# Assessment of the Sequential Principal Component Analysis chemometric tool to identify the soluble atmospheric pollutants in rainwater

Rocío Montoya-Mayor, Antonio José Fernández-Espinosa*, Miguel Ternero-Rodríguez

Department of Analytical Chemistry, Faculty of Chemistry,
University of Sevilla,
C. Profesor García González 1, Campus of Reina Mercedes,
41012 Sevilla, Spain
A.J. Fernández-Espinosa (*) (email: anjose@us.es)

**Abstract-** In this study a new method of PCA performance, the Sequential Principal Component Analysis (SPCA), was proposed and assessed on real samples. The aim was to identify the atmospheric emission sources of soluble compounds in rainwater samples, and the sample collection was performed with an automatic sampler. Anions and cations were separated and quantified by Ion Chromatography, whereas trace metals and metalloids were determined by ICP-MS spectroscopy. SPCA results showed eight interfering principal components (PCs) and ten significant PCs. The interfering cases originated from different atmospheric sources such as resuspended crustal particles, marine aerosols, urban traffic and a fertilizer factory. The significant PCs explained 84.6 % of the total variance. 28.1 % accounted for the main contribution, which was resuspended industrial soil from a fertilizer factory containing $NO_2^-$, $NH_4^+$, $NO_3^-$, $SO_4^{2-}$, $F^-$, Al, $K^+$, Mn, Sb and $Ca^{2+}$ as indicators of the fertilizer factory. Another important source (15.0 %) was found for $Na^+$, $Mg^{2+}$, $K^+$, $Cl^-$ and $SO_4^{2-}$, which represents the marine influence from south and southwest directions. Emissions of $Ba^{2+}$, Pb, $Sr^{2+}$, Sb and Mo, which represent a traffic source deposited in soils, were identified as another abundant contribution (12.1 %) to the rainwater composition. Other important contributions to the rainwater samples that were identified through SPCA included the following: different urban emissions (Cu, As, Cd, Zn, Mo and Co, 18.1 %), emissions from vegetation (HCOO⁻, 7.7 %), and industrial combustions (Ni, V 15.6 %). The application of SPCA proved to be a useful tool to identify the complete information on rainwater samples as indicators of urban air pollution in a city influenced mainly by traffic vehicles and resuspended polluted soils.

**Introduction**

The study of the chemical composition of rainwater has its origins in the 1950s when Eriksson (1953) [1, 2] investigated the effects of acid deposition on ecosystems. Erikson found that the origin of the main damage to the environment was caused by acid emissions of $SO_2$ and $NO_x$ gases. Thus, rainwater composition provides crucial information on existing air pollution [3] and has signalled great interest for two main reasons:

1) Precipitation is a route of entry for nutrients and contaminants in all biogeochemical cycles of terrestrial and aquatic ecosystems [4].

2) Acid rain transports large amount of toxic elements and causes serious effects on aquatic and terrestrial ecosystems, [5].

Recently, a lot of information about the composition of rainwater has been gathered in America [6], Central and Northern Europe [7, 8] and other parts of the world (China [9], India [10], etc). In Spain, several authors have studied the chemistry of atmospheric precipitation [11, 12], however the results are practically absent in the South of Spain. Furthermore, the papers pertaining to rainwater composition are scarce when compared to those referring to atmospheric particles.

Both particulate and soluble content of precipitation depends on several factors such as pollutant emission sources from air, amount of rain, washout ability of drops etc [13]. In this paper we are concerned with soluble atmospheric pollutants in rainwater. The atmosphere contains gases and particles coming from natural or anthropogenic sources, which after a residence time are deposited in the earth's surface through dry or wet sedimentation [14]. The acidity of rain depends on contributions from hydrocarbon combustion emissions, industrial processes, reactions in the combustion engines of vehicles, etc. Neutralizing alkaline substances may have an anthropogenic origin due to the activities of cement factories and power plants [15], nevertheless the origin of natural sources of these alkaline substances such as soil dust are much higher in arid zones [16]. In many cases rainwater has high concentrations of sulphates and nitrates but a low acidity, due to the neutralization effect. So, the study of cations also is of immense importance as they represent the major neutralizing agents, such as calcium [17], sodium [18] and ammonium [19].

On the other hand, the continuous changes in land use and the deforestation in Africa have greatly increased the origin of natural sources of particles found in Southern Europe. This material is transported long distances by winds into a long-range transport [20].This

phenomenon is blurring the distinction between anthropogenic and natural alkaline particles. Also, chemical species found in rain samples come from airborne particles [21] and gases either incorporated in the clouds (rainout), or washed out by the rain droplets below the cloud (washout) [22]. In addition, meteorological conditions also play an important role in all these reactions [23].

Many studies have investigated the relationship between meteorological conditions and airborne particles [24, 25]. A similar study on particles has been carried out in Seville [26] using single and multiple correlations. The relation between meteorology and particle size distribution has also been the subject of numerous recent and former works [27, 28], all of them using single correlations on particles. The relationship between rainfall and other meteorological parameters and particle size distribution of trace metals has been also studied in Seville [29]. These studies concluded that rainfall was the meteorological parameter that most favoured the elimination of particles and metals by means of physical and chemical washing. Fine particles are removed from the atmosphere independently of the amount of rain, whereas the elimination of coarse particles is dependent on rainfall values.

The complex composition of rainwater samples requires the use of chemometric tools for the extraction of information [30]. In the air pollution studies, Principal Component Analysis (PCA), Cluster Analysis (CA) or regression modelling have been applied to determine the possible influence and contribution of natural and anthropogenic factors in data trends [31]. PCA is a typical display method, which is able to reveal the 'hidden' structure of the data set; it tries to explain the influence of latent variables on the data [32].

However, due to the complexity of the sample matrix and the handicaps in identifying the source of the pollutants in rainwater, common chemometric tools are unsuitable. Therefore, the aim of this work is to test a new statistical technique that has previously been used in total suspended particles (TSP), both in total metal contents and in chemical speciation, called Sequential Principal Component Analysis (SPCA) [33]. In the present work we apply this technique to study the chemical composition of rainwater and the origins of their chemical compounds, including all the experimental information obtained and even the outlier samples with environmental significance.

**Experimental**

Multivariate statistical technique (SPCA)

On the basis of the results, basic and multivariate statistical analyses were applied to the analytical and meteorological data. Single Linear Regression (SLR) and Principal Components Analysis (PCA) were the statistical techniques employed. The modifications applied to the PCA technique led to the development of the new SPCA methodology. The main technique used was the PCA, and it was used in combination with the Single Linear Regression (SLR). The SLR correlation between the parameters constituted its exclusive use as a tool for graphical confirmation and the detection of outliers. For the PCA analyses the STATISTICA software package StatSoft, Inc. version 7, 2004 was used. The STATISTICA version 5.5, 1999 was used for the SLR to make the graphical study easier.

PCA is the most widely used technique in science and engineering [34]. The PCA was performed with a Varimax rotation, extracting the factors or PCs. This explains more than 5–10 % of the total variance of the data set, and together accounts for more than 75–85 % of accumulated variance. As concerns the SLR, correlation coefficients (r) were checked for the probability value (p). When the corresponding p values were greater than 0.05, a two-tailed t-test was applied with a 95% confidence level. If the null hypothesis is $r = 0$, then the two-tailed tests assess whether r was significantly different from zero.

The Sequential Principal Component Analysis (SPCA) presented in the current work was developed from the following considerations. Sometimes, when the samples called 'outlier' or 'interfering case', which have high concentrations of several elements, appear in the system studied they cause a distorting effect on the PCA result because of its least-squares property. In this case, the sample is distant from the other points in the measurement space and it can pull the corresponding PC towards itself and away from the direction of maximum variance. These samples contain highly concentrated metals that usually do not appear in any other PCs. Therefore just one interfering case can create an important distortion in the PCA. Normally, the solution found by many authors is simply to discard these cases and to repeat the PCA without them [35]. Usually, they suppose that these samples constitute a sample contamination or a general systematic error. Conversely, if this data is a product of special conditions one cannot remove it before making an interpretation about its meaning. It could be that high concentrations may be due to emission sources located nearby. We must identify the outliers and extract as much

information as possible because it could be a recurring process and it's important to know their indicators. These considerations were the reasons why the PCA was modified adding some rules to the performance of the technique, and this was then applied to the present data. The modifications consisted of an iterative method of sequentially performed PCA; interpreting the origin of samples that constitute outliers before there elimination.

During the SPCA procedure, the tool was interpreting initially only the variables grouped into the PC where the outlier had appeared (SPCA1). At the same time the presence of the outlier is confirmed graphically by the SLR representation of a pair of variables grouped into this PC. Likewise, only the analytical variables grouped into the PC1 were interpreted; assigning and characterizing a type of source. Then, after the interfering case is removed, a second PCA (SPCA2) is performed on the new matrix. The same procedure is repeated if another outlier appears again, until no more outliers are found. At that moment, all the PCs obtained in the last PCA can be interpreted. Thus, this method characterises all the sources in the last PCA besides the individual sources of the outliers.

The detection of interfering cases was carried out by applying the simple linear regression (SLR) technique. When an outlier appears in the system the 2D-scatterplot representation for concentrations of some pairs of variables graphically show at what point a falsely high correlation is induced. Usually in environmental systems an interfering case shows high concentrations of many variables, so there is no need to test the remaining variables to detect a common outlier point. Other graphical tools provide this capacity of detection such as Box & Whiskers plots, which have representations of outlier and extreme values [36]. The procedure consisted of verifying the linearity by observing the linear profile of the points, i.e., the pairs of data x-y. When a high correlation coefficient does not give a real linearity, the regression line is formed by an accumulation of points at one extreme and a lone point at the other extreme. In these cases, the correlation coefficients are due only to one sample, the outlier. Figure 1 shows high regression coefficients of 0.847 (Figure 1a) and 0.816 (Figure 1c) respectively when considering all samples for the pairs $Cu$-$SO_4^{2-}$ and $Mn$-$NO_3^-$. Nevertheless, we can check graphically that they are due to an agglomeration of data that give a false linearity with a discrepant case. After removing these cases, the ratio drops dramatically, tending to go to zero (0.195 (Figure 1b) and 0.380 (Figure 1d), respectively). We can see that these examples of high correlations are due to the anomalous cases. Therefore, the SLR tool is useful to confirm the presence of outliers before its elimination in the SPCA procedure.

Using this multivariate technique it was possible to simplify the interpretation for complex systems such as those related to the chemical composition of rainwater originating from airborne particles. Additionally, the SPCA reduces the large number of variables to a few principal components.

Methodology for chemical analysis

The chemical composition determined in rainwater samples was done by analyzing soluble metals and metalloids, cations and anions, and P and N nutrients. The general parameters pH, electrical conductivity (EC) and redox potential (E) were measured *in situ* during the collection of samples from the automatic sampler. Experimental pH values were converted to $H^+$ concentrations for the statistical treatment. The rH value gives information on reducing or oxidizing nature of the medium. The rH parameter was calculated from the redox potential ($E_{exp}$) and the pH from the following equation:

$$rH = \frac{E_h}{0.029} + 2pH = \frac{E_{exp} + c + 0.197}{0.029} + 2pH$$

Where *c* is a correction for redox potential and $E_h$ is the hydrogen electrode potential.

Heavy metals (Ti, V, Cr, Mn, Co, Ni, Cu, Zn, Pb, Cd, Mo, Ag), light metals (Be, Al, B, Si, Cs) and metalloids (Sb, As) determination was carried out by ICP-MS (Thermo Elemental X-7 Pierce Biotech., Rockland, IL, USA), according to the 200.8 US EPA Methods [37] with some modifications related to the tuning solution and mass calibration solution. Also, nutrients ($NO_3^-$, $NO_2^-$, $NH_4^+$, $PO_4^{3-}$), major and minor cations ($Li^+$, $Na^+$, $K^+$, $Ca^{2+}$, $Mg^{2+}$, $Sr^{2+}$, $Ba^{2+}$) and organic ($HCOO^-$, $CH_3COO^-$, $C_2O_4^{2-}$, $C_7H_6O_2^-$) and inorganic ($F^-$, $Cl^-$, $Br^-$, $I^-$, $SO_4^{2-}$) anions were determined. Determination of ionic compounds was performed by Ion Chromatography (IC) with a conductivity detector and a chemical suppression for anions (Metröhm 792 Basic IC, AG, Herisau, Switzerland). All determinations in the multi-element analysis were done in triplicate for each sample.

Regarding the ICP-MS technique, an adaptation established in the `Getting Started Guidelines for ICP-MS of Thermo Elemental Corp.´ was performed on the 200.8 US EPA Method for the preparation of the tuning and mass calibration solutions. The tuning solution was used for instrument regulation prior to analysis. This solution was prepared by mixing Be, Co, In, Ce, Ba and U (from Cromlab, Barcelona, Spain) in 2% (v/v) nitric acid Suprapure to produce a concentration of 10 µg L$^{-1}$ of each element. The mass calibration

solution was used prior to analysis to carry out a new mass calibration. This solution was prepared from a multielemental standard of Be, B, Al, Ti, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Sr, Mo, Ag, Cd, Sb, Cs, Ba, Tl, Pb, Th and U between 25 and 200 $\mu$g L$^{-1}$ in nitric acid Suprapure 2%. Some analytes were determined with values below detection limits (Ti, Ag, Be, Si, Cs, Li$^+$, C$_2$O$_4^{2-}$, C$_7$H$_6$O$_2^-$, Br$^-$ and I$^-$) and they were excluded from statistical analysis.

With regard to the Ion Chromatography instrument, a column METROSEP A Supp5-250 was used and it was protected by a Metröhm precolumn module. It was employed for the determination of the anions according to the following conditions: mobile phase 3.2 mM Na$_2$CO$_3$/1.0 mM NaHCO$_3$; flow rate 0.7 mL min$^{-1}$ and injection volume 100 $\mu$L. Determination of cations was performed using a column Cation METROSEP C 2 150, protected by a guardpack module, it was used in the following conditions: mobile phase 4 mM tartaric acid/0.75 mM pyridin-2,6- dicarboxilic acid; flow rate 1.0 mL min$^{-1}$; and injection volume 100 $\mu$L. Software "792 PC Software 1.0".

Acids and standard solutions for metals were from Merck (Darmstadt, Germany). Reagents for mobile phases and standard solutions for cations and anions were of an analytical grade from Panreac (Barcelona, Spain). Ultra-pure water came from Waters-Millipore (Milli-Q-grade, Model Plus). Regarding the plastic and glass material, they were cleaned with 0.2 M nitric acid for 24 hours. Care was taken in handling the samples in order to avoid contamination problems, and all reagents, materials and samples were handled within a vertical laminar airflow cabinet; this ensured air cleanliness standards of class 100 according to Federal Standard 209E. The vertical laminar airflow cabinet with a HEPA filter came from Indelab (Model IDL-48V).

Accuracy and precision of the applied methods was established by the use of international Certified Reference Material (CRMs) of rainwater for two levels of concentration: Rain-97 certifies low contents, and BCR-409 high contents. They were supplied by the National Water Research Institute of Canada (NWRI) and the Institute for Reference Materials and Measurements (IRMM). Tables 1 and 2 show calculated recovery values corresponding to the Rain-97 and BCR-409 CRMs. Experimental concentrations were obtained from 18 replicates of cation and anion determinations during six days. Fisher-Snedecor tests at $p < 0.05$ for the precision study demonstrated that experimental standard deviations were similar to the typical deviations of the CRMs for all ions determined. All the results obtained for the analyses of these materials represent optimum accuracy [38].

Measurement site and sampling

This study was undertaken in Seville (38°12'–36°51'N, 4°39'–6°32'W, 141.3 km$^2$, 8 m a.s.l.), the largest city in Southern Spain. Seville is located in the centre of the Guadalquivir Valley, which opens toward the ocean at the base of the triangle that the valley forms. The valley forms a triangle of about 300x300x100 km (about 15,000 km$^2$) that encompasses the cities of Huelva, Cadiz and Cordoba in the vertexes and Seville in the centre.

The city has a warm and dry Mediterranean climate with mean annual temperature of 18°C, rainfall 600 mm, atmospheric pressure 1014 hPa, relative humidity 65%, and wind speed 2 m s$^{-1}$. Seville is characterized by high temperatures and low wind speeds. Prevalent air currents come from the SW–NE direction. All this data is averaged from the 1961–2000 period. Predominant winds proceed from the Atlantic Ocean (southwest). Therefore, the situation of this sampling network in Seville represents the meteorological effects of a Mediterranean zone influenced by African winds as in other cities of Southern Europe.

Rainwater samples (wet depositions) were collected with an automatic wet-only sampler TISCH ENVIRONMENTAL 78 100 APS. This type of sampler is designed to collect atmospheric depositions in two separate containers: one of these opens only during episodes of rain, detected by an electronic sensor, and the other container is always open during dry periods. This allows the collection of both wet deposition (rainwater) and dry deposition (sedimentable particles). The results on rainwater samples collected during one year of sampling –from September 2006 to September 2007- are the ones presented in this current work.

Sample collection of rainwater was carried out in bottles of low density polyethylene (LDPE) of half-litre capacity located inside a refrigerator. The refrigerator power was optimized to achieve an average temperature between 4ºC and 6ºC.

To ensure the optimal preservation of samples, some modifications were carried out in the automatic collector. Firstly, a Teflon-coated funnel was inserted on the container where rainwater drops were collected. Secondly, a refrigerator system was placed under the sampler where the LDPE bottles collected rainwater samples. And thirdly, a waste system was connected to the funnel to prevent the bottle overflowing. All connections were made with Teflon tubings.

Meteorological data

Primary meteorological data was recorded. Meteorological data was provided daily by our local service of the National Institute of Meteorology (AEMET) in Seville. This data corresponds to the same periods of each sampling. These values for each sample correspond to the average hourly data provided throughout the sampling periods. Rainfall values were summed with the hourly data during the sampling period. Total rainfall (PP) in mm, wind direction (WD) in degrees, wind speed (WS) in m s$^{-1}$, ambient temperature (AT) in degrees Celsius, relative humidity (RH) in percentage and evapotranspiration (ET), were the meteorological variables used (Table 3).

**Results and discussion**

Multivariate statistical technique (SPCA)

In this study, the results from the SPCA were obtained for the 40 rainwater samples collected during the annual period 2006-07. According to the SPCA procedure mentioned earlier, the initial result is the first Principal Component Analysis (SPCA1, Table 4). The results show that 8 PCs are responsible for up to 84.6 % of the total variance, whereas PC1 accounts for 38.7 %. However, the PC1 is represented by only one case with an extremely high value of scores (se34). This sample corresponds to an episode with very high concentrations of many variables measured ($NH_4^+$, $F^-$, $SO_4^{2-}$, $NO_3^-$, V, $Cl^-$, $Ca^{2+}$, $K^+$, Co, Cu, Mn, Al, $Mg^{2+}$, $Ba^{2+}$, B),  as well as the highest EC and pH values of 147.3 µS cm$^{-1}$ and pH 7.7 respectively. These variables originated mainly from resuspended industrial soil from a fertilizer factory. The high concentrations were obtained probably due to the short rain episode that occurred in March after a long dry period characterized by a low value of rainfall (low rainwater volume collected).

PC2 explicated 12.0% of the variance, and it is also represented by one interfering case with a high weight (se04). In this case, the sample corresponds to an episode with high concentrations of variables (Mo, Zn, B, Cd, $NO_3^-$, Al, Cu) originating from urban traffic sources. In addition this sample was characterized by its extremely low pH value as an acid rain episode. Nevertheless, the sum of PC1 and PC2 account for 51 % of the variance of all cases, in other words, a half of all samples are represented by only two cases. The rest of the samples in bold in Table 4 are samples confirmed also as outliers by SLR (Figure

2). Consequently, this fact proves that the first PCA is not valid to explain the variance and main sources of the forty samples of rainwater studied. Therefore, according the present methodology of SPCA described earlier, it is necessary to remove the outliers and repeat the PCA sequentially until no more outliers appear. The results are shown in Table 5 where SPCA1 to SPCA8 only show the PCs related to straggler observations and the final stage corresponds to SPCA9 where outliers were removed. It can be noted that in Table 4 and Table 5 the first PC coincides in both cases, but the rest of PCs differ because the sequentially extreme cases were removed in Table 5. The above mentioned groups can be observed graphically in a score plot (Figure 3).

In SPCA1, sample se34 exhibits the highest concentrations practically in all the analysed elements. Apart from the characteristics mentioned above, the rain event also shows the highest values of atmospheric pressure and evapotranspiration with southern winds of low speeds. Thus, this sample shows the composition of rain taken in March 2007; it is of strong basic behaviour during a dry and anticyclonic weather period prior to spring rains.

After removing se34 and repeating SPCA (SPCA2) the result showed that the sample se53 is also an outlier. This is because it is representative of the principal component PC1 accounting for 29.1% of total variance and confirmed also by SLR (Figure 2). This sample is related to a marine influence, because its characteristics are high electrical conductivity, 81.5 $\mu S\ cm^{-1}$, and high values of $SO_4^{2-}$, $Mg^{2+}$, $Cl^-$, $K^+$, $F^-$ and $Ca^{2+}$ [39, 40]. This case corresponds to isolated rains at the beginning of June before the summer season but with southern winds at high speeds after a dry period of 20 days.

This sample is related to a marine influence, because its characteristics are high electrical conductivity, 81.5 $\mu S\ cm^{-1}$, and high values of $SO_4^{2-}$, $Mg^{2+}$, $Cl^-$, $K^+$, $F^-$ and $Ca^{2+}$ [39, 40]. This case corresponds to isolated rains at the beginning of June before the summer season but with southern winds at high speeds after a dry period of 20 days.

In SPCA4, the case se55 is the only representative case (Figure 2) of the principal component PC2, consisting of $Sr^{2+}$ and $Ba^{2+}$ from earth mixed with Mo, Pb and Sb from traffic sources: gasoline and vehicle brakes. This source represents the traffic emissions deposited formally in the surrounding lands of the city and latterly it is resuspended in the air. This sample was also related to a long dry period, but in this case it deals with a summer storm with high humidity and temperatures, and a scarce amount of precipitation (3 L $m^{-2}$ against 15-24 L $m^{-2}$ for the remainder).

For SPCA5 the most important elements found for the PC3 were mainly Sb, Mo Co and Cd, from an anthropogenic source [41], attributed to traffic emissions. In this case the outlier was the case se06, collected in very similar meteorological conditions to sample se04. The only difference being it was the second rain event of September 2006, only after a week from the se04 episode.

With respect to SPCA6, the straggled sample was found for se50 in PC1, explaining 23.5 % of the total variance. It was collected when it rained after a long dry period of 20 days in May 2007. The PC1 is characterized by an electrical conductivity of 51.1 $\mu$S cm$^{-1}$, and by high values of $NO_3^-$, $Ca^{2+}$, $NH_4^+$, $PO_4^{3-}$, $SO_4^{2-}$ coming from the contamination of soil by a fertilizer plant.

In the SPCA7, PC2 explains a group of variables represented only by the se45 case. This outlier originated from the resuspension of soils contaminated by the fertilizer factory. This sample collected in April 2007 featured a very slight precipitation (4 L m$^{-2}$) and a low wind speed. The group of variables was scarcely different from the previous resuspension sources; in soils contaminated in SPCA1 and SPCA6. The common variables that were most representative were $NH_4^+$, $SO_4^{2-}$ and $Ca^{2+}$. Meanwhile, in SPCA7 the common variables that were most representative were $K^+$, $F^-$, $Ca^{2+}$ and $PO_4^{3-}$.

For SPCA8 we found the interfering case se47 collected in May and it was characterized by the highest values of rainfall of 85 L m$^{-2}$, and a low value for electrical conductivity of 7.7 $\mu$S cm$^{-1}$. The variables grouped Cu, $F^-$, V, Co, $K^+$, $PO_4^{3-}$, $SO_4^{2-}$, $Ca^{2+}$ and Cd showed a mixture of origin from both anthropogenic and natural sources.

Once all these outliers were removed during the SPCA performance, the last PCA (SPCA9) could not reveal the presence of any discrepant PC. Therefore, we found ten significant PCs that explicated up to 84.6% total variance in SPCA9.

PC1 explains 23.8 % of variance and it represented the influence of the resuspension of industrial soil from the fertilizer plant, featured mainly by $NO_2^-$ and $NH_4^+$, and additionally $NO_3^-$, $SO_4^{2-}$, $F^-$, Al, $K^+$, Mn, Sb and $Ca^{2+}$. This type of source is represented by samples collected in winter as well as in the spring seasons, indistinctly [42]. PC2 is formed by the elements from marine aerosols characterized mainly by $Na^+$, $Cl^-$ and $Mg^{2+}$, and also by $SO_4^{2-}$ and $K^+$. The representative cases corresponded to rainwater samples collected during winter days with winds coming from southern directions, with moderate wind speeds and extremely high values of air humidity. As in PC1, PC3 is related to urban earths characterised mainly by the elements $Ba^{2+}$ and Pb and also by, $Sr^{2+}$, Sb and Mo [43]. The pair of $Ba^{2+}$-$Sr^{2+}$ is frequently attributed to a terrestrial source, although $Ba^{2+}$ is also

attributed to vehicle traffic [44]. This type of source already mentioned above was represented by samples collected in days with extremely high temperatures in September 2007. PC4 and PC7 correspond to combustion exhausts from industries or mainly from the urban traffic, featured by Co, Sb, As, Mo, $NO_3^-$ and $SO_4^{2-}$ and mainly by Ni and V. Samples represented in this PC were collected in days with low values of air humidity. In PC5 typical emissions were recognized to be coming from vehicle brakes, they were mainly the elements Cu, As and Sb [45] and also Mo, and the samples were collected during days with high values of air humidity. PC6 is attributed to vegetation emissions characterized by the presence of formate [46, 47] in episodes during November 2006 that had high rainfall values. PC8 shows another indicator of industrial emissions from the fertilizer factory of the city, which are characterized mainly by the presence of $PO_4^{3-}$ , but also by $NO_3^-$, $F^-$ and $K^+$. These emissions come from the factory through its industrial processes to manufacture the solid fertilizers such as potash ($K_2O$), phosphates $Ca_5(PO_4)_3F$, mineral rocks as raw materials, and nitric and hydrofluoric acids. PC9 is related to industrial emission from foundries with Cd, Zn, Co, Sb and Mn [33]. PC10 represents variables grouped in a similar way to SPCA7, indicating emissions coming from resuspension soils where $F^-$, $Ca^{2+}$ and also $PO_4^{2-}$ were the best indicators due to the fluorapatite mineral used in the fertilizer plant.

This SPCA shows that the main sources of soluble atmospheric pollutants in rainwater samples were the urban vehicle traffic emissions and the particles originating from industrial soils and lands which were resuspended in the air over the city.


**Conclusions**


In this work results concerning the rainwater composition for a one year period obtained in Seville, SW Spain, have been statistically studied by multivariate techniques. The aim was to resolve the complex origin of the chemical composition. The capability of a new approach was tested and it was based on the utilization of a chemometric technique (PCA) that was modified for its use on the complete data set (SPCA). The modification consisted of sequentially performing the PCA and interpreting the origin of outlier samples. This rendered a more complete set of information than if the outliers were discarded. This allowed the different outliers to be detected and the suitable information to be extracted.

The SPCA that was performed enabled the extraction of 8 interfering PCs from the atmospheric information corresponding to different types of special meteorological conditions less. These episodes can be recurring: acid rainfalls from urban traffic ($NO_3^-$, B,

Zn, Mo), industrial soils (EC, $NH_4^+$, $F^-$, $SO_4^{2-}$, V, $Ca^{2+}$, $K^+$), marine aerosols ($SO_4^{2-}$, $Cl^-$, $Mg^{2+}$, $K^+$, $Na^+$) and residues from traffic deposited on land ($Ba^{2+}$, $Sr^{2+}$, Mo, Pb, Sb). Additionally, the final PCA (SPCA9) was able to explain almost 85 % of the total variance of the complete system. It identified soluble atmospheric pollutants by a group of indicators such as: the main pollution source from contaminated industrial soils from a fertilizer factory ($NO_2^-$, $NH_4^+$, $NO_3^-$, $SO_4^{2-}$, $F^-$, Al, $K^+$, Mn, Sb and $Ca^{2+}$), another important source from marine aerosols ($Na^+$, $Mg^{2+}$, $K^+$, $Cl^-$, $SO_4^{2-}$), the pollutants originating from traffic, the resuspended particles from the land containing traffic residues ($Ba^{2+}$, Pb, $Sr^{2+}$, Sb, Mo), the combustion processes from industries and traffic (Ni, V, Co, $NO_3^-$, Sb, As, Mo, $SO_4^{2-}$, Zn), and the vehicle braking processes (Cu, As, Sb, Mo, Co). SPCA was also able to detect the minor sources coming from some metal foundries of the city (Cd, Zn, Co, Sb, Mn). Consequently, SPCA9 has accounted for some sources that have been already detected in the interfering PCs, but at more normal levels as opposed to the outlier high levels.

The statistical technique demonstrated that the outliers found were characterized by very high concentrations in many of the analytes measured. It was established that various meteorological factors caused these unusual levels. Almost all interfering samples corresponded to rain events where the wind direction was from the south/southwest and when the wind speed was quite low. This favoured the accumulation and concentration of atmospheric suspended particles in the air. In addition, most samples were collected after a long dry period; for example the first rains in the autumn. In addition, the amount of precipitation was also very low in almost all interfering cases. Therefore, it is supposed that particulate matter in the atmosphere was solubilised in a small volume of precipitation thus causing higher concentrations of pollutants in the rainwater.

In brief, it is concluded that the interfering cases originated from short rain episodes. These cases had high atmospheric particle content due to the accumulation of a lot particles in the atmosphere over a long period of time combined with low volumes of water. The application of the SPCA was proved to be a useful tool to identify all complete information on rainwater samples as indicators of urban air pollution in a city influenced mainly by traffic vehicles and resuspended polluted soils.

Based on the results of this preliminary study, there is a need for a longer investigation that would enable a time data record. Establishing a new rainwater quality network in SW Spain is a preliminary step in obtaining a historic knowledge about the chemical composition of rainwater in this area. Systematic measurements of rainwater quality are necessary in order to explore the relationships between pollutants and their principal sources as well as to validate specific deposition models for SW Spain.

**References**

1. Eriksson E (1952) Tellus 4:215–232

2. Eriksson E (1952) Tellus 4:280, 303

3. Vázquez A, Costoya M, Peña RM, García S, Herrero C (2003) Chemosphere 51:375, 386

4. Weathers K C, Lovett GM, Likens GE (1995) Atmos Environ 29:665, 72

5. Shih, Chieh Hsu, George T.F. Wong, Gwo, Ching Gong (2010) Mar Chem 120:116–127

6. Báez, A., Belmont, R., García, R., Padilla,H., Torres, M.C. (2007) Atmos Res 86, 61–75

7. Arsene C, Olariu RJ, Mihalopoulos N (2007) Atmos Environ 41:9452, 9467

8. Celle, Jeanton H, Travi Y, Loÿe, Pilot MD, Huneau F, Bertrand G (2009) Atmos Res 91:118, 126

9. Wei Hong, Wang Jian, li, Li Xu, guang (2005) Xinan Shifan Daxue Xuebao, Ziran Kexueban 30:725, 729

10. Momin GA, Ali K, Rao PSP, Safai PD, Chate DM, Praveen PS, Rodhe H, Granat L (2005) J Geo Res, [Atmospheres], 110:D08302/1, D08302/10

11. Alastuey A, Querol X, Chaves A, Ruiz, Carratala A, Lopez, Soler A (1999) Environ Poll 106:359, 367

12. Hontoria C, Saa A, Almorox J, Cuadra L, Sánchez A, Gasco JM (2003) Water, Air, and Soil Poll, 146:35, 54

13. Morselli L, Olivieri P, Brusori B, Passarini F (2003) Environ Poll 124:457–469

14. Juan M, González Benítez J, Neil Cape, Mathew R (2009) Atmos Environ 43:4087–4094

15. Demirak A (2007) Environ Mon Ass 129:189, 196

16. Alastuey A, Querol X, Chaves A (1999) Environ Poll 106:359, 367

17. Zheng Yan, ming, Zhong Wei, Peng Xiao, ying, Xue Ji, bin, Ma Qiao, hong, Cai Ying (2009) Huanan Shifan Daxue Xuebao, Ziran Kexueban 1:111, 115

18. Flues M (2002) Atmos Environ 36:2397, 2404

19. Pelicho AF, Martins LD, Nomi SN, Solci MC (2006) Atmos Environ 40:6827, 6835

20. Uygur N, Karaca F, Alagha O (2010) Atmos Res 95:55–64

21. Williams PT, Radojevic M, Clarke AG (1988) Atmos Environ 22:1433, 1442

22. Spanos T, Simeonov V, Andreev G (2002) Talanta 58:367–375

23. Nicolás JF, Yubero E, Pastor C, Crespo J, Carratalá A (2009) Atmos Res 94:330–337

24. Gimenez J, Pastor C, Castaner R, Nicolas J, Crespo J, Carratala A (2010) Atmos Environ 44:338, 346

25. Gonzalez E, Villasenor R, Gasca J, Lopez T, Mugica V (2002) Modeling of regional transport and dispersion of PM10 in central Mexico. Proceedings of the Air & Waste Management Association's. Annual Conference & Exhibition, 95th, Baltimore, MD, United States, June 23, 27

26. Melgarejo PL, Ternero M, Gracia I (1986) Int J Environ Anal Chem 24:283, 95

27. Despiau S, Cougnenc S, Resch F (1996) J Aerosol Sci 27:403, 415

28. Vakeva M, Hameri K, Puhakka T, Nilsson ED, Hohti H, Makela JM (2000) J Geo Res, [Atmospheres] 105:9807, 9821

29. Fernández AJ, Ternero M (2004) Idojaras 108:11, 32

30. Zhang P, Dudley N, Ure AM, Littlejohn D (1992) Anal Chim Acta 258:1, 10

31. Samara C, Kouimtzis T, Katsoulos GA (1994) Toxicol and Environ Chem 44:147, 60

32. Vanderginste B, Massart DL, Buydens L, De Jong S, Lewi P, Verbeke JS (1998) Handbook of Chemometrics and Qualimetrics. Elsevier

33. Fernández AJ, Ternero M, Fernández F (2004) Atmos Environ 38:873–886

34. Polkowska Z, Astel A, Walna B, Malek S, Medrzycka K, Gorecki T, Siepak J, Namiesnik J (2005) Atmos Environ 39:837, 855

35. Prendes P, Andrade JM, Lopez, Mahia P, Prada D (1999) Talanta 49:165, 178

36. Bouza Deaño R, Ternero Rodríguez M, Fernández Espinosa AJ (2008) Journal of Hydrology 361: 227-239

37. USEPA, US Environmental Protection Agency, 1994. Method 200.8. Determination of trace elements in waters and wastes by Inductively Coupled Plasma-Mass Spectrometry. Environmental Monitoring Systems Laboratory Office of Research and Development

38. AOAC, Association of Analytical Communities, 2010. Guidelines for Single Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals.

39. Panyakapo M, Onchang R (2008) J Environ Sci (Beijing, China) 20:441, 448

40. Cerón RMB, Padilla HG, Belmont RD, Torres MCB, García RM, Báez AP (2002) Atmos Environ 36:2367, 2374

41. Patriarca M, Menditto A, Rossi B, Lyon TDB, Fell GS (2000) Microchem J (2000) 67: 351, 361

42. Al, Momani IF (2003) Atmos Environ 37:4507–4515

43. Halstead MJR, Cunninghame RG, Hunter KA (2000) Atmos Environ 34 :665, 676

44. Fernández AJ, Ternero M (2004) Anal Bioanal Chem 379: 684–699

45. Fernández AJ, Ternero M, Barragán FJ, Jiménez JC (2000) Chemosphere , Global Change Science 2:123, 136

46. Peña RM, García S, Herrero C (2002) Atmos Environ 36:5277, 5288

47. Brooks A, Kieber RJ, Witt M, Willey JD (2006) Atmos Environ 40:1683–1693

**Table 1** Results on measured cation and anion levels in the BCR409 reference material

| Ion | Concentration obtained (mg L$^{-1}$) conc ± sd (n=18) | CRM concentration (mg L$^{-1}$) conc ± sd (N) | Recovery (%) | AOAC criteria[*] (%) |
|---|---|---|---|---|
| Na$^+$ | 1.850 ± 0.035 | 1.906 ± 0.028 (18) | 97.1 | 80-115 |
| NH$_4^+$ | 1.856 ± 0.044 | 1.591 ± 0.030 (14) | 116.6 | 80-115 |
| K$^+$ | 0.186 ± 0.011 | 0.166 ± 0.009 (12) | 112.0 | 75-120 |
| Ca$^{2+}$ | 0.561 ± 0.017 | 0.621 ± 0.012 (18) | 90.3 | 75-120 |
| Mg$^{2+}$ | 0.306 ± 0.035 | 0.299 ± 0.049 (18) | 102.3 | 75-120 |
| Cl$^-$ | 3.966 ± 0.083 | 4.006 ± 0.070 (20) | 99.0 | 80-115 |
| NO$_3^-$ | 4.859 ± 0.088 | 4.842 ± 0.062 (19) | 100.3 | 80-115 |
| SO$_4^{2-}$ | 5.092 ± 0.066 | 5.110 ± 0.067 (20) | 99.6 | 80-115 |

* AOAC (2010) [38]

**Table 2** Results on measured cation and anion levels in the RAIN97 reference material

| Ion | Concentration obtained (mg $L^{-1}$) conc ± sd (n=18) | CRM concentration (mg $L^{-1}$) conc ± sd (N=52) | Recovery (%) | AOAC criteria[*] (%) |
|---|---|---|---|---|
| $Na^+$ | 0.239 ± 0.034 | 0.298 ± 0.029 | 80.2 | 75-120 |
| $NH_4^+$ | 0.179 ± 0.030 | 0.186 ± 0.024 | 96.2 | 75-120 |
| $K^+$ | 0.146 ± 0.037 | 0.166 ± 0.026 | 88.0 | 75-120 |
| $Ca^{2+}$ | 2.733 ± 0.225 | 2.860 ± 0.203 | 95.6 | 80-115 |
| $Mg^{2+}$ | 0.909 ± 0.086 | 1.000 ± 0.078 | 90.9 | 75-120 |
| $Cl^-$ | 0.542 ± 0.067 | 0.574 ± 0.058 | 94.1 | 75-120 |
| $NO_3^- + NO_2^-$ | 10.373 ± 0.134 | 2.220 ± 0.120 | 81.5 | 80-115 |
| $SO_4^{2-}$ | 1.210 ± 0.114 | 5.680 ± 0.350 | 83.8 | 75-120 |

* AOAC (2010) [38]

**Table 3** Values of the meteorological variables corresponding to the sampling period 13-09-2006 to 12-09-2007

| Variable | Mean | Range (min-max) | Units | Quartile (Q1–Q3) |
|---|---|---|---|---|
| PP | 18.2 | 2.0–85.0 | mm | 6.8 – 20.8 |
| WS | 9.3 | 1.4–21.9 | m s$^{-1}$ | 5.1 – 12.8 |
| WD | 183º (S) | 106º-266º (E–W) | degree | 157º-202º (SE – S) |
| AT | 15.5 | 8.2–23.7 | °C | 12.0 – 18.5 |
| AH | 83.7 | 65.8–98.8 | % | 75.9 – 92.1 |
| AP | 1010.2 | 1001.0–1019.6 | hPa | 1007.1 – 1012.7 |
| ET | 24.7 | 4.0–58.0 | mm/day | 13.8 – 31.5 |

**Table 4** Result of the first Principal Component Analysis (SPCA1). Matrix including outliers (34 variables x 40 cases)

| SPCAn | PCn | Eigen. | % Variance/ % Cum. Var. (Rel. %) | Sign | Variables* | Cases | Source |
|---|---|---|---|---|---|---|---|
| SPCA1 | PC1 | 13.1 | 38.7 / 38.7 (45.7) | + | NH$_4$, Ca, F, K, SO$_4$, Co, V / EC, <u>Cu</u>, Mn / <u>Al</u> / <u>Mg</u>, NO$_3$ // Ba, Cl, B | **se34** | highest concentrations-mix sources |
| | PC2 | 4.1 | 12.0 / 50.7 (14.2) | + | <u>Mo</u>, Zn / B, Cd / <u>NO$_3$</u>, Al / Cu | **se04** | acid rain (pH=4) |
| | | | | - | H | se57, se21 | basic pH |
| | PC3 | 3.3 | 9.6 / 60.3 (11.3) | + | rH, <u>PP</u> / HCOO / Cr | **se47** // se20, se08, se22 / se08 | highest rainfall |
| | PC4 | 2.5 | 7.4 / 67.7 (8.7) | + | Sr, Pb, Ba // <u>Sb</u> / Mo | **se55** // se58, se57 | mix terrestrial-traffic |
| | PC5 | 1.8 | 5.2 / 72.9 (6.1) | + | Na, Cl / Mg / <u>PO$_4$</u>, CH$_3$COO / EC | **se53** // se10, se26, se22 | marine |
| | PC6 | 1.5 | 4.5 / 77.4 (5.3) | + | E | se23**,** se21, se24 / se 31 | oxidizing |
| | | | | - | <u>H</u>, As | se35 / se45, se44 | reducing |
| | PC7 | 1.3 | 3.8 / 81.2 (4.5) | + | NO$_2$ // Sb / Ni | **se06** // se18 / se09 / se28, se26 | traffic combustion |
| | | | | - | PO$_4$ | **se50** | fertilizers |
| | PC8 | 1.1 | 3.4 / 84.6 (4.0) | + | <u>HCOO</u>, <u>CH$_3$COO</u> // <u>Cr</u> | **se45 //** se20, se10 | vegetation |
| | | | | - | <u>Ni</u> | - se09 / se38, se58, se41, se08 | industrial combustión |

\* : Variables are in decreasing order of loading > 0.5
__ : Variable with the highest loading between PCs
/, //: Degree of differences in the loading. Ex: 0.8/0.7  0.9//0.6
**Bold**: Cases with high scores, outliers

**Table 5** Sequential Principal Component Analysis (SPCA) of 9 iterations. SPCA 1-8: outliers. SPCA 9: final PCA

| SPCAn | PCn | % Variances* | Sign | Variables | Cases | Origin |
|---|---|---|---|---|---|---|
| SPCA1 | PC1 | 37.4 / 86.9 (43.0) | + | $NH_4$, Ca, F, K, $SO_4$, Co, V / EC, Cu, Mn / Al / Mg, $NO_3$ // Ba, Cl, B | *se34* | highest concentrations. resuspended industrial soil |
| SPCA2 | PC1 | 29.1 / 87.8 (33.1) | + | $SO_4$ / Mg, Cl, K, Na / F, Ca / EC | *se53* | marine aerosol |
| SPCA3 | PC1 | 26.4 / 90.1 (29.3) | + | $NO_3$, B, Zn / Mo, Al, Mn / Cu / Cd / EC | *se04* | acid rain from traffic |
| SPCA4 | PC2 | 13.0 / 85.4 (15.2) | + | Ba, Sr / Mo, Pb / Sb // EC | *se55* | resuspended lands from traffic |
| SPCA5 | PC3 | 9.6 / 85.4 (11.2) | + | Sb, Mo, Co, Cd // B, Zn | *se06* | urban traffic |
| SPCA6 | PC1 | 23.5 / 87.6 (26.8) | + | $NO_3$ / Ca / $NH_4$, $PO_4$ / $SO_4$, EC // V, F, K | *se50* | resuspended industrial soil |
| SPCA7 | PC2 | 13.4 / 85.7 (15.6) | + | K, $PO_4$, F / Ca, HCOO, $SO_4$, $CH_3COO$ // Cu, $NO_3$ | *se45* | resuspended soil |
| SPCA8 | PC1 | 26.1 / 86.2 (30.3) | + | Cu, F, V, Co, K / rH, $PO_4$ / $SO_4$, PP, Ca, Cd // EC, Zn | *se47* | highest rainfall. mix source |
| SPCA9 | PC1 | 23.8 / 23.8 (28.1) | + | $NO_2$, $NH_4$ / <u>Al</u>, <u>K</u>, <u>$NO_3$</u>, <u>Mn</u>, <u>Sb</u> / $SO_4$, F / Ca, <u>EC</u> | se18 / se26 / se28, se39, se41, se20 | resuspension of industrial soil |
| | PC2 | 12.7 / 36.5 (15.0) | + | Na, Cl, Mg / EC, <u>$SO_4$</u> / K | se22, se26 / se10 / se14 | marine aerosol |
| | PC3 | 10.2 / 46.7 (12.1) | + | Ba, Pb, Sr // E, Sb / Mo | se58, se57 | resuspended lands from traffic |
| | PC4 | 7.4 / 54.1 (8.7) | + | Ni // <u>Co</u> / Sb | se09, se08 / se18 | combustions |
| | PC5 | 6.7 / 60.8 (7.9) | + | Cu / <u>H</u>, <u>As</u> // Sb, Mo | se35 / se57, se38 | urban traffic, brakes. reducing |
| | | | - | <u>E</u> // Al, Co, B, rH | se23, se21 | oxidizing rainwaters |
| | PC6 | 6.5 / 67.3 (7.7) | + | HCOO // Cr, <u>rH</u>, <u>B</u>, <u>PP</u>, Mn, Al | se20, se22 / se08 | high rainfalls. vegetation, terrestrial elements |
| | PC7 | 5.8 / 73.1 (6.9) | + | <u>V</u> // $NO_3$, As // EC, Mo, $SO_4$ | se32, se39, se35, se17, se44, se23 | combustions |
| | PC8 | 4.7 / 77.8 (5.6) | + | $PO_4$ // $NO_3$, K, F | se41, se43, se38, se39, se58, se10 | resuspensión of industrial soil, fertilizer factory |
| | | | - | <u>Mo</u> // As, Co | se28, se35, se36 / se23 | urban traffic, brakes |
| | PC9 | 3.9 / 81.7 (4.6) | + | Cd / <u>Zn</u> / Co // Sb, Mn | se33 / se07, se25, se38, se44, se26 | emissions from foundries |
| | PC10 | 2.9 / 84.6 (3.4) | + | <u>F</u>, <u>Ca</u> / B // H, EC, $PO_4$ | se38, se22, se36, se41, se58 | resuspended soil |
| | | | - | $CH_3COO$ // E, PP, As, Cr | se10 / se15, se12, se44, se20 | resuspended soil |

\* : In SPCA1 to 8 the column represent: Variance of PCi / total variance (percentage of variance respect total variance); in SPCA9 the column represent:
Variance of PCi / cumulated variance (percentage of variance respect total variance)

**Fig. 1** False high r coefficients in a) and c). Without outliers in b) and d) r trends to zero
**Fig. 2** Examples of correlations in outlier components (Tables 1 and 2)
**Fig. 3** Plot of loadings, PC1 vs. PC2 corresponding to SPCA9

CE vs. NH4 (n=40)
CE = 13.170 + 0.034 * NH4
Correlation: r = 0.937

SO4 vs. CE (n=39)
CE = 3.138 + 0.013 * SO4
Correlation: r = 0.840

NO3 vs. CE (n=38)
CE = 17.572 + 0.004 * NO3
Correlation: r = 0.679

Ba vs. Sr (n=37)
Sr = -13.640 + 1.479 * Ba
Correlation: r = 0.967

a) SO4 vs. Cu (n=24)
Cu = -1.373 + 0.003 * SO4
Correlation: r = 0.847

b) SO4 vs. Cu (n=18)
Cu = 1.332 + 0.000 * SO4
Correlation: r = 0.195

c) NO3 vs. Mn (n=32)
Mn = 0.474 + 0.001 * NO3
Correlation: r = 0.816

d) NO3 vs. Mn (n=29)
Mn = 0.416 + 0.001 * NO3
Correlation: r = 0.380