

Environmental Monitoring and Assessment

Application of multivariate statistical methods in determining the spatial water quality changes in the Austrian part of Neusiedler See

--Manuscript Draft--

Manuscript Number:	EMAS10450
Full Title:	Application of multivariate statistical methods in determining the spatial water quality changes in the Austrian part of Neusiedler See
Article Type:	Original Research
Keywords:	Cluster analysis, Neusiedler See, principal component analysis, spatial variability, water quality
Corresponding Author:	Norbert Magyar, B.Sc. Eotvos Lorand University Budapest, HUNGARY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Eotvos Lorand University
Corresponding Author's Secondary Institution:	
First Author:	Norbert Magyar, B.Sc.
First Author Secondary Information:	
Order of Authors:	Norbert Magyar, B.Sc.
	Istvan Gabor Hatvani, M.Sc.
	Ilona Kovacsne Szekely, Ph.D.
	Alois Herzig, Prof.
	Maria Dinka, Ph.D.
	Jozsef Kovacs, Ph.D.
Order of Authors Secondary Information:	
Suggested Reviewers:	Antal Füst, Prof. , D.Sc. University professor, Szent István University antalfust@gmail.com Prof. Füst is an expert on geostatistics and monitoring assesment. He is familiar with both the methods and the problems dealt with in the manuscript.
	Mária Nagy Reskóné, Ph.D. Natural Protection and Water Management reskone.maria@gmail.com Since Dr. Reskóné works at the water authorities on W Hungary she is familiar with the problems concerning the management and the water quality of Lake Neusiedl, while she is an expert on water biology and protection.
	Martin Dokulil, Prof. Univ. Prof., Austrian Academy of Sciences martin.dokulil@oeaw.ac.at Prof. Dokulil is an expert in limnology and hydrobiology. He is a member of the journal Limnologica's advisory board.He has many publication on Lake Neusiedl.
	János Korponai, Ph.D. associate professor, University of West Hungary korponai.janos@nyuduvizig.hu Dr. Korponai is an expert on water biology.
	He has published many studies on Hungarian water systems and lakes which present

methods like the ones applied in our manuscript.

Application of multivariate statistical methods in determining the spatial water quality changes in the Austrian part of Neusiedler See

Norbert Magyar^{a*}, István Gábor Hatvani^a, Ilona Kovácsné Székely^b, Alois Herzig^c, Mária Dinka^d, József Kovács^a

^a*Eötvös Loránd University, Department of Physical and Applied Geology, H-1117 Budapest, Pazmany Peter stny. 1/C, Hungary*

^b*Budapest Business School, Institute of Methodology, H-1054 Budapest, Alkotmány u. 9-11., Hungary*

^c*Biological Station Neusiedler See, A-7142 Illmitz, Austria*

^d*Danube Research Institute Research Centre of Ecology of the Hungarian Academy of Sciences, H-2163 Vácrátót, Alkotmany u. 2-4.*

Corresponding author: e-mail: magyarnorbert87@gmail.com , tel.: +36202936611, fax: +3613812130

Abstract

The aim of the present study was to distinguish areas with different chemical properties within Neusiedler See and to determine which background processes are responsible for this pattern. Uni- and multivariate data analysis was applied to the data concerning 13 mainly chemical and some biological parameters for the time period 2000-2009 from 33 sampling sites. The sampling locations were first clustered then grouped. Besides reed belt and open water areas, smaller local sections influenced by water inputs (the treatment plant, the river Wulka, the channels of weekend houses) were separated as well. Using Wilks' Lambda distribution it was determined that the main ions have a greater effect on forming the cluster groups than those parameters in close relation to the biological processes. These results concurred with those obtained from the principal component analysis (PCA) conducted on the whole lake and on the groups as well. It can be stated that most of the variance in the dataset can be explained by the main ions. The spatial distribution of the principal component scores was visualized with isoline maps. The result of this research was that Neusiedler See cannot be treated as one homogeneous system. This exceptional variability originates from the lake's shallow water depth, unstable water balance, and anthropogenic activity (agriculture, tourism, sewage treatment) in the lake's surroundings.

Keywords: cluster analysis, Neusiedler See, principal component analysis, spatial variability, water quality

Introduction

Surface water quality is a very sensitive issue, which transcends national boundaries (Yidana et al. 2008). One of the best examples is Lake Fertő/Neusiedler See, which lies on the Hungarian-Austrian border. Because of the Iron Curtain for decades neither country had any information concerning the state of the lake on the other side. Nowadays, however, with both countries being member states of the European Union, not only does the possibility exist for the two countries to assess the true state of the whole lake, but they are also obliged to do so (Water Framework Directive 2000/60/EC). It is recognized that water quality monitoring is one of the highest priorities in environmental protection policy (Simeonov et al. 2003), because to determine the quality of any body of water, its physical, chemical and biological parameters must be measured systematically (Sargaonkar and Deshpande 2003). Precise and effective monitoring is, however, not enough, if the acquired data is not analyzed thoroughly. In the case of the area studied in this research, because of its fragmented habitat structure and unstable water balance (Löffler 1979), the data sets deriving from it are complex, comprising mostly chemical and biological parameters which are often difficult to interpret and draw meaningful conclusions from. The application of different multivariate statistical methods, like cluster analysis (CA), discriminant analysis (DA) and principal component analysis (PCA) helps in the interpretation of such complex data matrices as are found in the case of Neusiedler See in order to obtain a better understanding of water quality and the ecological state of the study area (Zhou et al. 2007). Parallel to the fragmented habitat structure the study area is daily exposed to severe anthropogenic activity, mostly on the Austrian side, in the areas used for recreation. Anthropogenic influences such as recreational use, the influence of

1 agriculture, and changes in precipitation inputs degrade surface waters which,
2 because of their constantly changing surroundings, need to be monitored even
3 more accurately.
4
5
6

7 Using explorative multivariate data analysis methods on the time series of
8 the Austrian part of Neusiedler See the lake's heterogeneity was examined, and
9 the question was raised whether it is possible to distinguish areas with different
10 chemical properties within the lake, and if yes, what those properties are and
11 which background processes might result in this pattern.
12
13
14
15
16
17
18
19
20
21
22

23 **Materials and methods**

24 **Study area description**

25 Neusiedler See (47° 42' N, 16° 46' E) is the largest lake in Austria (309 km²,
26 75 km² of which lies in Hungary) (Dinka et al. 2010). It is situated in the lowest
27 plain of the country, at 116 m above sea level. It is a shallow (avg. 1.1 m), well-
28 mixed lake with regulated outflow, and large water level fluctuations (Dokulil and
29 Herzig 2009) (Fig. 1). The Neusiedler See regarding its water balance is
30 considered to be somewhere between astatic and semiastatic (Padisák J. 2005)
31 During the centuries the lake has dried out many times and its water level varied
32 in a wide interval, sometimes resulting in enormous floods. Its waters can be
33 described as salty (sodium and magnesium salts). Neusiedler See is Europe's
34 western-most and largest steppe lake, and is considered to be a wetland. Despite
35 the growth in awareness of their importance as ecosystems, these wetlands are
36 often endangered ecosystems. Their importance lies within their biological
37 productivity, hydrological and biogeochemical cycling and as habitats both for
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

humans and wildlife (Mitsch and Gosselink 1986). In 1977 the Austrian, and in 1979 the Hungarian part of the lake was taken under UNESCO protection as a biosphere reserve (Dinka et al. 2004; Dinka et al. 2010). Ten years later in 1989 this highly protected nature conservation area was brought under the de-jure protection of the Ramsar Convention. From 1991 the Fertő-Hanság National Park with the Nationalpark Neusiedlersee-Seewinkel became a part of the IUCN's¹ protection zone, while in 2001 it gained the World Heritage status.

Two main habitats can be identified in the lake, the reed (*Phragmites australis*) belt (174 km²)² and the open water area. A third habitat can be distinguished as well which are the channels in the reeds. The water balance of the lake is unstable mostly because its main water input comes from precipitation and its main output is evapotranspiration. Besides precipitation the Wulka and Rákos streams contribute to the lake's water input. The Wulka reaches the lake between Purbach and Oggau (AU), while the Rákos reaches it at Fertőrákos (HU) (Löffler 1979). The air is almost constantly in motion, because of orographic effects the prevailing wind direction is NW, which is followed (according to frequency) by SE. Thermal stratification can only be observed in summer under no-wind conditions and only for short periods, because the 2 ms⁻¹ wind is sufficient to mix the water column and stir up the sediment (Löffler 1979).

Data and sampling sites

The database used was acquired from the Neusiedler See Biological Station in Illmitz. This station is part of the national Water Inspectorate (AU) which

¹ International Union for Conservation of Nature and Natural Resources

² 110 km² on the Austrian, 64 km² on the Hungarian part of the lake

monitors the lake at 38 sampling sites both from the reed belt and the open water as well. It is worth mentioning that the samples from the reed belt were taken from channels within the dense reed stands. These reed channels form a transition zone between the reed and the open water habitats.

Fig. 1 Location of Neusiedler See and its sampling sites

The sampling was conducted monthly, though in certain cases, weekly. The sampled parameters which suited the conditions of the methods applied³ are as follows: pH, dissolved O₂, K⁺, Na⁺, Ca²⁺, Mg²⁺, Cl⁻, SO₄²⁻, HCO₃⁻ (mg l⁻¹), total phosphorus (TP), NH₄-N, NO₃-N, chlorophyll-a (Chl-a) (µg l⁻¹). The descriptive statistics of the parameters can be seen in Table 1.

Table 1 Descriptive statistics of the parameters used

CO₃²⁻ was not used because its values are highly dependent on pH⁴, which in certain sampling sites was below 8.3. If the zero values had been used the correlation coefficients between the parameters would have changed at those sampling sites. Because the same laboratory methods were employed in gathering the data, the results were comparable and comprehensive. The analyses were conducted for the time interval 2000-2009 because before the turn of the century the parameters in question were not measured frequently enough for these statistical methods to be applied. During the winter months and occasionally even

³ Parameters have to be measured at all sampling sites at all investigated time points

⁴ Under pH 8.3 there is likely to be no carbonate in the water (Daniel 1999)

1 in March most of the lake is covered with ice, so at many monitoring sites no
2 sampling was conducted during the cold period, as a result these dates have been
3
4 left out of the analyses. Specifically, 34 time points from 33 sampling stations
5
6 were examined, altogether 14,586 data (Fig. 1). The detailed description of the
7
8 sampling sites can be found in A1.
9
10

11 12 13 14 15 16 **Piper diagram**

17 “Piper diagrams were used (Piper 1944; Fig. 2) to display the relative
18 concentrations of the major cations and anions on two separate trilinear plots,
19
20 together with a central diamond plot where the points from the two trilinear plots
21
22 are projected. The central diamond-shaped field (quadrilateral field) is used to
23
24 show the overall chemical character of the water (Hill 1940; Piper 1944).” (Güler
25
26 et al. 2002).
27
28
29
30
31
32
33
34
35

36 37 **Cluster analysis**

38 Cluster analysis (CA) is a multivariate method which groups variables or
39
40 cases based on their similarity. The most common type is hierarchical CA, where
41
42 each case starts in a separate cluster and joins up to the other clusters as the
43
44 linkage distance grows and only one cluster remains (Lattin et al. 2003; McKenna
45
46 2003). This is visualized on a dendrogram. CA was conducted on 34 points and 13
47
48 parameters using Ward’s method (Ward 1963), and squared Euclidean distance. A
49
50 similar approach has been successfully applied to the assessment of water quality
51
52 in: Wunderlin et al. (2001), Simeonov et al. (2003), Singh et al. (2004), Shrestha
53
54 and Kazama (2007), Zhou et al. (2007), Filik Iscen et al. (2008), Juahir et al.
55
56 (2011).
57
58
59
60
61
62
63
64
65

Discriminant analysis

To verify the accuracy of the cluster results, discriminant analysis (DA) was used, because the data were not normally distributed. The analysis was carried out in order to show to what extent the planes separating the groups could be distinguished by building a predictive model for group membership. The model is composed of a discriminant function (for more than two groups a set of discriminant functions) based on linear combinations of the predictor variables that provide the best discrimination between the groups (Johnson and Wichern 1992). The result of the discriminant analysis is visualized on the surface stretched between the first two discriminating planes (function 1 and function 2) (Ketskeméty and Izsó 2005).

Wilks' lambda distribution

After the verification of the cluster groups, the role of each parameter was analyzed in determining the formation of the cluster groups. Using Wilks' λ distribution (Wilks 1932) a Wilks' λ quotient was assigned to every parameter for each sampling, where the quotient is:

$$Wilks' \lambda = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2}{\sum_i \sum_j (x_{ij} - \bar{x})^2}$$

where x_{ij} is the j^{th} element of the i^{th} group, $\bar{x}_i$ the i^{th} group's mean, and $\bar{x}$ the total mean.

The value of λ is the ratio of the within-group sum of squares to the total sum of squares. It is a number between 0 and 1. The smaller the quotient is, the more it determines the formation of the cluster groups (Hatvani et al. 2010).

Principal component analysis (PCA)

The final method was PCA, which was used to determine the reduced number of latent variables (principal components) which explain most of the variance in the experimental data set in the groups obtained by cluster analysis, and for the whole lake as well. In other words it is a powerful pattern recognition technique that attempts to explain the variance of a large dataset of intercorrelated variables with a smaller set of independent variables (principal components) (Simeonov et al. 2003). The reason why PCA was applied instead of factor analysis (FA) is that, while the PCA uses a correlation matrix, the FA uses an adjusted correlation matrix. In this matrix the elements on the diagonal (commonalities) can be estimated in different ways, as a result different solutions (of the FA) may be acceptable. The principle component (PC) can be expressed as

$$Z_{ij} = a_{i1}X_{1j} + a_{i2}X_{2j} + \dots + a_{im}X_{mj}$$

where z is the component score, a is the component loading, x is the measured value of the variable, i is the component number, j is the sample number, and m is the total number of variables (Juahir et al. 2011). A similar approach has been successfully applied to the assessment of water quality in Simeonov et al. (2003), Shrestha and Kazama (2007), Yidana et al. (2008), Filik Iscen et al. (2008), Kazi et al. (2009).

Data representation techniques

As discussed before, the Piper diagram in general represents the whole system; however knowledge regarding the inter-group differences is also needed. The most common tool to represent these differences is the box-and-whiskers

plot. More than one statistic can be visualized on them, where the boxes show the interquartile interval and the black line in the box is the median. Two upright lines can be drawn from the boxes if the database requires them. The data from which these lines are drawn are between 1.5 and 3 times the interquartile interval. Data with even higher values are indicated with a circle (offsets), and the extreme data with an asterisk (Norušis 1993; Vega et al. 1998).

Software used and data handling

All maps in the study were prepared using Golden Software Surfer 9 and kriging as a linear estimation method. Interpolation was applied between the sampling sites, while beyond the sampling sites (near the shore of the lake) extrapolation was applied. The statistical evaluations were made using IBM SPSS Statistics 19 and Microsoft Office Excel 2010. Using cluster analysis the data were z-scale standardized.

Results and discussion

Piper diagram

All the measured values of the ions were presented on Piper-diagrams (Fig. 2) (including CO_3^{2-})⁵, in order to determine which facies the lake belongs to, based on the dominant ions. However, in the case of cations no dominant type was determined in about 40% of the samples. In about 60% of the samples only the sodium type was identified. In the case of anions, again no dominant type was

⁵ A Piper diagram cannot be produced without carbonate

determined. This is based on the works of Morgan and Winner (1962) and Back (1966) quoted in Freeze and Cherry (1979).

Fig. 2 Piper diagram for anion and cation facies derived from major ion percentages

It is clear that no groups can be identified on the basis of the Piper diagram, because the samples were located close to each other, unlike in the study by Güler et al. (2002). Despite the lack of groups in the Piper diagram, the lake is known to be highly variable/unstable (Dinka et al. 2004) so it was expedient to use multivariate methods for further investigation.

Spatial variations in water quality

Clustering the 34 time points, eight representative groups were identified. In the following these are considered to be relatively similar water quality areas. Their sampling stations can be seen in Table 2, with their spatial locations in Fig. 3.

Table 2 Sampling stations forming the cluster groups

Fig. 3 Spatial distribution of the separated groups

Along with the three large, two smaller areas could be distinguished in the NW and SE regions, and three sampling sites (29, 34, 38) formed separate “groups” in themselves. The characteristics of the groups will only be interpreted after their statistics are discussed.

Discriminant analysis pointed out that the original groups were correctly identified in 89.0% of the cases. This did not change much even after cross validation (86.6%). The result of the DA can be seen in Fig. 4.

Fig. 4 Cross-validated discriminant results

Parameters influencing the grouping of the sampling locations

A Wilks' lambda quotient was calculated for each of the 34 time points, for every parameter examined, and their averages were determined. The smallest (0.2-0.3) quotient belongs to the main ions. These are followed by pH (0.38), dissolved oxygen (0.42), with even higher values coming after them TP (0.53), NO₃-N (0.53), NH₄-N (0.53) and chlorophyll-a (0.62). As a next step the obtained Wilks' lambda quotients (34 x the number of parameters) were clustered.

It is clear that the parameters can be distributed into two distinct groups according to their Wilks' lambda quotients on the dendrogram (Fig. 5). The main anions and cations with pH and dissolved oxygen (average 0.29) influenced the formation of the cluster groups the most, despite the fact that on the Piper diagram (Fig. 2) no groups were distinguishable. Parameters in close relation to the nutrient cycle, however, had a smaller influence (average 0.55).

Fig. 5 Dendrogram of Wilks' lambda quotients

Because Neusiedler See is a sodic lake the input waters, which are low in ions, may produce great differences between the areas. This may be the reason why ions have a much more determining role in forming the cluster groups than

the other parameters. Aside from the ions, pH and dissolved oxygen were among the more determining parameters. Their concentrations show big differences in the two main habitats of the lake (reed belt, open water).

Group statistics

The groups obtained from the 34 time points were compared using boxplots. The median of pH (Fig. 6a) in Groups 1 and 4 is higher than in the others, the smallest values can be observed in Groups 7 (Station 34) and 8 (Station 38). In Group 6 (Station 29) the interquartile interval of pH was the widest, this means that this is the place where its values varied between the widest boundaries. The concentrations of dissolved oxygen (Fig. 6b), - as in the case of pH - were higher in the eastern part of the lake (Groups 1 and 4) than elsewhere. The lowest values were observable in Group 7 (Station 34).

The ions, except for calcium, behave similarly, so sodium was chosen to represent these ions (Fig. 6c). The smallest concentrations were observable in Groups 5 and 6, while the highest values were in Groups 4, 7 and 8. The median of total phosphorus (Fig. 6d) is higher in Groups 5, 6, and 8 than in the others. In Groups 1, 2 and 4, which cover most of the eastern and central part of the lake, 45 extreme values were recorded, 27 of them in April and May 2004. The chlorophyll-a concentrations in Groups 1, 2, 4 and 6 (the middle and eastern parts of the lake) are lower than in the western, Groups 3, 5 and 7. In Group 8 (Station 38) values clearly higher than the average were registered on many occasions (Fig. 6e).

Fig. 6 Box-and-whiskers plot: pH (a); dissolved oxygen (b); sodium (c); TP (d); Chl-a (e)

Differences became obvious between the groups based on coefficients of variation (CV). These CVs were determined for the whole lake and the distinct groups. The groups' CVs were compared to each other and the whole lake as well. In the following, the CVs of a few exceptional parameters will be discussed (Fig. 7). H^+ concentration⁶ in Group 6 for example shows extremely high variability, while if the other groups are compared, the western part of the lake (Groups 3, 5, 7 and 8) is still more variable than the part comprising Groups 1, 2 and 4. The CV of dissolved oxygen is similar to that of the pH, which is lower in Groups 1, 2 and 4; on the contrary the CV of Group 7 (Station 34) is three times the average of the whole lake. The CVs of the ions are highest in Groups 5 and 6. In the case of total phosphorus there are again great spatial differences between the groups. In the western groups (3, 7 and 8, but not Group 5), the CV of TP is half the average for the lake and one-third of that of Groups 1 and 4. Last but not least the variability of chlorophyll-a is higher in Groups 1, 4 and 7 than the average for the lake.

Fig. 7 CVs of the parameters in each cluster group

Detailed group descriptions based on the basic statistics

The eight groups individually cover certain (smaller or larger) areas in the Austrian part of the lake. Referring to the covered areas each group received a name.

Group 1 ("open water") contains most of the lake's sampling sites. Here the pH and dissolved oxygen values are higher, while the parameters in close relation

⁶ pH was converted to hydrogen ion concentration because of its logarithmic scale

to the nutrient cycle show lower concentration in comparison to most of the other groups.

Group 2 (“transient”) is considered to be a transition zone between the reed belt and the open water area; it covers the SW part of the lake and consists mostly of the sampling sites located within a 100m range in front of the reed belt.

Group 3 (“the reed belt”) sampling sites are mainly the ones located in the channels that are used to gain access to the reeds in the case of cutting or mowing and fishing. In these 5-6 m wide channels pH and dissolved oxygen values are lower, while chlorophyll-a shows higher concentrations than at most of the sampling sites.

Group 4 (“the bay”) consists of the two sampling sites (Stations 3 and 4) located in the SE part of the lake. This area often “behaves” differently from the open water, which has its own general current system operating clockwise, in most cases independently of the wind direction. Here the retention time of ions increases. The connection between Group 4 and the area with the general current is ensured by a couple of narrow channels. In bays like Group 4 mixing of the water only occurs when there are suitable wind conditions, and in certain cases currents can even developed which are opposite to the general current (counterclockwise) of the open water (Löffler 1979). However if the wind is strong enough, intensive exchange can be observed between the “separated” bay and the open water. This may be the reason for the separation of sampling sites 3 and 4. In this particular group the concentration of dissolved oxygen and the main anions and cations (except calcium) is noticeably high and the chlorophyll-a is low.

Group 5 (“the Wulka”) is influenced by the inflow of the river Wulka. It enters the lake in a diffuse way, through a 5 km long reed belt. The low ion

1 concentrations are caused by the diluting effect of the Wulka (originating from the
2 Lajta Mountains) (Grosina 1990). From the high phosphorus values it became
3 obvious that this input source can be responsible for a considerable part of the
4 lake's phosphorus loads; it transports the phosphorus washed into its water from
5 agricultural land. This frequently happens during heavy rains.
6
7
8
9

10
11 Group 6 ("the sewage treatment plant") consist of only one sampling site
12 (Station 29). It is located near the sewage treatment plant situated in the vicinity
13 of Podersdorf. The ion concentrations at sampling site 29 are lower than in its
14 surroundings and the interquartile interval of pH is quite wide. This may happen
15 because of the chemical composition of the treated sewage reaching the lake
16 causing a local effect (dilution). It must be stated that the local currents carry these
17 extra inputs towards the inner parts of the lake.
18
19
20
21
22
23
24
25
26
27
28

29 Group 7 ("the weekend houses") consists of sampling site 34 alone. In the
30 group's surroundings many weekend houses can be found, all connected to
31 communal sewage system. Despite this fact, extremely low dissolved oxygen
32 values were occasionally recorded. The reason behind this is the oxygen
33 consumption of the organic substances oxidation.
34
35
36
37
38
39
40

41 Last but not least Group 8 ("the hotel") again contains only one sampling
42 site (Station 38), which is located near a hotel. The water in front of the hotel has
43 good light conditions because it is shielded from the wind by the nearby buildings.
44 Furthermore it has a high nutrient content (like all the areas near the shore). Again,
45 the conditions are optimal for algae to reproduce. This may be the reason behind
46 the high chlorophyll-a content of the water along its low CV. The concentrations
47 of pH and dissolved oxygen were low, while the values for ions showed high
48 concentrations. This sampling location separated from its neighboring ones
49 because of the high ion concentration.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Differences between the groups

The differences between the groups can be seen in the boxplots. For example the chemical, biological processes in the reed belt produce humic substances which originate mostly from the decomposition of reed detritus. This, and the diluting effect of the inflowing water results in lower pH in Groups 3, 5, 7 and 8 (Fig. 6_a). These inflowing waters can be considered to be truly representative of the lake only after flowing slowly through the reed belt. In the open water, because of the frequent mixing, and low nutrient input, the oxygen conditions are good. On the contrary, in the reed belt oxygen concentrations are low (Fig. 6_b) or in certain cases even anaerobic conditions may occur (Löffler 1979; Dinka and Szeglet 2001). The lack of mixing and the oxidization of organic substances decreases dissolved oxygen content.

In the reed channels (Groups 3, 7 and 8, but not Group 5) the CV of total phosphorus is lower than average (Fig. 7), because the reed stands have a role in the exchange of nutrients, absorbing nutrients from the sediment and water, and returning these through the processes of excretion and mineralization (Dykyjová and Úlehlová 1998), thereby forming a crucial link in the detritus food chain (Wetzel 2001; Dinka et al. 2010). The reed belt functions as a sink for a significant percentage of the phosphorus coming from diffuse sources (agriculture, livestock, etc.), dampening its effect in the channels.

The chlorophyll-a parameter, an interesting situation should be discussed, as in the open water (Groups 1, 2, 4 and 6) the chlorophyll-a content is much lower than in the channels and in front of the reed belt in the western part of the lake (Groups 3, 5, 7 and 8) (Fig. 6_c). This happens because the reeds protect the water

from the wind and as a result mixing is less frequent, while the width of the channels (5-6 m) enables good light conditions and the area is rich in nutrients providing good conditions for algae growth. On the contrary in the open water wind frequently stirs up the sediment and light conditions limit algae growth, in this way prohibiting their multiplication.

Principal component analysis

In determining the latent variables present over a long time period in the lake and the distinct groups, PCA was applied to the data for the years 2000-2009⁷ (Table 3), except for the year with extreme droughts (2004) (Table 4). Only component loadings with an absolute value higher than 0.70⁸ were considered to be a part of further investigations (in bold). Only the first two principal components were evaluated because these were the only ones explaining a significant part of the data's variance.

Latent variables present in the whole lake for the time interval 2000-2009

Table 3 1st and 2nd principal components of the PCA results (2000-2009) except 2004

Two principal components were identified as responsible for 57.6% of the data's variance. The first one explained 40.3% of the total variance of the data and in it only the ions with the exception of calcium had significant component

⁷ From this point on the time interval 2000-2009 only refers to the years without 2004.

⁸ The decision regarding the margin where a component score is considered to be relevant or not was made based on the study of Cloutier et al. 2008.

loadings, while the second explained 17.3% of total variance where only the pH and dissolved oxygen parameters had high component loadings. The component scores were determined for each sampling site; as a result, we were able to produce isoline maps describing their spatial changes. Three areas were distinguishable on this map regarding the first PC (Fig. 8_a): two areas with smaller component scores in the NE and NW parts of the lake, and one with high component scores on the western side. The second PC's component scores gradually increase from NW to SE (Fig. 8_b), suddenly reaching an explicit boundary (open water), indicating a change in habitat. The anomaly seen in the NW part of the lake in the case of the first PC cannot be observed in the case of the second PC.

Fig. 8 Spatial distribution of the 1st (a), and the 2nd PC's component scores (b) for the years 2000-2009 except 2004

Latent variables present in the whole lake for the exceptional year 2004

Table 4 1st and 2nd principal components of the PCA results (2004)

Regarding the exceptional year 2004, the first PC explained 43.8% and the second 18.6% of the total variance in the data. As in the case of the PCA conducted on the longer datasets (2000-2009) the ions were present in the first PC except for calcium. The difference was that in the case of 2004 HCO₃⁻ component loadings did not reach the critical 0.70 value. In the second PC pH lost its determining role, while TP and calcium gained their position next to dissolved oxygen as a determining parameter. In the first PC's map (Fig. 9_a) there is no

noticeable difference between 2004 and years 2000-2009, however in the second PC's map (Fig. 9_b) the component scores show a gradual decrease from their maximum in the NW part of the lake towards its center. This picture is highly different from the one presented in Fig. 8_b. The PCA maps were prepared for each individual year as well, these concurred with the map of the whole time interval (2000-2009) except of course for the year 2004.

Fig. 9 Spatial distribution of the 1st (a), and the 2nd PC's component scores (b) for the year 2004

Given that Neusiedler See is a salt lake, during PCA in the first PC only the inorganic parameters were dominant (except for calcium). It can be stated that ions determine the variance of the dataset the most, while calcium and biogenic carbonate precipitation play a less significant role. In the second PC the reason behind the determining role of pH and dissolved oxygen is the presence of humic substances and the oxidation of organic matter respectively. These are responsible for the water's brown color in the reed belt (Dokulil and Herzig 2009).

From the component loading maps (2000-2009) (Figs. 8_a and 8_b) it became clear that the Wulka, the sewage treatment plant, the hotel and the weekend houses can have a pronounced and local impact on the processes evolving in Neusiedler See. Because all of these sampling sites exercise influence on the ion content of the water, all of their locations can clearly be seen on the first PC's maps (Figs. 8_a and 9_a)s. On the maps with the values of the second PC (Fig. 8_b and 9_b) the effect of the Wulka cannot be seen because it does not have a serious effect on the pH and oxygen conditions, unlike the other three sampling sites (29, 34, 38). The key to the year 2004 being separate lies in the unusually high phosphorus values. This was caused by the unusually strong wind conditions which stirred up

the lakes sediment this way increasing the suspended solids. This may have happened in 2003 as well when the water in the reed belt almost fully evaporated, so its role as a significant sink factor for TP decreased.

Latent variables present in the different cluster groups

Because cluster analysis pointed out that there is more than one distinguishable area in the lake, PCA was conducted on the individual groups as well. Only those groups (Groups 1, 2, 3) were part of the analysis which had enough samples to conduct PCA on. The component loadings of the three groups can be seen in Table 5.

Table 5 Component loadings in Groups 1, 2, 3

The first PC explained 37-43% of the total variance of the data (Table 5) and the second PC 13-16%. As in the case of the previous analyses (2000-2009) the ions were present in the first PC except for again calcium. It is worth mentioning that the component loading of HCO_3^- in Groups 2 and 3 was lower than 0.70. The second PC does not show homogeneity, as did the first PC. In the second PC of Group 1 calcium and TP, in Group 2 dissolved oxygen and in Group 3 pH and dissolved oxygen had component loadings higher than 0.70.

Regarding the results obtained from the PCA conducted on the groups it can be said that the differences in the second PC originate from the different habitats' processes.

Conclusions

This study demonstrated that uni-and multivariate data analysis methods are excellent tools in interpreting complex datasets, as in this case in which data were retrieved from 33 sampling locations of Neusiedler See, regarding 13 chemical and biological parameters for the time interval 2000-2009. With cluster analysis the sampling sites were grouped and the grouping verified by discriminant analysis. At the different time points the sites showed different group patterns which indicate the variability of the lake's habitats and processes (just as the CVs did). Eight different areas were distinguished. Using Wilks' lambda statistics it was verified that the ions are most responsible for the conformation of the previously discussed cluster groups. The parameters closely connected to the nutrient content were less relevant in this question.

From the data of the selected 34 time points boxplots were produced. The parameters were then compared regarding each group. The anthropogenic activity in the lake's surroundings has a great and localized effect on the lake's water chemistry. The Wulka and the sewage treatment plant both have a clear effect on ion variability. The local (treated sewage at Station 29) and diffuse (materials with high nutrient content washed in from agricultural areas; Stations 13, 16 and 17) phosphorus loads are both considerable.

The final approach was to identify the parameters accounting for most of the variance of the dataset. It became clear that the ions are most determining. The component score maps proved the suspicion that the inflows and reeds have a great effect on the pH and oxygen conditions along with concentrations of the main anions (Cl^- , SO_4^{2-} , HCO_3^-)' and cations (K^+ , Na^+ , Mg^{2+})', in the area. The determining component scores proved to be stable throughout the investigated ten years. In space the first principal component shows a stable picture while in the

second principal component the component loadings of the variables differ to high extent between the reed belt and the open water.

In summary it can be said that Neusiedler See cannot be treated as one homogeneous system. The parameters are highly variable in space and time, to a greater degree than was suspected. This exceptional variability originates from the lake's shallow water depth, unstable water balance, and anthropogenic activity (agriculture, tourism, sewage treatment etc.) in the lake's surroundings. Based on the information obtained it is possible to optimize the current management even further, in this way increasing its cost efficiency, which is a yet greater question in this period of financial crisis.

Acknowledgments

We the authors would like to thank the Biologische Station Neusiedler See in Illmitz for providing the datasets; furthermore we would like to thank Georg Wolfram and the colleagues working at the Department of Physical and Applied Geology for the stimulating discussions.

References

Back, W. (1966). *Hydrochemical facies and ground-water flow patterns in northern part of Atlantic Coastal Plain*. U. S. Geol. Surv. Prof. Paper 424-D, Washington, DC: United States Government Printing Office.

Cloutier, V., Lefebvre, R., Therrien, R., & Savard, M. M. (2008). Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *Journal of Hydrology*, doi: 10.1016/j.jhydrol.2008.02.015

Daniel, J. J. (1999). *Introduction to Atmospheric Chemistry*. New Jersey: Princeton Univ. Press.

Dinka, M., & Szeglet, P. (2001). Some characteristics of reed (*Phragmites australis*/CAV./TRIN ex STEUDEL) that indicate different health between vigorous and die-back stands. *Verh. Internat. Ver. Limnol.*, 27, 3364-3369.

- Dinka, M., Ágoston-Szabó, E., & Szeglet, P. (2010). Comparison between biomass and C, N, P, S contents of vigorous and die-back reed stands of Lake Ferto/Neusiedler See. *Biologica*, doi: 10.2478/s11756-010-0006-x
- Dinka, M., Ágoston-Szabó, E., Berczik, Á., & Kutrucz, Gy. (2004). Influence of water level fluctuation on the spatial dynamic of the water chemistry at Lake Fertő/Neusiedler See. *Limnologica*, 34(1–2), 48–56.
- Dobesch, H. & Neuwirth, F. (1979). Climatic conditions. In H. Löffler (Ed.), *Neusiedlersee. The limnology of a shallow lake in Central Europe* (pp. 47-64). Håga: Dr. W. Junk bv. Publishers.
- Dobesch, H. & Neuwirth, F. (1979). Water balance. In H. Löffler (Ed.), *Neusiedlersee. The limnology of a shallow lake in Central Europe* (pp. 79-84). Håga: Dr. W. Junk bv. Publishers.
- Dokulil, M. T., & Herzig, A. (2009). An analysis of long-term winter data on phytoplankton and zooplankton in Neusiedler See, a shallow temperate lake, Austria. *Springer Science Aquat. Ecol.*, doi: 10.1007/s10452-009-9282-3
- Dykyjová, D. & Úlehlová, B. (1998). Mineral economy and cycling of minerals in wetlands. In D. F. Westlake, J. Kvet, & A. Szczepansky (Eds.), *The production ecology of wetlands* (pp. 319-366). Cambridge: Cambridge University Press.
- Filik Iscen, C., Emiroglu, Ö., İlhan, 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.*, doi: 10.1007/s10661-007-9989-3
- Freeze, R. A., & Cherry, J. A. (1979). *Groundwater*. New Jersey: Prentice-Hall.
- Grosina, H. (1990). *Forschungsbericht 1987-1989*. Eisenstadt: Rötzer-Druck.
- Güler, C., Thyne, G. D., McCray, J. E., & Turner, A. K. (2002). Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology Journal*, doi: 10.1007/s10040-002-0196-6

Hatvani, I. G., Kovács, J., Kovácsné Székely, I., Jakusch, P., & Korponai, J. (2010). 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.*, doi:10.1016/j.ecoleng. 2010.12.028

Hill, R. A. (1940). Geochemical patterns in Coachella Valley, California. *Trans. Am. Geophys. Union*, 21, 46–49.

Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis (3rd ed.)*. New Jersey: Prentice Hall.

Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. I. T., Armi, A. S. M., Toriman, M. E., & Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environ. Monit. Assess.*, doi: 10.1007/s10661-010-1411-x

Jungwirth, M. (1979). Currents. In H. Löffler (Ed.), *Neusiedlersee. The limnology of a shallow lake in Central Europe* (pp. 85-88). Håga: Dr. W. Junk bv. Publishers.

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. *Ecotox. Environ. Safe*, doi:10.1016/j.ecoenv.2008.02.024

Ketskemény, L. & Izsó, L. (2005). *Introduction into the SPSS system (in Hungarian: Bevezetés az SPSS programrendszerbe)*. Budapest, Hungary: ELTE Eötvös Kiadó.

Lattin, J. M., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. New York: Duxbury Press.

McKenna, J. (2003). An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environmental Modelling and Software*, doi: 10.1016/S1364-8152(02)00094-4

Mitsch, W. J., & Gosselink, J. G. (1986). *Wetlands*. New York: Van Norstand Reinhold Company.

Morgan, C. O., & Winner, M. D. Jr. (1962). *Hydrochemical facies in the 400 foot and 600 foot sands of the Baton Rouge Area, Louisiana*. U. S. Geol. Surv. Prof. Paper 450-B. Washington, DC: United States Government Printing Office.

- Neuhuber, F., & Hammer, L. (1979). Oxygen conditions. In H. Löffler (Ed.), *Neusiedlersee. The limnology of a shallow lake in Central Europe* (pp. 121-130). Håga: Dr. W. Junk bv. Publishers.
- Norušis, M. J. & SPSS Inc. (1993). *SPSS for Windows Professional Statistics Release 6.0*. Englewood Cliffs, New Jersey: Prentice Hall.
- Padisák J. (2005). *General limonology (in Hungarian: Általános limnológia)*. Budapest: ELTE Eötvös Kiadó.
- Piper, A. M. (1944). A graphic procedure in the geochemical interpretation of water-analyses. *Trans. Am. Geophys. Union*, 25, 914–923.
- Sargaonkar, A., & Deshpande, V. (2003). Development of an overall index of pollution for surface water based on a general classification scheme in Indian context. *Environ. Monit. Assess.*, doi: 10.1023/A:1025886025137
- Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, doi:10.1016/j.envsoft.2006.02.001
- Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsas, D., Anthemidis, A., Sofoniou, M., & Kouimtzis, Th. (2003). Assessment of the surface water quality in Northern Greece. *Water Research*, doi:10.1016/S0043-1354(03)00398-1
- Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study. *Water Research*, doi: 10.1016/j.watres.2004.06.011
- 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*, doi: 10.1016/S0043-1354(98)00138-9
- Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236-244.

1 Wetzel, R.G. (2001). *Limnology: lake and river ecosystems, 3rd edition*. San Diego: Academic
2 Press.

3
4
5 Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, 24, 471–494.
6

7
8
9 WFD, Directive of the European Parliament and of the Council 2000/60/EC Establishing a
10 Framework for Community Action in the Field of Water Policy, European Union, Luxembourg,
11 2000, PE-CONS 3639/1/00 REV 1.
12

13
14
15
16 Wunderlin, D. A., Diaz, M. P., Ame, M. V., Pesce, S. F., Hued, A. C., & Bistoni, M. A. (2001). Pattern
17 recognition techniques for the evaluation of spatial and temporal variations in water quality. A
18 case study: Suquia river basin (Cordoba-Argentina). *Water Research*, doi: 10.1016/S0043-
19 1354(00)00592-3
20

21
22
23
24 Yidana, S. M., Ophori, D., & Banoeng-Yakubo, B. (2008). A multivariate statistical analysis of
25 surface water chemistry data - The Ankobra Basin, Ghana. *Journal of Environmental*
26 *Management*, 86(1), 80-87.
27

28
29
30
31 Zhou, F., Liu, Y., & Guo, H. (2007). Application of multivariate statistical methods to water quality
32 assessment of the watercourses in Northwestern New Territories, Hong Kong. *Environ. Monit.*
33 *Assess.*, doi: 10.1007/s10661-006-9497-x
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Cover Letter

Dear Madam, Sir,

I am enclosing herewith a manuscript entitled “**Application of multivariate statistical methods in determining the spatial water quality changes of Lake Neusiedl**” submitted to “**Environmental Monitoring and Assessment**” for possible evaluation.

The reason why we chose this paper to submit our manuscript to is that reading previous papers of your journal we truly feel that it fits in its scope perfectly.

In the study we use uni- and multivariate data analysis methods (boxplot, Piper diagram, cluster analysis, discriminant analysis, principal component analysis) on 13 mainly chemical and some biological parameters for the time period 2000-2009 from 33 sampling sites of Lake Neusiedl to distinguish areas with different chemical properties within the lake and to determine which background processes are responsible for this pattern.

We believe that the results obtained may help scientists to optimize the monitoring of Lake Neusiedl this way ensuring the long term and cost efficient management of both Hungary's and Austria's natural heritage, Lake Fertő/Lake Neusiedl.

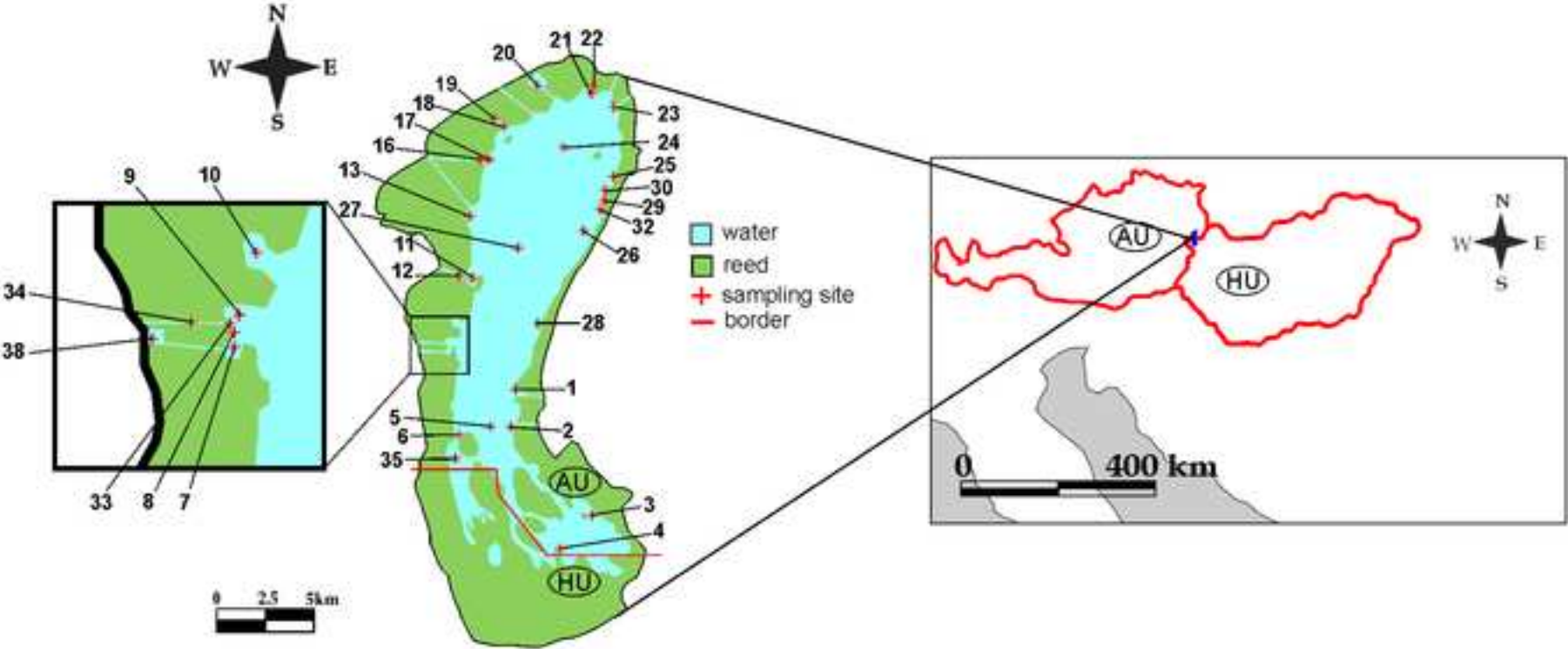
With the submission of this manuscript I would like to undertake that the above mentioned manuscript has not been published elsewhere, accepted for publication elsewhere or under editorial review for publication elsewhere; and that my Institute's (Eötvös Loránd University, Department of Physical and Applied Geology) representative is fully aware of this submission. We would prefer to have the figures produced in color only in the on-line version of the paper. No printed color artwork is needed and no hard copies of the paper either.

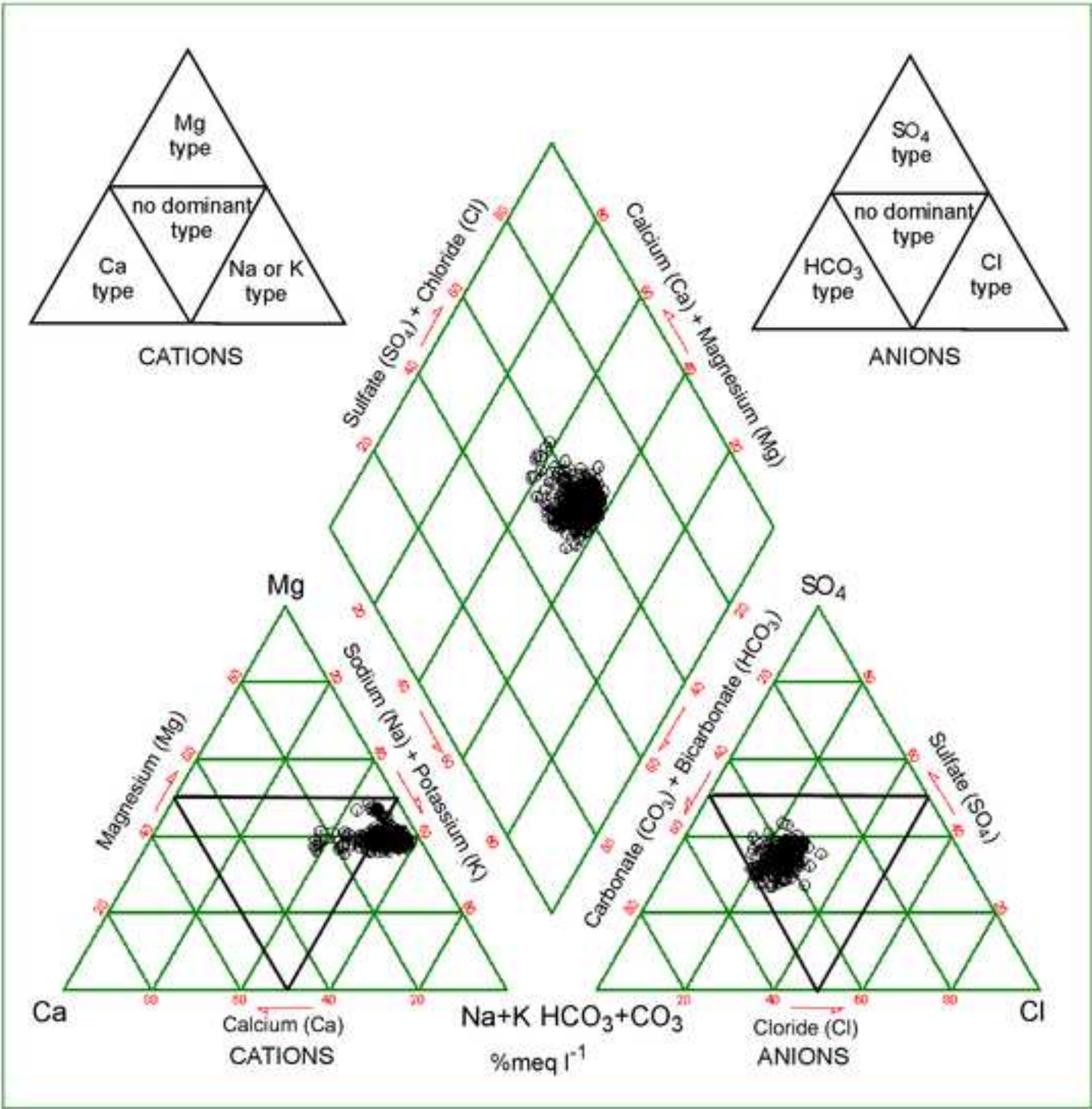
We are looking forward to reading your opinion of the article.

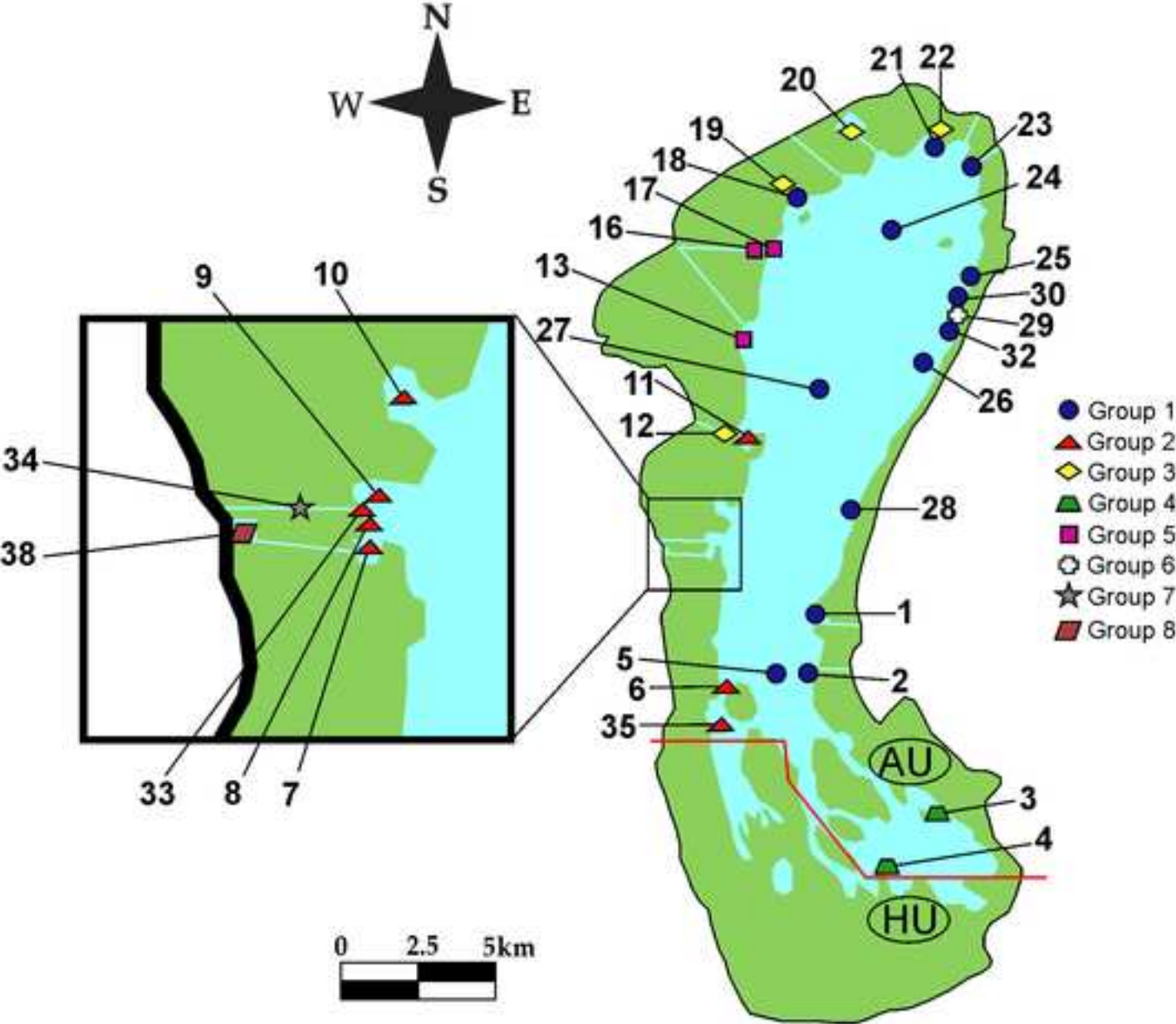
Yours faithfully,

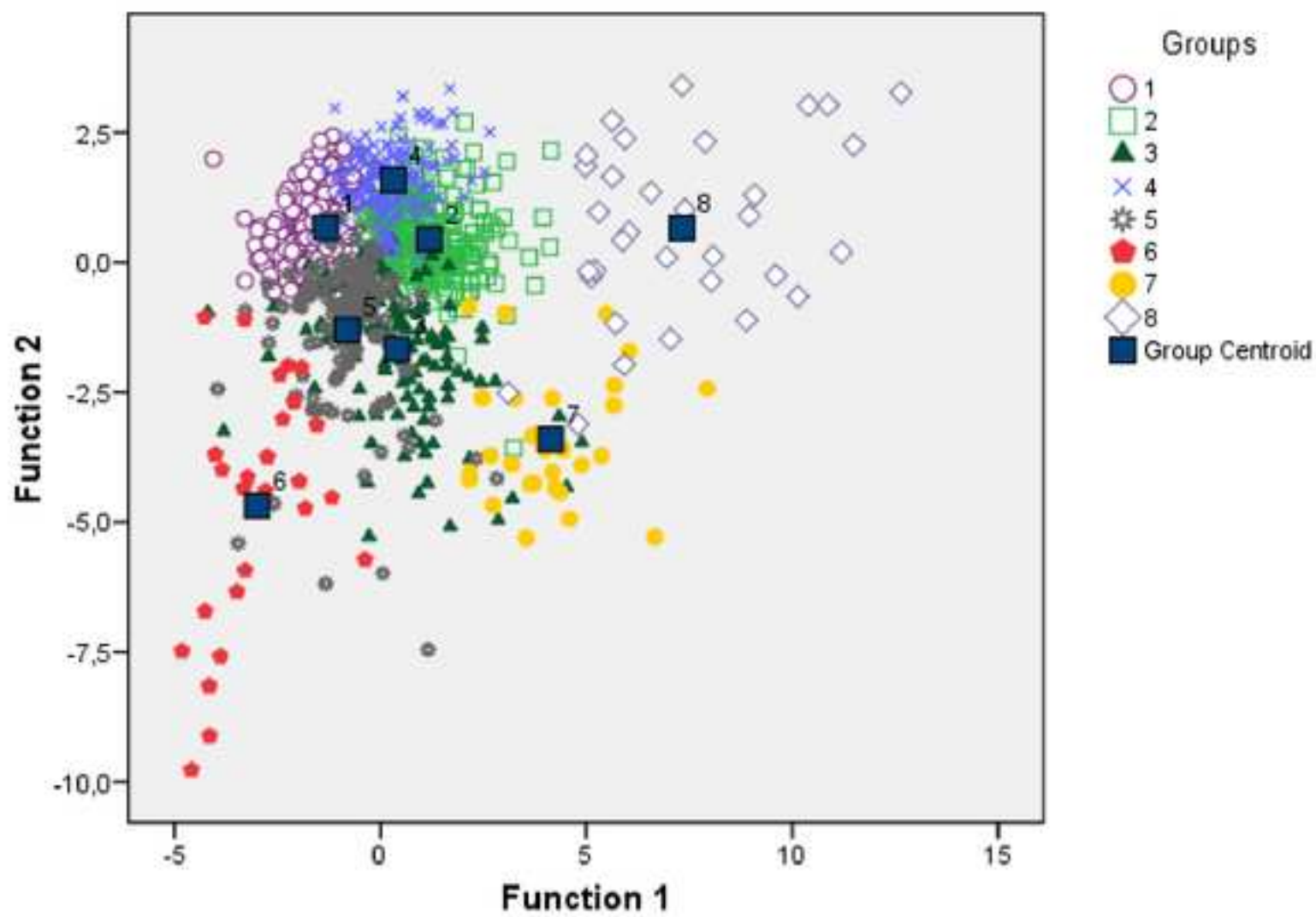
Budapest
10.02.2012

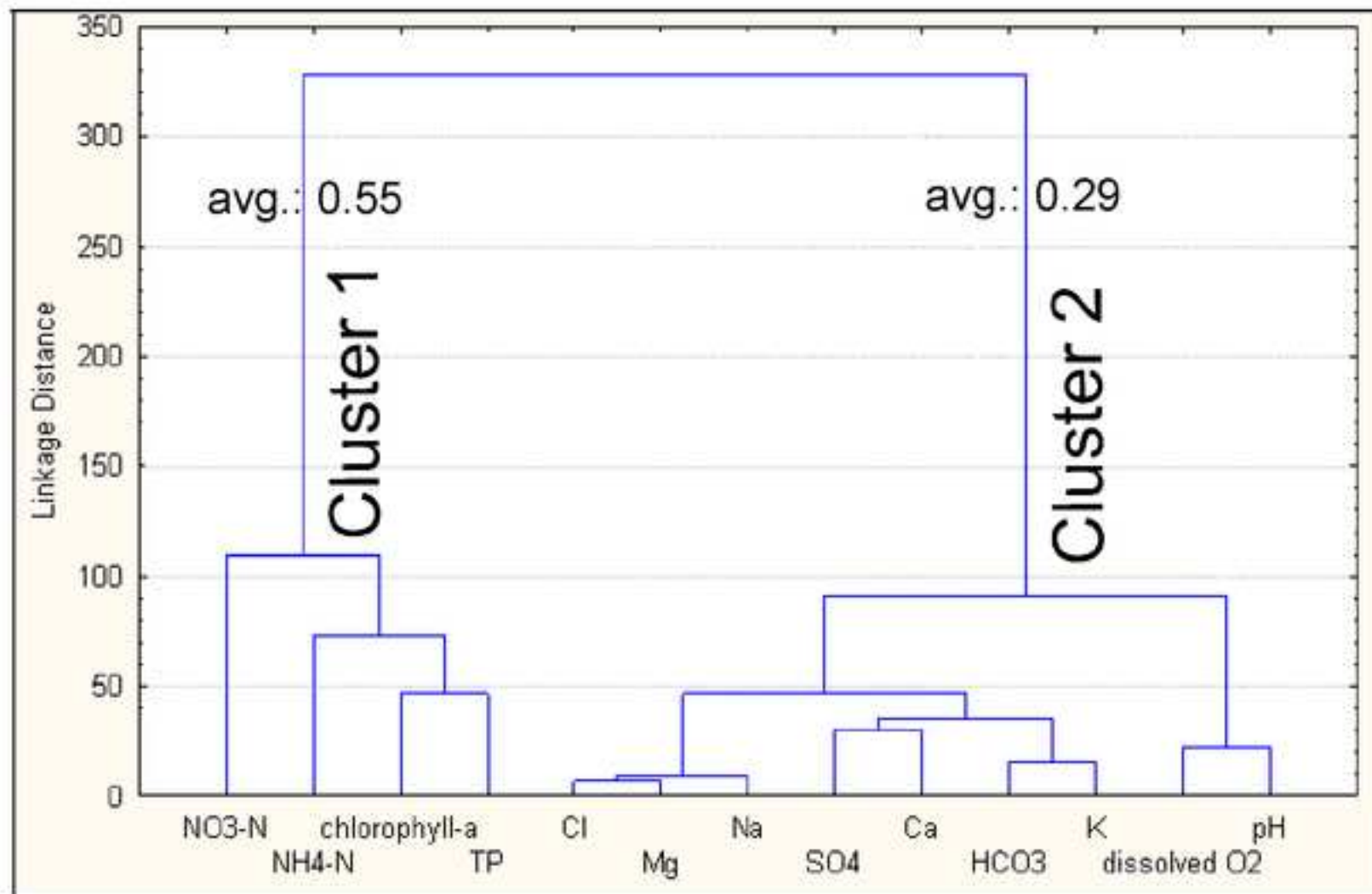
Norbert Magyar

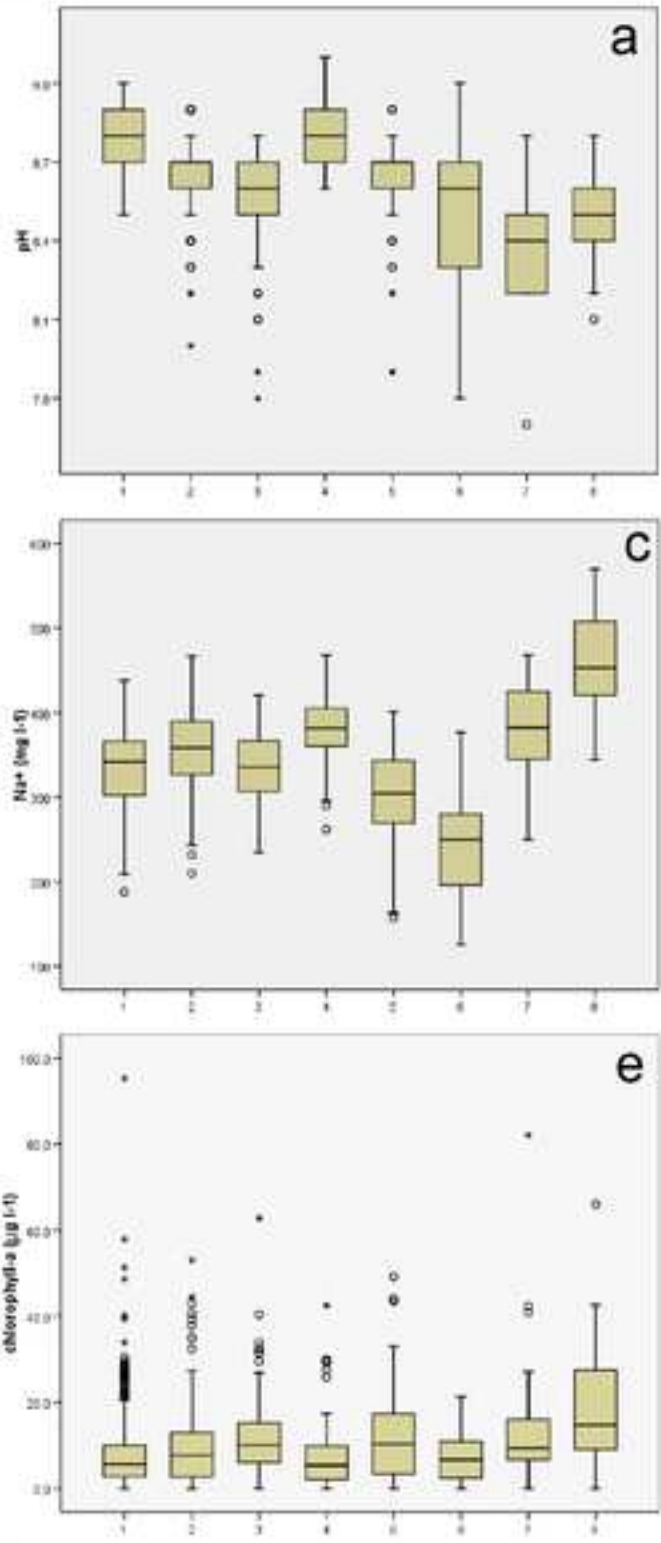
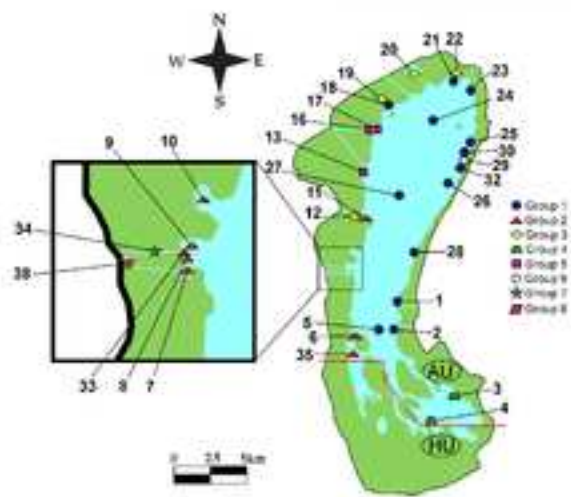




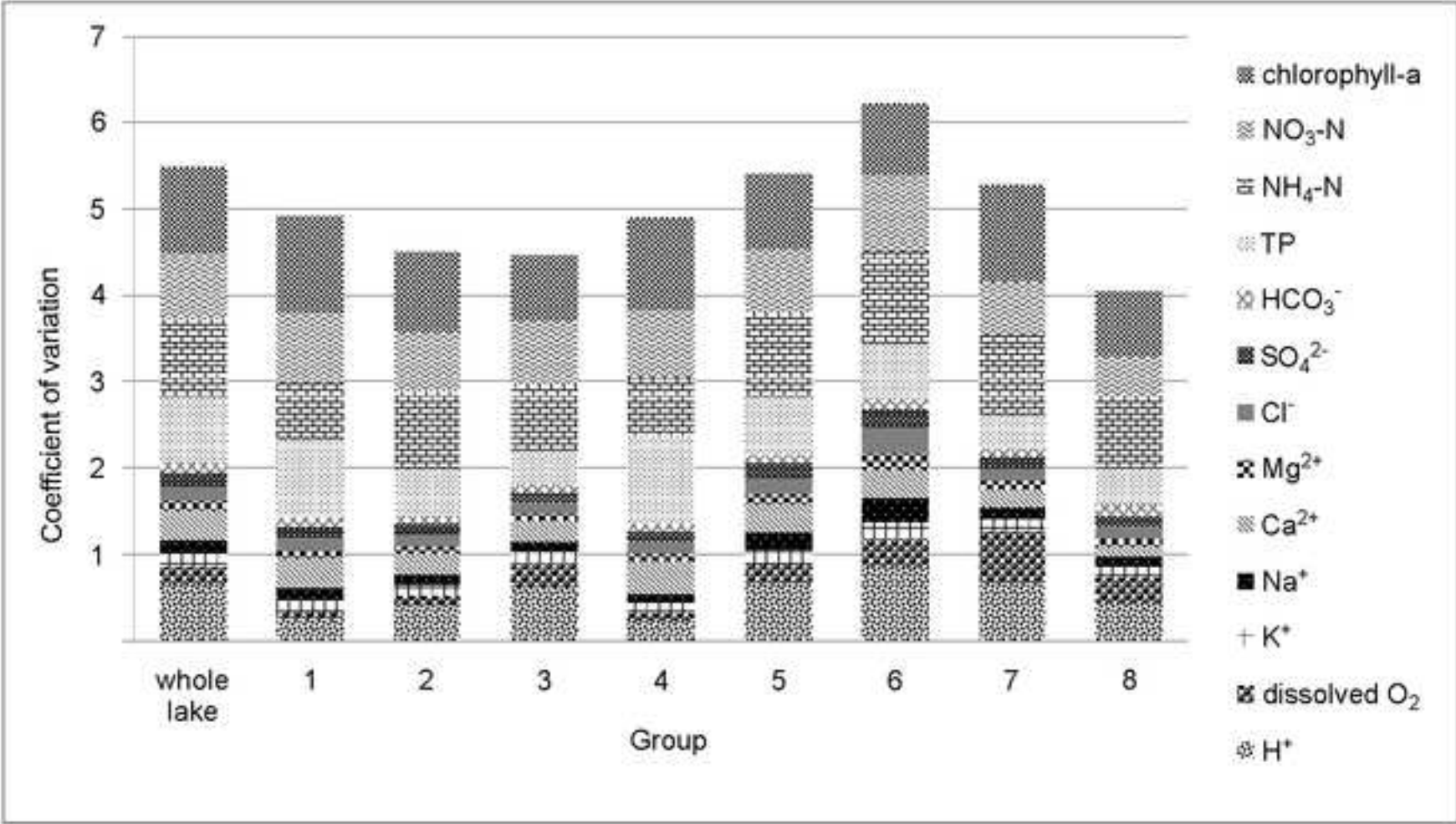


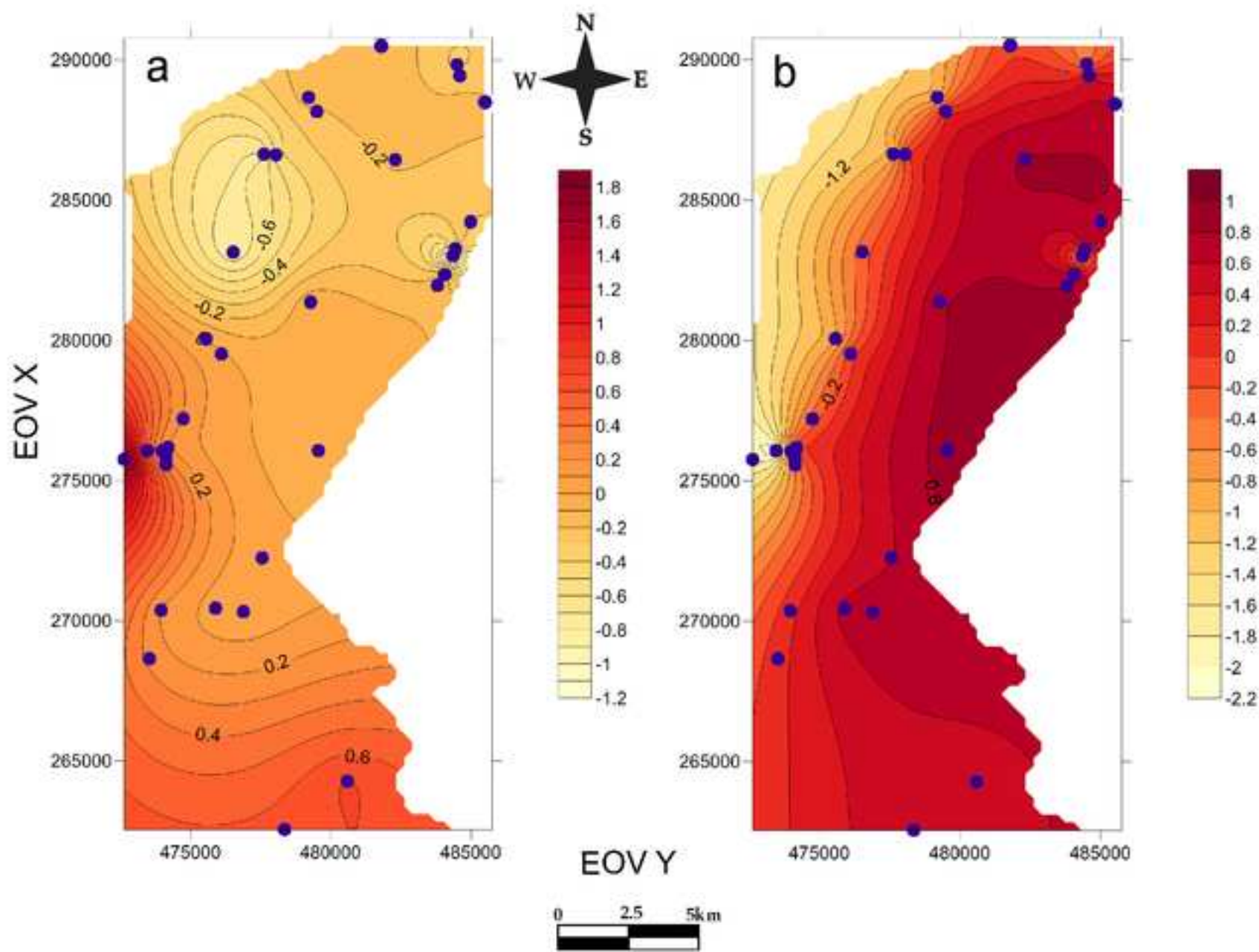




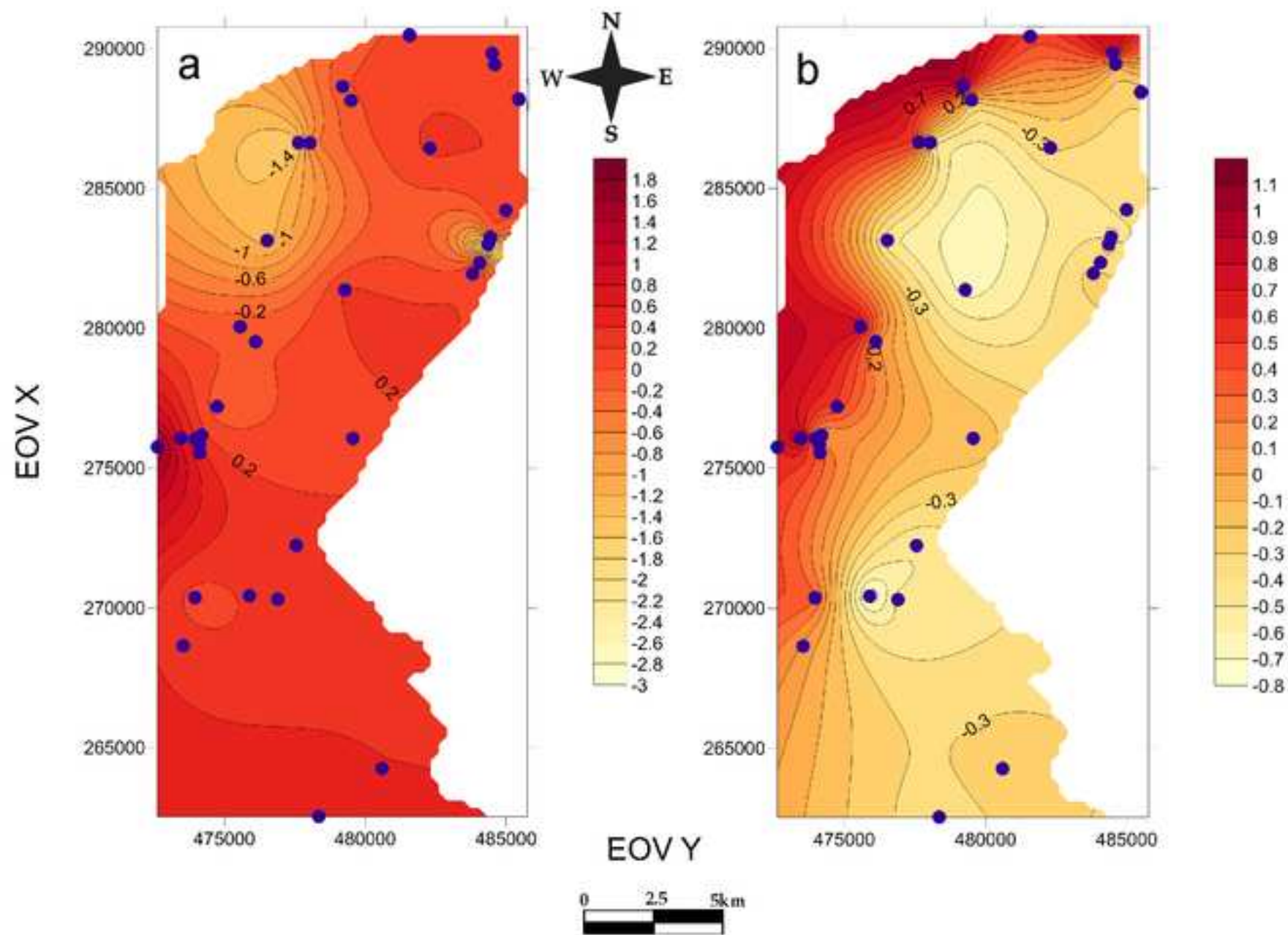


Group





Distribution of the 1st (a), and the 2nd PC's component
[Click here to download high resolution image](#)



	<i>pH</i>	<i>dissolved O₂</i> <i>(mg l⁻¹)</i>	<i>K⁺</i> <i>(mg l⁻¹)</i>	<i>Na⁺</i> <i>(mg l⁻¹)</i>	<i>Ca²⁺</i> <i>(mg l⁻¹)</i>	<i>Mg²⁺</i> <i>(mg l⁻¹)</i>	<i>Cl⁻</i> <i>(mg l⁻¹)</i>	<i>SO₄²⁻</i> <i>(mg l⁻¹)</i>	<i>HCO₃⁻</i> <i>(mg l⁻¹)</i>	<i>TP</i> <i>(µg l⁻¹)</i>	<i>NH₄-N</i> <i>(µg l⁻¹)</i>	<i>NO₃-N</i> <i>(µg l⁻¹)</i>	<i>Chl-a</i> <i>(µg l⁻¹)</i>
<i>Average</i>	8.68	8.36	37.44	343.55	36.40	140.95	244.01	466.14	607.86	82.11	36.19	201.67	9.56
<i>Median</i>	8.7	8.8	38	347	34	141	245	473	600	68	28	152.5	7.1
<i>Standard deviation</i>	0.18	1.79	5.56	55.95	12.10	16.59	42.10	66.48	76.90	62.28	32.98	151.35	9.62
<i>Minimum</i>	7.7	0.2	13	126	14	72	94	248	375	3	2	0	0
<i>Maximum</i>	9.1	11.2	55	570	119	202	422	661	1061	552	333	872	95.3
<i>Range</i>	1.4	11.0	42	444	105	130	328	413	686	549	331	872	95.3

<i>Group</i>	<i>Number of sampling stations</i>
<i>1</i>	1, 2, 5, 18, 21, 23, 24, 25, 26, 27, 28, 30, 32
<i>2</i>	6, 7, 8, 9, 10, 11, 33, 35
<i>3</i>	12, 19, 20, 22
<i>4</i>	3, 4
<i>5</i>	13, 16, 17
<i>6</i>	29
<i>7</i>	34
<i>8</i>	38

	1st PC	2nd PC
pH	0.29	0.81
dissolved O₂	-0.07	0.78
K⁺	0.85	0.03
Na⁺	0.96	-0.06
Ca²⁺	-0.31	-0.55
Mg²⁺	0.94	-0.05
Cl⁻	0.95	-0.06
SO₄²⁻	0.92	0.04
HCO₃⁻	0.73	-0.51
TP	-0.08	-0.37
NH₄-N	-0.12	-0.47
NO₃-N	-0.39	-0.17
chlorophyll-a	0.19	-0.16
Explained variance	5.23	2.25
Proportion of total	0.403	0.173

	1st PC	2nd PC
pH	0.54	-0.34
dissolved O₂	0.17	-0.70
K⁺	0.96	0.07
Na⁺	0.98	0.06
Ca²⁺	-0.42	-0.73
Mg²⁺	0.91	-0.23
Cl⁻	0.95	0.22
SO₄²⁻	0.90	0.32
HCO₃⁻	0.67	-0.15
TP	0.07	-0.87
NH₄-N	-0.35	0.39
NO₃-N	-0.34	0.03
chlorophyll-a	0.29	-0.36
Explained variance	5.69	2.42
Proportion total	0.438	0.186

	Group 1		Group 2		Group 3	
	1 st PC	2 nd PC	1 st PC	2 nd PC	1 st PC	2 nd PC
pH	0.59	0.44	0.37	0.64	0.33	0.78
dissolved O ₂	-0.22	0.00	-0.16	0.75	-0.03	0.88
K ⁺	0.85	0.08	0.84	-0.02	0.88	0.04
Na ⁺	0.96	0.02	0.93	0.01	0.97	-0.02
Ca ²⁺	-0.05	-0.80	-0.07	-0.55	-0.07	0.02
Mg ²⁺	0.93	-0.05	0.93	0.06	0.92	0.02
Cl ⁻	0.95	0.06	0.94	0.01	0.90	-0.08
SO ₄ ²⁻	0.95	0.00	0.94	-0.02	0.89	0.14
HCO ₃ ⁻	0.78	-0.43	0.69	-0.36	0.60	-0.66
TP	-0.02	-0.75	0.14	-0.37	0.08	-0.07
NH ₄ -N	-0.02	0.04	-0.13	-0.56	-0.01	-0.39
NO ₃ -N	-0.50	0.06	-0.48	0.05	-0.29	0.06
chlorophyll-a	0.17	-0.37	0.14	0.48	0.27	0.36
Explained variance	5.58	1.74	5.15	2.08	4.80	2.14
Proportion of total	0.429	0.134	0.396	0.16	0.369	0.164

Number of sampling site	Location of sampling site	Sampling site represents
1	N 47°46'16"; E 16°44'47"	water in front of the reed belt (E)
2	N 47°45'13"; E 16°44'18"	water in front of the reed belt (SE)
3	N 47°42'01"; E 16°47'24"	water in the "bay" near the boarder
4	N 47°41'03"; E 16°45'39"	water in the "bay" near the boarder
5	N 47°45'16"; E 16°43'30"	open water
6	N 47°45'12"; E 16°41'57"	water in front of the reed belt (SW)
7	N 47°48'00"; E 16°41'58"	water in front of the reed belt (W)
8	N 47°48'10"; E 16°41'57"	water in front of the reed belt (W)
9	N 47°48'20"; E 16°42'00"	water in front of the reed belt (W)
10	N 47°48'54"; E 16°42'25"	water in front of the reed belt (W)
11	N 47°50'10"; E 16°43'27"	water in front of the reed belt (W)
12	N 47°50'27"; E 16°43'00"	water in the channel
13	N 47°52'08"; E 16°43'42"	water in front of the reed belt (W)
16	N 47°54'02"; E 16°44'50"	water in the channel
17	N 47°54'02"; E 16°44'30"	water in front of the reed belt (NW)
18	N 47°54'53"; E 16°45'58"	water in front of the reed belt (NW)
19	N 47°55'09"; E 16°45'43"	water in the channel
20	N 47°56'20"; E 16°47'44"	water in the channel
21	N 47°55'39"; E 16°50'02"	water in front of the reed belt (N)
22	N 47°55'52"; E 16°49'57"	water in the channel
23	N 47°55'08"; E 16°50'59"	water in front of the reed belt (N)
24	N 47°54'00"; E 16°48'15"	open water
25	N 47°52'51"; E 16°50'28"	water in front of the reed belt (E)
26	N 47°51'36"; E 16°49'34"	open water
27	N 47°51'13"; E 16°45'57"	open water
28	N 47°48'22"; E 16°46'18"	water in front of the reed belt (E)
29	N 47°52'11"; E 16°49'59"	water in the vicinity of the sewage treatment plant
30	N 47°52'19"; E 16°50'03"	water in front of the reed belt (E)
32	N 47°51'49"; E 16°49'46"	water in front of the reed belt (E)
33	N 47°48'16"; E 16°41'50"	water in front of the reed belt (W)
34	N 47°48'16"; E 16°41'25"	water in the vicinity of the weekend houses
35	N 47°44'16"; E 16°41'39"	water in front of the reed belt (SW)
38	N 47°48'05"; E 16°40'45"	water in front of the Hotel

Table captions

Table 1	Descriptive statistics of the parameters used
Table 2	Sampling stations forming the cluster groups
Table 3	1 st and 2 nd principal components of the PCA results (2000-2009) except 2004
Table 4	1 st and 2 nd principal components of the PCA results (2004)
Table 5	Component loadings in Groups 1, 2, 3

Figure captions

- Fig. 1 Location of Neusiedler See and its sampling sites
- Fig. 2 Piper diagram for anion and cation facies derived from major ion percentages
- Fig. 3 Spatial distribution of the separated groups
- Fig. 4 Cross-validated discriminant results
- Fig. 5 Dendrogram of Wilks' lambda quotients
- Fig. 6 Box-and-whiskers plot: pH (a); dissolved oxygen (b); sodium (c); TP (d); Chl-a (e)
- Fig. 7 CVs of the parameters in each cluster group
- Fig. 8 Spatial distribution of the 1st (a), and the 2nd PC's component scores (b) for the years 2000-2009 except 2004
- Fig. 9 Spatial distribution of the 1st (a), and the 2nd PC's component scores (b) for the year 2004