

ENVIRONMETRICS IN RISK ASSESSMENT

Pavlina Simeonova¹, Vasil Lovchinov¹, Vasil Simeonov²

¹*Institute of Solid State Physics, Bulgarian Academy of Sciences, 1784 Sofia,
72, Tzarigradsko Chaussee Blvd., Bulgaria poly-sim@issp.bas.bg lovchinov@issp.bas.bg*

²*Faculty of Chemistry, University of Sofia "St. Kl. Okhridski", 1164 Sofia,
1, J. Bourchier Blvd., Bulgaria vsimeonov@chem.uni-sofia.bg*

Key words: risk assessment, aerosol, multivariate statistics, chemical mass balance

Recently, environmetrics turned to get a very important tool in assessment of ecological risk and of estimation of human life quality. It is based on the application of multivariate statistical methods like cluster and factor analysis, apportioning of the pollution sources contribution, trend and time-series analysis, etc, to classify, model and interpret large data sets from environmental monitoring.

The aim of the present study is to demonstrate the results obtained by the use of various environmetric approaches in assessment of the air quality in several big Austrian cities like Vienna, Linz, and Graz. The data collection consists of aerosol samples (PM 2.5 and PM 10) analysed for major, organic and trace constituents from many sampling sites in the cities mentioned. By the use of cluster analysis, principal components analysis (factor analysis) and chemical mass balance modeling an attempt is made to detect the latent sources of air pollution and to determine their contribution to the formation of the total particle mass or to the total concentration of each constituent. The advantages and the disadvantages of each environmetric approach to solve the problem are stressed and discussed. It is convincingly shown that the traditional projection and classification methods like cluster or factor analysis lack real "resolution" ability in spontaneous source identification but they offer a simple way to find hidden pollution factors. On contrary, the chemical mass balance method allows better apportioning of pollution sources but it requires a preliminary determined set of source profiles of the really existing pollution emitters, which requires additional experimental work and, hence, financing.

Introduction

Suspected adverse health effects of even low levels of airborne particulate matter have led to increased concern over how many fine particulate might best be controlled [1]. The development of effective control strategies for fine particulate air pollution abatement in turn requires a knowledge of the relative importance of the various sources that contribute to the particulate matter concentrations at ambient air monitoring sites [2, 3]. In principle, this approach of the risk assessment strategy can employ three modes of evaluation depending on the monitoring data available. All of them belong to environmetrics, a relatively new branch in the environmental physics and environmental chemistry, which deals with application of multivariate statistical methods for classification, projection, modeling, and interpreting of environmental data. The first mode involves typical statistical spontaneous modeling using the complete data sets (e.g. cluster analysis or principal components analysis, the latter being very similar to factor analysis for environmental studies [4,5]). In this case no preliminary information about the possible pollution impact of the local or secondary emission sources is needed. If source emissions

data and ambient monitoring data are available, then the other two approaches could be employed: source-oriented models and receptor-oriented models. In the first case emission data and fluid mechanically explicit transport calculations to predict pollution concentrations at specific receptor monitoring locations are used [6]. Receptor-oriented models infer source contributions by determining the best-fit linear combination of emission source chemical composition profiles needed to reconstruct the measured chemical composition of ambient samples. The approach is known as “chemical mass balance” (CMB) method [7].

The aim of the present study to compare different assessment modeling approaches to ambient aerosol data from urban sites in Austria.

Experimental

The sampling sites were located in three major Austrian cities: Vienna (6 sites), Graz (1 site), and Linz (1 site).

The aerosol data collection was gathered in the period between October 2000 and September 2001 for aerosol particles of the classes PM₁₀ and PM_{2.5}. The sampling was performed by the use of a high – volume sampler (Digital DHA-80), which is a completely automated device. The aerosol particles on quartz fiber filters (QAT-UP, Pallflex, USA) allowing in this way determination of the carbon content.

The particle total mass was determined by weighing of the sampling filters before and after sampling. The determination of the water-soluble ions (cations: sodium, ammonium, potassium, magnesium and calcium; anions: chloride, nitrate, sulfate) was performed by the use of two ion-chromatographic systems after extraction of the filters by deionised water in ultrasonic bath for 20 min. The concentration of the heavy metals was determined by the use of atomic absorption spectrometry. One quarter of the filter was cut by a ceramic scissor and the sample was weighted and extracted with 10 mL 10 % HNO₃.

The analytical procedure for determination of carbon (total carbon, TC, black carbon, BC and organic carbon, OC) used the developments of the well-established approaches for sample burning in oxygen atmosphere (TC), optical determination (BC) and the difference between TC and BC for OC determination.

In the data treatment three approaches of the environmetrics were used: cluster analysis, principal components analysis, and CMB.

Cluster analysis is a well-known and widely used classification approach. In order to cluster objects characterized by a set of variables one has to determine their similarity. A preliminary step of data scaling is necessary, where normalized dimensionless numbers replaces the real data values in order to eliminate dimension differences. Then, the similarity (or the distance) between the objects in the variable space can be determined. Very often the Euclidean distance is used for clustering purposes. Thus, from the input matrix (raw data) a similarity matrix is calculated. There is a wide variability of hierarchical algorithms but the typical ones include the single linkage, the complete linkage and the average linkage methods. The representation of the results of the cluster analysis is usually performed by a tree-like scheme called dendrogram comprising a hierarchical structure (large groups are divided into small ones).

Principal components analysis (PCA) is a typical display method, which allows estimating the internal relations in the data set. There are different variants of PCA but basically, their common feature is that they produce linear combination of the original columns in the data matrix (data set) responsible for the description of the variables characterizing the objects of observation. These linear combinations represent a type of abstract measurements (factors, principal components) being better descriptors of the data structure (data pattern) than the original (chemical or physical) measurements. Usually, the new abstract variables are called latent factors and they differ from the original ones

named manifest variables. It is a common finding that just a few of the latent variables account for a large part of the data set variation. Thus, the data structure in a reduced space can be observed and studied. The new coordinates are called factor scores and the regression coefficients from the linear combination of the old variables – factor loadings.

As already mentioned in the CMB approach the chemical composition of the emissions from individual sources can be used to estimate source contributions to atmospheric samples taken at receptor air monitoring sites. A mass balance is constructed in which the concentration of specific chemical constituents in a given ambient sample is described as arising from a linear combination of relative chemical compositions of the contributing sources. The concentration of chemical constituent i at receptor site k , c_{ik} , can be expressed as:

$$c_{ik} = \sum f_{ijk} a_{ij} s_{jk}$$

Where s_{jk} is the increment to total particulate mass concentration at receptor site k originating from source j , a_{ij} is the relative concentration of chemical constituent i in the emissions from source j , and f_{ijk} is the coefficient of fractionation that represents the modification of a_{ij} during transport from source j to receptor k . Very often this coefficient is accepted to be near to 1.

Results and discussion

The real monitoring carried out ensures a very large data collection. In order to demonstrate the different environmetric approaches for pollution source identification and mass (concentration) partitioning we have chosen data only from one sampling site (the typical urban site AKH in Vienna, almost downtown location) and only for one aerosol particle size – PM_{2.5} (fine particles). We start with cluster analysis of the chemical variables where a typical hierarchical dendrogram (Ward's method of linkage, squared Euclidean distance as similarity measure, autoscaled input data for 22 chemical constituents) is presented in Fig. 1. The clusters obtained reveal a specific relationship between the chemical parameters, which could be an important indication for the possible pollution sources in the neighborhood. The similarity groups are marked on the plot and one could distinguish 4 clusters informing on 4 possible emission sources:

K1: K⁺, OC, BC, Na⁺, Mg²⁺, Cl⁻, Ni, Cd, Co, Mn;

K2: NH₄⁺, NO₃⁻, V, Zn, As, Pb, Cu

K3: Ca²⁺, Fe;

K4: SO₄²⁻, Cr, Oxal (oxalate).

It has to be immediately stated that the source identification is not easy.

The first possible source (related to K1) is very complex and includes probably mineral salt, vehicle traffic and combustion emitters. The second one (related to K2) is also quite undefined and contains possible industrial impacts, traffic, and secondary emission contributions. The third and the fourth sources (K3, K4) could be related to street dust and cooking impacts.

If the same data set is a subject to factor analysis, the final result does not seem more encouraging. In Table 1 the factor loadings of the original variables are presented and the significant ones are marked. Again, four latent factors are responsible for the data set structure and they explain almost 90 % of the total variance of the system. In this interpretation the first latent factor (45.8 % explained variance) is very complicated in structure since it includes with high factor loadings carbon content (BC, OC) major ions (K⁺, NH₄⁺, Cl⁻, NO₃⁻, SO₄²⁻), heavy metals (As, Cu, Ni, Pb, V, Zn). Probably, this latent factor is related strongly to secondary emission sources, combustion sources (vehicle, heating, cooking) and even mineral salt sources. Therefore, this identified factor is of mixed origin. The second factor, which explains 17.6 % of the total variance, is probably related to street and soil dust emitters (Ca²⁺), the third (with 13.5 % explanation) and the

fourth (with 11.7 %) indicate high loadings for Mg²⁺ and Cd, on one hand, and for Co, Mn and Fe, on the other. It is quite difficult to identify acceptable pollution sources only with these tracer components (probably street dust).

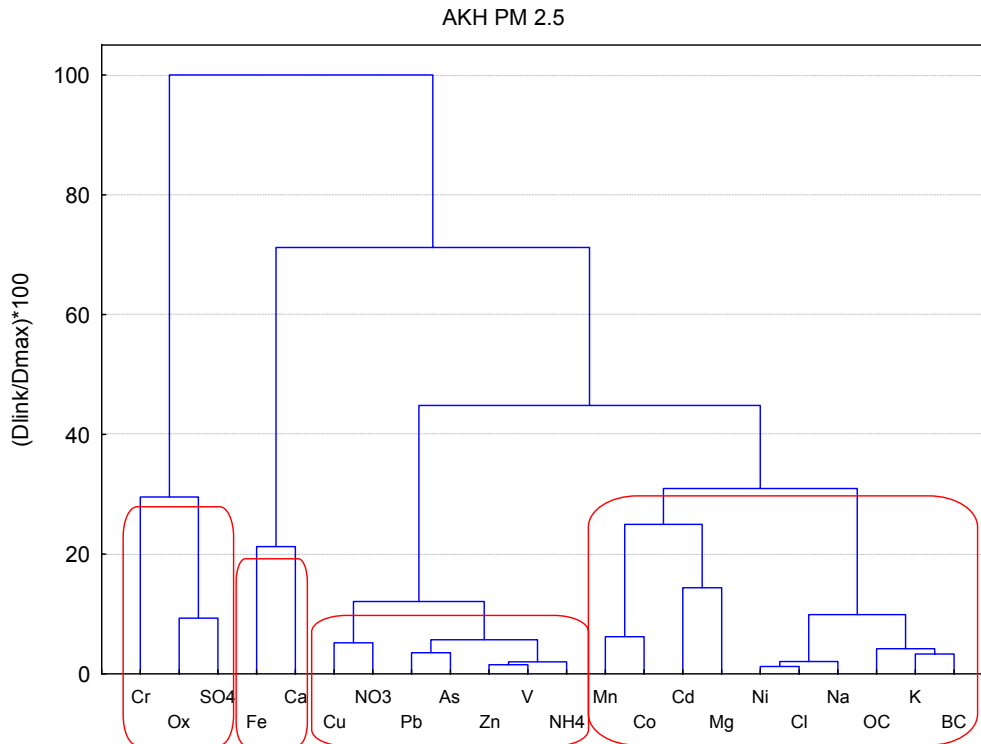


Fig. 1. Hierarchical dendrogram for chemical components clustering

The “classical” multivariate statistical methods for creation of statistical models directly from the ambient data face serious problems in pollution sources identification (and respective source apportioning). Their resolution scale is quite coarse and need additional refining. In the next figure (Fig. 2) the results of the application of CMB method are illustrated. The same ambient data set is used but data from pollution sources are

Table 1. Factor loadings AKH PM_{2.5} data set

Species	PC1	PC2	PC3	PC4
BC	0.712	0.416	0.374	0.121
OC	0.664	0.185	0.627	0.294
NA	0.582	0.682	0.311	0.223
NH ₄	0.977	-0.039	0.164	-0.028
K	0.676	0.276	0.603	-0.086
CA	-0.240	0.866	-0.024	0.315
MG	0.277	-0.153	0.809	0.131
CL	0.653	0.581	0.408	0.159
NO ₃	0.823	0.330	0.253	0.156
SO ₄	0.602	-0.571	-0.104	-0.209
AS	0.879	-0.0478	0.168	-0.099
CD	0.047	0.325	0.665	0.582
CO	0.574	0.413	0.046	0.663

CR	-0.212	0.878	0.086	0.207
CU	0.890	0.251	-0.133	-0.010
FE	-0.298	-0.012	0.109	0.807
MN	0.515	0.065	0.185	0.725
NI	0.754	0.517	0.297	0.183
PB	0.884	-0.023	0.346	-0.029
V	0.924	-0.100	0.301	0.114
ZN	0.906	-0.006	0.289	0.207
Expl.Var%	45.8	17.6	13.5	11.7

involved for the apportioning. The source profiles are taken from the Speciate 3.2 data bank for literature source profiles (mainly from USA sources). The CMB models obtained are summarized for each trimester of the sampling year. The source profiles used are: wood combustion (WOOD), steel production (STEEL), petrol combustion (PETROL), paved road dust (PAVE), Diesel combustion (DIESEL_C), earth's crust (CRUSTAL) and coke combustion (COKE). In the figure the apportioning of the total mass is indicated but the same modelling could be performed for each of the chemical constituents. It is readily seen what part of the total mass remains unexplained by the model (due to lack of appropriate local sources or due to the fact that not real local source profiles are used), what is the contribution of each source in each season to the particulate mass.

Conclusion

The environmetric modeling for the aim of risk assessment and air quality assessment has shown that the most reliable source apportioning could be obtained by application of chemical mass balance approach. It proves to be a finer partitioning instrument as compared to the spontaneous modeling by multivariate statistical methods,

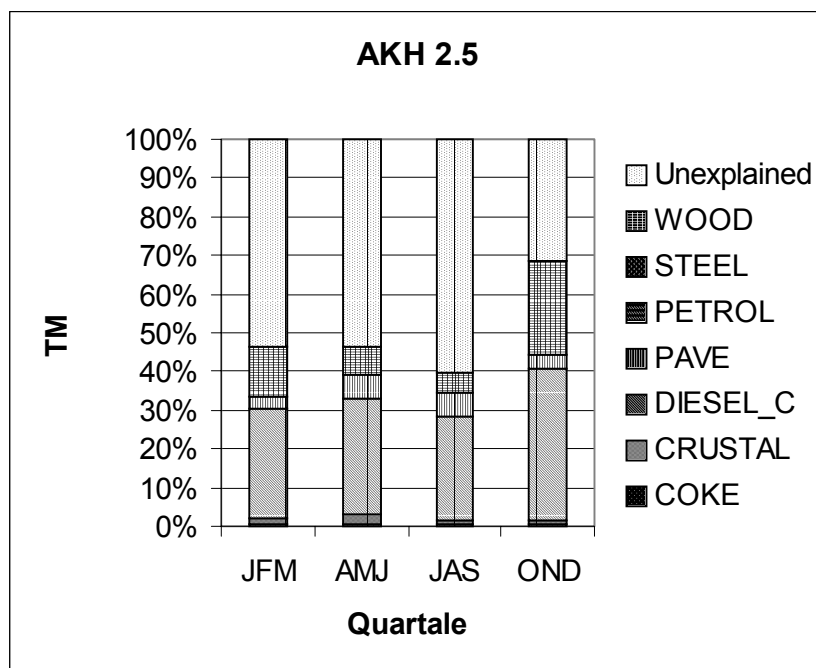


Fig. 2 CMB apportioning of total mass for quarter (the letters indicate the name of the months)

where the complex character of the pollution impacts in a certain environment often deteriorates source identification. However, the CMB modeling requires measurement of own local source profiles, which is an expensive and analytically complicated Procedure.

The reliable assessment of the pollution risks needs all environmetric approaches in order to collect full information on the systems of interest.

References

1. D.W. Dockery, C. A. Pope, X. Xu, J. D. Spengler, J. Ware, E. Martha, B. Ferris, F. Speizer. 1993. *N. Engl. J. Med.* **329**, 1753.
2. E. Atkinson, D. Lewis. 1974. *J. Envir. Econ. Management* **1**, 237.
3. R. Harley, S. Hunts, G. Cass. 1989. *Envir. Sci. Technol.* **23**, 1007.
4. V. Simeonov, S. Tsakovski, T. Lavric, P. Simeonova, H. Puxbaum. 2004. *Microchim. Acta* **148**, 293.
5. P. Simeonova, V. Simeonov, L. Lux, I. Dakova, T. Spanos. 2005. *Ecol. Chem. Eng.* **12(7)**, 727.
6. K. Benkala, J. seinfeld. 1979. *Atmospheric Environment* **13**, 1181.
7. J. Watson. 1984. *JAPCA* **34**, 619.