 occasionally separated from the others, since it lies apart from the system as pointed out by CA, and was the most eutrophic of all. Due to the seasonal reproduction of algae and the cyclic planktonic eutrophic processes Phase I was the "compartment" where the parameters displayed the highest autocorrelation and followed annual periodicity to the highest extent.

This algae-rich water enters the wetland, which after its inundation retained Chl-a content from entering Lake Balaton, more specifically, Keszthely Bay, by reducing the nutrients possibly available for algae (Van Donk et al., 1990). It should be stated that the border between the two Phases dynamically changed over the years, mainly because of the water level management of the system, as expressed in the CA result. One marginal extra input reached the system at SS 205 leading to its unstable state regarding its similarity (or lack of similarity) to the other SSs of Phase II, as shown in CA. The more significant extra loads reach the KBWPS after Kb210. These, besides bringing approximately 25-30% extra nutrient input to Keszthely

Bay, further degrade the already decreased capability (because of the shading provided by the macrophytes) of the wetland to follow the seasonality characteristic of the moderate climate. It is known that macrophytes are usually beneficial because they take up large amounts of nutrients during the year. However with regard to the periodic behavior of the KBWPS, their shading results in smaller diurnal-, and seasonal- temporal fluctuations, so every process that is light- or temperature - controlled loses something of its periodicity. In general this also adversely affects the otherwise characteristic ability of the parameters to show high autocorrelation. *It is presumed that with the additional areas the wetland will be less vulnerable and become a better and extended buffer zone for these nutrient loads, and will consequently be able to follow seasonality to higher degree.*

The waters “treated” in the KBWPS finally leave the area through Z27, entering Keszthely Bay. The statistical evaluation presented in this research confirms and supports the observation that nutrients have less effect on the background processes – the stochastic relationships pointed out by PCA - of Keszthely Bay nowadays than in the past. From the management point of view this is an important conclusion, although it is obvious that besides the inundation of the KBWPS, additional measures have contributed to decreasing the external nutrient loads to the bay and eventually to its oligotrophication processes. With PCA and CA and the additional input-output analysis it has been demonstrated that the construction of the KBWPS has facilitated the establishment of an ecological integrity between the reservoirs and Keszthely Bay, resulting in a new, alternative stable (independent) state (May, 1977; Scheffer et al., 1993). This equilibrium can still be influenced by external effects (Scheffer et al., 2001), in this case runoff and external loads entering the system, although only to a moderate degree, and then not in such an immediate way. It can be described by:

(i) a decreased inter-annual variability, which does not cause significant growth in macrophyte cover because of sediment resuspension and winds (Istvánovics et al., 2008). (Here it should be remarked that, although there have not been direct measurements, the public have reported the increase of submerged macrophytes in the shoreline);

(ii) decreased external nutrient loads

(iii) a decrease in internal nutrient loads

  a. the rapid immobilization of mobile P in the sediment (Istvánovics & Somlyódy, 2001),

    b. among many others, sediment containing high amounts of P is covered by new layers, thus remobilization is unlikely.

(iv) low - or no - discharge from the KBWPS in dry years, especially in summer months . This is one reason why Kszt "behaves" differently in these situations. On the contrary, the background processes in the wetland area of the KBWPS seem to be highly influenced by runoff.

I was able to point out key changes in the trophic processes of Keszthely Bay over a time period of many decades, which lead Keszthely Bay from being hypertrophic in 1981-1985 and 1985-1992 to becoming mezo/eutrophic in 1993-2009. The KBWPS works as a buffer, lessening the impacts of extreme flow, and as consequence, possible extreme load events. It also retains the nutrients from entering the load-sensitive Keszthely Bay by providing a place for intensive planktonic eutrophication. These processes are therefore moved upstream to Phase I. This facilitates the oligotrophication of Keszthely Bay. The identification of these key changes in the system was the result of the interaction of the parameters, and could not have been made using solely univariate methods. The beneficial effect of the KBWPS on the oligotrophication of Lake Balaton has been proved in this work from the perspective of concentration and stochastic relationships as well.

In 2012 the completion of Phase II began, so I hope that my results may be of help to scientists in gaining a broader perspective on the processes evolving in the KBWPS, as also on the further oligotrophication on Keszthely Bay. To my knowledge, the present plans - in contrast to the previous ones – place greater emphasis the element of nature preservation in the KBWPS. The already-finished area of Phase II is going to be detached, and for nature preservation reasons the water level will be altered manually in harmony with the actual weather conditions. Measures such as this were successful in preserving and helping re-colonization by water-plants e.g. in the "Alte Donau" (Dokulil et al., 2006). On the contrary, the as-yet unfinished part will be managed with constant water level in order to retain the nutrients brought by the River Zala and the canals connecting to it from the South.

Naturally, the newly-finished areas will have to be monitored as well. From the VA results it became clear that, in the case of recalibration, it could be suggested that the current daily sampling be reduced to a maximum three days sampling frequency. This is valid in the case of nutrient load estimation for the whole system (present conditions). Although this sampling frequency is rarer than daily, it has been demonstrated that it is capable of providing comprehensive results, and in a more economical way, as the costs of the sampling could be reduced by approximately 50%. Based on this, the same sampling frequency may be applied to the unfinished part of Phase II as well. However a 26-week daily sampling is suggested for verification, because the characteristics of the examined and the as-yet unfinished area differ to a great extent.

The achievement of this study has been to (i) point out changes in the KBWPS over a long time period, and (ii) extract and discuss excess information from the data using various stochastic methods which have in most cases never before been applied to such a system. It has also underlined the importance of the completion of the unfinished constructional Phase II, so its buffer capacity could be enlarged. Hopefully my work will be of help to scientists who wish to gain a broader perspective on processes evolving in the KBWPS and on the further oligotrophication on Keszthely Bay.

8. List of theses (*new scientific results*)

1) The separation of point and non-point nutrient loads is extremely difficult. However, despite this, with the aid of dynamic factor analysis, changes in the governing processes of the River Zala were determined, specifically the decreased role of diffuse-, and the increased role of point source nutrient loads, as well as biological activity, as indicated by the role of water temperature.

2) By means of cluster analysis it was found that the border between Phase I and Phase II changes dynamically, the most explicit change being after 1997, when sampling site 202i disconnected from the cluster group covering Phase II and connected to the one covering most of the SSs of Phase I. This happened as a result of water management issues (constant water level), which degraded the macrophyte vegetation in the vicinity of SS 202i.

3) The use of input-output analysis enabled a discussion of the P retention capacity of the KBWPS, in which the decreased overall output to Lake Balaton could be demonstrated. This was due to the economic changes following 1989, the water quality ameliorating measures taken, and the beneficial effects of the KBWPS. Furthermore, it becomes clear that since the publication of the work of Somlyódy et al. (2003) the nutrient retention capacity of the system has not changed.

4)

  a) Using cluster- and principal component analysis with Wilks' lambda statistics, it was found that the adverse decrease of nutrients was responsible for Keszthely Bay forming a cluster group alone in the low runoff years from the perspective of the KBWPS.

  b) With reference to the determining role of runoff, it was found that as far as the stochastic relations are concerned, in the years with low runoff, Keszthely Bay was not the area where the most influential parameters underwent a change. On the contrary, before the year 2000 - when runoff was low - the inorganic parameters were those which dominated the water body of Phase II to the greatest extent, which after the year 2000 became co-dominated by the parameters describing eutrophication and the nutrient cycle. This was the period when Keszthely Bay formed a distinct area with regard to the KBWPS.

5) Using periodicity analysis a new perspective was gained on the vulnerability of wetlands. As seen from the input-output analysis, 40% extra runoff and approx. 25-30% extra nutrient input is brought by the canals coming from Somogy County. These disrupt the capacity of the water quality parameters measured in Phase II to follow annual periodicity. Therefore, if the buffer capacity of the wetland were to be increased against the extra nutrient loads with the inundation of the remaining areas, it would presumably be able to follow the seasonal changes (i.e. indicate annual periodicity) to a higher degree. In terms of periodic behavior it would be less influenced by sudden nutrient loads.

6) Based on the elementary time series and cluster results it has been suggested that, in order to preserve the wetland area of Phase II, its water-level management should be altered so that it is able to follow the seasonal fluctuations.

7) Using cluster-, principal component- and Wilks' lambda analyses the new stable state of Keszthely Bay was determined, where the KBWPS functions as the location of intensive planktonic eutrophication, which facilitates the movement of the eutrophication processes upstream to Phase I. (*After the after reaching the partially finished state of the KBWPS (1992), neither the changes in the background processes of Phase II nor the changes in runoff and loadings were able to influence the solid state of the stochastic interdependences at Keszthely Bay, while its trophic processes (observable in their continuous decrease 1981-1984 & 1985-1992) moved upstream from Keszthely Bay to Phase I.)*

8) Based on the coded cluster result it has been proven, from the perspective of Lake Balaton, that both the outlet of the KBWPS and Keszthely Bay form separate water quality areas (water-bodies) of Lake Balaton after 1998.

9) Using semivariograms it was determined that the present daily sampling could be rarefied to three days regarding TN & TP, if the aim is nutrient load estimation. Such a step would result in a reduction in water management costs and the more economic functioning of the monitoring network. This principle could have wider applicability in the management of other water bodies as well.

9 English abstract

Lakes are sensitive to environmental changes and anthropogenic effects, and Lake Balaton, the largest shallow freshwater lake in Central Europe, is no exception. To protect it (and primarily Keszthely Bay) against elevated nutrient loads, the Kis-Balaton Water Protection System (KBWPS) was constructed as a mitigation wetland at the mouth of the River Zala, the largest tributary of the lake.

This work aims to demonstrate the applicability of various multivariate state-of-the art geomathematical methods in water protection, and in parallel to explore the long-term time series of the KBWPS to describe its main processes related to water quality.

In the course of the research the time series of a total of 24 weekly and 3 daily sampled natural parameters (variables) were used from 14 sampling sites for various time intervals, out of which the longest was 1977-2009. The analyzed dataset comprised approx. 250,000 data. The methods used comprised the following: descriptive statistics, cluster-, discriminant-, principal component-, dynamic factor-, variogram-, wavelet spectrum and autocorrelation analysis.

First the descriptive statistics were formed, after which the sampling sites were clustered for each year, showing that the border between the two Phases changes dynamically, and furthermore, when the runoff in the system was low, Keszthely Bay behaved as a separate entity. In the years with high runoff, however, it displayed a similarity rather to the wetland area of the KBWPS. These results were verified in more than 95% of the cases with discriminant analysis. Secondly, PCA showed that, from the perspective of stochastic relations, the parameters of Keszthely Bay which stand in connection to eutrophication decreased in influence after the settling of the KBWPS. Furthermore, the eutrophication processes moved upstream from Keszthely Bay to the hypertrophic area of KBWPS. This study confirms that the effects of the KBWPS - together with other load reduction measures – have had a beneficial effect on the oligotrophication of Lake Balaton. As for the River Zala, the main input of the KBWPS, it was pointed out that the role of the diffuse and point-source nutrient loads governing most of the river's processes changed in 1991 because of the socio-economic changes and the measures taken to reduce point source nutrient loads. The penultimate step in the process was periodicity and autocorrelation analyses, which in fact pointed out that the incomplete Phase II is unable to conserve the periodicity of the system and sustain a high autocorrelation because of the shading provided by the macrophyte cover, as well as the extra inflows. As a final step the practice of temporal sampling frequency estimation was described in order to suggest a sampling frequency that is suitable for estimating annual nutrient loads and which nevertheless produces a dataset within which environmental processes can be followed. The results pointed towards a three-day sampling frequency for the whole KBWPS in the case of total phosphorus and total nitrogen (if the purpose of the monitoring is nutrient estimation). This reduction in sampling frequency would make an approx. 50% cost reduction possible.

The main achievement of this study was that it pointed out changes in the KBWPS over a long time period using various stochastic methods, most of which have individually never before been applied to such a system, much less in an integrated way. It also underlines the importance of the completion of the unfinished constructional Phase II, so its buffer capacity could be enlarged. Hopefully my work will be of assistance to scientists who wish to gain broader perspectives on both the processes evolving in the KBWPS, and also on the further oligotrophication of Keszthely Bay.

10 Magyar nyelvű összefoglalás

A felszíni vizek között a sekély tavak fokozottan érzékenyek a környezeti változásokra és antropogén beavatkozásokra. Ez alól Közép-Európa legnagyobb sekély vizű tava, a Balaton sem kivétel. A XX. század második felében tapasztalt vízminőség romlás megállítására létrehozták a Kis-Balaton Vízvédelmi Rendszert (KBVR), melynek célja, hogy két konstrukciós ütemével visszatartsa a Zala folyó által szállított növényi tápanyagokat a tó nyugati medencéjétől, a Keszthelyi-öböltől.

A dolgozat célja a KBVR-ben zajló vízminőségi folyamatok és arra vonatkozó új ismeretek feltárása modern geomatematikai módszerek segítségével.

A kutatás során 24 heti, és három, napi mintavételezésű paraméter (valószínűségi változó) idősora képezte a vizsgálat tárgyát 14 mintavételi pontról, 1977-2009-es időintervallumból. A felhasznált adatok száma közel 250 000 mért adat volt. Az elengedhetetlen adatelőkészítés és a leíró statisztikák kiértékelése után, klaszter-, diszkriminancia-, főkomponens-, dinamikus faktor-, variogram-, wavelet spektrum analízist és az autokorreláció függvények becslését végeztem el.
A mintavételi pontokat klaszter analízissel évenként csoportosítottam. Ennek eredményeiből megállapítható, hogy a két Ütem közötti határ dinamikusan változott a vizsgált évek során, továbbá, hogy az alacsony vízhozamú években a Keszthelyi-öböl különálló csoportban helyezkedett el, ami arra utal, hogy a belső anyagforgalmi folyamati eltérnek a KBVR anyagforgalmától. Ugyanakkor, a Keszthelyi-öböl, a nagy vízhozamú években a KBVR vizes élőhelyéhez (II. Ütem) hasonlított, ezekben az években azonban a KBVR területéről érkező tápanyagterhelésnek hatása van az öböl belső anyagforgalmi folyamatira. A klaszter analízissel kapott csoportok létezését a diszkriminancia analízis erősítette meg. A főkomponens analízis eredményei megmutatták, hogy a Keszthelyi-öböl eutrofizációhoz köthető paraméterei az évek során folyamatosan kikerültek a varianciát leginkább meghatározó főkomponens(eke)t befolyásoló paraméterek köréből, valamint az eutrofizációs folyamatok átkerültek Keszthelyről az I. Ütembe. A kutatási eredmények bizonyították, hogy a KBVR üzembe helyezése és további tápanyagterhelés csökkentő intézkedések jótékony hatással voltak a Keszthelyi-öböl oligotrofizációjára. A KBVR befolyója, a Zala meghatározó folyamataiban 1991 után a diffúz terhelések szerepe lecsökkent és a pontforrás eredetű tápanyagterhelésé megnőtt. A folyamatok periodicitása szempontjából a II. Ütemen kapott eredmények a leglényegesebbek. Ezek rámutattak arra, hogy a wetland nem követi az évszakos váltakozást (éves periódust), mint ahogy azt az eutróf I. Ütem teszi. Azt is sikerült kimutatni, hogy a wetland jelenlegi állapotában nem képes „tompítani” a beérkező vízfolyásokból származó többletterheléseket. Végül, az időbeli mintavételezési gyakoriság becslés alapján kijelenthető, hogy a jelenlegi napi mintavételezési gyakoriság (összes foszfor- és nitrogén esetén) három napra ritkítható jelentős információ veszteség nélkül. Ez hozzávetőleg 50%-os költségcsökkenést eredményezhet.

Eredményeim, melyeket sokváltozós feltáró adatelemzéssel értem el, úgy vélem, hozzájárulhatnak annak megértéséhez, hogyan viselkednek a különböző vízi ökoszisztémák egy rekonstrukciós beavatkozás során. A KBVR hatásvizsgálatával sikerült bemutatnom, hogy a Balatonban tapasztalt kedvező vízminőség javulásban milyen jelentőséggel bír a Kis-Balaton Vízvédelmi Rendszer.A gyakorlati hidrobiológia egy átfogóbb képet kaphat a Kis-Balaton Vízvédelmi Rendszerben zajló folyamatokról, így remélem, érdemben hozzá tudok járulni a Kis-Balaton befejezésének munkálataihoz.

11 Deutsche Zusammenfassung

Seichte Gewässer reagieren auf ökologische Schwankungen sowie auf menschliche Einflüsse besonders sensibel. Dies gilt auch für den Plattensee (ungarisch: Balaton), welcher den größten seichten See in Zentraleuropa darstellt. Um die Verschlechterung der Wasserqualität zu verhindern, welche in der zweiten Hälfte des zwanzigsten Jahrhunderts entdeckt wurde, ist das Kis-Balaton Wasserschutzsystem (ungarisch: KBVR, Kis-Balaton Védelmi Rendszer) etabliert worden. Das Hauptziel war es die pflanzlichen Nährstoffe, welche durch den Fluss Zala zufließen, von der Bucht Keszthely fernzuhalten. Geplant ist, dass das Wasserschutzsystem an der Mündung des Flusses einen eutrophen Lebensraum sowie ein „klassisches“ Feuchtgebiet ermöglicht. Die Durchführung der Bauarbeiten wurde in zwei Phasen geplant. Das Gebiet der Phase 1 konnte im Jahr 1985 in 5 Schritten beflutet werden, während das Gebiet der Phase 2 den Zielzustand noch nicht erreicht hat: Im Jahr 1992 konnte man nur einen Bereich von 16 km2 übergeben, wobei die Fertigstellung aktuell für den Zeitraum von 2013-2014 geplant ist.

Das Ziel dieser Arbeit ist einerseits eine Darstellung über die Anwendung moderner geomathematischer Methoden im Bereich des Wasserschutzes, anderseits die Verarbeitung der langfristig erhobenen Daten im Prozess des Wasserschutzsystems. Zur Untersuchung wurden Daten von 24 Wochen und 3 Tagen, von 14 unterschiedlichen Standorten, in einem Zeitintervall von 1977-2009 herangezogen. Die analysierte Datenquelle umfasst in etwa 250 000 Dateneinträge.

Nach den notwendigen Datenvorbereitungen und den deskriptiv-statistischen Analysen habe ich folgende Methoden angewandt: Cluster-, Diskriminanz-, Hauptkomponenten-, dynamische Faktoren-, Variogramm-, Wawelet Spektrum- und Autokorrelationsanalyse.

Die Standorte der Proben wurden zunächst durch Clusteranalysen in einer jährlichen Abbildung gruppiert. Aus den Ergebnissen kann deutlich festgestellt werden, dass sich die Grenze zwischen den 2 Gebieten dynamisch verändert hat. Ferner wurde eine Clusteranalyse bei der Auswertung der Wirkung des Wasserschutzsystems auf die Wasserqualität des Balatons und der Bucht Keszthely herangezogen. Die Ergebnisse zeigen, dass die Bucht Keszthely in den Jahren, in welchen eine niedrige Durchflussmenge verzeichnet wurde, im Vergleich zum Gebiet des Wasserschutzsystems, unterschiedliche Nährstoffwerte aufweist und daraus gefolgert werden kann, dass in solchen Jahren zwei unterschiedliche Systeme bestehen.

Betrachtet man dagegen die Jahre mit einer hohen Durchflussmenge, ähneln die Werte in der Bucht dem Lebensraum des Wasserschutzsystems (Gebiet #2). In solchen Perioden hat die Nährstoffbelastung des Wasserschutzsystems eine ausgeprägte Wirkung auf die internen Prozesse der Bucht.

Die Diskriminanzanalyse hat die Gültigkeit der Gruppen, die durch die Clusteranalyse erstellt worden sind, in 95% der Fälle bewiesen. Die Ergebnisse der Hauptkomponentenanalysen zeigen die Tendenz, dass die Parameter der Eutrophierung der Bucht Keszthely im Laufe der Jahre immer weniger an Bedeutung als Parameter haben, wenn es um die Beeinflussung der Varianz geht. Die Prozesse der Eutrophierung haben sich aus Keszthely in das Gebiet #1 verlagert.

Die Ergebnisse dieser Forschung haben damit bewiesen, dass die Herstellung des Wasserschutzsystems und die Etablierung weiterer Maßnahmen zur Nährstoffminderung, eine positive Gesamtwirkung auf die Oligotrophisierung der Bucht Keszthely haben.

Nach 1991 konnten wesentliche Veränderungen in der Wasserqualität des Flusses Zala festgestellt werden, welcher das Wasserschutzsystem nährt. Die Rolle der diffusen Belastung ist gesunken während die Nährstoffbelastungen welche aus Punktquellen stammen zugenommen haben.

Aus der Sicht der Periodizität sind die Ergebnisse des Gebietes #2 am bedeutendsten. Diese verdeutlichen, dass sich das Feuchtgebiet ganz anders verhält, als die Lebensräume in offenen Gewässern da sie keinen Jahreszeiten folgt, wie dies in eutrophen Gebieten wie #1 zu beobachten ist. Die Resultate stellen auch dar, dass das Feuchtgebiet die zusätzlich hereinströmenden Nährstoffbelastungen nicht ausreichend mindern kann.

Eine Frequenzanalyse zeigt auf, dass die Häufigkeit der Stichprobenziehung (Phosphor und Stickstoff betreffend) von einer täglichen Erhebung auf 3mal pro Woche verringert werden kann, ohne einen wesentlichen Informationsverlust in Kauf nehmen zu müssen. Diese Maßnahme könnte des Weiteren eine Kostreduzierung von 50% erwirken.

Meiner Meinung nach können die dargestellten Ergebnisse für das Verständnis bei Veränderungen in unterschiedlichen Wasserökosystemen während eines Rekonstruktionsprozesses von Bedeutung sein.

Durch die Wirkungsanalyse des Wasserschutzsystems soll die Bedeutung von diesem im Hinblick auf die Wasserqualitätsverbesserung im See Balaton verdeutlicht werden. Die Darstellung der Prozesse welche im Kis-Balaton Wasserschutzsystem ablaufen können der angewandten Hydrobiologie ein umfassendes Bild bieten. Hoffentlich wird durch meine Arbeit die Bedeutung der Fertigstellung des Gebietes am Kis-Balaton deutlich und trägt zu dieser positiv bei.

12 References

Afifi, A., Clark, V. A., May, S., Raton B., 2004. Computer-Aided Multivariate Analysis (4th ed.). Chapman & Hall/CRC, USA, p. 489.

Agterberg, F.P., 1974. Geomathematics. Mathematical Background and Geo-Science Applications. Elsevier Scientific Publishing Company, USA, 596p.

Albek, E., 2003. External Phosphorus Load Estimates and P-Budget for the Hydroelectric Reservoir of Bort-les-Orgues, France. Water, Air, and Soil Pollution 147, 229–243.

Asaeda, T., Trung, V.K., Manatunge, J., Bon, T.V., 2001. Modelling macrophyte–nutrient–phytoplankton interactions in shallow eutrophic lakes and the evaluation of environmental impacts. Ecol. Eng. 16, 341–357.

Baker, L.A., 1992. Introduction to non-point source pollution in the United States and prospects for wetland use. Ecol. Eng. 1, 1–26.

Bartram, J. & Chorus, I., (eds.) 1999., Toxic Cyanobacteria in Water: a guide to their public health consequences, monitoring and management. Geneva, World Health Organization.

Benedetto, J. & Frazier, M. (Eds.), 1994. Wavelets: Mathematics and Applications. Boca Raton, FL, USA: CRC Press, 575p.

Beran, J., 1994. Statistics for Long-Memory Processes. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 315p.

Boersen, P.M.T., 2006. Automatic autocorrelation and spectral analysis. Springer-Verlag London, UK, 298p.

Borg, I. & Groenen, P.J.F., 2007. Modern multidimensional scaling – Theory and applications. Springer Science + Business Media, USA, 613p.

Brigault, S. & Ruban, V., 2000. Estimation of Point and Diffuse Contaminant Loads to Streams by Non-Parametric Regression Analysis of Monitoring Data. Water, Air, and Soil Pollution 119, 91-103.

Chambers, J.M., Rigley, T.J., McComb, A.J., 1993. The potential use of wetlands to reduce phosphorus export from agricultural catchments. Fert. Res. 36, 157–164.

Chambers, R.M. & Odum, W.E., 1990. Pore-water oxidation, dissolved phosphate and the iron curtain. Biogeochemistry 10, 37-52.

Chan, H.K., Hayya, J.C., Ord, K.J., 1977. The case of Polinomial Regression versus Variate Differencing. Econometrica 45, 737-744.

Chilès, J-P. & Delfiner, P., 2012. Geostatistics. Wiley, Canada, 731p.

Clement, A., 2001. Improving uncertain nutrient load estimates for Lake Balaton. Water Science and Technology 43, 279-286.

Clement, A., 2013. Personal communication, 8. December 2013.

Clement, A., Somlyódy, L., Koncsos, L., 1998. Modeling the phosphorus retention of the Kis-Balaton Upper Reservoir. Wat. Sci. Tech. 37, 113-120.

Cleneghan, C., 2003. Phosphorus Regulations; National Implementation Report. Environmental Protection Agency, Wexford, Ireland.

Cleveland, W.S. & Devlin, S.J., 1988. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association 83, 596-610.

Cleveland, W.S., 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association 74, 829-836.

Cook, S. R. & Rubin, D.R., 2007. Imputation. Wiley Encyclopedia of Clinical Trials, Wiley Online Library.

Copeland, H.E., Tessman, S.A., Girvetz, E.H., Roberts, L., Enquist, C., Orabona, A., Patla, S., Kiesecker, J., 2010. A geospatial assessment on the distribution, condition, and vulnerability of Wyoming’s wetlands. Ecol. Ind. 10, 869-879.

Davis, J., Froend, R., Hamilton, D., Horwitz, P., McComb, A., Oldham, C., Thomas, D., 2011. Environmental flows initiative technical report, Report number 1: Environmental Water Requirements to Maintain Wetlands of National and International Importance. Environment Australia, Canberra.

Day, W.H.E. & Edelsbrunner, H., 1984. Efficient algorithms for agglomerative hierarchical clustering methods. Journal of Classification 1, 7–24.

Dianmo, L., Zhongwei, G., 2000. Some aspect of ecological modeling developments in China. Ecol. Model. 132, 3-10.

Dokulil, M.T., Donabaum, K., Pall, K., 2006. Alternative stable states in floodplain ecosystems. Ecohydrology & Hydrobiology 6, 37-42.

Dryden, I., Márkus, L., Taylor, C., Kovács, J., 2005. Non-Stationary spatio-temporal analysis of karst water levels. Journal of the Royal Statistical Society, Series C-Applied Statistics 54, 673-690.

Filik Iscen, C., Emiroglu, Ö., Ilhan, S., Arslan, N., Yilmaz, V., Ahiska, S., 2008. Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. Environ. Monit. Assess. 144, 269-276.

Freeden, W., Nashed, M.Z., Sonar, T., 2010. Handbook of geomathematics. Springer Verlag Berlin, Heidelberg, Germany 1350p.

Füst, A. & Geiger, J., 2010. Monitoring planning and evaluation using geostatistics, I. Geostatistical support for verification sampling based on professional opinion (In Hungarian). Földani Közlöny, 140, 303-312.

Füst, A., 2004. Short Course of Geostatistics (manuscript in Hungarian), Szent István Egyetem, Gépészmérnöki Kar, Informatika Tanszék, Gödöllő.

Füstös, L., Meszéna, Gy., Simonné M., N., 1986. Többváltozós adatelemzés (in English: The methods of multivariate data analysis), Akadémiai Kiadó, ISBN 963-05-3859-8, Budapest Hungary.

Gallant, S.I., 1995. Neural network learning and expert systems. Massachusetts Institute of Technology. USA 371p.

Geweke, J.F., 1977. The dynamic factor analysis of economic time series models. In Ainger, D.J. and Goldberger, A.S. (Eds.), Latent variables in Socio-economic Models (pp. 365-382). Amsterdam: North-Holland, Gilvear, D.J. & McInnes, R.J., 1994. Wetland Hydrological Vulnerability and the Use of Classification Procedures: a Scottish Case Study. Journal of Environmental Management 42, 403-414.

Golden, R.M., 1996. Mathematical methods for neural network analysis and design. Massachusetts Institute of Technology, USA 427p.

Hajnal E. & Padisák J., 2008. Analysis of long-term ecological status of Lake Balaton based on the ALMOBAL phytoplankton database. Hydrobiologia 599, 259-276.

Hammer, D.A., 1992. Designing constructed wetland systems to treat agricultural nonpoint source pollution. Ecol. Eng. 1, 49–82.

Han, J., Kamber, M. Pei, J., 2012. Data Mining - Concepts and Techniques (third edition). Elsevier, San Francsco, CA, USA, 744p.

Hanrahan, G., Gibani, S., Miller, K., 208. Multivariate chemometrical classification and assessment of Lake Tuendae: A Mojave desert aquatic environment housing the endangered Mojave Tui Chub. Ecol. Inf. 3, 334-342.

Hans, G., 2005. Univariate Time Series in Geosciences. Springer-Verlag, ISBN 3540238107, New York, USA, 718p.

Hattermann, F.F., Krysanova, V., Habeck, A., Bronstert, A., 2006. Integrating wetlands and riparian zones in river basin modeling. Ecol. Mod. 199, 379-392.

Hatvani, I.G., Clement, A., Kovács, J., Kovácsné, Sz.I., Korponai, J., 2014c. 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); DOI: http://dx.doi.org/10.1016/j.jglr.2013.12.010 (in press).

Hatvani, I.G., Kovács, J., Kovácsné Sz.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. Ecol. Eng. 37, 629-635.

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 Contaminant Hydrology (submitted to journal).

Hatvani, I.G., Magyar, N., Zessner, M., Kovács, J., Blaschke, A.P., 2014a. The Water Framework Directive: Can more information be extracted from groundwater data? A case study of Seewinkel, Burgenland, eastern Austria. Hydrogeology Journal, DOI: http://dx.doi.org/10.1007/s10040-013-1093-x (in press)

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice-Hall, New Jersey, USA 842p.

HCSO (Hungarian Central Statistical Office; 2012).

Herodek S., Istvánovics V., Jolánkai G., Csathó P., Németh T., Várallyay Gy., 1995. P-cycle in the Balaton catchment – a Hungarian case study. In: Thiessen H. (ed.) Phosphorus in the Global Environment. John Wiley and Sons Ltd. pp. 275-300.

Herodek, S., 1984. The eutrophication of Lake Balaton: measurements, modeling and management. Internazionale Vereinigung für Theoretische und Angewandte Limnologie

Istvánovics V. & Somlyódy, L., 1999. Changes in the cycling of phosphorus in the Upper Kis-Balaton Reservoir following external load reduction. Freshwater Biology 41, 1-19.

Istvánovics V., Somlyódi L., Clement, A., 2002. Cyanobacteria-mediated internal eutrophication in shallow Lake Balaton after load reduction. Water Research 36, 3314-3322.

Istvánovics, V. & Somlyódi, L., 2001. Factors influencing lake recovery from eutrophication - the case of Basin 1 of Lake Balaton. Water Research 35, 729–735.

Istvánovics, V., Clement A., Somlyody L., Specziar A., G.-Toth L., Padisak J., 2007. Updating water quality targets for shallow Lake Balaton (Hungary), recovering from eutrophication. Hydrobiologia 581, 305-318.

Istvánovics, V., Herodek, S., Szilágyi, F., 1989. Phosphate adsorption by different sediment fractions in Lake Balaton and its protecting reservoirs. Water Research 23, 1357-1366.

Istvánovics, V., Honti, M., Kovács, Á., Osztoics, A., 2008. Distribution of submerged macrophytes along environmental gradients in large, shallow Lake Balaton (Hungary). Aquatic Botany 88, 317-330.

Istvánovics, V., Osztoics, A., Honti, M., 2004. Dynamics and ecological significance of daily internal load of phosphorus in shallow Lake Balaton, Hungary, Freshwater Biology 49, 232-252

Istvánovits, V., Kovács, A., Vörös, L., Herodek, S., Pomogyi, P., 1997. Phosphorus cycling in a large, reconstructed wetland, the lower Kis-Balaton Reservoir (Hungary). Verh. Internat. Verein. Limnol. 26, 323–329.

Johnson, R.A. & Wichern, D. W., 1992. Applied multivariate statistical analysis 3rd ed. New Jersey, Prentice Hall, 642p.

Kárpáti, I., (Ed.) 1980. Latest research results of Lake Balaton (In Hungarian), Hungarian Academy of Sciences, Veszprém, HUN

Kazi, T.G., Arain, M.B., Jamali, M.K., Jalbani, N., Afridi, H.I., Sarfraz, R.A., Baig, J.A., Shah, A.Q., 2009. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. Ecotoxicology and Environmental Safety 72, 301-309.

Ketskeméty, L., Izsó, L., 2005. Introduction to the SPSS software. Methodical guide and task collection for statistical analysis (in Hungarian). Eötvös Kiadó, Budapest, 459p.

Korponai, J., Braun, M., Buczkó, K., Gyulai, I., Forró, L., Nédli, J., Papp, I., 2010. Transition from shallow lake to a wetland: a multi-proxy case study in Zalavári Pond, Lake Balaton, Hungary. Hydrobiologia 641, 225-244.

Korponai, J., Mátyás, K., Tátrai, I., Paulovits, G., Kovács, N.m 1997. Effect of fish community on the cladoceran plankton assemblages, in the Kis-Balaton Reservoir, Hungary. Hydrobiologia 360, 211-220.

Kovacic, D.A., Twaitb, R.M., Wallace, M.P., Bowlinga, J.M., 2006. Use of created wetlands to improve water quality in the Midwest—Lake Bloomington case study. Ecological Engineering 28, 258–270.

Kovács J., 2007. Modern geomatematikai módszerek alkalmazása hidrogeológiai feladatok megoldására, PhD értekezés, 112p.

Kovács, B. & Szanyi, J., 2005 Hydrodinamic and transport modeling in MODFLOW environment (in Hungarian: Hidrodinamikai és transzportmodellezés Processing MODFLOW környezetben) II., Miskolci Egyetem – Szegedi Tudományegyetem – GÁMAGEO, 160p.

Kovács, J., Hatvani, I.G., Korponai, J., Kovacsné, Sz.I., 2010. Morlet wavelet and autocorrelation analysis of long term data series of the Kis-Balaton Water Protection System (KBWPS). Ecol. Eng. 36, 1438-1440.

Kovács, J., Korponai, J., Kovácsné, Sz.I., Hatvani, I.G., 2012b. Introducing sampling frequency estimation using variograms in water research with the example of nutrient loads in the Kis-Balaton Water Protection System (W Hungary). Ecol. Eng. 42, 237-243.

Kovács, J**.**, Márkus, L., Halupka, G., 2004. Dynamic Factor Analysis for Quantifying Aquifer Vulnerability, Acta Geol. Hung. 47, 1–17.

Kovács, J., Nagy, M., Czauner, B., Kovácsné, Sz.I., Borsodi, A.K., Hatvani, I.G., 2012c. 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., Reskóné-Nagy, M., Kovácsné-Székely, I., 2005. Mintavételezés gyakoriságának vizsgálata tér-statisztikai függvénnyel a Velencei-tó példáján (in English: Sampling frequency analysis using GIS techniques on the example of Lake Velence). Hidrol. Közl. 85, 68-71.

Kovács, J., Tanos, P., Korponai, J., Kovácsné, Sz.I., Gondár, K., Gondár-Sőregi, K., Hatvani, I.G., 2012a. Analysis of Water Quality Data for Scientists In: Kostas Voudouris; Dimitra Voutsa (eds.) Water Quality and Water Pollution: Evaluation of Water Quality Data. Rijeka: InTech Open Access Publisher pp. 65-94.

Kuzma, J.W. & Bohnenblust, S.E., 2005. Basic Statistics for the Health Sciences, 5th ed. McGraw-Hill, Boston, MS, USA, 458 p.

Lafreniére, M. & Sharp, M., 2003 Wavelet analysis of inter-annual variability in the runoff regimes of glacial and nival stream catchments, Bow Lake, Alberta. Hydrological Processes 17, 1093–1118.

Landau, S. & Everitt, B.S., 2004. A handbook of statistical analyses using SPSS. CRC Press, FL, USA, 356p.

Láng, I., 1986. Impact on Policymaking: Background to a Government Decision. In Somlyódy, L.; L.G. Van Straten (eds.), Modeling and Managing Shallow Lake Eutrophication. Springer, Berlin, pp. 110-121.

Leadbetter, M.R., 1991. On a basis of ‘Peaks over Threshold’ modeling. Statistics and Probability Letters 12, 357-362.

Li, X., Xiao, D., Jongman, R.H., Harms, W.B., Bregt, A.K., 2003. Spatial modeling on the nutrient retention of an estuary wetland. Ecol. Mod. 167, 33-46.

Lomb, N.R., 1976. Least-squares frequency analysis of unequally spaced data. Astrophysics and space science 39, 447-462.

Lotz Gy., 1988. A Kis-Balaton Vízvédelmi Rendszer (in English: The Kis-Balaton Water Protection System). Hidrol. Tájékoztató. Október 20-22, Hungary.

Magyar, N., Hatvani, I.G., Kovácsné, Sz.I., Herzig, A., Dinka, M., Kovács, J., 2013. Application of multivariate statistical methods in determining spatial changes in water quality in the Austrian part of Neusiedler See. Ecol. Eng. 55, 82-92.

Márkus, L., Berke, O., Kovács, J., Urfer, W., 1999. Spatial prediction of the intensity of latent effects governing hydrogeological phenomena. Environmetrics 10, 633-654.

Matheron, G., 1965. Les Variables Regionaliées et leur Estimation. (in English: Regional variables and their distribution) MaTSSn at Cie. Editeurs, Paris, 305p.

May, R.M., 1977. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. Nature 269, 471-477.

McCosker, R.O., 1998. Methods addressing the flow requirements of wetland, riparian and floodplain vegetation. In Arthington, A.H. and J.M. Zalucki (eds.), Comparative Evaluation of Environmental Flow Assessment Techniques: Review of Methods. Land and Water Resources Research and Development Corporation, Canberra.

Mitch, W.J., 1992. Landscape design and the role of created, restored. and natural riparian wetlands in controlling non-point source pollution. Ecol. Eng. 1, 27–47.

Mitch, W.J., Zhang, L., Anderson, C.J., Altor, A.E., Hernandez, M.E., 2005. Creating riverine wetlands: ecological succession, nutrient retention, and pulsing effects. Ecol. Eng. 25, 510–527.

Molnár, S. & Füst A., 2002. Környezet-informatikai modellek I. (in English: Environment informatics models I.) Szent István Egyetem, Gépészmérnöki Kar, Informatika Tanszék, Gödöllő.

Molnár, S., 1987. On the Convergence of the Kriging Method, Ann., Univ., Sci. Budapest, Sect. Comp., 1987, tomes 8, 45-50.

Molnár, S., Füst, A., Szidarovszky, F., Molnár, M., 2010. Környezetinformatikai modellek II. (in English: Environment informatics models II.) Szent István Egyetem, Gödöllő.

Muñoz-Carpena, R., Ritter, A., Li, Y.C., 2005. Dynamic factor analysis of groundwater quality trends in an agricultural area adjacent to Everglades National Park. Journal of Contaminant Hydrology 80, 49-70.

Mustafa, A., Scholz, M., Harrington, R., Carroll, P., 2009. Long-term performance of a representative integrated constructed wetland treating farmyard runoff. Ecol. Eng. 35, 779-790

Norušis, M. J., 1993. SPSS for Windows Professional Statistics Release 6.0. SPSS Inc., USA, 385p.

NYUDUVIZIG, 2012; Nyugat-dunántúli Vízügyi Igazgatóság (in English: West Transdanubian Water Authority) http://www.nyuduvizig.hu/upload/Atadott_Vizmerleg-2012.pdf Cited on 10.11.2013.

Padisák J. & Reynolds, C. S., 2003. Shallow lakes: the absolute, the relative the functional and the pragmatic. Hydrobiologia 506-509, 1-11.

Padisák, J., 1997. Cylindrospermopsis raciborskii (Wołoszynska) Seenayya et Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology. Arch., Hydrobiologia, 107, 563–593.

Padisák, J., 2005. Általános limnológia (in English: General Limnology). ELTE Eötvös Kiadó, Budapest, 310p.

Parpala, L.,G.-Tóth, L., Zinevici, V., Németh, P., Szalontai, K., 2003. Structure and production of the metazoan zooplankton in Lake Balaton (Hungary) in summer, Hydrobiologia 506-509, 374-351.

Peltier, J., 2013. Personal communication, 8 November 2013; jon@peltiertech.com; www.peltiertech.com

Pernthaler, J., 2005. Predation on prokaryotes in the water column and its ecological implications. Nature Reviews Microbiology 3, 537-546.

Pilotto, L., Hobson, P., Burch, M.D., Ranmuthugala, G., Attewell, R., Weightman, W., 2004. Acute skin irritant effects of cyanobacteria (blue-green algae) in healthy volunteers. Aust N Z J Public Health 28, 220-224.

Pomogyi, P., (ed.) 1991. A Kis-Balaton Védőrendszer kémiai, biológiai, anyagforgalmi vizsgálatai, Összefoglaló jelentés az 1985-1990 közötti kutatásokról (in English: Analyses on the chemical, biological an nutrient characteristics of the The Kis-Balaton Water Protection System), Szombathely-Keszthely, 258p.

Pomogyi, P., Szeglet, P., Csató, É., 1996. The change in the reed community of the KBWPS, Fenéki pond, based on vegetation-mapping results (In Hungarian). In Pomogyi, P. (Eds.) Second Kis-Balaton conference: Summary of the research result on the KBWPS from between 1991-1995. Keszthely: University of Pannonia Georgikon Faculty of Agriculture, pp. 206-207.

Raisin, G.W., Mitchell, D.S., Croome, R.L., 1997. The effectiveness of a small constructed wetland in ameliorating diffuse nutrient loadings from an Australian rural catchment. Ecol. Eng. 9, 19-35.

Ramsar Convention, "Convention on Wetlands of International Importance especially as Waterfowl Habitat. Ramsar (Iran), 2 February 1971. UN Treaty Series No. 14583. As amended by the Paris Protocol, 3 December 1982, and Regina Amendments, 28 May 1987."

Rapcsák, T., 2002. On minimization on Stiefeld Manifolds. European Journal of Operational Research 143, 365-376.

Reddy, K.R., Kadlec, R.H., Flaig, E., Gale, P.M., 1999. Phosphorus Retention in Streams and Wetlands: A Review. Environmental Science and Technology 29, 83-146.

Rencher, A.C., 2002. Methods of Multivariate Analysis 2nd ed., Wiley-Intescience, Canada, 727p.

Ritter, A., Muñoz-Carpena, R., 2006. Dynamic factor modeling of ground and surface water levels in an agricultural area adjacent to Everglades National Park. Journal of Hydrology 317, 340-354.

Rogerson, P.A., 2001. Statistical methods for geography. SAGE Publications, London, UK, 304p.

Sas, H., 1989. Lake Restoration and Reduction of Nutrient Loading: Expectations, Experiences, Extrapolations. - Academia Verlag Richarz, St. Austin.

Scheffer, M., 1998. Ecology of Shallow Lakes. Chapman and Hall, London, UK, 358p.

Scheffer, M., Carpenter, S., Jonathan, A.F., Folke, C., Walker, B., 2001. Nature 413, 591-596.

Scheffer, M., Hosper, S.H., Meijer, M-L., Moss, B., Jeppesen, E., 1993. Alternative equilibria in shallow lakes. Trends in Ecology & Evolution 8, 275-279.

Schneider, T., 2001. Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values. J. Climate 14, 853-871.

Scholz, M., 2006. Wetland Systems to Control Urban Runoff. Elsevier, Amsterdam, The Netherlands, 360p.

Shannon, C.E., 1949. Communication in the presence of noise. Proc. IRE 37, 10 -21.

Sheng-Bing, H., Li, Y., Hai-Nan, K., Zhi-Ming, L., De-Yi, W., Zhan-Bo, H., 2007. Treatment Efficiencies of Constructed Wetlands for Eutrophic Landscape River Water. Pedosphere 17, 522–528.

Shumway, H.R. & Stoffer, S.D., 2000. Time Series Analysis and Its Application. Springer-Verlag New York, USA, 572p.

Sisák I., 1993. Analysis of diffuse loads on the western watershed of Lake Balaton using nutrient load estimation method, Candidate thesis. (In Hungarian) Pannon Agrártudományi Egyetem, Georgikon Mezőgazdaságtudományi Kar, Keszthely.

Sisák, I. & Pomogyi, P., 1994. Analysis of the loads arriving to the River Zala (In English). Vízügyi Közlemények 76, 417-434.

Slutsky E., 1937. The summation of random causes as the source of cyclic processes. Econometrica 5, 105-46.

Somlyódy, L., 1998. Eutrophication modeling, management and decision making: the Kis-Balaton Case. Wat. Sci. Tech. 37, 165-176.

Somlyódy, L., Clement, A., Istvánovics, V., G. Tóth, L., Jolánkai, G., Sisák, I., Padisák, J., Speczíár, A., 2003. Evaluation of the water quality of Lake Balaton, Study to provide scientific basis for the modification of the government regulation – 1038/1999 (IV.21) (in Hungarian) manuscript, Budapest University of Technology and Economics, Budapest.

Somlyódy, L., Van Straten, G., (eds.) 1986. Modeling and Managing Shallow Lake Eutrophication. Springer, Berlin. 386p.

Sprőber, P., Hesham, M. S., Présing, M., Kovács, A. W., Herodek, S., 2003. Nitrogen uptake and fixation in the cyanobacterium Cylindrospermopsis raciborkii under different nitrogen conditions. Hidrobiologia 506-509, 169-174.

Stämpfli, N., 2006. Tools for Phosphorus Management, Soil and Water Quality Management Course BREE 625. McGill University, Ste-Anne-de-Bellevue, QC.

Stein, M.L., 1999. Interpolation of spatial data: some theory for kriging. Springer-Verlag New York, USA, 257p.

Stenback, G.A., Crumpton, W.G., Schilling, K.E., Helmers, M.J., 2011. Rating curve estimation of nutrient loads in Iowa rivers. Journal of Hydrology 396, 158-169.

Stockburger, D. W., 1998. Multivariate Statistics: Concepts, Models, and Applications, Missouri State University, Cited on: 03.10.2013. http://www.psychstat.miTSSuristate.edu/multibook/mlt00.htm

Stodden, V., 2006. Model selection when the number of variables exceeds the number of observations, Ph.D. dissertation, Department of Statistics, Stanford University, USA Cited on 03.10.2013. http://www.stanford.edu/~vcs/thesis.pdf .

Tabachnick, B.G. & Fidell, L.S., 1996. Using Multivariate Statistics 3rd ed. HarperCollins Publishers, 880p.

Tátrai I., Kálmán, M., Korponai, J., Paulovits, G., Pomogyi, P., 2000. The role of the Kis-Balaton Water Protection System in the control of water quality of Lake Balaton, Ecological Engineering 16, 73–78.

Taylor, J.R., 1997. An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, second ed. Univ.Sci.Books, CA, USA, 327p.

Tunney, H., Carton, O.T., Brookes, P.C., Johnston, A.E., (Eds) 1997. Phosphorus loss from soil to water. CABI Wallingford.

Van Donk, E., Grimm, M.P., Gulati, R.D., Klein Breteler, J.P.G., 1990. Restoration by biomanipulation in a small hypertrophyic lake: First year results. Hydrobiologia 200/201, 275-289.

Van Straten, G. & Herodek, S., 1982. Estimation of algal growth parameters from vertical primary production profiles. Ecol. Mod. 15, 287-311.

Vega, M., Pardo, R., Barrado, E., Debán, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research 32, 3581-3592.

Verhoeven, J.T.A., Arheimer, B., Yin, C., Hefting, M.M., 2006. Regional and global concerns over wetlands and water quality. TRENDS in Ecology and Evolution 21, 96-104.

Vidakovic, B., 1999. Statistical Modeling by Wavelets. Wiley, New York, USA, 382p.

Virág, Á., 1998 A Balaton múltja és jövője (in English: The past and the present of Lake Balaton, Egri Nyomda, Hungary.

Vollenweider R.A. & Kerekes J. 1982: Eutrophication of waters. Monitoring, assessment and control. OECD Cooperative programme on monitoring of inland waters (Eutrophication control), Environment Directorate, OECD, Paris, 154 p.

Webster, R. & Oliver, M.A., 2007. Geostatistics for Environmental Scientists, 2nd ed. Wiley, USA, 330p. ISBN 978-0-470-02858-2.

Wetzel, R.G., 2001. Limnology: Lake and River Ecosystems, third ed. Academic Press, San Diego, CA, USA, 1006p.

Wilkinson, D.J., 2006. Stochastic Modelling for Systems Biology. Chapman & Hall/CRC, USA, 254p. ISBN 978-158488-540-5.

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

Wu, H., Zhang, J., Li, P., Zhang, J., Xie, H., Zhang, B., 2011. Nutrient removal in constructed microcosm wetlands for treating polluted river water in northern China. Ecol. Eng. 37, 560-568.

Xu, H., Yang, L-Z., Zhao, G-M., Jiao, J-G., Yin, S-X., Liu, Z-P., 2009. Anthropogenic Impact on Surface Water Quality in Taihu Lake Region, China. Pedosphere 19, 765-778.

Yanyou, G., Yijun, H.; Mingkui, L., 2006. Multi-scale wavelet analysis of TOPEX/Posseidon altimeter significant wave height in eastern China seas. Chinese Journal of Oceanology and Limnology 24, 81-86.

Yates, C.R., Prasher, S.O., 2009. Phosphorus reduction from agricultural runoff in a pilot-scale surface-flow constructed wetland. Ecol. Eng. 35, 1693-1701.

Yuan, Y.C., 2010. Multiple imputation for missing data: Concepts and new development (Version 9.0). SAS Institute Inc., Rockvikke, MD

Zhang, Q., Chen, J., Becker, S., 2007. Flood/drought change of last millennium in the Yangtze Delta and its possible connections with Tibetan climatic changes. Global and Planetary Change 57, 213-221.

Zhang, Q., Gemmer, M., Chen, J., 2008. Climate changes and flood/drought risk in the Yangtze Delta, China, during the past millennium. Quaternary International 167-177, 62-69.

Zhang, Q., Liu, C., Xu, C-y., Xu, Y., Jiang, T., 2006. Observed trends of annual maximum water level and stream flow during past 130 years in the Yangtze River basin, China. Journal of Hydrology 324, 255-265.

Zhao, Y., Xia, X.H., Yang,Z.F., Wang, F., 2012. Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques. Procedia Environmental Sciences 13, 1213-1226.

Ziermann, M. & Michaletzky, Gy., 1995. Factor analysis of time series. Sigma Journal of the Society for Economic Modeling 26, 77-91.

Zlinszky, A., 2011. A Balaton és a környző mocsarak történelmi rekonstrukciója térinformatikai módszerekkel. (in English: Reconstruction of Lake Balaton and it s surrounding swamps using GIS methods) Ecology of Lake Balaton/A Balaton ökológiája 1, 49-60.

Zlinszky, A., Timár, G., 2013. Historic maps as a data source for socio-hydrology: a case study of the Lake Balaton wetland system, Hungary. Hydrol. Earth. Sci. Discuss. 10, 7733-7781.

Zuur, A.F., Fryer, R.J., Jolliffe, I.T., Dekker, R., Beukema, J.J., 2003b. Estimating common trends in multivariate time series using dynamic factor analysis. Environmetrics 14, 665-685.

Zuur, A.F., Tuck, I.D., Bailey, N., 2003a. Dynamic factor analysis to estimate common trends in fisheries time series. Can. J. Fish. Aquat. Sci. 60, 542-552.

13 Appendices

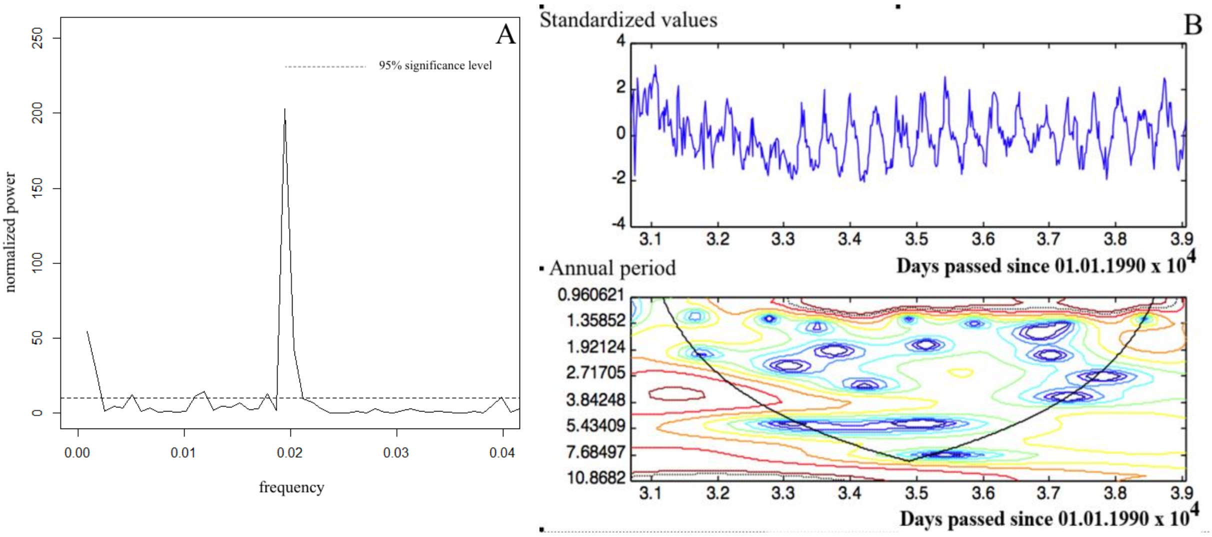


Figure A-1 Lomb-Scargle periodogram indicating a significant annual period A) and Morlet wavelet PSD graph B) of calcium on SS Z15 (1993 – 2006).

In the case of the Lomb-Scargle periodogram the annual period 1/52 = 0.019 is indicated, however as it becomes clear from the Morlet wavelet it was not present in the first two years of the investigated time period, only from 1995.

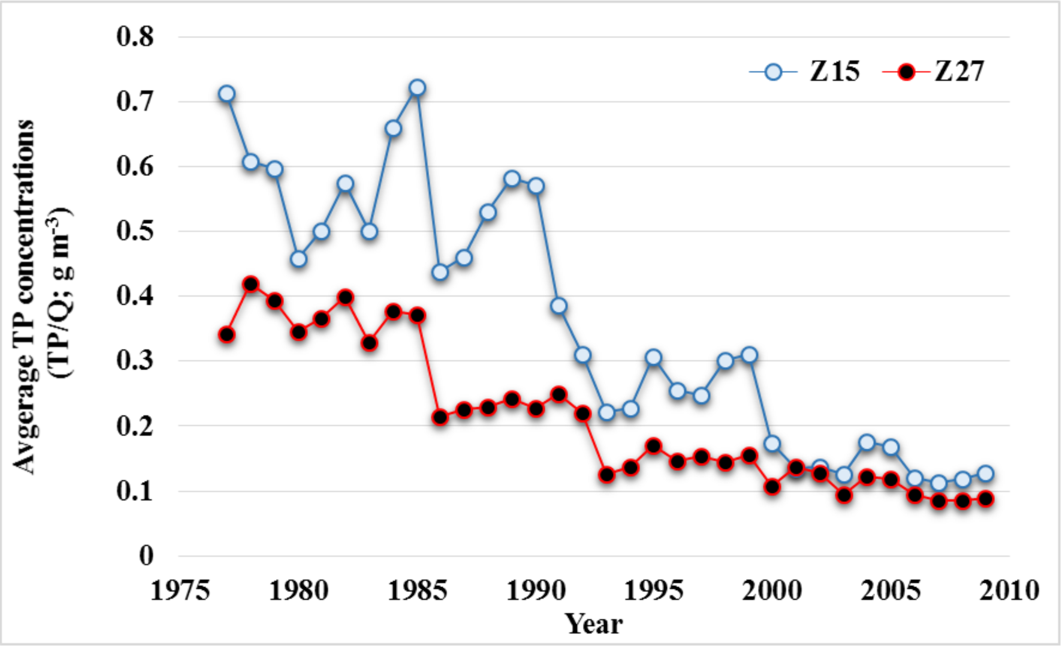


Figure A-2 Average TP concentrations at Z15 & Z27 (1975-2010)

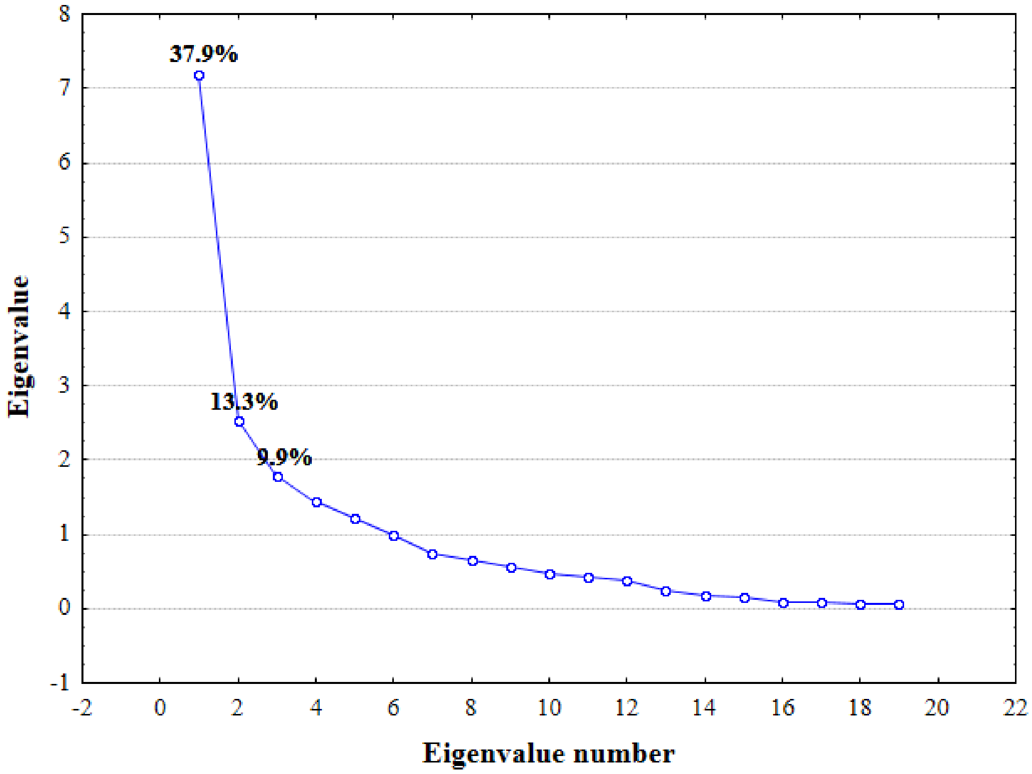


Figure A-3 Scree plot of SS Z11 (1994-2009), based on which the first principal component was considered

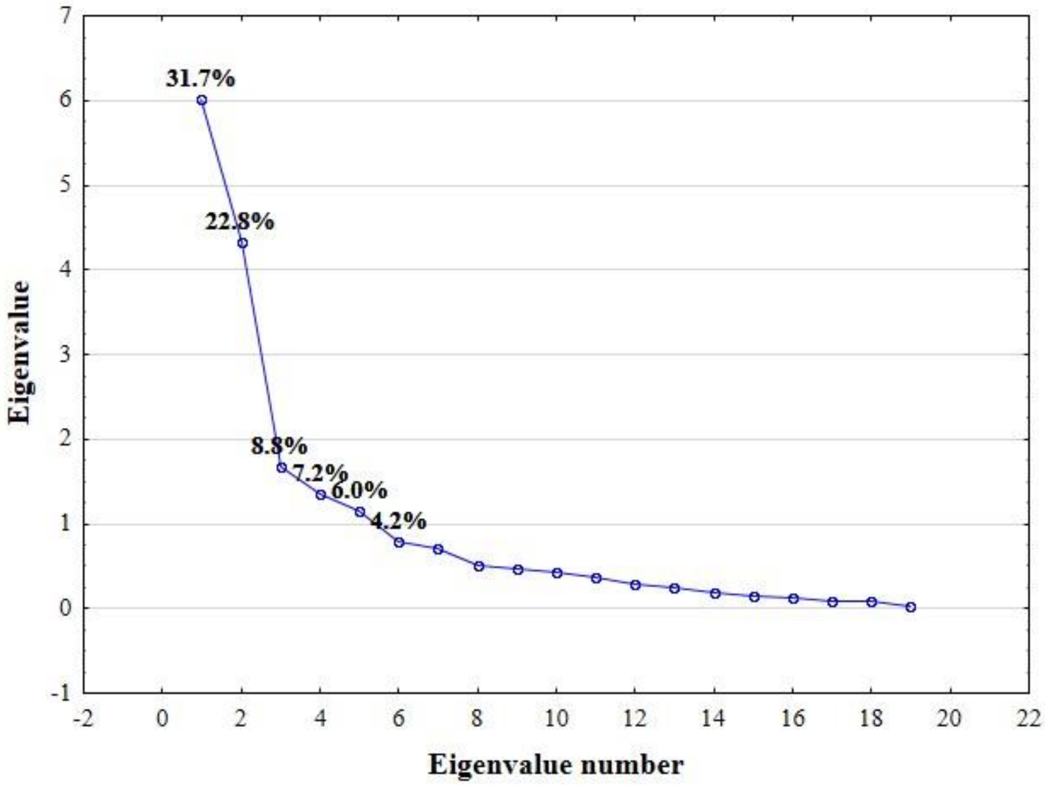
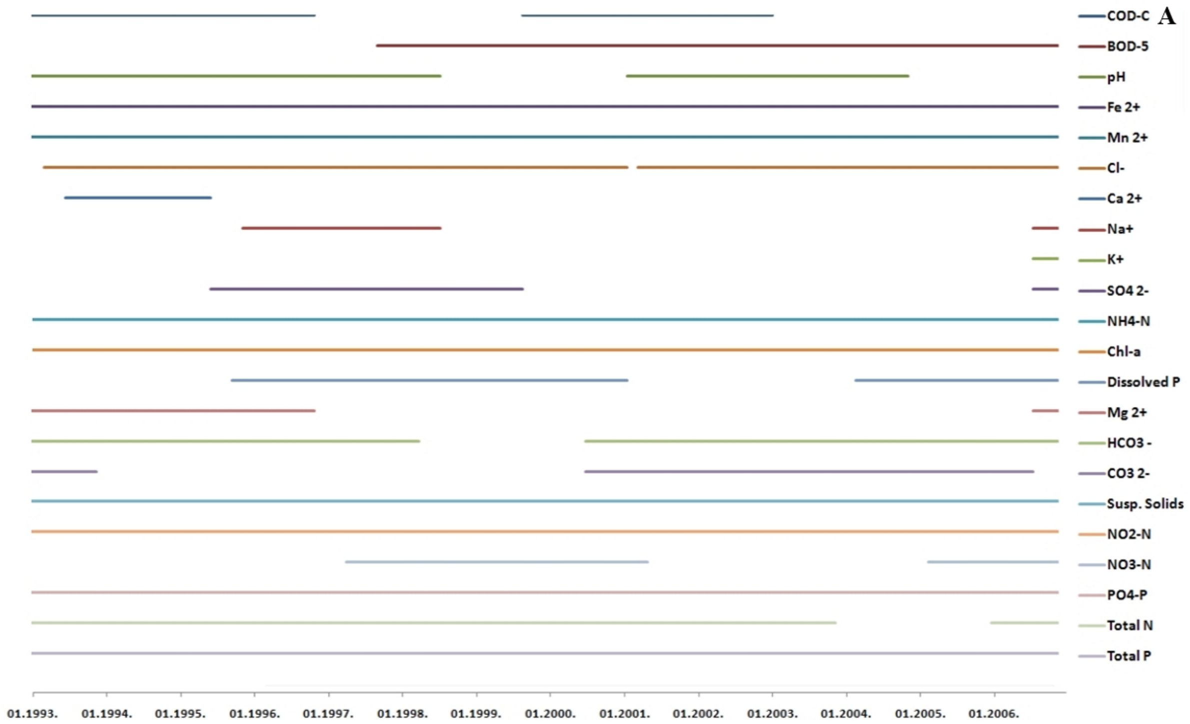
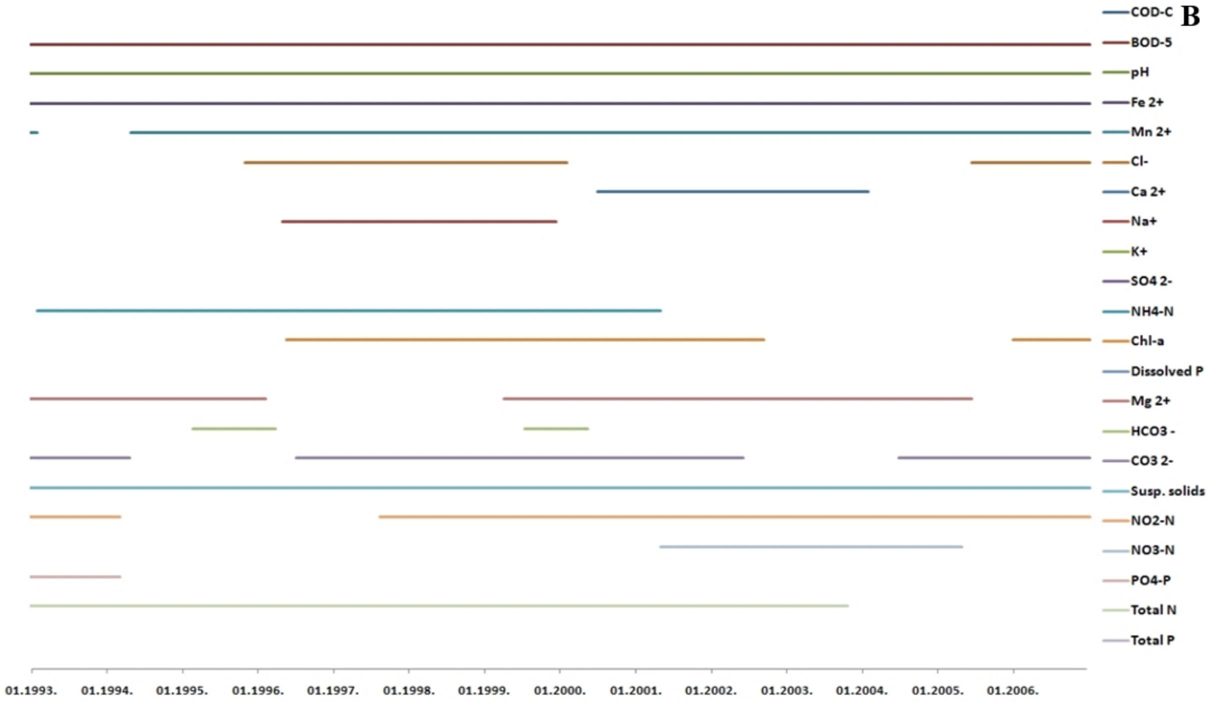


Figure A-4 Scree plot of SS Kszt (1985-1992), based on which the first and second principal components were considered


A
COD-C
BOD-5
pH
Fe 2+
Mn 2+
Cl-
Ca 2+
Na+
K+
SO4 2-
NH4-N
Chl-a
Dissolved P
Mg 2+
HCO3 -
CO3 2-
Susp. Solids
NO2-N
NO3-N
PO4-P
Total N
Total P
01.1993.
01.1994.
01.1995.
01.1996.
01.1997.
01.1998.
01.1999.
01.2000.
01.2001.
01.2002.
01.2003.
01.2004.
01.2005.
01.2006.


B
COD-C
BOD-5
pH
Fe 2+
Mn 2+
Cl-
Ca 2+
Na+
K+
SO4 2-
NH4-N
Chl-a
Dissolved P
Mg 2+
HCO3 -
CO3 2-
Susp. solids
NO2-N
NO3-N
PO4-P
Total N
Total P
01.1993.
01.1994.
01.1995.
01.1996.
01.1997.
01.1998.
01.1999.
01.2000.
01.2001.
01.2002.
01.2003.
01.2004.
01.2005.
01.2006.

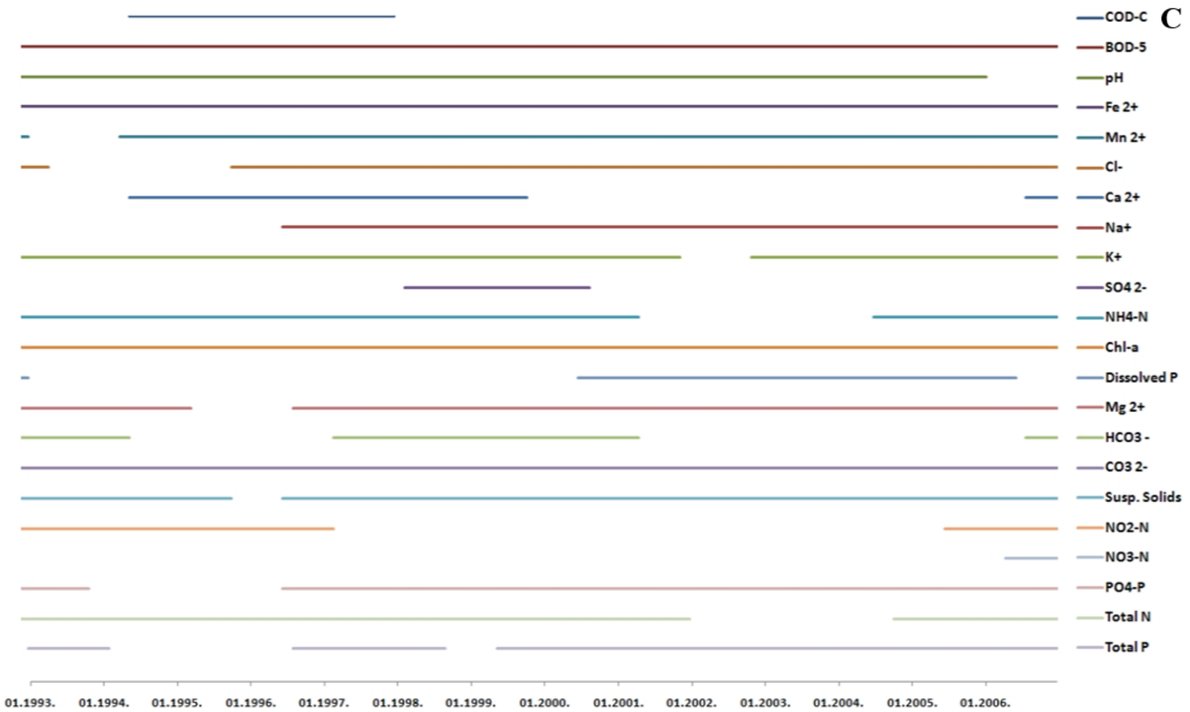


Figure A-5 Wavelet spectrum analysis results showing the absence of annual periods on Z15 A), Kb210 B) & Z27 C)

Declaration of Authorship

Last name: Hatvani

First name: István Gábor

I declare that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other University.

Formulations and ideas taken from other sources are cited as such. This work has not been published in its present form.

Location, Date
Budapest, 21.01.2014

Signature