Cluster analysis interpretation of aerosol samples

Pavlina Simeonova¹, Danail Simeonov², Lyubomir Spassov³, Vasil Simeonov⁴,*

¹ Laboratory of Environmental Physics, Georgi Nadjakov Institute of Solid State Physics, Bulgarian Academy of Sciences, Tzarigradsko Chaussee Blvd. 72, 1784 Sofia, Bulgaria
² Institute of General and Inorganic Chemistry, Bulgarian Academy of Sciences, Acad. G. Bontchev Str., Bl. 11, 1113 Sofia, Bulgaria
³ Institute of Catalysis, Bulgarian Academy of Sciences, Acad. G. Bontchev Str., Bl. 11, 1113 Sofia, Bulgaria
⁴ Laboratory of Chemometrics and Environmetrics, Faculty of Chemistry and Pharmacy, University of Sofia “St. Kl. Okhridski”, 1164 Sofia, J. Bourchier Blvd. 1, Bulgaria
* Corresponding author. E-mail: VSimeonov@chem.uni-sofia.bg

A total of 26 sets of sized aerosol samples were collected using a Berner low pressure impactor from an urban site in the center of Thessaloniki, Northern Greece, during the period of February-July 2007. The aerosol samples were chemically characterized for water soluble inorganic species (Cl⁻, NO₃⁻, SO₄²⁻, NH₄⁺, Na⁺, K⁺, Ca²⁺, Mg²⁺) and total organic carbon (TOC). The ten fractions analyzed (particle diameter between 0.03-16 μm) were conditionally divided into four categories: nano particles (less than 0.03 μm to 0.06 μm); ultrafine particles (0.06-0.50 μm); fine particles (1-4 μm) and coarse particles (8-16 μm).

Multivariate statistical analysis (cluster analysis) of the collected data was performed in order to reveal the specific data structure. Possible sources of air pollution were identified and an attempt is made to find patterns of similarity and dissimilarity between the different aerosol fractions. It was also checked if the sampling season affects the data structure. It was found that several factors are responsible for the data structure despite the size of the aerosols – mineral (soil) dust, sea sprays, secondary emissions, combustion sources and industrial impact. The seasonal separation proved to be not very specific.

**Keywords:** aerosol fractions; cluster analysis; data structure

**Introduction**

Recently, many epidemiological and toxicological studies have been focused on the importance of the aerosol particle size on human health [1-6]. Smaller aerosol particle sizes impose a higher hazard due to the size dependent deposition mechanism in the human respiratory system and different toxicities of smaller particles [7-9]. Current attention has been focused on the size and composition characteristics of fine and nano particles, as well as on toxicological mechanisms in laboratory scale, controlled human exposures and population-based epidemiology.

The size of primary aerosol particle is strongly related to the emission source, while various atmospheric processes, like nucleation, condensation and accumulation, play a major role in the size evolution of fine particles [10].
In urban environments, motor vehicle emissions are mainly responsible for the presence of nano and ultrafine particles [10,11]. A fine mode particle fraction, with diameter around 0.1 μm has been observed in aerosol samples obtained from "on road" sampling [12]. On the other hand, coarse mode particles have been reported for urban locations originating from crustal material, road dust and other automotive "non-exhaust" emissions and for sea salt particles[13].

The knowledge of the size distribution of the particles’ chemical components enhances useful information about their origin, formation and other chemical or physical properties [14-17]. The majority of the studies refer to PM$_{2.5}$ and PM$_{10}$ data in relation to the organic compounds and the inorganic chemical constituents. Only a few studies have been done on the size distribution of organic components [18,19] in various Portuguese areas (urban and rural).

The application of chemometric methods for classification and modeling of different sized aerosol particles monitoring data could also successively contribute to the interpretation of this complex environmental system [20,21].

The main aim of the present work is to interpret the monitoring data of each fraction by the use of cluster analysis in order to compare the data structure for each one of the fractions.

Experimental

Sampling site and aerosol collection

The aerosol sampling was carried out during the period of February to July 2007, at an atmospheric monitoring station in the center of Thessaloniki. The city of Thessaloniki (40° 62 ΄Ε, 22° 95 ΄Ν), with a population of one million, lies on the northern part of Thermaikos gulf and is surrounded by hills on its south-eastern side. On the northwestern part of the city various industrial activities operate like oil refining, petrochemical, iron/steel manufacturing, electrolytic MnO$_2$ production, aluminum production, etc.

The aerosol collection was conducted by a ten stage low pressure impactor (LPI 25/0.015, Berner type) with a size range of 0.015-16 μm. The sampling device was set on the roof of the station (3.0 m above ground). The air flow was adjusted at 25 L min$^{-1}$, while the air pump provided flow 12 m$^3$ h$^{-1}$ at pressure 150 mbar at the exit of the system which was periodically checking after two or three samplings. Celulose filters (Whatman 41) were used at each stage for the collection of the particles on 24 hour base, starting at 8.00 am. The collected particles were separated into 10 ranges with the following equivalent cut-off diameters: 0.015-0.03, 0.03-0.06, 0.06-0.125, 0.125-0.25, 0.25-0.50, 0.50-1.0, 1.0-2.0, 2.0-4.0, 4.0-8.0 and 8.0-16 μm. Totally, 26 aerosol sets, that is 26×10 fractions = 260 samples, were obtained with a frequency of 5-6 sets per month. The filters (before and after sampling) were conditioned for 48 h at standard conditions of relative humidity (50%) and temperature (20 °C) before weighing by using a Mettler Toledo (AB135-S/FACT) balance with readability of 0.01 mg/0.1 mg.

In order to check the quality of the weighing procedure, blank (N = 5) and loaded filters from each size (N = 3) were measured for five days. The standard deviation of the measurements for the blanks was ± 200 μg and for loaded filters it was between 300 and 600 μg for the various particle sizes. This check was carried out two times during the sampling period. The loaded filters, after their transfer to laboratory, were stored in refrigerator until extraction and analysis (within a week).

Chemical analysis

The collected filter samples were divided into two equal portions. Each portion was extracted with 10 ml distilled water, in an ultrasonic bath for 30 min and then the extract was filtered through a 0.45 μm pore membrane filter in order to proceed forward with the chemical analysis.

One of the above water extracts was used for anions (Cl$^-$, NO$_3^-$, SO$_4^{2-}$) and cations (K$^+$, Na$^+$, NH$_4^+$) determination using ion chromatography (IC). The cations Cu$^{2+}$, Mg$^{2+}$ were measured by atomic absorption spectroscopy.

The other half-filter extract was used for total organic carbon determination. This was conducted by using a Shimadzu TOC-V CPH liquid organic carbon Analyzer, applying the non-purgeable organic carbon (NPOC) technique [22].

Cluster analysis

In the present study cluster analysis was used as multivariate statistical method [39].

Cluster analysis is a widely used approach for environmetric purposes. In order to cluster objects characterized by a set of variables (e.g. sampling sites by chemical concentrations or pollutants), one has to determine their similarity. A preliminary step of data scaling is necessary (e.g. autoscaling or z – transform, range scaling, logarithmic transformation) where normalized dimensionless numbers replaces the real data values. Thus, even serious differences in absolute (concentration) values are scaled to similar ranges. Then, the similarity 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. Typical methods for linkage of similar objects into clusters include the single linkage, the complete linkage, the average linkage methods or, very often, the Ward’s method. The representation of the results of the cluster analysis is performed either by a tree-like scheme called dendrogram comprising a hierarchical structure.
Results and discussion

The main goal of the chemometric analysis was to identify the structure of the environmental sources responsible for the data structure for each one of the aerosol fraction analysed. Cluster analysis of the chemical parameters for each fraction (10 fractions) was carried out using the monitoring data for 26 sampling dates (winter and summer samples). All 10 chemical variables (ions and TOC) were included in the data interpretation.

Nano particles (0.03 μm)

In Figs. 1 and 2 the hierarchical dendrograms for chemical variables (Fig. 1) and sampling periods (Fig. 2) are presented for the most fine fraction (0.003 μm).

It is seen that three significant clusters determine the data classification (Ca, Mg, K) (Na, NH₄, SO₄, Cl) and (TOC, NO₃). F is a typical outlier. The sources associated with this grouping of variables are, respectively, (mineral and soil dust) (sea sprays and secondary emissions) and (combustion processes). Probably fluoride is associated with an anthropogenic source with specific composition (metallurgy of aluminum).

Since the object of interpretation are 26 weeks in spring and summer period it could be stated that no seasonal specificity is actually observed. Two clusters are formed but their division is not with respect to season. Only one week appears as outlier for this aerosol fraction – week 4 characterized by very high concentrations of Cl, Na, SO₄, NH₄, i.e. strong impact of sea spays (Cl, Na) and secondary emissions (SO₄, NH₄) a spring local event.

Nano particles (0.03-0.06 μm)

The next two figures (Figs. 3 and 4) demonstrate the data classification for the next fraction of aerosols both for the chemical components (Fig. 3) and the sampling season (Fig. 4).

For the clustering of the chemical parameters one detects the same significant clusters (Na, Cl, NO₃, SO₄), (Ca, K, Mg, NH₄) and two separate variables TOC and F. It could be assumed that the same sources and factors determine the data structure for the second category of fractioned aerosol – (sea sprays and secondary emissions) (mineral and soil dust) and combustion processes with the specific position of F as anthropogenic impacted variable. It proves that nano fractioned aerosols have one and the same origin.

The sampling period shows again homogeneity with one obvious outlier (object 4 – week 4) being discussed above.
Cluster analysis interpretation of aerosol samples

Ultra fine particles (0.06-0.125 μm)

Figs. 5 and 6 represent the classification of data for the next aerosol fraction of ultra fine particles. A slight difference is observed with respect to the identified sources affecting the data structure (Fig. 5) – the clusters formed include separate influence of marine sources or sea sprays (Na, Cl) and secondary emission (NH₄, SO₄, NO₃, Mg) which is an indication for the different way of aerosol formation by the transition from nano to ultra fine aerosols. Soil dust is present again as well as both outliers for combustion and anthropogenic sources.

The seasonal effect is also slightly transformed – outliers (Fig. 6) are now weeks 10 and 12. The input data information indicates that these weeks are characterized by serious increase of the level of secondary emission species (NH₄, SO₄, NO₃) as compared to the other sampling events. Probably, atmospheric processes are responsible for this situation. It is proved by checking the back trajectories for the period where a transatlantic transfer is proven.

Ultra fine particles (0.125-0.25 μm)

In Figs. 7 and 8 the clustering of chemical variables (Fig. 7) and sampling dates (Fig. 8) is presented. Again, a very well-expressed impact of sea sprays and secondary emissions (cluster containing Na, Cl, NH₄, SO₄), mineral dust (clustering of Ca, Mg, K) and combustion activities (linkage of TOC and NO₃) is observed for this aerosol fraction. The behavior of F is in any other case – outlier.

Sampling weeks 12 and 16 appear to be outliers in comparison with the homogeneity of all other sampling dates for this fraction. The situation for week 12 was discussed
above (impact of enhanced levels of secondary emissions). Additionally to this impact an increased level of combustion products is observed in week 16 due to traffic jams.

**Ultra fine particles (0.25-0.50 μm)**

In the third category of ultra fine aerosols the identified sources responsible for the data structure of the fraction is slightly different as sea sprays (Na, Cl) impact and secondary emission impact (SO₄, NH₄) are separated (Fig. 9).

![Hierarchical dendrogram, objects, ultra fine particles (0.125-0.25 μm)](image)

![Hierarchical dendrogram, variables, ultra fine particles (0.25-0.50 μm)](image)

The mineral and soil dust impact and the combination between combustion and anthropogenic influences are well defined.

For this fraction the sampling events seem similar and no substantial outliers are identified. For significance level of Sneath 2/3 D_max only week 7 and 12 differ insignificant from the rest of events due to enhanced impact of mineral and soil dust components (Ca, Mg).

**Fine particles (0.50-1.0 μm)**

In Fig. 11 the source identification for the fraction is well-defined. It seems that with the increase of the particle effective size the factors responsible for the data structure get more specified and selected. In this fraction five different sources are seen resembling, in principle, those identified in the other fractions: secondary emission (SO₄, NH₄), sea sprays (Na, Cl), mineral dust (Ca, Mg), agriculture soil impact (K, NO₃), combustion (TOC) and F as outlier (anthropogenic impact).

![Hierarchical dendrogram, variables, fine particles (0.50-1.0 μm)](image)
Weeks 10 and 12 are specific with respect to the composition of this aerosol fraction as seen in Fig. 12. Enhanced concentrations of nitrate and potassium (soil, fertilizer) impact determine this specificity in the fine fraction (0.50-1.0 μm).

**Figure 12.** Hierarchical dendrogram, objects, fine particles (0.50-1.0 μm)

---

**Fine particles (1.0-2.0 μm)**

The next two figures (Fig. 13 and 14) reflect the clustering of variables and objects for this fine particle fraction.

**Figure 13.** Hierarchical dendrogram, variables, fine particles (1.0-2.0 μm)

In this case there is no detailed separation between the traditional sources of aerosol formation – sea sprays and secondary emissions form a cluster (Na, Cl, NH₄, SO₄); dust contributions (soil, street, mineral, fertilizer) form another cluster (K, Ca, Mg, NO₃) and almost constant position of TOC and F as indicators for combustion and anthropogenic impacts.

The similarity between the sampling dates is affected only by two outliers – weeks 4 and 12 are characterized by heavy secondary emissions and sea sprays (week 4) and enhanced dust impact (week 12).

**Figure 14.** Hierarchical dendrogram, objects, fine particles (1.0-2.0 μm)

The only specific and non-traditional clustering is the formation of two subclusters (SO₄, Cl) and (NH₄, Na) for level of cluster significance 1/3 $D_{max}$ which is not quite logical. But the definition of clustering at level 2/3 $D_{max}$ these subclusters are joint together and form the sea sprays/secondary emissions linkage found in any of the previous aerosol fractions. TOC and F are the outliers; Ca and Mg form the dust impact (street, mineral) and K and NO₃ – the other dust fraction (soil, fertilizer).

**Figure 15.** Hierarchical dendrogram, variables, fine particles (2.0-4.0 μm)
The compact clustering of the sampling events is influenced only by three outliers – weeks 4, 10 and 12 (all of them were interpreted for previous aerosol fractions): impact of sea sprays and secondary emissions for week 4, soil dust impact for week 10 and mineral dust impact for week 12.

**Coarse particles (4.0 – 8.0 μm)**

In Fig. 17 the clustering of chemical components for the first coarse particle fraction is shown.

Obviously, with the increase of the aerosol size some changes of the linkage of the chemical variables is observed. It seems that the formation of this fraction is determined mainly by dust impact (Ca, Mg, NO₃). F is the typical outlier as in all other fractions. There is separation in the stable cluster for other fractions (sea sprays/secondary emissions) in two subclusters. For the first time the role of the combustion impact is also different – it is linked to part of the formation “sea sprays/secondary emissions”. Indeed, if the criterion $2/3D_{\text{max}}$ is applied, then one obtains one big cluster of the impact of simultaneously sea sprays, combustion and secondary emissions.

In Fig. 18 the linkage of the sampling periods is shown. Only two outliers (weeks 4 and 12) are obtained being characterized by strong secondary emissions impact.

**Coarse particles (8.0-16.0 μm)**

In the last aerosol fraction a well-defined separation of the sources is shown (Fig. 19). The interesting point is that combustion factor (TOC) is related to sea sprays/secondary emissions factor (SO₄, Na, NH₄, Cl). The dust impact is correctly defined (K, Mg, Ca, NO₃) and F is outlier.
There is no difference to the previous coarse fraction – homogeneous clustering with only two exceptions (weeks 4 and 12).

Conclusions

The major sources contributing to the formation of the different fractions of urban aerosol in City of Thessaloniki, Greece are sea sprays, secondary emissions, soil, street and mineral dust, combustion processes and anthropogenic activities.

A combination of local emissions and long-range transport has been found to influence the presence of particulate matter (coarse and fine particles respectively) in the study area.

The particles size is decisive for determination of the dominance of the pollutions source. For fine particles sea sprays and secondary emissions are favorite and for coarse fractions – mineral dust.

No specificity with respect to season of sampling is found. There are some specific time-dependent events (weeks 4, 10, 12 for almost all particle fractions) within the period of 26 sampling events (weeks) related to atmospheric processes (wind direction, long-distance transfer) characterized by enhanced impact of sea sprays, secondary emissions and dust impact.

The present study makes it possible to determine the relationship between particle size and the source responsible for the chemical composition of the aerosol system.

Acknowledgement. The authors would like to express their sincere gratitude to Prof. Dr. Roxani Tisouridou-Tzimou for offering the data set for statistical interpretation.

Цитирана литература

Интерпретация на аерозолни проби чрез кластерен анализ

Павлина Симеонова1, Данаил Симеонов2, Любомир Спасов3, Васил Симеонов4,*

1 Лаборатория по физика на околната среда, Институт по физика на твърдото тяло „Акад. Г. Наджаков”, БАН, Цариградско шосе 72, 1784 София, България
2 Институт по обща и неорганична химия, БАН, Акад. Г. Бончев, Бл. 11, 1113 София, България
3 Институт по катализ, БАН, Акад. Г. Бончев, Бл. 11, 1113 София, България
4 Лаборатория по хемометрия и екометрия, Факултет по химия и фармация, Софийски Университет „Св. Кл. Охридски”, 1164 София, Д. Боучер 1, България
* Автор за кореспонденция. E-mail: VSimeonov@chem.uni-sofia.bg

Получено: 22 февруари 2015
Рецензирана: 16 март 2015
Приета: 17 март 2015
Излязла online: 25 март 2015

Двадесет и шест набора от данни за разделени по размер аерозолни проби, събрани с колектор на Бернер от градска пробовземателна станция в град Солун, Гърция за периода февруари – юли 2007 г. са обект на изследването. Пробите са анализирани за водоразтворими неорганични компоненти (Cl−, NO3−, SO42−, NH4+, Na+, K+, Ca2+, Mg2+) и общ органичен въглерод (TOC). Десетте анализирани фракции (диаметър на частиците между 0.03-16 μm) са разделени условно на четири категории: наночастици (от по-малки от 0.03 μm до 0.06 μm); ултрафини частици (0.06-0.50 μm); фини частици (1-4 μm) и груби частици (8-16 μm).

Осъществен е кластерен анализ на данните с цел разкриване на специфичната структура на данните. Идентифицирани са възможни източници на атмосферно замърсяване за откриване на образци на подобие между различните аерозолни фракции. Бе проверено дали периодът на пробовземане е от значение за структурата на данните. Бе намерено, че отговорни за структурата на данните са няколко фактора, независимо от размера на аерозола – минерален (почвен) прах, морски аерозоли, вторични емисии, горивни източници и индустриално влияние. Сезонното разделяне не се оказва специфично.

Ключови думи: аерозолни фракции; кластерен анализ; структура на данни