# Chemometric Evaluation of Analytical Data Concerning Chemoecology of City Agglomeration

[a]M. MATHERNY*, [a]V. BALGAVÁ, and [b]J. EINAX

[a]*Department of Chemistry, Faculty of Metallurgy, Technical University, SK-042 00 Košice*

[b]*Friedrich Schiller University, Institute of Inorganic and Analytical Chemistry, D-077 43 Jena*

Received 13 September 1993

Chemoecological evaluation of a city agglomeration atmosphere is in the given case directed towards an application of mutually connected chemometric operations. Their sequence is as follows: cluster analysis (CA), principal component analysis (PCA), and multidimensional variance and discriminant analysis (MVDA). The results of evaluation rely on a two-year period of sampling of atmosphere and determination of some inorganic element pollutants in the gravitation dust sediments.

The atmosphere of the industrial city agglomeration of Košice was chemoecologically evaluated first of all by observing the chemical composition of the gravitation dust sediments [1] and the airborn dust [2]. The gravitation dust sediments evaluation was directed to the centre of city Košice. The results of chemometric evaluation of a three-year period of sampling revealed the classical communal and industrial heating by fosile fuels and repeatedly whirled and scattered city dust of anthropogenic origin as the most important dustiness forming factors. Other industrial immissions in the region show, according to the chemometric evaluation [1], only secondary influence.

Another investigation [3] was directed towards elucidation of chemoecological character of close surrounding of the city Košice with the aims to reveal disproportion in the so-called Košice-Valley, mainly when taking into account two typical different meteorological situations.

## THEORETICAL

The following methods in the given sequence are discussed in the theoretical part describing the principles of the applied chemometric operations: cluster analysis (CA), principal component analysis (PCA), and multidimensional variance and discriminant analysis (MVDA).

If there is no a priori information available the CA can serve as a useful chemometric tool to detect structure (or similarities) in a multidimensional data array. In investigated cases the data matrix is subjected to a hierarchical agglomerative CA in order to find out whether territorial structures with different multivariable pattern of pollutants exist within the area of investigation. Different algorithms of CA are described in the literature [4].

The principle of the PCA is a transformation of original features, for instance, the heavy metal concentrations in dustlike immissions into uncorrelated new variables (PC) by means of linear combinations. The first principal component PC-1 explains most of the variance within the data matrix and the second principal component PC-2 explains most of the remaining variance, *etc.* So, the representation of the scores of the two first PC (PCS-2 *vs.* PCS-1) can be used as a display method for the graphical illustration of multidimensional data. It means that a maximum of information of the data set is represented in the plane of two strongest PC. The mathematical principle of the PCA is the following: a data matrix $\mathbf{X}_{m,n}$ (*1*) is splitted in the product of two matrices

$$\mathbf{X}_{m,n} = \mathbf{A}_{m,s}\,\mathbf{F}_{s,n} \qquad (1)$$

where $m$ is the number of features, $n$ is the number of objects, and $s$ concerns the number of PC. The matrix of the PC loadings **A** has to be computed by solving the following "eigenvalue" problem (*2*)

$$(\mathbf{R} - \lambda_i \cdot \mathbf{U}) = \mathbf{Q} \qquad (2)$$

where **R** is the matrix of the correlation coefficients, $\lambda_i$ are the s.c. eigenvalues and **U** is the unity matrix. The obtained principal components PC-i are orthogonal, which means that they are uncorrelated. In the second step of the PCA the matrix of the principal component scores **F** has to be calculated in analogy to a multiple linear regression. The mathematical principles are described in Refs. [5, 6].

The principle of the MVDA is a separation of pretended classes s.c. a priori classes of objects. Un-

*The author to whom the correspondence should be addressed.

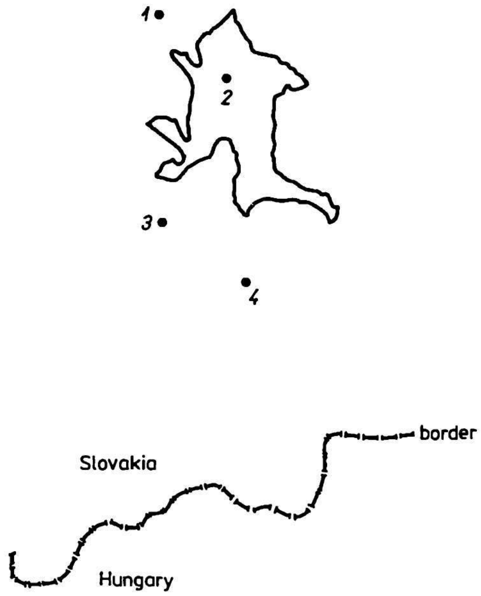


Fig. 1. The sampling points in the city agglomeration of Košice. *1.* Bankov, *2.* city centre, *3.* airport Košice, *4.* village Šebastovce.

der simultaneous consideration of all features observed and their interactions, the variance between classes is maximized and the variance within classes is minimized. Classification of new objects into a priori classes or a reclassification of the learning data set is carried out by means of the nonelementary discriminant functions. These nonelementary discriminant functions DF-i are linear combinations of an optimum separation set of original features. The first nonelementary discriminant function DF-1 has the strongest separation power. Mathematical principles are described in Ref. [7].

## EXPERIMENTAL

The experimental technique applied is based on the aimed combination of several different analytical methods [8—10].

Critical evaluation of the method [8] and the results reached [1] confirmed the necessity of both determination of important major elements iron and titanium and observing the concentration changes of chemoecologically important minor and trace elements. The results of earlier investigations [1, 3] were also taken into account, in addition to analytical-methodological factors when determining the final number of analyzed elements.

Optimization of the energy dispersive fluorescence X-ray spectroscopy [9, 10] was first of all directed to obtaining high detectability and acceptable precision values at the same time. The accuracy of the analytical results was tested by help of ENO standards [11], where $c(X_{st})$ is the declared concentration of standards. The relative accuracy values (RAV) [3] were derived (*3*) from the differences (*4*) between experimental concentration value $c(X_i)$ and declared concentration value of standard $c(X_{st})$

**Table 1.** Evaluating Parameters of the Spectroscopical Results

| Element | Limit of detection $c(X_L)$/ppm | Relative precision $s(c_{X,r})$/% | Relative accuracy RRW/% |
|---|---|---|---|
| As | 2.7 | 3.5 | 3.0 |
| Co | 5.8 | 2.2 | 2.0 |
| Cr | 15.1 | 3.9 | 4.2 |
| Cu | 4.3 | 1.0 | 1.5 |
| Fe | 21.0 | 0.7 | 5.0 |
| Ni | 5.0 | 1.4 | 1.6 |
| Pb | 7.3 | 2.7 | 3.2 |
| Rb | 6.0 | 1.7 | 2.9 |
| Sr | 5.5 | 1.3 | 1.7 |
| Ti | 10.0 | 3.8 | 5.0 |
| Zn | 9.0 | 1.6 | 1.4 |
| Zr | 4.7 | 1.8 | 5.3 |

$$\text{RAV} = (c(X_{st})/\Delta c) \cdot 100 \qquad (3)$$

$$\Delta c = |c(X_{st}) - \bar{c}(X_i)| \qquad (4)$$

The Bergerhoff method [12] with 28 days sampling frequency was used for collecting the gravitation dust sediment from a region of approx. 450 km$^2$ (Fig. 1). The samples contained a mixture of rain water and dust fraction. Samples were first vacuum-concentrated and then evaporated to dryness at 105 °C. The final product was used for analysis [9].

The number of elements determined (Table 1) was reduced in comparison with the previous work [1], since most of the concentration values for cobalt, chromium, nickel, and zinc were below the detection limit $c(X_L)$ of the given method and elements. Finally, only the concentration values for arsenic, copper, lead, rubidium, and zirconium were used for chemometric evaluation, further the concentration values of major and minor elements as iron and titanium and, of course the total amount of the monthly fraction of gravitation sediment, $m_{gds}$.

The chemometric evaluation of the total chemoecological situation of the city Košice was based on the array of concentration data originated from 240 numerical data. The original matrix was finally reduced to 216 numerical data.

The parameters of meteorological situation in the two-year sampling period, namely frequential wind direction frequency were for the purpose of chemometric evaluation abandoned as constant in the given region. However, in the interpretation of the chemometric evaluation of individual sampling points these data could be used.

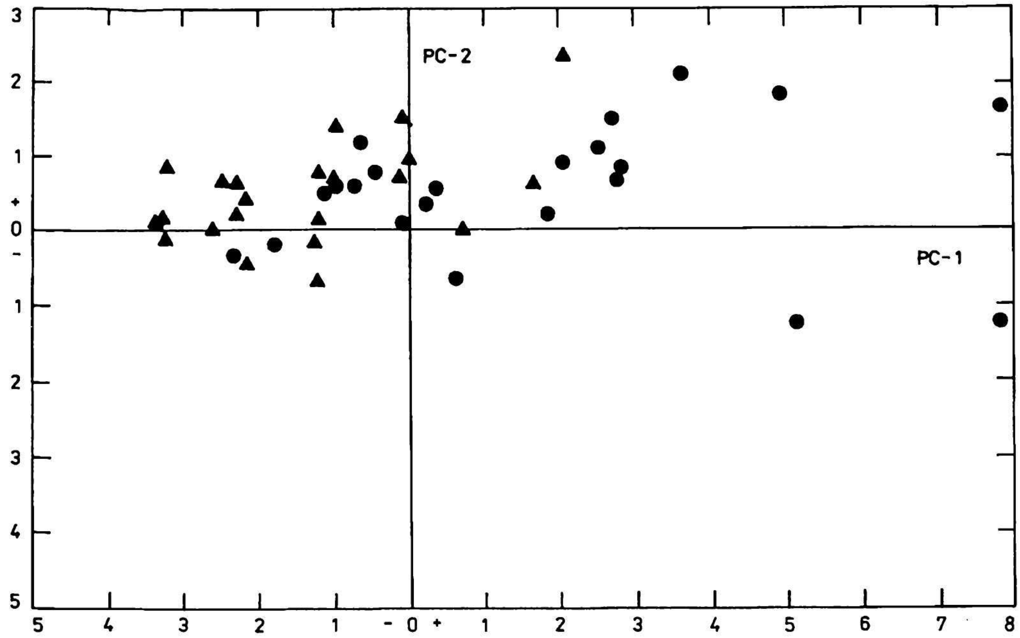


**Fig. 2.** Scattering of the points PC-2/PC-1 for the observed city agglomeration. ▲ Locality Bankov, ● city centre.

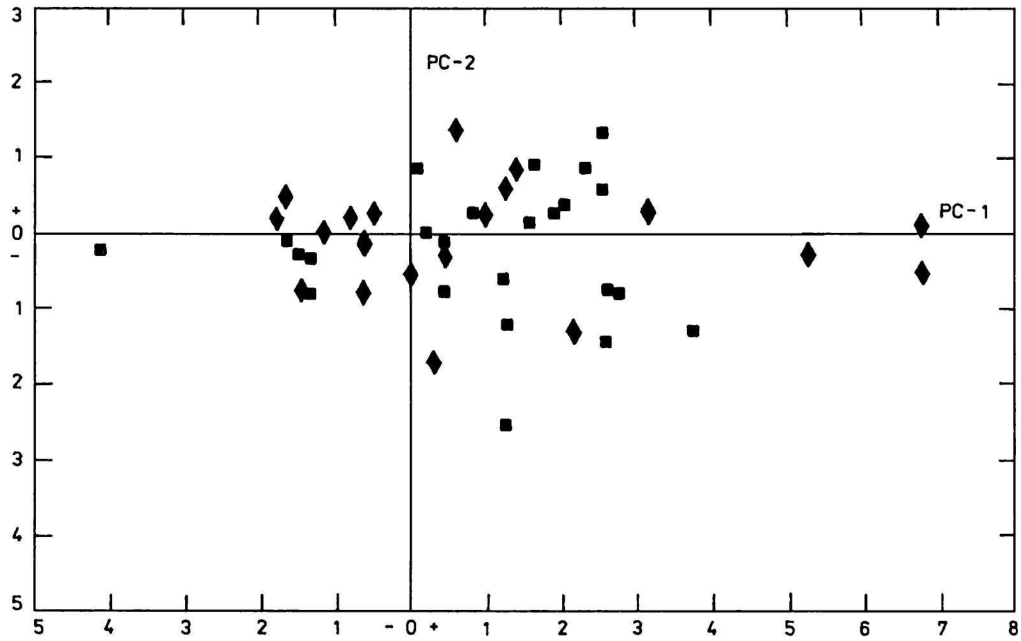


**Fig. 3.** Scattering of the points PCS-2/PCS-1 for the observed city agglomeration. ◆ Airport Košice, □ village Šebastovce.

## DISCUSSION

The sampling point 1 was placed above the city level in the forest at hill Bankov. This sampling position provided samples with relatively low immission pollution. The sampling point 2 was situated in the centre of the city and was more or less identical with the sampling point used in the previous work [1]. The point 3 was placed in the area of airport Košice, and point 4 at the southern end of the village Šebastovce.

The chemometric evaluation of the above defined set of analytical data consisted of the following partial steps: data pretreatment consisting of recalculating the element immission values (mg $m^{-2}$ $(28\ d)^{-1}$); deleting the 12 objects for the inconvenient values ($c(X_i) \leq c(X_L)$), and substituting these values by arithmetic means from the corresponding elemental set with randomly generated 10 % scattering; autoscaling of all values of the input array.

The cluster analysis from the whole readjusted data

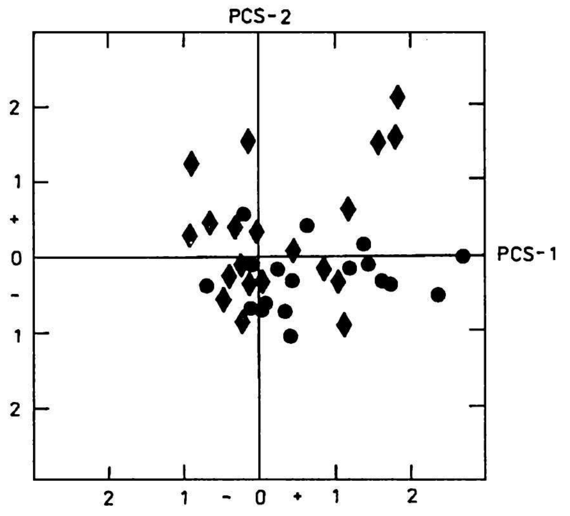


**Fig. 4.** Scattering of the points PCS-2/PCS-1 for the observed city agglomeration. ■ Village Šebastovce, ▲ locality Bankov.

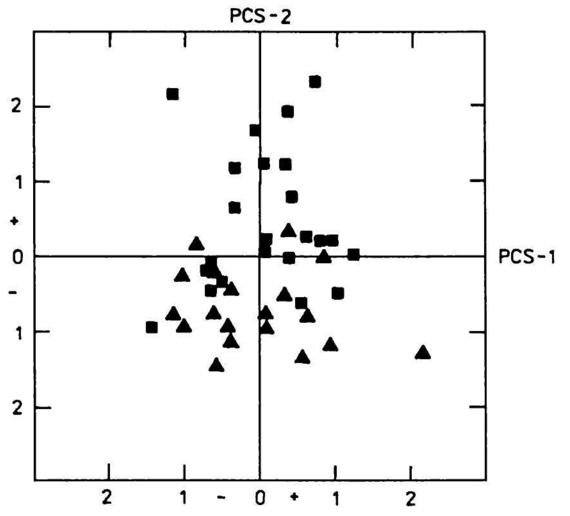


**Fig. 5.** Scattering of the points PCS-2/PCS-1 for the observed city agglomeration. ◆ Airport Košice, ● city centre.

set provided, as expected, no specifically interpretable results since the differences between the chosen sampling points were not significant, or the similarities were very high. Two factors could be extracted as regards the major and minor elements, iron and titanium, respectively. In Figs. 2 and 3, the stochastic dependence PC-2 = *f*(PC-1) is presented for pairs of sampling points. The derived chemometric data from the individual sampling points show, however, significant scattering. Therefore, only qualitative responses may be obtained in this way. The coincidence of points for the localities Bankov and centre of the city Košice in the quadrants PC-2/PC-1 and -PC-2/-PC-1 shown in Fig. 2 confirms the close chemoecological character of both these regions mainly in case of dominant northern winds [3]. The extension of points corresponding to the centre of the city in the quadrant PC-2/PC-1 is most likely given by the dominant southern winds influence period. Such influence is more remarkable in the centre of the city than at the locality Bankov [3].

Factor analysis enables extraction of two factors. The most important result is the separation of PCS-values for the village Šebastovce (see quadrants PCS-2/± PCS-1 in Fig. 4) from the values for Bankov -PCS-2/± PCS-1 and thus, for the centre of the city. The factor PCS-2 is dominant for characterization of chemometrical situation in this case. The differences may be explained by the fact that the village Šebastovce and the airport Košice are placed in an open plane whereas Bankov is placed above the city level in a forest country. Here the atmosphere is less burdened by inorganic element pollutants due to prevailing northern winds with relatively lower immissions. The points in the PCS-2 = *f*(PCS-1) dependence for the localities Košice airport and city centre (Fig. 5) are quite uniformly scattered, which proves certain degree of similarity of chemoecological character, but, at the same time, confirms the differences between these two sampling places.

The most convincing chemometrical results were provided by multidimensional variance and discriminant analysis. Two extracted discrimination functions represent 82.4 % of separable variances of the whole data set, which means that multidimensional immission conditions are satisfyingly considerable in a plane. The class differences defining the immission properties differences in the whole region investigated are not significantly distinguished, which is proved by the value 27.8 % for the mean reclassification error. The above value is, however, also a proof of similarity of the given regions expressed in addition by partial but remarkable overlaps of the scattering field of DF-2 = *f*(DF-1) dependence (Fig. 6).

The immission situation for Šebastovce and Košice airport is relatively inhomogeneous. Overlap of the scattering fields for both the regions is more than 90 % whereas for Bankov and city centre it is only 30 %. The immission situation for these localities is more homogeneous and more different at the same time. The differences are caused preferentially by local meteorological variation at constant immission situation.

## CONCLUSION

The investigated region of industrial-city agglomeration Košice represents a circle area of a radius approx. 12 km. The atmospheric pollution in this region may be characterized as medium. The most

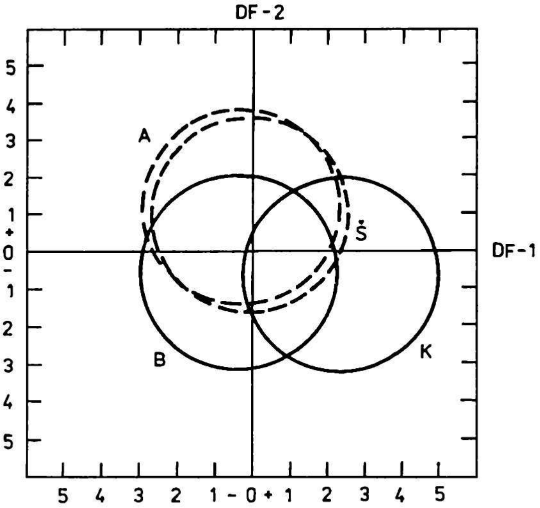


**Fig. 6.** Scattering of the points DF-2/DF-1 for all observed localities.

significant elemental toxic contamination is represented by arsenic and lead. Arsenic is transported to the atmosphere of city Košice mainly as emission from burning the fossil fuels whereas lead from burning petrols with antidetonation additives. The total pollution of this region is not remarkably different. The main part of the inorganic element pollutants is originated in the city centre, which is generally valid for all city agglomerations in Slovakia not disregarding the local specificities.

## SYMBOLS

| | |
|---|---|
| CA | cluster analysis |
| PCA | principal component analysis |
| PCS | principal component score |
| FA | factor analysis |
| DF | differential factor analysis |
| MVDA | multidimensional variance and discriminant analysis |
| RAV | relative accuracy value |
| $\bar{c}(X)$ | arithmetical mean of the $c(X_i)$ concentration values |
| $c(X_{st})$ | declared concentration value of the element (X) of the used standard |
| $c(X_L)$ | detection limit of the element (X) |
| $m_{gds}$ | mass of the monthly sampled fraction of the gravitation dust sediment |
| $\mathbf{X}_{m,n}$ | basic data matrix |
| $\mathbf{A}_{m,s}$, $\mathbf{F}_{s,n}$ | splitted matrices of the basic matrix |
| $\lambda_i$ | s.c. "eigenvalue" (proper value) matrix |
| $\mathbf{U}$ | unity matrix |

## REFERENCES


1. Einax, J., Danzer, K., and Matherny, M., *Int. J. Environ. Anal. Chem. 44*, 185 (1991).
2. Matherny, M. and Ondášová, M., *Transaction of the Technical University of Košice 2*, 102 (1992).
3. Flórián, K., Matherny, M., Ondášová, M., and Pliešovská, N., *Chem. Listy 86*, 617 (1992).
4. Henrion, G., Henrion, A., and Henrion, R., *Beispiele zur Datenanalyse mit BASIC-Programmen*. Deutscher Verlag der Wissenschaften, Berlin, 1988.
5. Malinovski, E. R., *Factor Analysis in Chemistry.* 2nd Edition. Wiley, New York, 1991.
6. Weber, E., *Einführung in die Faktorenanalyse*. Fischer Verlag, Jena, 1974.
7. Ahrens, H. and Läuter, J., *Mehrdimensionale Varianzanalyse*. Akademie-Verlag, Berlin, 1981.
8. Flórián, K., Gálová, M., Koller, L., Krakovská, E., Lux, L., Matherny, M., Nickel, H., and Pliešovská, N., *Acta Chim. Hung.—Models in Chemistry 129*, 611 (1992).
9. Balgavá, V. and Matherny, M., *J. Radioanal. Nucl. Chem. 170*, 171 (1993).
10. Matherny, M. and Balgavá, V., *Fresenius J. Anal. Chem. 346*, 162 (1993).
11. Kalinčak, M., *Standard ENO*, No. 12-1-02. Institute of Radioecology and Nuclear Technique, Košice, 1985.
12. VDI: Messung partikelförmiger Niederschläge. Bestimmung der partikelförmigen Niederschläge mit dem Bergerhoff-Gerät; Standardverfahren VDI 2119 Bl. 2. VDI Kommission Reinhaltung der Luft, Stuttgart, 1972.


Translated by M. Matherny