ECOLOGICAL CHEMISTRY AND ENGINEERING A

Vol. 18, No. 2 2011

Stefan TSAKOVSKI[1] and Vasil SIMEONOV[2*]

CLASSIFICATION OF DIFFERENT-SIZED AEROSOL MONITORING DATA

KLASYFIKACJA DANYCH MONITORINGOWYCH FRAKCJI AEROZOLU O RÓŻNYCH ROZMIARACH CZĄSTEK

Abstract: The present study deals with the application of self-organizing maps (SOM) of Kohonen for the classification of aerosol monitoring data sets from two sampling points (Arnoldstein and Unterloibach) located close to the border between Austria and Slovenia. The goal of the chemometric data treatment was to find some specific patterns in the classification maps for five different aerosol fractions collected in four different seasons of the year. The results obtained indicated a distinct separation of the ultrafine particles (PM 0.01–PM 0.4) from the other fractions which underlines their specific effect on human health. Seasonal separation but only between summer and winter sampling is also observed.

Keywords: chemometrics, classification, self-organizing maps, aerosol fraction, seasonal sampling

Assessment of the air pollution at sites of interest requires consequent and constant monitoring of carefully selected parameters. This usually results in a dataset with complex and multiway structure. For instance, consider a data set obtained as a result of monitoring air pollution described by several parameters measured at different sampling sites over 10 years. Possible multivariate statistical approaches to interpret and model such a data set are PARAFAC [1, 2], PARAFAC2 [3, 4] and Tucker3 [5, 6]. Such data can be seen as 3-way data array arranged as *sampling site* × *parameters* × *time* and thus explored with approaches dialing with 3-way data structure. Another opportunity for interpretation and classification of such type of data seems to be the application of self-organizing maps (SOM) of Kohonen which offer significant advantages in treating multidimensional data sets [7–12].

In the present study a more complex dataset is analyzed. It has been of a great interest [13] to monitor the concentration of different chemical components in the

[1] Chair of Physical Chemistry, Faculty of Chemistry, University of Sofia "St. Kl. Okhridski", J. Bourchier Blvd. 1, 1164 Sofia, Bulgaria.

[2] Chair of Analytical Chemistry, Faculty of Chemistry, University of Sofia "St. Kl. Okhridski", J. Bourchier Blvd. 1, 1164 Sofia, Bulgaria.

* Corresponding author.

aerosol samples collected at two sampling sites scattered in the region of the Austrian province Carynthia close to the Slovenian border during spring, summer, autumn and winter. Additionally, the samples were classified in five particle size fractions in order to reveal differences, if they exist, in behavior of fine and coarse particles with respect to the concentrations of chemical components measured in the samples over the four seasons of the monitored year. It is substantial to note that industrial activity could be found on the Slovenian territory since on the Austrian side no industry is available.

The aim of the present study is to classify the monitoring data set from industrial region of Austria by the use of self-organizing maps of Kohonen in order to reveal the relationship between particle size and seasonality with respect to the air quality of the region in consideration. It is our conviction that this study will bring the better understanding of the complex processes that take place in the local area of interest. Besides, the increasing intensity of production of nanomaterials put on the agenda of the society the problem of possible hazards due to the emission of ultrafine particles in the atmosphere. Therefore, the main tasks of this model study are to find out if the different aerosol fractions and their seasonal distribution form different patterns in the classification schemes. No similar study has been ever performed up to our knowledge.

Experimental

The data set contains concentrations of 16 chemical components (Na, NH_4^+, K, Ca, Mg, Cl^-, NO_3^-, SO_4^{2-}, C, Cd, Cu, Fe, Mn, Pb, V and Zn) and the amount of dust [$\mu g/Nm^3$], measured in triplicate in spring, summer, autumn and winter, in five different particle size fractions (PM0.04-PM0.1, PM0.1-PM0.4, PM0.4-PM1.6, PM1.6-PM6.4, PM6.4-PM25) at the two sampling sites Unterloibach and Arnoldstein, Carynthia, Austria.

Sampling sites

Two sampling sites are considered both located in the Austrian province Carynthia near to the Austrian-Slovenian border.

The sampling site Unterloibach is located in the Austrian province Carynthia at height 629 m a.s.l. (latitude 46° 32′18″ and longitude 14° 48′52″). The site is a typical rural one and is located near to the border with Slovenia. At a distance (northeast direction) of nearly 15 km a lead smelter is still active (on Slovenian territory) and in the eastern direction, again in Slovenia, a steel work is producing special quality steels. In southern direction one could detect the biggest Slovenian coal power station, which delivers over 75 % of the electricity for the country.

The sampling site Arnoldstein is also located in the Austrian province Carynthia at height 564 m a.s.l. (latitude 46° 33′31″ and longitude 13° 42′12″). The site is close to a small settlement with 700–800 m distance from an industrial region. At present three different industrial enterprises – for waste recycling, for polymer production, and for steel finishing are active.

Sampling procedure

The aerosol data collection was gathered in the period between March 1999 and February 2000. The sampling was performed by the use of a high – volume sampler (Digitel DHA-80), which is a completely automated device. The aerosol particles of the class PM_{10} are collected on a daily basis on quartz fiber filters (QAT-UP, Pallflex, USA) allowing in this way determination of the carbon content. Additionally to the sampling by high-volume sampler four seasonal sampling campaigns by low-pressure impactor (LPI 75/0.04 according to Berner) were carried out.

Analytical procedure

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 cm^3 10 % HNO_3.

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 of [14] for sample burning in oxygen atmosphere (TC), optical determination (BC) and the difference between TC and BC for OC determination.

The complete description of the sampling devices, the pre-sampling preparation of the filters and the chemical analysis procedure could be found in [13].

Statistical data analysis

The SOM is an algorithm used to visualize and interpret large high-dimensional data sets [15]; it is an unsupervised pattern cognition method similar to cluster analysis. The main advantage of SOM is the simultaneous classification of variables and objects (sampling locations). Typical applications are visualizations of process states or financial results by representing the central dependencies within the data on the map. The map consists of a regular grid of processing units called neurons.

A model of some multidimensional observations, possibly a vector consisting of features (variables), is associated with each unit. The map attempts to represent all available observations with optimal accuracy using a restricted set of models. At the same time the models become ordered on the grid so that similar models are close to each other and dissimilar models far from each other. Fitting of the model vectors is usually carried out by a sequential regression process, where t = 1, 2, ... is the step index. For each sample $x(t)$, the winner index c (best matching unit – BMU) is first identified by the condition:

$$\forall i, \| x(t) - m_c(t) \| \leq \| x(t) - m_i(t) \|$$

When the BMU has been found, the weight vectors of the SOM are updated so that the BMU is moved closer to the input vector in the input space.

Then, all the model vectors or a subset of them belonging to the nodes centered around node $c = c(\mathbf{x})$ are updated as:

$$m_i\ (t + 1) = m_i(t) + h_{c(x),i}(x(t) - m_i(t))$$

Here, $h_{c(x),i}$ is the "neighborhood function", a decreasing function of the distance between the i-th and c-th nodes on the map grid. This regression is usually reiterated over the available objects.

The trained map can be graphically presented by 2D planes for each variable, with the variable distribution values being indicated by different colors on the different regions of the map. Additionally, the node "coordinates" (vectors) can be clustered by the non-hierarchical K-means classification algorithm.

All calculations concerning SOM classification were performed by a free Teuvo Kohonen toolbox (SOM Toolbox 2.0), which can be downloaded together with documentation from http://www.cis.hut.fi/projects/somtoolbox/. Matlab 6.5 software as an environment was used. The statistical tests were performed by the use of software package STATISTICA 6.0.

Results and discussion

Since the main target of the study was to find specific relationships between the season of sampling and the aerosol fraction, the data set was divided into two major parts: Arnoldstein monitoring results and Unterloibach monitoring results.

Site Arnoldstein

In Fig. 1 the self-organizing maps for all aerosol fractions and all sampling seasons are shown.

Several typical patterns could be detected from the classification maps. There is a group of classification objects (seasons, fractions) which are characterized by highest concentrations of calcium, sodium, iron, manganese, lead, copper and to some extent chloride (these objects are located at the right down corner of the SOMs). It might be assumed that this pattern reflects the effects of industrial impact around the sampling site. Another pattern is formed by the maps similarity of vanadium, zinc, carbon and to some extent potassium (highest concentrations of the analytes on the left down corner of the SOMs). This pattern is probably a reflection of the impact of combustion processes (oil, gasoline, coal, wood burning) in the neighborhood. The role of the atmospheric

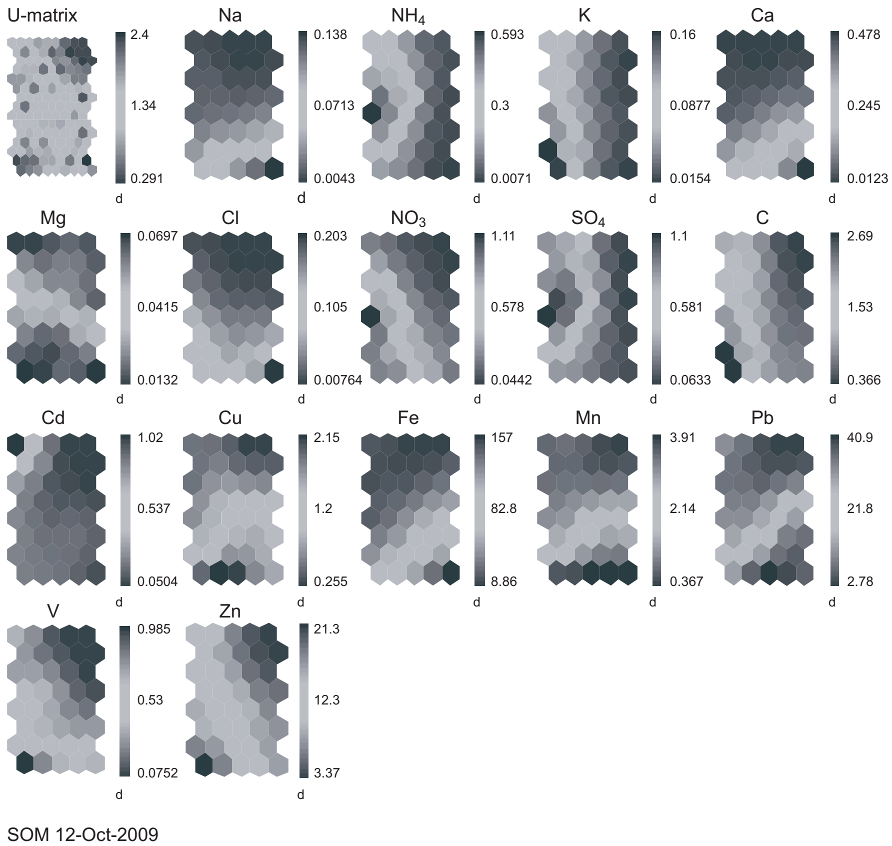

Fig. 1. Self-organizing maps for site Arnoldstein (all fractions, all seasons)

transfer of secondary aerosol is indicated by the formation of the third pattern where a resemblance is found in the maps of nitrate, sulfate, and ammonium (the highest concentrations of these major components of the aerosol are found on the middle of the left side of the SOMs). Magnesium and cadmium indicate a different distribution and do not resemble any of the patterns already defined. It may be assumed that magnesium shows this specificity due to the crustal impact of the Alpine region and cadmium – due to the ore content for the lead smelter production.

In Fig. 2 the grouping of the variables is indicated.

The variables classification panes show a slightly different grouping as compared by the empirical comparison of patterns of similarity between the general SOMs. Calcium, iron and sodium form the group probably related to the metallurgical impact of the Slovenian neighborhood. This group is very close to the one to which anthropogenic origin could be attributed – lead, manganese, copper. The effect of the secondary aerosol transport is indicated by the link between sulfate and ammonium and the burning sources are presented by the formation of group

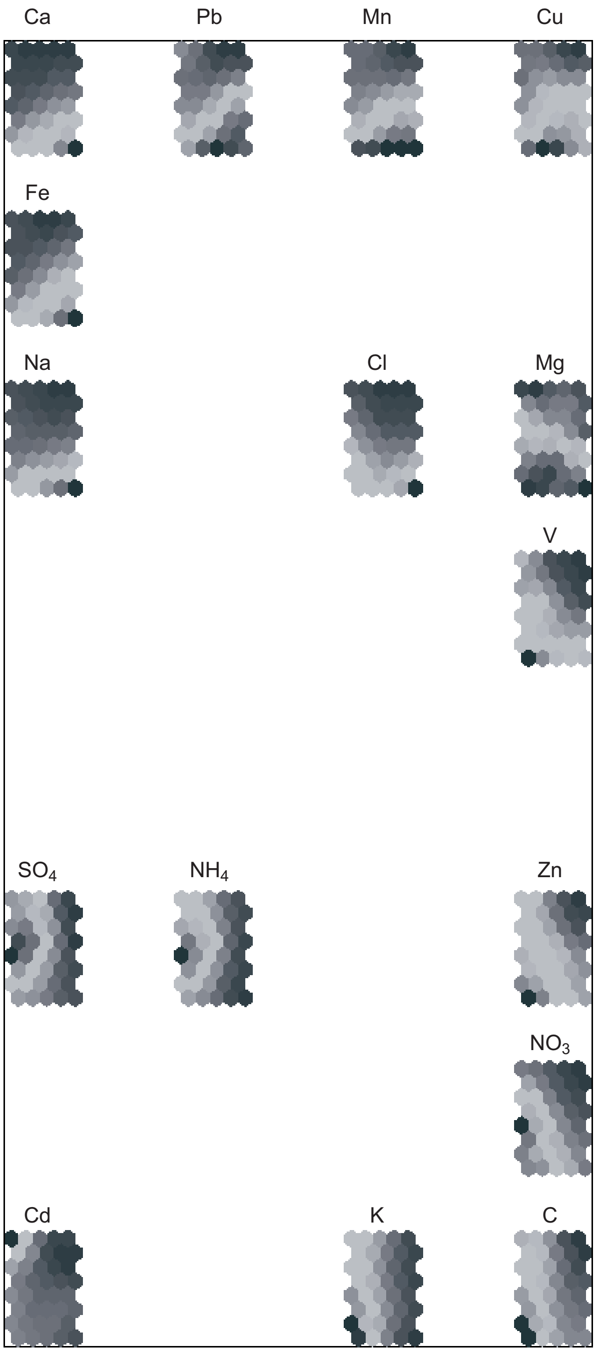

Fig. 2. The classification planes of the variables (Arnoldstein site)

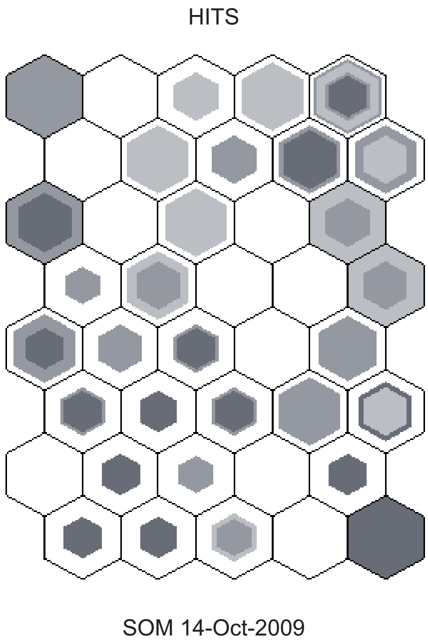

Fig. 3. Formation of seasonal clusters (Arnoldstein site)

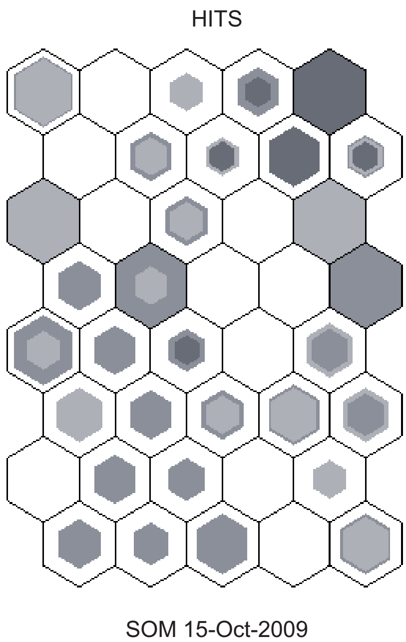

Fig. 4. Classification of the aerosol fractions (Arnoldstein site)

of similarity between potassium and carbon. Again, cadmium is located apart from the other chemical parameters reflecting the strong and specific impact of the lead smelter.

One important goal of the study was to try to find seasonal effects from the monitoring data collected in all seasons of the year. In Fig. 3 the classification of all aerosol fractions with respect to the seasonal distribution of the dust is shown.

It is obvious that the separation between summer and winter samples at Arnoldstein site is well expressed. The typical summer samples are dominantly located in the bottom part of the hit diagram since the winter samples are concentrated in the upper part. The other two seasons (spring and autumn) are not so distinctly separated from the rest of the samples. Therefore, one could introduce individual summer and winter pattern of the samples since the third one (for the spring and the autumn) is of non-specific character.

It is important to note that in the winter period the total amount of particulate matter, of the pollutant concentrations (arsenic, cadmium, chromium, copper, iron, manganese, nickel, lead, vanadium, zinc), and the carbon content all of them marking the anthropogenic impact have their highest concentrations. During the summer period these concentrations are lowest. Higher winter concentrations show also sodium, potassium, calcium and magnesium being related to the crustal impact to the total amount of the aerosol. The spring-autumn pattern is characterized by highest levels of the ions related to the secondary aerosol transfer – nitrate, sulfate and ammonium.

Since a more or less specific classification was found for the seasonal factor, it was interesting to check if such specificity could be found for the separate aerosol fractions. The results could be use when assessing the effects of different-sized aerosol particles on human welfare and health. This problem is quite interesting with respect to the increasing role of the nano-sized technologies.

In Fig. 4 the hit diagram with classification results for all five fractions for all seasons (Arnoldstein site) is presented.

The finest aerosol fraction (PM0.04-PM0.1) is clearly separated from the other four and forms a well-defined cluster (upper right corner of the diagram). The "ultrafine" pattern is discriminated from the others by the lowest concentrations of chemical species in this particular fraction.

The next two fractions (PM0.1-PM0.4 and PM0.4-PM1.6) form a "fine mixed" pattern characterized by increasing level of concentrations reaching its maximum in fraction (PM0.4-PM1.6). A similar mixing ("coarse mixture") is observed for the coarse fractions (PM1.6-PM6.4 and PM6.4-PM25). In the latter case the concentration of the chemical components starts decreasing after reaching the maximum of the "fine mixed" pattern.

Site Unterloibach

Next four figures (Figs. 5–8) illustrate the classification results for the different seasons and fractions of the aerosol samples collected at the Unterloibach site.

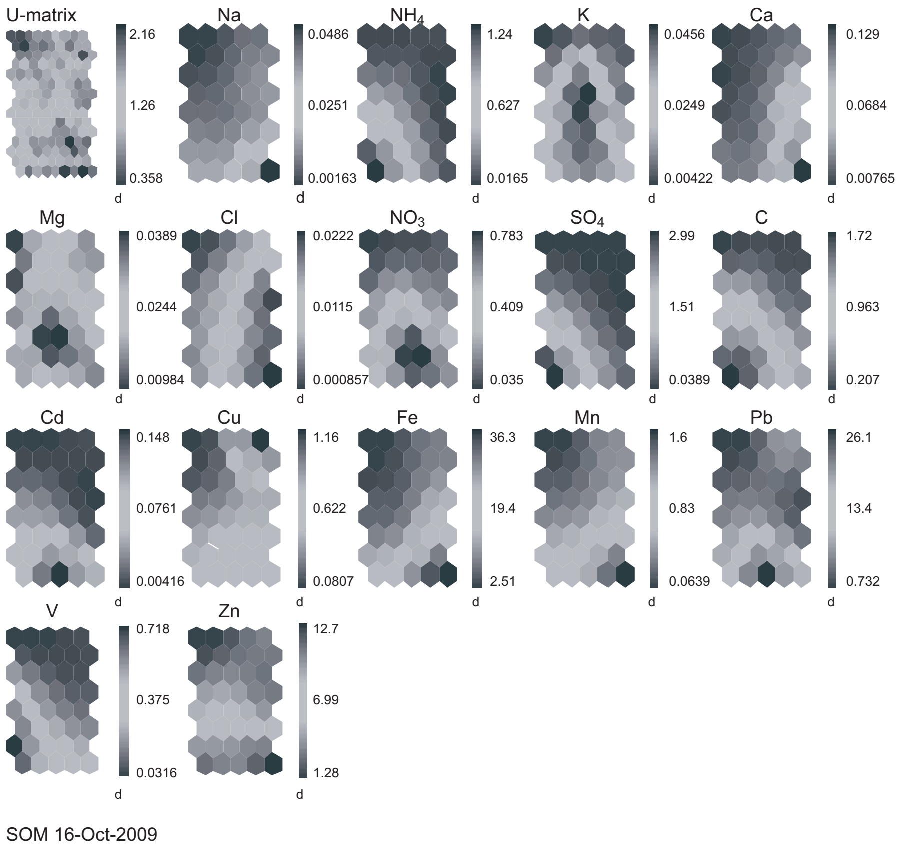

Fig. 5. Self-organizing maps for site Unterloibach (all fractions, all seasons)

From the general maps (Fig. 5) the patterns formed are including the pattern of the industrial impact (sodium, calcium, chloride, manganese, iron, zinc), the second mixed pattern of combustion processes and secondary emissions (ammonium, carbon, sulfate and vanadium) and the third combined pattern reflecting anthropogenic and crustal influences (nitrate, lead, cadmium, potassium and magnesium). The specific position of copper on the general maps location reflects site specificity related to copper ore treatment.

Almost the same configuration could be derived from the classification planes of the variables (Fig. 6).

It is interesting to note that for this particular site the seasonal classification differs substantially from that of Arnoldstein site (Fig. 7). No distinct division between seasons is found and it proves the significant difference of the air quality and its seasonal parameters at the two sites.

However, the classification of the aerosol fractions resembles almost completely the separation at site Arnoldstein (Fig. 8). A very clear separation of the “ultrafine” fraction

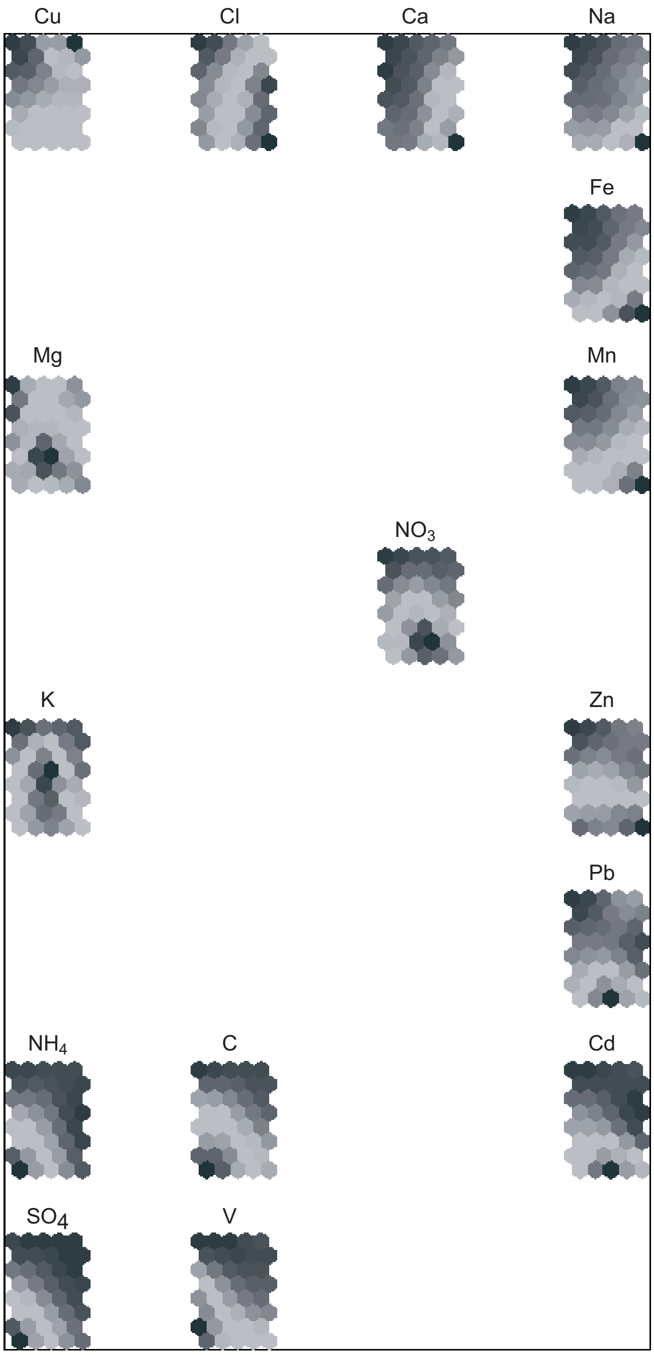

Fig. 6. The classification planes of the variables (Unterloibach site)

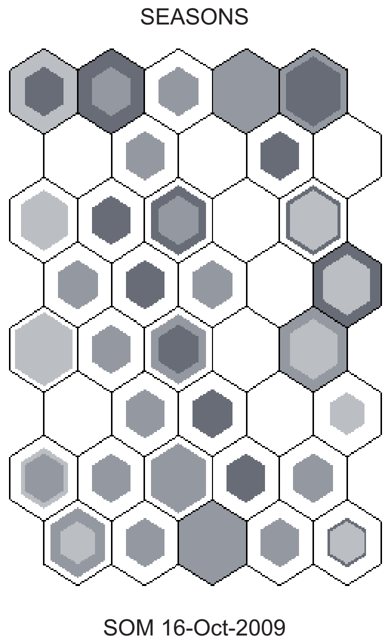

Fig. 7. Formation of seasonal clusters (Unterloibach site)

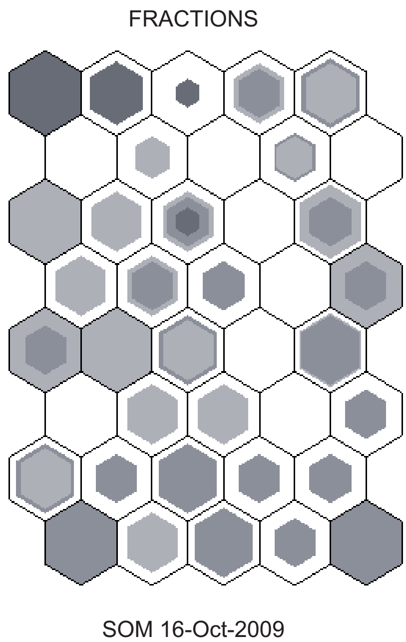

Fig. 8. Classification of the aerosol fractions (Unterloibach site)

from the rest of the fractions is observed since the other four fractions are of mixed character.

Conclusions

The most significant results from the study carried out was the classification of the aerosol fractions and separating a “ultrafine” fraction pattern characterized by lowest concentrations of the chemical parameters but a very stable and not mixable with the other fraction patterns system. This is an indirect proof of the special physicochemical and probably health-affecting role of the fine aerosol fractions in the ambient atmosphere. Besides, there is a clear separation of the aerosol effects between winter and summer seasons which indicates that the health hazards could be different in different seasons. Finally, the site location is also an separate factor which should be interpreted in all possible air quality assessments.

Acknowledgement

The authors would like to express their sincere gratitude to the National Science Fund – Ministry of Education and science (Project UNION DO 02-82) for the financial support which made this study possible. Thank are also due to Prof. Dr. H. Puxbaum, Technical University of Vienna and Dr. Tomaz Lavric for offering the data set.

References

[1] Bro R.: Chemomet. Intell. Lab. Syst. 1997, **38**, 149–171.
[2] Faber N., Bro R. and Hopke P.: Chemomet. Intell. Lab.Syst. 2003, **65**, 119–137.
[3] Kiers H., Ten Berge J. and Bro R.: J. Chemomet. 1999, **13**, 275–294.
[4] Bro R., Anderson C.A. and Kiers H.A.L.: J. Chemomet. 1999, **13**, 295–309.
[5] Henrion R.: Chemomet. Intell. Lab. Syst. 1994, **25**, 1–23.
[6] Geladi P.: Chemomet. Intell. Lab. Syst. 1989, **7**, 11–30.
[7] Tsakovski S., Simeonova P., Simeonov V., Freitas M.C., Dionosio I. and Pacheko A.M.G.: J. Radioanal. Nucl. Chem. 2009, **281**, 17–22.
[8] Wen G., Guo X., Huang D. and Liu K.: IEEE Int. Conf. on Neural Networks – Conf. Proc. Art. 4371093, 2007, 991–996.
[9] Stanimirova I. and Simeonov V.: Chemomet. Intell. Lab. Syst. 2005, **77**, 115–121.
[10] Astel A., Tsakovski S., Barbieri P. and Simeonov V.: Water Res. 2007, **41**, 4566–4578.
[11] Tsakovski S., Kudlak B., Simeonov V., Wolska L. and Namiesnik J.: Anal. Chim. Acta 2009, **631**, 142–152.
[12] Kolehmainen M., Martikainen H., Hiltunen T. and Ruuskanen J.: Environ. Monit. Assess. 2000, **65**, 277–286.
[13] Lavric T.: Composition and Sources of Aerosols (PM10) in the border Region Carynthia – Slovenia – Italy. Ph.D. Thesis, Technical University of Vienna, Vienna 2002.
[14] Puxbaum H. and Rendl J.: Microchim. Acta 1983, 263–272.
[15] Kohonen T.: Self-Organizing Maps. Springer, New York 2001.

KLASYFIKACJA DANYCH MONITORINGOWYCH FRAKCJI AEROZOLU O RÓŻNYCH ROZMIARACH CZĄSTEK

Abstrakt: Przedstawiono wyniki badań monitoringowych próbek aerozolu atmosferycznego pobranych z dwóch punktów pomiarowych (Arnoldstein i Unterloibach) z pobliża granicy między Austrią i Słowenią.

Dane zinterpretowano z wykorzystaniem samoorganizujących się map (SOM) Kohonena. Celem chemometrycznej interpretacji danych było znalezienie charakterystycznych struktur na mapach klasyfikacji dla pięciu różnych frakcji aerozoli, zebranych w czterech różnych porach roku. Uzyskane wyniki wskazują na wyraźne oddzielenie najdrobniejszych cząstek (PM 0,01 – PM 0,4) od innych frakcji, co wskazuje na ich specyficzne działanie na zdrowie człowieka. Obserwuje się również zmiany sezonowe, ale tylko między próbkami pobranymi latem i zimą.

Słowa kluczowe: chemometria, klasyfikacja, mapy samoorganizujące się, frakcja aerozolowa, próbkowanie sezonowe