CHAPTER 18
APPLICATION OF CHEMOMETRIC TECHNIQUES IN ANALYTICAL EVALUATION OF BIOLOGICAL AND ENVIRONMENTAL SAMPLES

Piotr Szefer
Department of Food Sciences, Medical University of Gdańsk, Al. Gen. J. Hallera 106, 80-416 Gdańsk, Poland

ABSTRACT
The advantages of using multivariate methods in evaluation of environmental and biological data with respect to mineral composition of the objects studied are demonstrated. Special attention is paid to interpret concentration data concerning environmental samples (biological and geological material) coming from the marine environments. The applicability of advanced statistical chemometric techniques such as principal component analysis (PCA), factor analysis (FA), linear discrimination analysis (LDA), canonical discriminant analysis (DA), end-member analysis, cluster analysis (CA) and neural network (NN) and their capability in interpreting the complex environmental data are discussed. These statistical techniques appeared to be very useful in solving many environmental problems and are helpful in both the classification of different features of the examined environmental compartments and the identification of existing pollution pattern. Their application is helpful in a deeper understanding of the situation of selected metals and metalloids, including toxic trace elements, in the marine ecosystems. Multivariate methods are recommended in monitoring studies leading to reduction of reducing the expense for further environmental survey. The chemometric approach is also a powerful tool in analytical evaluation of foodstuff quality.
1 INTRODUCTION
Multivariate data analysis has been presented extensively by several authors (1-5).
The aim of multivariate data analysis is to decompose mixed data structure into its components. In order to reduce relatively large number of variables to a smaller number of orthogonal factors, the concentration data are treated by multivariate statistical methods, i.e. principal components analysis (PCA), factor analysis (FA), linear discrimination analysis (LDA), canonical discriminant analysis (DA), end-member analysis, cluster analysis (CA), neural network (NN), etc.
PCA creates “new” dimensions of the data (6) and evaluates a reduced number of independent factors or principal components describing the information included in a system of characteristic but partly dependent variables. The aim of PCA and FA is to find a few components or factors that explain the major variations within the data matrix. Each component or factor in PCA or FA respectively, are a weighted linear combination of the original variables. Components or factors only with eigenvalues higher than unity should be preferably considered (7, 8). The factor loading quantities the individual variables contribution to the respective factor. The ranking of the factors is characterised by the amount of variance which they explain (9). The main criticism towards PCA is associated with the difficulties in interpreting the components because of the lack of information about their meaning in either a physical or chemical sense. Moreover, in reduction of all the original variables to only a few factors, a relatively small number of components are used to describe a large part of the variation; hence some information is omitted (8). However, according to Kuik et al. (10) also this unexplained variance can be taken into account resulting in improvement of the reliability of this approach.
Cluster analysis consists of a number of various techniques (11). In clustering the objects are grouped so that ‘similar’ objects fall into the same class. Objects in one cluster should be homogenous in relation to some characteristics explaining within cluster properties; they also should be well separated from other elemental groupings (8). Cluster analysis assigns particular variables with similar courses to clusters of variables (9). Clustering techniques are divided into two basic groups, namely hierarchic and non-hierarchic methods. It is important to decide which clustering procedure is the most suitable. According to Sharma (11), Wards’s minimum variance technique was superior because of providing a larger amount of correct classified observations as compared to most other methods, although it is not always better than average linkage clustering. This finding was supported by Massart and Kaufman (12). One of the major difficulties and criticisms of the technique is defining of objectivity (8, 13). Clustering always produces some clusters, even if the results are completely random and that most methods are biased towards finding spherical and elliptical shaped clusters. When another shape of cluster is obtained, these clusters are not always found to cause a loss of information and sometimes even misleading data (3, 13, 14).
Discriminant analysis determines variations between groups of nominal ‘elements’ which are characterised by numerical variables. Discriminant functions depending linearly on the element concentration studied are formed. The numerical values of the discriminant functions are the coordinates of the locations in a plane described by the two discriminant functions (9). The particular endmember analysis undertaken on the sediment dataset is described by Renner et al. (15). Objectives definition of external endmembers in the analysis of mixtures was given by Full et al. (16). In general, there are
indefinitely many sets of extreme points for a particular set of exact mixtures. However, since associations between elements of a geochemical dataset are not arbitrary, a conservative strategy is to seek extreme compositions (datapoints) that are geometrically close to the data and therefore close to observed reality (16-18). A detailed examination of the multivariate analyse was performed by Renner (18-21) and Renner et al. (22). The abundances of the endmember estimates for any sample are non-negative mixture proportions and therefore also sum to one. Endmember compositions include extreme values for all the elements studied. Depending on the number of endmembers, they are represented geometrically by extreme points or vertices of simplexes (line segments, triangles, tetrahedra etc.). All the datapoints must lie within such a simplex.

Applications of chemometric approach in characterisation of environmental samples
Multivariate statistical techniques have been frequently applied to quantitative evaluation of the distribution of trace, minor, major and macroelements in environmental and biological samples as well as in food products. They have been successfully used for processing concentrations data concerning biological material, e.g. plankton (23), phytobenthos (9), zoobenthos (24-37), fish (28, 38-40), marine mammals (41-43) and mushrooms (44).
The distribution of metals and metalloids in abiotic environmental samples has been also interpreted in view of PCA, FA, DA, etc., including atmospheric fallout and marine aerosols (23, 45-47), suspended matter (23, 48-51), soils (52), stream and marine sediments (8, 13, 53-69), metalliferous sediments (70) and ferromanganese nodules (15, 53, 71).
The statistical analysis of compositional data sets is complicated by the non-negativity and constant-sum constraints, as has been thoroughly documented by Aitchison (72) and others. Ehrlich and Full (17) discussed use of statistical methods in the earth sciences. Q-mode factor analysis of compositional data (especially geochemical and petrologic) has been also presented (73-75).
Spatial, interspecies, inter-size and seasonal and other environmental variations in elemental concentrations are tested by analysis of variance (ANOVA) and the multiple comparison test of Tukey (76-77). Beebe et al. (7) in their book showed how to solve different problems using the most widely available chemometric methods.
Chemometric techniques have been used for:
- identification of sources of chemical pollutants (23, 25-27, 53, 78-81),
- analytical evaluation of quality of some food products (82-93),
- evaluation of the analytical potential of coupled methods with regard to future international standardised methods (94),
- evaluation of effectiveness of sample pre-treatment methods for trace elements determination (83).
**Chemometric evaluation of analytical methods**

PCA appeared to be a useful tool to estimate effectiveness of different sample pre-treatments for seafood products analysed for concentration As, Cd, Cr, Cu, Fe, Mg, Mn, Ni, Pb, Se and Zn (83) by flame atomic absorption spectrometry (FAAS) and electrothermal atomic absorption spectrometry (ETAAS). Comparing classic pre-treatments with modern procedures the authors pointed out that, in contrast to ultrasound assisted-acid leaching (C) and enzymatic hydrolysis (E), methods such as microwave assisted-acid digestion (A), microwave assisted-acid leaching (D) and slurry sampling procedures (B) appeared as adequate techniques for the multi-element determinations in seafood samples. The examination of the two-dimensional scatterplot informs us about systematic error of the employed pre-treatment procedures by considering the distance of data attributable to each procedure from the reference material. Precision of each group of the data can be estimated by examining the spread of each group (83). Cluster of data corresponding to pre-treatment D is more closely located to certified concentration data (X) as compared to C grouping. Since scores of A and B are strictly close to the score X it appears that microwave-acid digestion (A) and the slurry sampling technique (B) offer the more accurate results. The precision of the various pre-treatment procedures assayed by PCA is satisfactory, although data obtained after using both the B and D methods appear to be more spread (83).

An example of successful use of chemometric approach, as powerful methodological tool, for validation of GC-ICP-AES has been presented by Aguerre et al. (94).

**Chemometric evaluation of environmental and biological samples**

**Biological samples**

The above mentioned statistical analyses were all applied to estimate data obtained for samples collected in the marine environments with respect to spatial, species, age or seasonal trends. This approach concerned element concentrations in invertebrates, i.e. *Cerastoderma glaucum* from the Baltic Sea and other regions (29), *Mytilus edulis* and *Fucus vesiculosus* from the Baltic Sea, North Sea and the Hardangerfjord, Norway (9, 28, 32, 33). Representative vertebrates of three species of fish perch *Perca fluviatilis*, *Gadus virens* and *Platichthys flesus* from the Baltic Sea and the coasts of Norway were also analysed with this respect (28, 40). The distribution patterns of trace metals in sea mammals, i.e. *Phocoena phocoena* from coastal areas of the Polish, Danish and Greenland, harp seals (*Pagophilus groenlandicus*) and hooded seal (*Cystophora cristata*) from the Greenland Sea were also studied in view of FA or PCA (41, 42).

**Seaweeds**

Since macroalgae can accumulate elements from surrounding water, groups of locations were formed according to the course of salinity. Discriminant analysis (DA) indicated that Baltic and North Seas locations are clearly separated as in the case of cluster analysis (CA). DA was also performed for trace element concentrations (9) resulting in reduced number of location groups in comparison with the DA of the concentration patterns of macroelements for the seaweed. This distribution pattern is in agreement with CA and indicates the reduced influence of trace-element-independent ecosystem parameters on the uptake of trace elements as compared to the uptake of macroelements.
DA of trace-element concentrations in seaweed collected at Eckwarderhörne made possible the identification of the pollutants emitted by industrial activity in Wilhelmshaven (9).

**Molluscs**

Distribution of trace elements in soft tissue of molluscs from different geographical location has been discussed by Struck et al. (9), Szefer and Wołowicz (29) and Szefer et al. (31-34). Based on the chemometrically processed concentration data it was possible to characterise and classify all the samples analysed with respect to their species features and provenience. For example, the observed differentiation between two groups of object samples attributable to 2 species (Acanthopleura haddoni and Ostrea cucullata) of molluscs from the Gulf of Aden could be explained by the differences in their feeding habits (34).

An example of spatial differentiation of mineral composition of mollusc tissue obtained by FA is an influence of industrial activity of man on heavy metal distribution in Mytilus galloprovincialis from coastal waters of the Korean Peninsula (Szefer et al., 2003b). Interesting results have been also obtained for molluscs coming from different aquatic regions all over the world in view of chemometric data. Szefer and Wołowicz (29) processed statistically the concentration data (Cd, Cu, Fe, Mn, Ni, Zn) for the soft tissue of Cerastoderma glaucum from four geographical regions, i.e. the Gulf of Gdańsk (Baltic Sea), Marennes-Oleron Bay, Arcachon Bay (French Atlantic coast) and Embiez Islands (Mediterranean Sea) (Fig. 1). About 74% of the total variance is explained by the first three factors. Both score and loading data are presented on the first two principal vectors by means of a biplot (Fig. 2). Three-dimensional scatter-plot in space determined by PC1, PC2 and PC3 is shown in Fig. 3. It follows from comparison between the distribution of the object scores and the loading vector direction (Fig. 2a) that mainly Mn and Fe concentrations in the mollusc analysed are responsible for differentiation between populations from Marennes-Oleron Bay and Arcachon Bay. Zn, Cd and partly Ni have a main contribution in distinguishing the Gulf of Gdańsk cluster from the others (Figs. 2 and 3). Such distribution pattern suggests that anthropogenic sources may be responsible for elevated levels of Cd and Zn in C. glaucum inhabited the coastal and industrialised area of the Gulf of Gdańsk. The PCA data display that both spatial and seasonal factors are of great importance in the distribution of the metals studied in the mollusc tissues (29).
Figure 1. Sampling sites of *Cerastoderma glaucum* populations; a - the Gulf of Gdańsk (Baltic Sea), b - Marennes-Oleron Bay (Atlantic), c - Arcachon Bay (Atlantic), d - Embiez Islands (Mediterranean Sea), R - Rzucewo, M - Mechelinki, S – Sopot.

Special attention has been paid for applying multivariate technique in classification of Mytilids from different geographical zones (30-32). The concentration data for the soft tissue and byssus obtained from ca. 10 000 specimens of the mussels collected in temperate, subtropical and tropical zones were processed by FA. As can be seen in Figs. 4 and 5, after removing extreme values (attributable to extremely contaminated samples in highly industrialised areas of Saganoseki, Japan and Öxelosund, Sweden) it is possible to distinguish selected mussel populations with respect to their geographical provenience (Fig. 5).

Interesting factorial data, with respect to mineral composition, have been obtained for the soft tissue and byssus of *Mytilus trossulus* taken from three areas of the southern Baltic (33). The first three factors described 44.0% (for the soft tissue) and 46.1 % (for the byssal threads) of the total variance with corresponding eigenvalues greater than unity. As can be seen in Fig. 6 the objects corresponding to the soft tissue of molluscs inhabiting the Pomeranian Bay and the Słupsk Bank region show the highest values of F2 and form a cluster which is separated from that corresponding to tissue samples coming from the Gulf of Gdańsk, described by the lowest values of F2. A biplot for trace elements illustrates the distribution of loading of Ag, Fe, Co, and partly Pb, Cr and Hg which is attributable to the Gulf of Gdańsk samples characterised also by the lowest values of F2. It is well isolated from loading of other metals, especially Ni, Zn, Cd and Cu referring to the Pomeranian Bay and the Słupsk Bank specimens (described by the highest values of F2).
Figure 2. Bi-plot for object scores of the first two principal vectors of 50 mollusc samples: a - regional differences are illustrated by clusters of points corresponding to samples from the Gulf of Gdańsk (○), Marennes-Oleron Bay (●), Arcachon Bay (■) and Embiez Islands (Δ). Association between principal components (PCI x PC2) and variable (metals) vectors are also indicated; b - season dependent variations are illustrated by clusters of points corresponding to samples collected during January-May. These groupings are indicated by shaded areas.

Figure 3. Three-dimensional scatterplot in the space determined by PC1, PC2 and PC3: a - for object scores; samples from the Gulf of Gdańsk (○) and Marennes-Oleron Bay (●) are encircled; b - for loadings (metals).
Figure 4. Biplot of scores and loadings (metals) corresponding to Mytilidae soft tissue from the Baltic Sea and other geographical regions.
Figure 5. Biplot of scores and loadings (metals) corresponding to Mytilidae byssus from the Baltic Sea and other geographical regions.
Figure 6. Biplot of scores and loadings (metals) corresponding to *Mytilus* soft tissue from the three areas of the Southern Baltic.
The distribution pattern of byssus samples from the Pomeranian Bay region have the highest values of F1; it means that the byssal threads are the most influenced by input of the fluvial material (Oder River, Fig. 7). Loadings such as Cd, Mn, Cu, Ni and Zn are described by the highest values of F1. Samples from the Gulf of Gdańsk with low values of F1 may reflect, in part, the high levels of Hg, Cr, Ag, Co, Pb and Fe; this pattern suggests that the trace metals are preferably accumulated in byssus of specimens inhabited the area adjacent to the Vistula River estuary. Factors 1 and 2 show clear separation of both the byssi and tissue samples, respectively based on their geographic distribution, possibly reflecting a different rate of deposition of clay minerals at the head of the Pomeranian Bay and the Gulf of Gdańsk. Such differentiation between these two groups could be explained by the differences in environmental parameters in the geographical sectors, e.g. food supply for *Mytilus*, metal runoff, geochemical composition of the adjacent sediments, etc. Moreover, various sources of metallic pollutants are specific for each sector (33).

In order to verify the regional influences of seawater on the biochemical composition of *Mytilus edulis* from the Baltic Sea and North Sea, which are independent of the presence of trace elements, DA was performed for macroelement concentrations in the mussel as variables (9). This distribution pattern allows to distinguish Baltic and North Sea locations such as in the case of *F. vesiculosus* in spite of different food habits between these two zoobenthic organisms. Location groups based on the trace-element concentration patterns showed a less distinctive geographical arrangement in comparison of the location clusters based on macroelement concentration pattern. This picture suggests modified conditions for the accumulation of trace elements in *M. edulis* as in *F. vesiculosus* as compared to the uptake of macroelements (9).

![Figure 7. Biplot of scores and loadings (metals) corresponding to Mytilus byssus from the three areas of the Southern Baltic.](image-url)

365
Figure 7 (continued). Biplot of scores and loadings (metals) corresponding to *Mytilus byssus* from the three areas of the Southern Baltic.

*Fish*

The first two factors (eigenvalues > 1.0) account for ca. 70 % (for the liver) and 62 % (for the muscle) of the total variance (40). In order to clarify the graphical presentation of the seasonal and age-related variations in several metal clusterings, selected plot scores, corresponding to winter and summer catches (Fig. 8a) as well as I and III age-groups of fish are specified (Fig. 8b). Figure 8a illustrates the object score distribution pattern for the liver of perch characterising by the highest values of F1 for the summer season and the lowest ones for the winter season. Fig. 8b shows clustering of points represented by old specimens (III age-group, quantified by the highest values of F2 and the lowest values of F1) clearly isolated from that identified as young fish (I age-group with the lowest values of F2 and the highest values of F1). In order to demonstrate which metals control the clustering of the samples described by object scores (Figs. 8a and 8b), a biplot for loadings (metals) is presented in Figure 8c. F1 displays loading of hepatic Pb and Cd and corresponds to the winter samples described by the lowest values of F1. It is well distinguished from loadings of hepatic Cu and especially Zn, and refers to the summer samples (with the highest values of F1). Hepatic Cd allows identification of the oldest specimens, while hepatic Cu and Zn are responsible for selection of points corresponding to younger ones (40).

A biplot of the object scores shows a grouping of the muscle samples (Figs. 9a and 9b). Seasonal differences, similarly to hepatic objects, are also well visualised. Muscle samples corresponding to the summer season are clearly separated from those attributable to winter (Fig. 9a); however age-related differences (Fig. 9b) show no such regular distribution pattern as in the case of hepatic samples. As can be seen in the distribution pattern of loadings (Fig. 9c) the winter muscle samples are generally loaded with Cd, Pb and Hg while both muscle Zn and Cu are mainly determinants of summer objects.
Figure 8 Biplot of scores reflecting seasonal (A) and age (B) differences of metals (C) in the liver of *Perca fluviatilis* from the Southern Baltic.
Figure 9. Biplot of scores reflecting seasonal (A) and age (B) differences of metals (C) in the muscle of *Perca fluviatilis* from the Southern Baltic.
The observed seasonal variations in levels of selected metals in perch (40) are reflected by different metal bioavailabilities depending on the ligands present in the biotopes and the chemical speciations between the dissolved and particulate phases (39). Fish metabolism may be dependent on food supply and the stage of the cycle reproduction as well as the abiotic conditions (39, 95, 96).

**Sea mammals**

Nyman et al. (43) have estimated contaminant exposure and effects in Baltic ringed and grey seals using various biomarkers. Besides hepatic, renal and muscle metals, other blood and chemical parameters, e.g. PCBs, DDT, vitamins E and A, cytochrome P4501A activity were processed with PCA. The first two components amounted to 49% of the total variance. PCA showed the specific relationships between each potential biomarker and the pollutants for grey seal.

The concentration data obtained for *Phocoena phocoena* from the Baltic and Danish waters and other northern areas such as the Greenland were processed using FA (42). The first three factors for hepatic and renal distributions of metals describe 67.9 and 72.8% of the total variance, respectively. Eigenvalues amounted to 2.56, 1.77 and 1.11 (for liver) and 3.17, 1.56 and 1.10 (for kidney). As can be seen in Figs. 10 and 11 the hepatic and renal samples attributable to specimens of old harbour porpoises show the highest values of F1 and form a grouping of points that is separated from that consisting of young specimens (with the lowest values of F1).

![Figure 10](image_url)

**Figure 10. Biplot of scores (a) and selected metals (b) corresponding to the liver of Phocoena phocoena from 3 geographical regions.**

Factor 2 describes spatial differentiation between harbour porpoise populations; specimens from the southern Baltic are identified by object scores (liver and kidney) in the left...
part of the scatter-plot (lowest values of F2) and the Greenlandic cluster is characterised by higher values of F2.

The third group corresponding to Danish specimens overlaps these two extremely situated clusters. The Danish group confirms the close association of samples corresponding to Greenlandic and Baltic populations. In order to demonstrate which metals are responsible for the grouping of the object scores, corresponding biplots for loadings (metals) are presented in Figs. 10 and 11. From this distribution pattern results that loadings of hepatic Cd (Fig. 10) as well as renal Cd and Zn (Fig. 11) accompanied by age and weight are marked by the highest values of F1 and are also clearly distinguished from loadings of other metals, especially hepatic Cr, Cu and Fe (Fig. 10) and renal Mn and Fe (Fig. 11) described by the lowest values of F1. A distinct differentiation of the object score groupings is also associated to the geographical distribution of the trace elements studied; for instance both hepatic and renal Fe and Cr allow identification of samples represented by Baltic specimens (quantified by lower values of F2).

Other metals found to the right of the plot (described by higher values of F2), especially Cd, Mn, Zn and Cu, are helpful in recognition of samples represented by Greenland specimens (Figs. 10 and 11). Geographical variations in hepatic and renal metals support the above suggestion that the differences in metal bioaccumulation are mainly caused by specific feeding habits of the porpoises living in southern Baltic and southwest Greenland (42).
Geological samples
Several authors utilised FA or PCA for quantitative evaluation of both the horizontal and vertical distributions of different elements in geological material, e.g. bottom sediments from the marine and lacunal ecosystems (62, 63, 65-69, 97).

Suspended matter
Multivariate techniques could be useful in identification of various geochemical phases present in particles of particulate material.
A PCA was performed using a data matrix including the hydrographical data and the relative abundance of the particle types (48). The first four PCs represent 70% of the total variance. The first PC described differences between oxygenated surface samples, relatively rich in aluminosilicates, and poor in the oxygen deep water samples containing higher levels of Fe- and Mn-particle concentrations. The second PC is mainly loaded by salinity, Ca-rich particle type and temperature, i.e. it possibly describe the differences occurring in the mixing area between the Baltic Sea and the North Sea waters. The third PC distinguished the barite particle type and temperature from the depth, suspension content, nitrate concentration and the Fe-rich aluminosilicate particle type (48).

Surface sediments
In the chemometric classification of marine bottom sediments in relation to their chemical composition several digestive procedures are used based on different extrac-tability of selected chemical elements from particular granulometric fractions. The fol-lowing sediment fractions and corresponding chemical reagents have been used (8, 63, 66, 67, 98, 99):
1. bulk sediments (< 2 mm) digested using mixture concentrated HNO₃, HClO₄ and HF,
2. sediment fraction (<80µm) digested using mixture concentrated HNO₃ and HClO₄,
3. sediment fraction (<63 µm) digested mixture concentrated HNO₃, HClO₄ and HF,
4. sediment fraction (<80 µm) leached using 1 M HCl,
5. fusing with LiBO₂ and dissolution in HNO₃.

Among the above given approaches, techniques 1-3 were the most frequently used.

Approach 1.
The first three factors with eigenvalues > 1.0 were extracted from the data set studied. These accounted for 79.5% of the total variance with F1 contributed 62.2% of the total data variance (98). The first factor is mainly influenced by sediment grain size characteristics (i.e. negative loadings are attributed to coarse-grained sediments), i.e. it describes different granulometric structure of geological material (98). This is undesirable arrangement because coarse-grained structure (sandy) pattern significantly masks the geochemical composition of elements concentrated in clay material (< 80 µm). Therefore concentration data corresponding to these bulk sediments (< 2 mm) were processed by endmember analysis (71). Use of endmember analysis has enabled to refinement of the results of previous studies to show different pathways for the introduction of Cu, Zn, and Ag and Cd, and Pb respectively into the Gulf of Gdansk. It is sup-
posed that this difference reflects the fact that Cu, Zn and Ag are introduced into the sediments of the Gulf of Gdansk principally from the Vistula River whereas Cd and Pb are introduced, in part, by atmospheric transport. Renner et al. (71) identified origin of selected heavy metals in bulk sediments of the southern Baltic using endmember analysis.

**Approach 2.**
The first three factors with eigenvalues $> 1.0$ accounted for 73.3% of the total variance. F1 explained 28.6% of the total data variance was associated with mineralogical composition of the samples studied. It is postulated that F2 (23.0%) corresponded to terrigenic and biogenic phases while F3 is related to (21.7%) geochemical composition of estuarine and open sea sediments coming from sampling sites shown in Fig. 12 (63).

![Figure 12. Location of sediment sampling sites in the southern Baltic.](image)

Figure 12a illustrates factorial distribution of object scores in the three-dimensional scatter plot. It is well visualised that open-sea samples (Nos. 25-29) form a separate cluster which is well isolated from the grouping of points represented by typical estuarine samples (Nos. 6, 7, 16 and 17). The remaining samples are located in mid-distance between the estuarine and open-sea clusters. It is concluded that this region is less exposed to the influx of material of the fluvial origin. Fig. 13b shows distri-
bution of loadings in the three-dimensional scatter-plot which is similar to that of the object scores (Fig.13a). Elements such as K, Mg, Ca, Na, Sr as well as physical parameters like salinity and depth of water form a cluster which is distinguished from the grouping of points corresponded to Zn, Cd, Ag, Cu, Cr, Pb, chlorophyll-a and Fe (63). The localisation of the latter (described by lower values of F3) substantiates identification of samples originating from the Vistula River’s mouth (characterised also by lower values of F3). The upper cluster (described by higher values of F3) identifies samples of typical open-sea provenience (having also higher values of F3).
Figure 13. Three-dimensional scatterplot of object sample scores (a) and loadings by individual variables (b) obtained for acid (concentrated HNO₃-HClO₄) leachates of sediment (fraction < 80 µm) sample data. Samples are numbered as in Fig. 5.9 and Table 6.1; Sa = salinity; D = depth of water; Ch = chlorophyll-a. Samples originating close to the subarea of the Vistula estuary (open circles) and from the open-sea region (filled circles) are indicated.
Approach 3.

FA was also applied (66) for evaluation of the distribution of As, Cd, Co, Cr, Cs, Cu, Ba, Bi, Ga, In, Ni, Pb, Rb, Sb, Se, Sr, Th, Ti, Tl, U, V, Zn, RRE (Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sm, Tb, Tm, Yb) in surface sediments of the Vistula Lagoon sediments (Polish sector). This approach was also used for classification of chemical elements such as Al, As, Be, Cd, Co, Cr, Cu, Fe, Mn, Mo, Pb, Sb, Se, Sn, Sr, Ti, V, Zn, REE in three sectors of a southern Baltic Sea (67). Figure 14a illustrates factorial distribution of the object scores in three dimensional scatter plot.

It can be seen that the Szczecin Lagoon samples form a separate cluster, closely fitted to the Pomeranian Bay samples. On the other hand, the Vistula Lagoon object scores are neighbouring to those represented by the Gulf of Gdansk (properly Puck Bay) sediments. This scatter-plot clearly illustrates a great dissimilarity between geochemical composition of the Szczecin and Vistula Lagoon sediments. The Pomeranian Bay with the Szczecin Lagoon belong to the Odra River estuary, while the Gulf of Gdansk with the Vistula Lagoon, excluding its western inner part named the Puck Bay, are supplied with the Vistula River. Bereft of topographical barriers, the Pomeranian Bay is exposed to permanent, intensive water exchange between itself and the neighbouring Arkona and Bornholm Basins and its exchange with the Central Baltic represented by the Slupsk Bank takes only about 3 weeks.

Figure 14. Scatterplots of object sample scores (A) and loadings by individual variables (B) in space spanned by axes F1 and F2 obtained for sediments (fraction < 63 µm) sample data from different areas of the Southern Baltic.
Such long-distance water exchange is possibly reflected by overlapping of the object scores corresponding to the Pomeranian Bay and the Slupsk Bank (67). Fig. 14b shows distribution of loadings (variables) in two-dimensional scatter plot which is similar to that of the object scores presented previously. Elements such Zn, Cu, As, Pb, Cd, Sb, Mn, Fe and Sr (described by high values of F1) are isolated from the groupings of points represented by Al, V, Tl, Be and Yb (characterised by lower values of F1). Since Yb and Al are typically terrigenic elements in origin, and Cd and Pb belong to anthropogenically derived metals, it means that the Pomeranian Bay with especially the Szczecin Lagoon are the most polluted areas of the southern Baltic (67).

**Sediment cores**
The vertical profiles of selected chemical elements in sediment cores from southern Baltic have been processed by PCA (62, 65). The first three factors described 75% of the total variance with eigenvalues greater than unity. The sample numbers and depths of samples taken from four sediment cores collected at sampling sites shown in Fig. 15 are listed in Table 1.

Figure 15. Map of the Southern Baltic region indicating the locations of sampling stations of the cores studied.
TABLE 1.
The object numbers and depths of corresponding core segments

<table>
<thead>
<tr>
<th>Object number*</th>
<th>Sample depth in core [cm]</th>
<th>Object number*</th>
<th>Sample depth in core [cm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core P-2</td>
<td></td>
<td>Core P-10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7-2.0</td>
<td>1</td>
<td>0.0-1.6</td>
</tr>
<tr>
<td>4</td>
<td>3.0-4.0</td>
<td>3</td>
<td>2.6-3.9</td>
</tr>
<tr>
<td>6</td>
<td>5.0-6.0</td>
<td>5</td>
<td>5.5-7.4</td>
</tr>
<tr>
<td>7</td>
<td>6.0-7.0</td>
<td>6</td>
<td>7.4-9.2</td>
</tr>
<tr>
<td>8</td>
<td>7.0-8.0</td>
<td>7</td>
<td>9.2-11.4</td>
</tr>
<tr>
<td>10</td>
<td>10.0-12.0</td>
<td>8</td>
<td>11.4-14.1</td>
</tr>
<tr>
<td>11</td>
<td>12.0-15.0</td>
<td>9</td>
<td>14.1-17.8</td>
</tr>
<tr>
<td>12</td>
<td>15.0-20.0</td>
<td>10</td>
<td>17.8-21.5</td>
</tr>
<tr>
<td>13</td>
<td>20.0-25.0</td>
<td>11</td>
<td>21.5-25.8</td>
</tr>
<tr>
<td>14</td>
<td>25.0-30.0</td>
<td>12</td>
<td>25.8-31.3</td>
</tr>
<tr>
<td>Core G-2</td>
<td></td>
<td>13</td>
<td>31.3-34.8</td>
</tr>
<tr>
<td>1</td>
<td>0.0-1.0</td>
<td>Core P-38</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.0-2.0</td>
<td>1</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>4</td>
<td>3.0-4.0</td>
<td>2</td>
<td>0.7-1.8</td>
</tr>
<tr>
<td>5</td>
<td>4.0-5.0</td>
<td>5</td>
<td>5.0-6.2</td>
</tr>
<tr>
<td>6</td>
<td>5.0-6.0</td>
<td>6</td>
<td>6.2-8.1</td>
</tr>
<tr>
<td>10</td>
<td>9.0-10.0</td>
<td>7</td>
<td>8.1-9.7</td>
</tr>
<tr>
<td>11</td>
<td>10.0-12.0</td>
<td>8</td>
<td>9.7-12.1</td>
</tr>
<tr>
<td>12</td>
<td>12.0-15.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>15.0-20.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>20.0-25.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* corresponds to the number in Figure 15.

The distribution of PC scores (Fig. 16) is comparable to that of PC loadings (Fig. 17). Since Al is a typical element of crustal origin and C\text{org} represents organic matter molecules, localisation of the two antipathetic PC1 clusters (Fig. 17) indicates that Al, Fe, Ti, K, Mg, Th, Co and Ni (as positive values of PC1) in the sediment cores are terrigenous, whereas C\text{org}, N, Cu, Pb, Zn, Cd and possibly P (as negative values) are biogenic. These two main groupings of elements let us to identify elements anthropogenic in origin (Pb-Cd) accumulated in recently formed top layers of sediments as well as elements terrigenic in origin (Al) deposited in deeper “background” segments corresponding to the precivilisation era (65).
The concentration data of As, Cd, Co, Cu, Fe, Hg, Mn, Mo, Ni, Pb, V and Zn in sediment cores from 59 stations of the Baltic Proper have been processed using cluster analysis (61). The sampling sites represented different geochemical facies which were divided into three clusters. Spatial trends can be explained mainly by only PC1 (60% of the total variance). Factorial distribution pattern of metals with sediment depth is recognised from their correlation with PC1 which explained 81% of the stratum effect (61). Elements with an increasing of their concentrations towards the surficial layers of the core exhibit a significant correlation with curvature, i.e. Cd, Cu, Mo, Ni, Pb, V and Zn.
This is either can be due to a diagenetic processes taking place at that depth or anthropogenic impact of these metals.

The geochemical fluctuations in deposition in the most polluted part of the sediment core of the Bay of Töölönlahdi, southern Finland, was interpreted in view of PCA (100). The first two components accounted for 65.4% of the variance. A cluster containing organic, carbonate and hydroxide Ca together with total P was located in the middle of the sediment sequences, representing an increase in the nutrient concentrations in the beginning of the eutrophication, i.e. to an expansion in diatom populations. Other cluster attributing to a depth of 30-40 cm, representing segments deposited during 1930-1950, was dominated by Al, Fe and Mn (bound to organic matter), Cu (carbonates, bound to Fe-Mn oxides and organic material), Zn (exchangeable, carbonates, bound to Fe-Mn oxides and organic material), LOI, and total S (100).

Chemometric evaluation of food quality

Based on numerous literature data it should be emphasised that chemometric methods have been an effective statistical tool in analytical evaluation of food quality, i.e. seafood products (83), tea beverages (90), commercial wines and must (82, 85, 86, 89, 93), Sherry wine vinegars (87), brandies (84), beer (88) and rice (91).

Seafood

PCA has been successfully applied in analytical evaluation of a degree of seafood pollution by toxic trace elements. This technique makes possible the identification of edible mussel population, characterised by elevated concentration of heavy metals such as Hg, Cd and Pb. Popham and D’Auria (24) have used PCA for deciding if blue mussels have been collected from the coastal waters of British Columbia, polluted with trace metals such as Pb, Zn, Cu, Mn and Fe. These metals were determined by X-ray energy spectroscopy (XES). The authors identified areas in which mussels contained very high levels of tissue heavy metals reflecting industrial pollution of the ambient water. An interesting aspect of their study is that it was possible to identify regions characterised by less Zn and Pb pollution or much higher their levels in relation to the reference area. These areas have been identified in spite of the fact that specimens of the mussels analysed were different size (age) and were collected at an different time period of the year. In order to characterise and differentiate polluted and unpolluted edible mussels taken from waters of the Adriatic Sea, two multivariate techniques such as linear discriminant analysis (LDA) and linear principal component analysis (LPCA) have been successfully applied by Favretto et al. (38) after determination of Mn, Fe, Co, Ni, Cu, Zn, Cd, Hg and Pb in mussel samples by atomic absorption spectrometry (AAS). Using LPCA it was also possible to distinguish some elements of terrigenous origin (Pb, Cd).

Bechmann et al. (36) have reported spatial-related trends of metal concentrations in edible mussels from Limfjord, Denmark. High resolution inductively coupled plasma spectrometry (HR-ICP-MS) was used for determination of metals such as V, Cr, Co, Ni, Cu, Ga, Rb, Cd, Ba and Pb in the soft tissue of blue mussel (Mytilus edulis). PCA appeared to be useful technique when decision should be made upon which regions are appropriate for mussel fishery.

It results from the above mentioned studies that PCA is powerful tool to identify polluted areas inhabited by edible mussels which can be simultaneously used as useful
biomonitors (25-27). Mussels transplanted from an aquaculture farm in a clean open bay to the Bays of NW Mediterranean were analysed for several biomarkers as well as Cd, Cu and Zn concentrations. PCA was applied to discriminate the different transplantation sites and appeared to be a useful tool for assessing water quality in this area (37).

Tea beverages
Based on measurements of different physical and chemical parameters and after applying PCA and especially LDA, Fernández et al. (90) have concluded that the metals profile studied (Cu, Zn, Mn, Fe, Sr, Ba, Al, Ca, Mg, K, Na determined by ICP-AES) is a very good descriptor for the distinguishing of the common types of tea beverages such as infusions, instant teas and soft drinks. From statistical comparison clearly results that infusions prepared with black teas are enriched in Mn, Mg, Al and K and that there are significant differences between mineral composition of infusions prepared with different time of extraction.

Alcoholic beverages
Several authors (82, 85, 86, 93) processed chemometrically concentration data concerning different kinds of wines including must (82). Concentrations of Cu, Zn, Fe, Mn, Li, Rb, Sr, Ca, Mg, Na and K were determined by AAS or AES (85, 86) and Al, B, Ca, Cr, Cu, Fe, I, K, Li, Mg, Mn, Na, P, Pb, Rb, Si, Sn, Sr, Zn, As, Ba, Be, Cd, Ce, Co, Cs, Ga and Ge by ICP-MS (93). Cluster analysis (CA) and PCA showed differences in mean concentrations in wines with respect to their geographical origin and the ripening state of the grapes. The results reported indicate that the determined elemental composition of Italian wines could be a fast way to characterise and classify their samples with discriminating two high quality wines from three less fine wines (93). According to Frías et al. (86), a high degree of the sensitivity and specificity has been obtained for classification of commercial Spanish wines using SIMCA as a modelling multivariate technique. Based on chemical measurements, González and Peña-Méndez (82) have applied PCA, cluster analysis (CA), biplot analysis (BA) and factor analysis (FA) which allowed the classification and differentiation of must and Canarian wine samples. Elements such as Cd, Cu, Zn, Pb, Mn, Fe, Al, Mg, Ca, K and Na were determined by flame and graphite furnace AAS (FAAS and GFAAS) in samples of Sherry brandies and chemometric approach was followed to differentiate of brandies according to the metal profile (84). Three chemometric approaches were applied, i.e. PCA, DA and BPNN. The use of classification procedure based on neural networks leads to ca. 90% hits in prediction ability.

Using a chemometric approach, the mineral composition of beer provided a valuable information on differentiation of its samples (88). Zn, Mn, B, Fe, Al, Ba, P, Sr, Ca, Mg, Na and K were considered as chemical descriptors and determined by ICP-ES in 32 beer samples. It is recommended to apply supervised learning PR methods, such as LDA, based on artificial neural networks (BP-MLP).

Rice
Isotopic ($\delta^{18}$O, $\delta^{13}$C) and elemental (B, Se, Rb, W, Gd, Ho, Mg) analyses using Isotope Ratio Mass Spectrometry (IRMS) and ICP-MS, respectively, have been applied to de-
termine the geographical origin of premium long grain rice (91). The maximum discrimination between rice samples from regions such as America, Europe and India/Pakistan was clearly identified by canonical discriminant analysis (CDA).

**Mushrooms**
Chemometrical techniques such as CA and DA have been applied to the concentration data for fruiting bodies of mushroom *Xerocomus badius* (caps and stalks) and the underlying soil substratum collected from north-eastern Poland (44). In the caps, three first functions explained 75.8% of the total variance. The lowest values of F1 (Fig.18) and the highest values of F2 (Fig. 19) discriminate cap and stalk samples, respectively coming from Trójmiejski Landscape Park, adjacent to the Tricity agglomeration.

![Figure 18. Scatterplot of object scores of the two first discriminant functions of 166 samples of caps. (b). Location loadings for 14 metals in the caps.](image)
Figure 19. Scatterplot of object scores of the two first discriminant functions of 166 samples of stalks. (b). Location loadings for 14 metals in the stalks.

The CA data (hierarchical clustering, Ward’s method) for the sampling sites as objects are shown in Figure 20.
Figure 20. Hierarchical dendrogram for 90 sampling sites as objects (1-16 Augustowska Forest, 17-27 Białowieska Forest, 28-42 Borecka Forest, 43-58 Adjacent area of Morąg, 59-68 Ilawskie Lake district, 69-83 Wdzydzki Landscape Park, 139-145 Adjacent area of Kołobrzeg).
The dendrogram is built up of two main clusters. The first one contains two subclusters with the objects from Ólaskie Lake district (C59-C68) and the adjacent area of Morag (C43-C58), the most similar regions by environmental conditions. A second cluster contains five subclusters with the objects from Augustowska Forest (C1-C16), Białowieska Forest (C17-C27), Borecka Forest (C28-C42), Wdzydzki Landscape Park (C69-C83) and the adjacent area of Kołobrzeg (C139-C145). The most similar areas were distinguished, e.g. Augustowska and Białowieska Forests are protected areas, deprived of industrial or urban influences (observed for Trójmiejski Landscape Park). The principal factor governing the accumulation of trace elements in mushrooms is pollution via atmospheric deposition. Apart from anthropogenic factor, the natural effects such as geochemical composition of bed rock, pH and granulometric structure of soil, genetic potential of fungi, ectomycorrhizal occurrence and the kind of undergrowth (mosses, lichens, ferns) have a real influence on some metal concentrations in mushrooms (44). From the dendrogram presented results that CA made it possible to separate the most similar forest areas with relation to chemical composition of both the underlying soil substratum and biomass overgrown with *X. badius* in the studied areas (44).

As shown in this overview, multivariate methods are very useful in solving many environmental problems and performing analytical evaluation of food quality. They are recommended in monitoring coastal and estuarine areas with respect to metal pollution in order to reduce the expense for further more extended environmental survey. Application of these statistical techniques is also helpful in deeper understanding behaviour of toxic trace elements in the marine and terrestrial environments.

REFERENCES


Chapter 18


