Available online at www.sciencedirect.com
SCIENCE DIRECT®
Environmental Modelling & Software 21 (2006) 61–68
Environmental Modelling & Software
www.elsevier.com/locate/envsoft

Study of the evolution of air pollution in Salamanca (Spain) along a five-year period (1994–1998) using HJ-Biplot simultaneous representation analysis

J.M. González Cabrera[a], M.R. Fidalgo Martínez[b], E.J. Martín Mateos[a,*], S. Vicente Tavera[c]

[a]Departamento de Química Analítica, Nutrición y Bromatología, University of Salamanca, E-37008 Salamanca, Spain
[b]Departamento de Física General y de la Atmósfera, University of Salamanca, E-37008 Salamanca, Spain
[c]Departamento de Estadística, University of Salamanca, E-37008 Salamanca, Spain

Received 4 April 2003; received in revised form 24 September 2004; accepted 22 October 2004
Available online 11 January 2005

Abstract

Five-year data of SO_2, PM10, NO, NO_2, O_3 and CO levels recorded at two air pollution monitoring stations in the city of Salamanca were analysed using a HJ-Biplot Simultaneous Representation analysis method in order to determine the possible relationships among the different pollutants and to gain insight into the temporal evolution of pollution in this city. Use of this statistical method proved to be very useful for such purposes. Current trends in the evolution of atmospheric pollution in Salamanca are shown, pointing to an important decrease in SO_2 and PM10, and an increase in NO_x and photochemical pollution, as indicated by O_3 levels. Additionally, the increase in photochemical pollution has meant that this is more preponderant in the summer. In general, at the start of the study period pollution was mainly characterised by pollutants typically found in winter, thereafter evolving towards pollution characterised by nitrogen oxides and ozone towards the end of the period studied.

Keywords: Air pollution; Urban pollution; HJ-Biplot simultaneous representation analysis

1. Introduction

Study of the relationships among air pollutants, in which a large number of experimental data are involved, is very complicated. Multivariate Data Analysis, Correspondence Analysis (Greenacre, 1984), and Principal Component Analysis (Flocchini and Myrup, 1990; Statheropoulos et al., 1998; Ionescu et al., 2000; Yu and Chang, 2000; Stella et al., 2002; Park et al., 2002; Sánchez-Ccoyllo and Andrade, 2002) are just some of the statistical techniques used to handle and analyse huge amounts of such data. On the other hand, modelling pollutants concentrations is other way to evaluate the evolution of the air pollution; by using diverse statistical tools (e.g. Hurley et al., 2005; Nunnari et al., 2004; Lim et al., in press), models are designed to study the urban pollution or to prognost future evolution or to develop systems to assess in this scientific field.

In the present study, five-year data on SO_2, PM10, NO, NO_2, O_3 and CO levels recorded at two air pollution monitoring stations situated in the city of Salamanca were analysed by using the HJ-Biplot simultaneous representation analysis. The aim of this was to uncover possible relationships among the pollutants and to attempt to reach conclusions about the temporal evolution of pollution in the city.

* Corresponding author. Fax: +34 923 294483.
E-mail address: ejmm@usal.es (E.J. Martín Mateos).

1364-8152/$ - see front matter
doi:10.1016/j.envsoft.2004.10.009

Most of the city of Salamanca is situated on the right bank of the River Tormes and is spread over almost 50 km^2. The standard coordinates of the city are $5° 41'$ W $40° 58'$ N, at an altitude of 808 m a.s.l. with a maximum difference in altitude of 40 m. In view of its geographic situation, the characteristics of the surrounding terrain, and its altitude, the climate of Salamanca is continental, due to the proximity of the Atlantic Ocean, although this is attenuated by the relief. The climate is dry, the warmest month being July, with a mean temperature of 21.9 °C, and the coldest one January with 3.4 °C. Thus, the climate of the city can be characterised as being between moderate and extreme in accordance with the climatic classification based on temperature (and given that the difference between the mean temperature of the hottest month and that of the coldest month is almost 20 °C).

Precipitation mean value in the city is around 400 mm $year^{-1}$. April and May show the highest levels, followed by October, while the driest months are usually August and November, followed in this sense by July and March. In general winds are weak; the highest annual and monthly frequencies of wind direction correspond to the SW, followed by the NE and W.

The population of the city, including the metropolitan zone, is about 250,000 inhabitants. Salamanca has a low level of industrialization; atmospheric pollution is of clearly urban nature, the main sources of emission being combustion (domestic heating systems and traffic).

Fig. 1. Location for air quality monitoring stations in the city of Salamanca. E.1: *Avenida de Portugal*; E.2: *Paseo de la Marina*; E.3: *Paseo del Líbano*; E.4: *Plaza del Maestro Turina.*

2. Experimental

2.1. Sampling site

The automatic network for the monitoring and control of the atmospheric pollution in the city of Salamanca comprises four monitoring stations (Fig. 1). Their basic characteristics are as follows: E.1 is located on the *Avenida de Portugal*, a mostly residential area with a heavy traffic density in the mornings and evenings; E.2 is on the *Paseo de la Marina*, a peri-urban zone with a lower density of buildings and traffic and close to a public garden; E.3 is located on the *Paseo del Líbano*, an open zone with a high traffic density and also close to a public garden; E.4 is in the *Plaza del Maestro Turina*, an urban zone with a low density of buildings, close to a light industry estate and to several of the main access roads to the city.

2.2. Data acquisition

To carry out this work, we relied on the data collected at stations E.1 and E.2, these being the first ones installed in the city. The pollutants measured were SO_2, PM10, NO, NO_2 and CO, at E.1, and SO_2, PM10, NO, NO_2 and O_3 at E.2. Analytical determinations were carried out on SO_2 (fluorescence UV), PM10 (attenuation of β radiation), NO_x (NO and NO_2) (chemiluminescence), CO (IR absorption spectrometry), O_3 (UV absorption spectrometry). Values were collected at 15-min intervals.

2.3. Methods

Having observed in an previous study (González Cabrera, 2002) that the type of day of the week (weekday versus weekend) strongly affects pollutant concentrations, we established weekly reference periods, for which pollutant concentrations show a certain trend, with more generalised decreases at the weekend and a return to higher levels during the week. Thus, starting from 15-min values for the period from the second week of March 1994 (when the E.2 station was put into operation) to the last week of December of 1998, the mean weekly values were obtained. This allowed us to avoid the use of daily values, because the main objective was to interpret the evolution of air quality over a relatively long period (several years). In addition, the use of weekly values avoids the generation of complex plots that would hinder the interpretation of results.

HJ-Biplot is a statistical method, which allows simultaneous representation of all the variables considered. "*A Biplot is a graphic plot of a matrix* $\boldsymbol{X}$ $(n \times p)$ *by markers* $\boldsymbol{g}_1,\ldots,\boldsymbol{g}_n$ *for its rows and markers* $\boldsymbol{h}_1,\ldots,\boldsymbol{h}_p$ *for its columns, chosen in such a way that the internal product* $\boldsymbol{g}_i'.\ \boldsymbol{h}_j$ *represents the element* x_{ij} *of the matrix* $\boldsymbol{X}$" (Gabriel, 1971). The Biplot method (Gabriel, 1971) allows joint plotting of the rows (samples) and columns (variables) of a data matrix in a reduced subspace dimension.

The initial matrix can be written according to the singular value decomposition:

$$\mathbf{X} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{Q}'$$

where $\mathbf{P}$ is the matrix of eigenvectors of the matrix $\mathbf{XX}'$; $\boldsymbol{\Lambda}$ is the matrix of eigenvalues of the previous matrix ordered from the largest to the smallest, and $\mathbf{Q}'$ is the matrix of eigenvectors of the matrix $\mathbf{X}'\mathbf{X}$. Suitable choice of markers for the rows and for the columns in a low-dimension subspace affords the Biplot representation.

The approach (s) at low range: $s < r$ (where r is the X range) would be done merely by taking the first columns ($X(s)$) of the factorial decomposition:

$$\mathbf{X}_{(s)} = \mathbf{P}_{(s)}\boldsymbol{\Lambda}_{(s)}\mathbf{Q}'_{(s)}$$

In 1990, Gabriel and Odoroff considered the following assignation from the decomposition in singular values and low-range approach ($s = 2$)

$$\mathbf{X}_{(2)} = \mathbf{P}_{(2)}\boldsymbol{\Lambda}_{(2)}\mathbf{Q}'_{(2)}$$

where $\mathbf{P}_{(2)}$ and $\mathbf{Q}_{(2)}$ are the first two columns of $\mathbf{P}$ and $\mathbf{Q}$, respectively and $\boldsymbol{\Lambda}_{(2)}$ is the diagonal matrix formed by the first two elements of the matrix $\boldsymbol{\Lambda}$.

The markers for the rows $\mathbf{G}_{(2)}$ and for the columns $\mathbf{H}_{(2)}$ are chosen as follows:

$$\mathbf{G}_{(2)} = \mathbf{P}_{(2)}\boldsymbol{\Lambda}^{\boldsymbol{\alpha}}_{(2)}; \quad \mathbf{H}_{(2)} = \mathbf{Q}_{(2)}\boldsymbol{\Lambda}^{(1-\boldsymbol{\alpha})}_{(2)}$$

where $\boldsymbol{\alpha}$ is a constant that can take different values.

Gabriel and Odoroff (1990) reported three values for this constant, a notation also followed by other authors (Gower, 1990 or Jackson, 1991), giving rise to three types of Biplot of special interest:

$$\text{If } \boldsymbol{\alpha} = 0: \quad \mathbf{G}_{(2)} = \mathbf{P}_{(2)} \text{ and } \mathbf{H}_{(2)} = \mathbf{Q}_{(2)}\boldsymbol{\Lambda}_{(2)}$$

the Biplot representation obtained is called the GH′-Biplot or CMP-Biplot.

$$\text{If } \boldsymbol{\alpha} = 1: \quad \mathbf{G}_{(2)} = \mathbf{P}_{(2)}\boldsymbol{\Lambda}_{(2)} \text{ and } \mathbf{H}_{(2)} = \mathbf{Q}_{(2)}$$

the Biplot representation obtained is called the JK′-Biplot or RMP-Biplot.

$$\text{If } \boldsymbol{\alpha} = 1/2: \quad \mathbf{G}_{(2)} = \mathbf{P}_{(2)}\boldsymbol{\Lambda}^{1/2}_{(2)} \text{ and } \mathbf{H}_{(2)} = \mathbf{Q}_{(2)}\boldsymbol{\Lambda}^{1/2}_{(2)}$$

The Biplot representation obtained for this decomposition was called the SQRT-Biplot by Gabriel.

Thus, Gabriel and Odoroff essentially described two types of Biplot: the CMP-Biplot (Column Metric Preserving), which leads to high quality for variables and the RMP-Biplot (Row Metric Preserving), which leads to high quality for rows.

The HJ-Biplot (Galindo, 1986) is a symmetric simultaneous representation technique that to a certain extent resembles Correspondence Analysis, but is not restricted to frequency data. It takes as markers for the columns of the matrix $\mathbf{X}$:

$$\mathbf{H} = \mathbf{Q}\boldsymbol{\Lambda}$$

and as markers for the rows:

$$\mathbf{J} = \mathbf{P}\boldsymbol{\Lambda}$$

Both markers can be represented in the same reference system.

Indeed, taking into account the relationship between $\mathbf{P}$ and $\mathbf{Q}$ ($\mathbf{P} = \mathbf{XQ}\boldsymbol{\Lambda}^{-1}$):

$$\mathbf{P}\boldsymbol{\Lambda} = \left(\mathbf{XQ}\boldsymbol{\Lambda}^{-1}\right)\boldsymbol{\Lambda} = \mathbf{XQ}$$

The markers chosen for the rows coincide with the projection of the n points in the space of the principal components of the column.

Likewise, $\mathbf{Q} = \mathbf{X}'\mathbf{P}\boldsymbol{\Lambda}^{-1}$, and hence:

$$\mathbf{Q}\boldsymbol{\Lambda} = \left(\mathbf{X}'\mathbf{P}\boldsymbol{\Lambda}^{-1}\right)\boldsymbol{\Lambda} = \mathbf{X}'\mathbf{P}$$

That is, the markers chosen for the columns coincide with the projection of the p points in the space of the principal components of the rows.

Both representations are related. Let:

$$\mathbf{A} = \mathbf{XQ} \text{ and } \mathbf{B} = \mathbf{X}'\mathbf{P}$$

$$\mathbf{B} = \mathbf{X}'\mathbf{P} = \mathbf{X}'\mathbf{XQ}\boldsymbol{\Lambda}^{-1} = \mathbf{X}'\mathbf{A}\boldsymbol{\Lambda}^{-1}$$

$$\mathbf{A} = \mathbf{XQ} = \mathbf{XX}'\mathbf{P}\boldsymbol{\Lambda}^{-1} = \mathbf{XB}\boldsymbol{\Lambda}^{-1}$$

One has that the hth coordinate of the jth column can be expressed as a function of the hth coordinates of the n-rows.

According to Martín-Rodríguez et al. (2001), the RMP and CMP represent the $\mathbf{X}_{(2)}$ approximation of the matrix $\mathbf{X}$ whose goodness of fit is ($\boldsymbol{\lambda}_i$ are the eigenvalues):

$$\frac{\lambda_1^2 + \lambda_2^2}{\sum_i \lambda_i^2}$$

The RMP and CMP Biplots each have a factor in standard form and another factor in principal form. The goodness of fit for the principal factors is quantified thus:

$$\frac{\lambda_1^4 + \lambda_2^4}{\sum_i \lambda_i^4}$$

The factors in standard form of the RMP and CMP have a drawback in that their goodness of fit is in the order of:

$$\frac{\lambda_1^0+\lambda_2^0}{\sum_i \lambda_i^0}=\frac{2}{r}$$

In HJ-Biplot, the goodness of fit is:

$$\frac{\lambda_1+\lambda_2}{\sum_i \lambda_i}$$

The method achieves an optimum quality of representation for both rows and columns, since both are represented on the same reference system. The method is closely related to principal component analysis (CA), since variance and covariance matrices are plotted on planes that account for most of the inertia.

The distance between row points (samples) is interpreted as similarity, and the angle formed by the vectors (variables) is interpreted as correlation; if a row point (sample) is close to a column point (variable), this is interpreted as preponderance.

Several measures are essential for a correct HJ-Biplot interpretation (Galindo and Cuadras, 1986). Thus, the Relative Contribution of the Factor to the Element (C.R.F.E.) expresses the part of the variability of the element (row or column) explained by the factor (axis). The Quality of Representation (C.L.R.) is the sum of the C.R.F.E. of the factors considered. Only the points with good quality of representation can be interpreted correctly in the subspace observed.

Finally, the statistical software employed was Matlab because of the simplicity of the operating procedures.

3. Results and conclusions

On the factorial planes, we have shown the markers of the variables with poor representation quality in lighter print in order to show that marker is not clearly interpretable. We only plotted the markers of the weeks with good representation quality. We considered as points with poor representation quality those whose C.L.R. did not surpass half the C.L.R. of the best represented point.

Here we worked with a matrix of 237 rows and 10 columns. The rows refer to the weekly means of the concentrations of the pollutants and cover the period studied. We took into account the mean weekly values of the weeks for which at least 50% of the experimental values possible was available, discarding the weeks in which this did not occur.

The rows are represented by an alphanumerical symbol that includes a letter and a number or two letters. In the first case, the letter indicates the month of the year, according to the following:

E	January	m	May	S	September
F	February	J	June	O	October
M	March	j	July	N	November
A	April	a	August	D	December

The number corresponds to the complete week in the month in question. If there are two letters, this refers to an intermediate week between two months, which includes days from two contiguous months.

The matrix was used to apply the HJ-Biplot representation method, obtaining the results shown in Table 1.

Seventy percent of the total inertia is absorbed by only the first three factorial axes, indicating that 70.0% of the total information is present on these three axes.

The Factorial Plane 1–2 absorbs 59.0% of the total inertia. Fig. 2 shows the graphic plot corresponding to the set of all the years, and Fig. 3 shows the same plane but signalling only the markers of each year: five separate plots that, if they were grouped, would give the above Plane 1–2.

Axis 1 separates the O_3 variable (a pollutant that reaches its highest levels in summer and that in winter has much lower values) from all the other variables: SO_2 E.1, SO_2 E.2, PM10 E.1, CO E.1, NO E.1, NO E.2, NO_2 E.1 and NO_2 E.2, all of them pollutants with high concentrations in winter.

Axis 2 separates two groups of variables: on one hand, a first group formed by the variables SO_2 E.1, SO_2 E.2, PM10 E.1 and CO E.1 (ordered according to their projection on axis 2), which are highly correlated, and another group, with lower correlations among the components NO_2 E.1, NO E.2, NO_2 E.2 and NO E.1 (ordered according to their projection on axis 2). Between these two groups, correlation is fairly low. The pollutants of the first group are more typical of winter, more affected by the fixed combustion of coal and other fossil fuels (except CO, whose main source is

Table 1
Results of the inertia for each axis

Axis	Absorbed inertia	Accumulated inertia
1	38.2	38.2
2	20.8	59.0
3	11.0	70.0
4	7.8	77.8
5	6.2	84.0
6	5.9	89.9
7	3.7	93.7
8	2.6	96.3
9	2.1	98.3
10	1.6	100

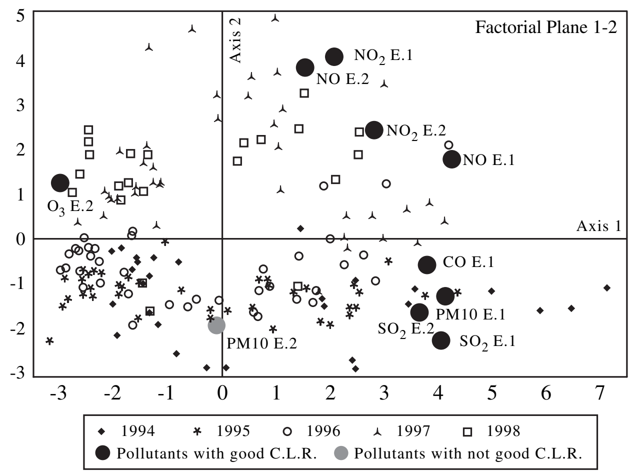

Fig. 2. Factorial Plane 1–2.

traffic). Those of the second group are pollutants emitted by fixed combustion and above all by traffic, such that they are present at appreciable and considerably more similar levels than the others along the year.

Regarding the C.R.F.E., the most important variables for the axis 1 are NO E.1, PM10 E.1, SO_2 E.1, CO E.1 and SO_2 E.2 (Table 2), and to a lesser extent O_3 E.2. The most important variables for the axis 2 are NO_2 E.1 and NO E.2. The quality of representation for the PM10 E.2 variable is not good on this plane (grey colour point instead of black colour point).

Fig. 3 shows that axis 1 discriminates the points in a temporal manner: at the left are the weeks corresponding to the summer period together with the weeks in which the meteorological conditions resemble those of this period and on the other side are the weeks in which the meteorological conditions are typically those of winter. The intermediate part is occupied by weeks corresponding to the spring and autumn periods.

From Figs. 2 and 3 the following may be deduced:

The shift occurring in the group of summer weeks. These take increasingly higher values in the O_3 variable and their preponderance also increases, whereas they take low values in the other variables. Also, in 1997 some of these weeks have an important value in the nitrogen oxide group of variables.

Regarding the group of winter weeks – at the start of the period studied, 1994 – this begins with high preponderance in the variables SO_2 E.1, SO_2 E.2, PM10 E.1 and CO E.1), apart from having high values in other variables, particularly NO E.1 and NO_2 E.2. The group ends in 1997 and 1998, thus lying in the intermediate zone between both groups of pollutants.

Whereas the summer weeks remain in a stable zone with respect to axis 1, the weeks of the winter periods are located in a relatively more extensive zone. This is because, in general, in summer the concentrations of nearly all the pollutants except O_3 (and to a lesser extent NO_2), decrease with respect to winter levels and this period is characterized by such a decrease. In winter, by contrast, pollutants in general increase, although since their concentrations in the atmosphere largely depend on the prevailing atmospheric conditions a broad range of situations is seen. Additionally, the variability in the concentrations of the pollutants is affected by changes in the atmospheric conditions and these in turn – in Salamanca – vary more in winter than in summer.

Two episodes of appreciably higher pollution can be identified. One of them occurs at the end of 1994 and is characterised by an increase in the concentrations of primary pollutants and is located, from the graphic point of view, in the direction of the variables SO_2, CO and PM10 of station E.1. The other is located at the end of summer and at the start of autumn 1997 and is due to an increase in the concentration of NO_2 E.1, NO E.2 and of O_3 (Fig. 3d). It is worth mentioning that these episodes involve two different types of pollution. There is a very prominent correlation between NO E.2 and NO_2 E.1. In view of the location of both stations and the nature of the second episode, it is possible to identify the existence of a localised focus of emission, a fixed industrial (a fertilizer factory) source that may emit of NO and of other chemical species (ammonia and ammonium salts) that could be converted into NO by photochemical oxidation, apart from that coming from mobile sources (traffic) or from domestic heating systems. Knowledge of the industrial processes carried out at the fertilizer factory, the oxidation-reduction reactivity of the ammonia compounds and verification of the emissions coming from that industrial source (in situ verification, smell, concentrations emitted, etc.) allows this hypothesis to be proposed. These localised emissions later affect the generation of NO_2 recorded at station E.1; this pollutant becomes increasingly important within the set of those measured in the city and increasingly so as from that year.

The Factorial Plane 1–3 (Fig. 4a) absorbs 49.2% of the total inertia (Table 1). The variables defining axis 3 are PM10 E.2 and NO_2 E.2 (particularly the former, Table 2). On this plane, the C.L.R. of PM10 E.2 and of PM10 E.1 is good. Accordingly, Fig. 4a shows the different origins of the particulate matter measured at both stations. The weeks taking the lowest values for PM10 at E.2 are the winter weeks, when the highest values at E.1 are reached. The location of E.2, in an open zone close to a park, explains the increase in the concentration of the natural particles with the dry season.

The Factorial Plane 2–3 (Fig. 4b) encompasses 31.8% of the total inertia (Table 1). It is possible to note the high correlation between NO E.2 and NO_2 E.1. It is also

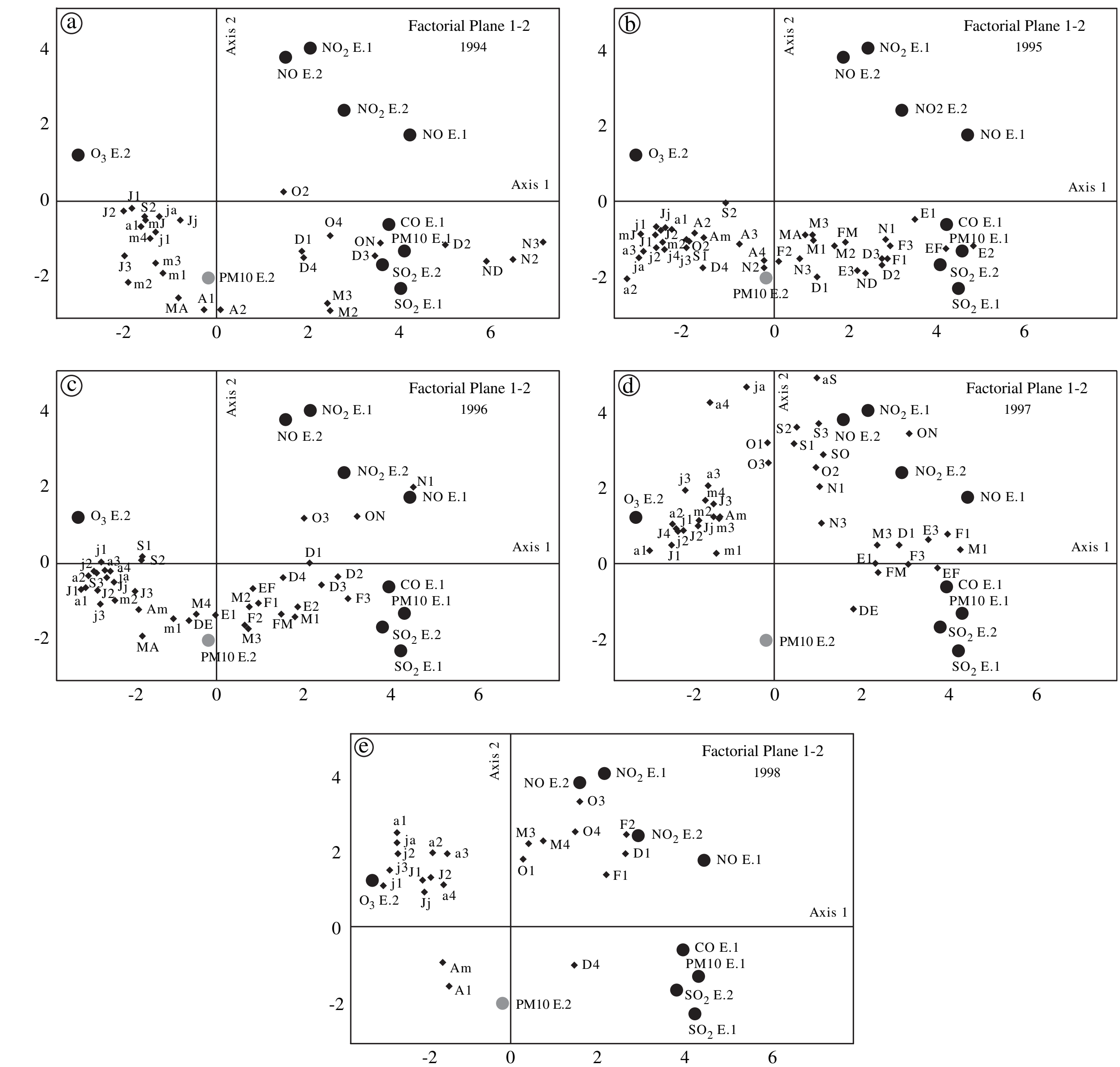

Fig. 3. Factorial Plane 1–2. Representations of the five years: (a) 1994; (b) 1995; (c) 1996; (d) 1997; (e) 1998.

possible to appreciate the general evolution of the pollution: from a small effect of NO_x at the beginning of the study to the situation in which these pollutants define the pollution of nearly the whole year. The episode of pollution by NO_x in 1997 is best seen on this plane.

As general conclusions, the following are pertinent:

The HJ-Biplot analysis shows the evolution of urban pollution. On using the HJ-Biplot method instead of employing other perhaps more conventional ones, such as Principal Component Analysis or Correspondence Analysis (CA), important advantages can be gained. When Principal Component Analysis is used, the axes are combinations of the variables, but these do

Table 2
Relative contribution of the factor to the element (C.R.F.E.) for the three first axes

Variable	Axis 1	Axis 2	Axis 3
SO_2 E.1	605	214	36
PM10 E.1	634	74	53
NO E.1	670	108	6
NO_2 E.1	155	601	0
CO E.1	534	17	30
SO_2 E.2	494	116	52
PM10 E.2	1	157	700
NO E.2	83	531	6
NO_2 E.2	290	209	207
O_3 E.2	354	51	15

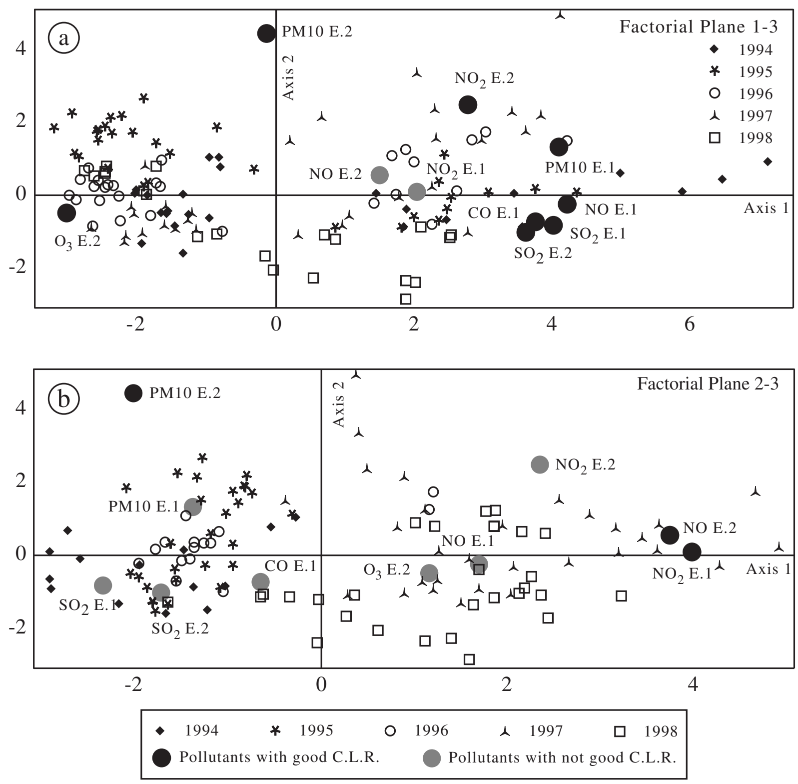

Fig. 4. (a) Factorial Plane 1–3; (b) Factorial Plane 2–3.

not appear on the plots, such that very important information concerning the correlations among them is lost, as is the information that tells one about the relative situation of the points with respect to the variables, which in the HJ-Biplot method is interpreted in terms of greater or lesser preponderance.

Also, the results obtained using Correspondence Analysis could be considered similar to those obtained with the HJ-Biplot method, although CA tends to use frequency data (contingency tables) and, additionally, it is only possible to work with real positive integers. With data such as those used in the present work, it is better to use the HJ-Biplot method for the above-explained reasons and also because better values are obtained for parameters such as C.L.R., C.R.F.E., etc.

In the years studied, in Salamanca there was an important decrease in the pollutants SO_2 and PM10, associated with fixed combustion. To a large extent, this situation was due to the decrease in the use of coal in domestic heating systems and to the decrease in the content of S in fuels. In turn, nitrogen oxides and photochemical pollution became increasingly important.

Regarding pollution in the city, in general there are two clearly differentiated seasons – summer and winter – each of them with variables that characterise them: in summer, O_3, and in winter the rest of the variables. Spring and autumn show greater variability.

Acknowledgements

The authors wish to express their gratitude to the Municipal Council of the city of Salamanca for allowing them to make use of data from the Local Air Pollution Control Network in this work.

References

Flocchini, G., Myrup, O., 1990. A principal component analysis of visibility and air pollution in six California cities. Atmosfera 3, 127–141.

Gabriel, K.R., 1971. The Biplot-graphic display of matrices with application to Principal Component Analysis. Biometrika 58, 453–467.

Gabriel, K.R., Odoroff, Ch.L., 1990. Biplots in biomedical research. Statistic in Medicine 9, 469–485.

Galindo, M.P., 1986. Una alternativa de representación simultánea: HJ-Biplot. Qüestió 1, 13–23.

Galindo, M.P., Cuadras, C.M., 1986. Una extensión del método Biplot y su relación con otras técnicas. Publicaciones de Bioestadística y Biomatemática. Universidad de Barcelona 17.

González Cabrera, J.M., 2002. Evaluación de la contaminación atmosférica en Salamanca. Periodo 1993–1998. Tesis Doctoral. Universidad de Salamanca.

Gower, J.C., 1990. Three-dimensional biplots. Biometrika 77, 773–785.

Greenacre, M.J., 1984. Theory and Applications of Correspondence Analysis. Academic Press Inc., London.

Hurley, P.J., Physick, W.L., Luhar, A.K., 2005. TAPM: a practical approach to prognostic meteorological and air pollution modelling. Environmental Modelling & Software 20, 737–752.

Ionescu, A., Candau, Y., Mayer, E., Colda, I., 2000. Analytical determination and classification of pollutant concentration fields using air pollution monitoring network data. Methodology and application in the Paris area, during episodes with peak nitrogen dioxide levels. Environmental Modelling & Software 15, 565–573.

Jackson, J.E., 1991. A User's Guide to Principal Components Analysis. J. Wiley & Sons Publication, New York.

Lim, L.L., Hughes, S.J., Hellawell, W.E. Integrated decision support system for urban air quality assessment. Environmental Modelling & Software, in press, doi: 10.1016/j.envsoft.2004.04.013.

Martín-Rodríguez, J., Galindo-Villardón, M.P., Vicente-Villardón, J.L., 2001. Comparison and integration of subspaces from a biplot perspective. Journal of Statistical Planning and Inference 102, 411–423.

Nunnari, G., Dorling, S., Schlink, U., Cawley, G., Foxall, R., Chatterton, T., 2004. Modelling SO_2 concentration at a point with statistical approaches. Environmental Modelling & Software 19, 887–905.

Park, S., Kim, Y., Kang, C., 2002. Atmospheric polycyclic aromatic hydrocarbons in Seul, Korea. Atmospheric Environment 36, 2917–2924.

Sánchez-Ccoyllo, O.R., Andrade, M.F., 2002. The influence of meteorological conditions on the behavior of pollutants concentrations in São Paulo, Brazil. Environmental Pollution 116, 257–263.

Statheropoulos, M., Vassiliadis, N., Pappa, A., 1998. Principal component and canonical correlation analysis for examining air pollution and meteorological data. Atmospheric Environment 32, 1087–1095.

Stella, A., Picardo, M.T., Coradeghini, R., Redaelli, A., Lanteri, S., Armanino, C., Valerio, F., 2002. Principal component analysis application in polycyclic aromatic hydrocarbons "mussel watch" analyses for source identification. Analytical Chimica Acta 461, 201–213.

Yu, T.Y., Chang, L.F., 2000. Selection of the scenarios of ozone pollution at southern Taiwan area utilizing principal component analysis. Atmospheric Environment 34, 4499–4509.