

# Neuro-Fuzzy Analysis of Atmospheric Pollution

Ángel Arroyo<sup>1</sup>(✉), Verónica Tricio<sup>2</sup>, Emilio Corchado<sup>3</sup>,  
and Álvaro Herrero<sup>1</sup>

<sup>1</sup> Department of Civil Engineering, University of Burgos, Burgos, Spain  
{aarroyop, ahcosio}@ubu.es

<sup>2</sup> Department of Physics, University of Burgos, Burgos, Spain  
vtricio@ubu.es

<sup>3</sup> Departamento de Informática y Automática,  
University of Salamanca, Salamanca, Spain  
escorchado@usal.es

**Abstract.** Present study proposes the application of different soft-computing and statistical techniques to the characterization of atmospheric conditions in Spain. The main goal is to visualize and analyze the air quality in a certain region of Spain (Madrid) to better understand its circumstances and evolution. To do so, real-life data from three data acquisition stations are analysed. The main pollutants acquired by these stations are studied in order to research how the geographical location of these stations and the different seasons of the year are decisive in the behavior of air pollution. Different techniques for dimensionality reduction together with clustering techniques have been applied, in a combination of neural and fuzzy paradigms.

**Keywords:** Hybrid systems · Clustering techniques · Air quality · Statistical models · Artificial neural networks

## 1 Introduction

In recent years, our knowledge of atmospheric pollution and our understanding of its effects have advanced greatly. It has been accepted for some years now that air pollution not only represents a health risk. Other serious consequences may be mentioned such as acid rain, corrosion, climate change and global warming. Thus, all efforts that are directed towards studying these phenomena [1] may improve our understanding and help us to prevent the serious problematic nature of atmospheric pollution. Systematic measurements in Spain, which are usually taken within large cities, are fundamental due to the health risks caused by high levels of atmospheric pollution. Recent trends point to the benefits of continuing to extend the network of atmospheric pollution measurement stations. These measurement stations acquire data continuously. Thanks to the open data policy promulgated by the public institutions [2] these data are available for further study and analysis.

Dimensionality reduction techniques [3] do the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. These techniques have been previously applied to the field of Environmental Conditions (EC) [4–6]. There is a wide range of dimensionality reduction techniques, such as Principal Component

Analysis (PCA) [7], Local Linear Embedding (LLE) [8], Isometric Mapping (ISO-MAP) [9] and Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [10], which have previously achieved very good results in the field of Environmental Conditions (EC) [11]. One characteristic of these techniques is that clusters are identified in the graphical representation with the naked eye, and there is not any label or assignment of each sample to a certain group of data. These techniques are very useful as pre-processors for the application of other machine learning paradigms, or as the final step in the visualization task.

On the other hand, clustering or grouping techniques [12] divide a given dataset into groups of similar objects, according to several different “similarity” measures. In most cases, the number of desired clusters and the function determines the assignment of each sample to a certain cluster is defined. These sets of techniques have been previously applied to EC [13, 14]. Most of the previous studies apply clustering techniques and only a few dimensionality reduction techniques. So, to the best of the authors’ knowledge, this is the first time that those techniques are applied in unison, and its results are compared.

The main idea of present work is to analyse data describing air pollution from a case study associated to the region of Madrid (Spain). Firstly, some neural dimensionality reduction techniques are applied to project the data and get an intuitive visualization of the data structure. In a second step, clustering techniques are applied to the original data set in order to find the best possible clustering of data. For that, the output of first step is taken into account and fuzzy logic is applied for the clustering.

The rest of this study is organized as follows. Section 2 presents the techniques and methods that are applied to analyse the data. Section 3 details the real-life case study that is addressed in present work, while Sect. 4 describes the experiments and results. Finally, Sect. 5 sets out the conclusions and future work.

## 2 Techniques and Methods

In order to analyse data sets with atmospheric pollution information, several dimensionality reduction techniques and clustering methods are applied. Those methods are described in this section.

### 2.1 Neural Projection Techniques

To reduce the cost of the dimensionality of the representation space, the use of feature selection procedures was proposed [15], with the aim of reducing this dimensionality. The problem of dimensionality reduction can be expressed as follows: for each sample  $i$  determine a selection or transformation of attributes so that:

$$x_{ij} \longrightarrow y_{ik}, j = 1, \dots, n; k = 1, \dots, l, l < n \quad (1)$$

where  $x_{ij}$  represent each vector in the input space,  $y_{ik}$  represent each vector in the output space,  $n$  and  $l$  are the number of dimensions in the input and output spaces, respectively.

From the wide range of neural projection techniques, two different ones have been applied in present work. Principal Component Analysis has been applied as a standard projection technique. At the same time, a more advanced technique (Cooperative Maximum Likelihood Hebbian Learning) is also applied for comparison purposes.

**Principal Component Analysis.** Principal Component Analysis (PCA) [16] is a well-known method that gives the best linear data compression in terms of least mean square error by addressing the data variance. Although it was proposed as an statistical method, it has been proved that it can be implemented by several Artificial Neural Networks [17, 18].

**Cooperative Maximum Likelihood Hebbian Learning.** Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [19] is an extended version of Maximum Likelihood Hebbian Learning (MLHL) [20] that incorporates lateral connections, which have been derived from the Rectified Gaussian Distribution. The resultant net can find the independent factors of a data set but does so in a way that captures some type of global ordering in the data set.

Consider an  $N$ -dimensional input vector  $x$ , an  $M$ -dimensional output vector  $y$  and a weight matrix  $W$ , where the element  $W_{ij}$  represents the relationship between input  $x_j$  and output  $y_i$ , then as it is shown in [19], CMLHL can be carried out as a four-step procedure:

Feed-forward step, where outputs are calculated according to:

$$y_i = \sum_{j=1}^N W_{ij}x_j, \forall i \quad (2)$$

Lateral activation passing step, where lateral connections of output neurons are given by:

$$y_i(t+1) = [y_i(t) + \tau(b - Ay)]^+ \quad (3)$$

Feedback step:

$$e_j = x_j - \sum_{i=1}^M W_{ij}y_i, \forall j \quad (4)$$

Weight update step, where the following learning rule is applied:

$$\Delta W_{ij} = \eta \cdot y_i \cdot \text{sign}(e_j) |e_j|^{p-1} \quad (5)$$

Where  $\eta$  is the learning rate,  $\tau$  is the “strength” of the lateral connections,  $b$  the bias parameter,  $p$  a parameter related to the energy function, and  $A$  is a symmetric matrix used to modify the response to the data.

## 2.2 Clustering Techniques

Cluster analysis organizes data by abstracting underlying structure either as a grouping of individuals or as a hierarchy of groups. The representation can then be investigated to see if the data group according to preconceived ideas or to suggest new experiments [12]. Intuitively, two elements belonging to a valid cluster should be more similar to each other than those which are in different groups. As in the case of projection techniques, only two clustering methods, from the wide range of clustering techniques, have been applied in present work. The k-means method has been selected as a standard clustering technique and fuzzy c-means as a more advanced technique, comprising Fuzzy Logic, to compare clustering outputs.

**K-means.** The standard *k*-means [21] is an algorithm for grouping data points into a given number of clusters. Its application requires two input parameters: the number of clusters, *k*, and their initial centroids, which can be chosen by the user or obtained through some pre-processing. Each data element is assigned to the nearest group centroid, thereby obtaining the initial composition of the groups. Once these groups are obtained, the centroids are recalculated and a further reallocation is made. The process is repeated until the centroids do not change. Given the heavy reliance of this method on initial parameters, a good measure of the goodness of the grouping is simply the sum of the proximity Error Sums of Squares (SSE) that it attempts to minimize:

$$SSE = \sum_{j=1}^k \sum_{x \in G_j} \frac{p(x_i, c_j)}{n} \tag{6}$$

where *p* is the proximity function, *k* is the number of the groups, *c<sub>j</sub>* are the centroids and *n* the number of rows.. In the case of working with Euclidean distance, the expression is equivalent to the global mean square error.

**Fuzzy c-means.** Conventional clustering approaches (such as *k*-means) assume that an object can belong to only one cluster. In practice, the separation of clusters is a fuzzy notion. Fuzzy clustering algorithms, that combine Fuzzy Logic and cluster analysis techniques, outperform conventional clustering algorithms by allowing each object to be assigned to one or more clusters. According to Fuzzy Logic, the membership function is not a crisp value (0, 1) but it is now defined as a percentage.

The *c*-means clustering algorithm [22] is a clustering method that lets one piece of data belong to two or more clusters. Let  $X = [x_1, x_2, \dots, x_m]$  be a set of numerical data in  $\mathbf{R}^N$ . Let *c* be an integer,  $1 < c < M$ . Given *X*, it can be said that *c*-fuzzy subsets  $\{u_k; X \rightarrow [0, 1]\}$  are partitions of *X* if the following conditions are satisfied:

$$0 \leq u_{kj} \leq 1 \quad \forall k, j \tag{7}$$

$$\sum_{k=1}^c u_{kj} = 1 \quad \forall j \tag{8}$$

$$0 < \sum_{j=1}^M u_{kj} < n \forall k \quad (9)$$

Where  $u_{kj} = u_k(X_j)$ ,  $1 \leq k \leq c$  and  $1 \leq j \leq M$ .

### 3 Real-Life Case Study

This study is focused on the analysis of air pollution data recorded in the region of Madrid [23] and Madrid city (central Spain) [24].

Three reasons determine the selection of the stations under study: the same pollutants are acquired and these pollutants are the most interesting ones in order to study the atmospheric conditions [25]. The second reason is that the areas, where the stations are located, have strong differences, such as traffic density, population or protection of the vegetation. This fact is important for comparing the levels of air pollution between areas with high population density and others with low population density.

From the timeline point of view, data are selected from years 2007 and 2008 because that was the time for the beginning of the economic crisis in this country. This fact implies a significant variation in the levels of air pollution compared with other periods of time. Along these two years the economic activity in strategic fields like building, was significantly reduced. Further studies will compare these results with other ones from subsequent years.

The data was gathered in stations located in the following areas:

1. Aranjuez: this is an urban background station. The full station specification can be found in [23].
2. Atalaz: this is a vegetation protection station. Full station specification can be found in [23].
3. Madrid city center: it represents the average values for the stations located in the downtown area of the Madrid city. Full station specification can be found in [24].

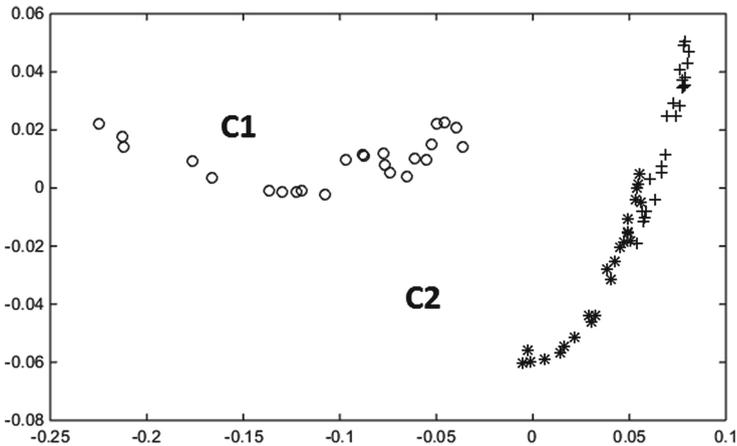
There are a total of 72 samples for the twelve months of 2007 and 2008 and the 3 stations analyzed in this study, 12 samples for each station, one sample per month (monthly daily average). The following parameters (four air pollutants and two meteorological variables) were analyzed:

1. Nitric Oxide (NO) -  $\mu\text{g}/\text{m}^3$ , primary pollutant. NO is a colorless gas which reacts with ozone undergoing rapid oxidation to  $\text{NO}_2$ , which is the predominant in the atmosphere.
2. Nitrogen Dioxide ( $\text{NO}_2$ ) -  $\mu\text{g}/\text{m}^3$ , primary pollutant. From the standpoint of health protection, nitrogen dioxide has set exposure limits for long and short duration.
3. Particulate Matter (PM10) -  $\mu\text{g}/\text{m}^3$ , primary pollutant. These particles remain stable in the air for long periods of time without falling to the ground and can be moved by the wind important distances.
4. Ozone ( $\text{O}_3$ ) -  $\mu\text{g}/\text{m}^3$ , secondary pollutant. Ozone is an odorless, colorless gas composed of three oxygen atoms.

5. Wind Speed Module - m/s.
6. Air Temperature - °C.

### 4 Results and Discussions

As previously stated, many dimensionality reduction methods have been firstly applied. In this step a possible structure in the data is seek. PCA is the first technique applied in order to find a possible structure in the data. Figure 1 shows the two principal components of the data through which the whole dataset is projected.

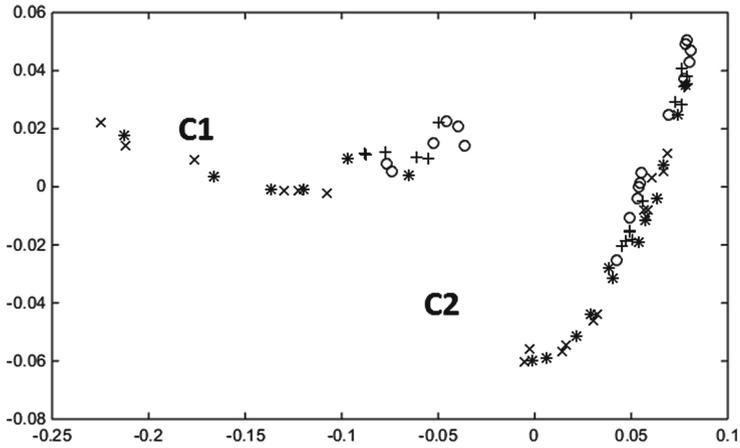


**Fig. 1.** PCA projection. Number of dimensions: 2. Label representation: ‘\*’ urban area, ‘+’ vegetation protected area, ‘o’ madrid center.

In Fig. 1 data samples are depicted according to their geographical location (data acquisition station). PCA depicts two main groups of data (labeled in Fig. 1 as C1 and C2). Group C1 comprises all the samples corresponding to the mean values of the stations located in the city center of Madrid. This area is related to the highest values of air pollution, given mainly by a very large increase of the values of NO. Group C2 contains the samples corresponding to the urban area and vegetation protection area. So, data from these two stations are merged in the same group.

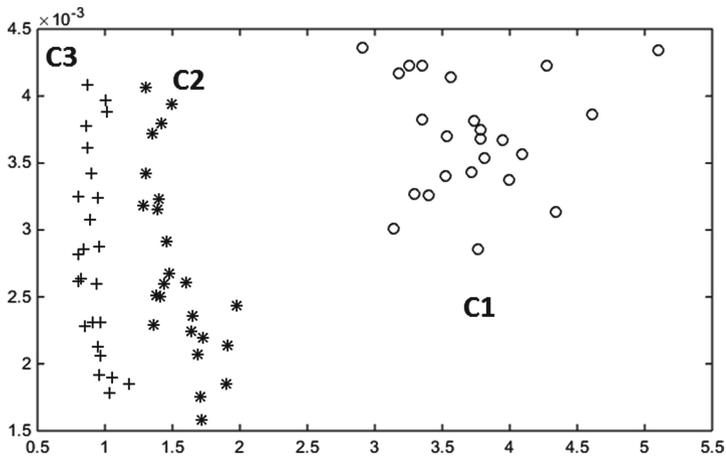
PCA projection is also depicted in Fig. 2 but, in this case, data are projected according to the season of the year they belong to (winter, spring, summer and autumn). The obtained projection is not as clear as the previous one (Fig. 1). Reviewing the original data set, it can be observed an evolution of the atmospheric pollution along the year, as Winter and Autumn data contains higher levels of air pollution in Madrid center (cluster C1). In cluster C2, it is also observed an increase of atmospheric pollution especially in the station of Atalaz (protected vegetation area).

Figure 3 shows the CMLHL projection where data are depicted according to their geographical location (place of the data acquisition station). Three main clusters of data



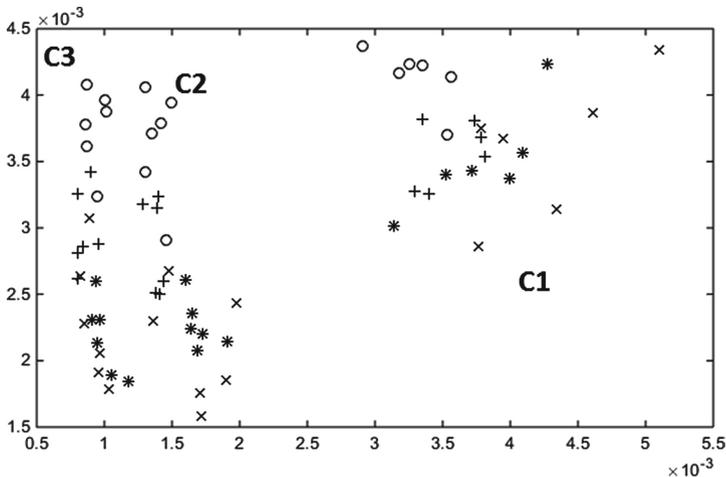
**Fig. 2.** PCA projection. Number of dimensions: 2. Label representation: ‘\*’ winter, ‘+’ spring, ‘o’ summer, ‘x’ autumn.

can be easily identified in this projection (labelled as C1, C2 and C3 in Fig. 3). CMLH outperforms PCA as it is able to separate the samples of the urban area (all of them contained in group C2), from the samples of the vegetation protection area (all of them contained in C3). In a subsequent analysis, it has been found that in C3, where there is little traffic or pollutant emission sources of industrial activity in comparison with others areas, the pollution values decrease significantly, especially the values of the NO variable, with a major increase in the values of O<sub>3</sub>, especially during Summer. This may be because the NO pollutant is an ozone precursor gas. For high air temperatures (in the presence of sunlight) there is ozone formation in the study area.



**Fig. 3.** CMLHL projection. Number of dimensions: 2, number of iterations: 1000, learning rate: 0.001,  $p$  parameter: 2.5,  $\tau$ : 2.5. Label representation: ‘\*’ urban area, ‘+’ vegetation protected area and ‘o’ madrid center.

As for PCA, Fig. 4 shows the CMLHL projection, where data samples are depicted according to associated season of the year. As for previous figure, CMLH projection outperforms that from PCA (Fig. 2) as in the three clusters (C1, C2 and C3), it is observed a clear direction in the data, being the seasons of Winter and Autumn those with higher levels of air pollution. C1 groups the samples corresponding to the city of Madrid in the seasons of Winter and Autumn. This area offers the highest values of air pollution, given mainly by a very large increase of the values of NO, reaching values of  $150 \mu\text{g}/\text{m}^3$  in the autumn and winter seasons. C1 also groups the samples of Madrid center in Summer and Spring, where the air pollution is still high. A certain ordering of the data (from north-west to south-east) can be identified in the group visualization through CMLH. C2 groups the samples corresponding to Aranjuez (urban areas), these samples shows the highest values of air pollution in urban areas in season of Autumn and Winter and a decrease in the other two seasons of the year. In cluster C3, which groups the most of samples acquired in the protected vegetation area, the pollution values fall significantly, especially the values of NO variable, with a notable increase in the values of  $\text{O}_3$ . In this case, the pollution values have fallen significantly, especially in relation to values of NO variable, not significantly increasing its value in any season of the year. In the case of groups C2 and C3, a certain ordering of data can be seen as well: from top to bottom, data are ordered according to the season.



**Fig. 4.** CMLHL projection. Number of dimensions: 2, number of iterations: 1000, learning rate: 0.001,  $p$  parameter: 2.5,  $\tau$ : 2.5. Label representation: '\*' winter, '+' spring, 'o' summer and 'x' autumn.

In this first step two dimensionality reduction techniques are applied in order to visually identify the structure of data, that leads to an analysis of the environmental conditions in the case study. CMLHL obtains better graphical results in order to identify clusters of data. By applying these dimensionality reduction methods, several groups of data can be easily observed with the naked eye.

In a second step, some clustering methods are applied in order to compare the results of the two techniques and the validity of both techniques in this case of study. The results of such techniques are shown in Table 1.

**Table 1.** Clustering results for the atmospheric pollution dataset.

| Method        | K | Metric | Cluster samples allocation (%) |               |               |
|---------------|---|--------|--------------------------------|---------------|---------------|
|               |   |        | Madrid center                  | Atalaz        | Aranjuez      |
| K-means       | 2 | 0.1704 | [100 0]                        | [0 100]       | [0 100]       |
| Fuzzy c-means | 2 | 0.1443 | [100 0]                        | [0 100]       | [0 100]       |
| K-means       | 3 | 0.1191 | [100 0 0]                      | [0 75 25]     | [0 0 100]     |
| Fuzzy c-means | 3 | 0.0854 | [100 0 0]                      | [0 4 96]      | [0 41 39]     |
| K-means       | 4 | 0.0622 | [0 0 29 71]                    | [80 20 0 0]   | [4 96 0 0]    |
| Fuzzy c-means | 4 | 0.0432 | [75 0 0 25]                    | [0 26 74 0]   | [0 96 4 0]    |
| K-means       | 5 | 0.0935 | [100 0 0 0]                    | [0 33 0 0 67] | [0 0 41 59 0] |
| Fuzzy c-means | 5 | 0.0285 | [88 12 0 0 0]                  | [0 0 75 25 0] | [0 0 25 75 0] |

In Table 1, column *k* represents the number of clusters specified for the algorithm in advance. Metric is the output value of the method to assess the quality of the clustering: in fuzzy c-means it is the distance from any given data point to a cluster center weighted by that data point’s membership grade; in k-means it is the total quantization error for the data set. Columns Madrid center, Atalaz and Aranjuez represents the percentage of samples from this location that are allocated to each one the clusters. e.g. [100 0] represents 100 % of samples allocated in the first cluster and 0 % allocated in the second one. In fuzzy c-means is assumed that each data is assigned to the cluster that score for the highest percentage of membership.

Selected values for parameters of k-means are: method: ‘seq’, D: original dataset, k: number of clusters selected, epoch: number of iterations (max 100).

Selected values for parameters of fuzzy c-means are: method: ‘seq’, D: original dataset, k: number of clusters selected, exponent for the matrix U: 2.0, maximum number of iterations: 100, minimum amount of improvement:  $1 \cdot e^{-5}$ .

In Table 1, it can be seen that although PCA clearly defines two clusters of data, (in Figs. 1 and 2) and CMLHL clearly defines three clusters (in Figs. 3 and 4), clustering techniques give substantially more optimal values for higher values of clusters, see column ‘metric’.

Analyzing the distribution of samples in clusters, it is observed that increasing the number of selected clusters, this subdivision of clusters of data occurs at the station of Aranjuez and to the station of Atalaz in a minor way, while most of the Madrid center samples stay in the same cluster. This fact reveals that Madrid city center has very particular atmospheric conditions and highly differentiated from the other two locations.

## 5 Conclusions and Future Work

Conclusions can be divided into two parts, firstly with regard to the analysis of air pollution in the analysed case study. On the other hand, the behaviour of the combined methods of clustering and dimensionality reduction in the process of data grouping.

The conclusions about the quality of the air pollution in Madrid in this case study reflects the fact that there are very high values of NO in the center of Madrid, especially during the Autumn and Winter seasons. This accounts for an increase in air pollution. The O<sub>3</sub> concentration and the wind speed are lower than in the other two seasons, which affects air quality in a negative way. Background urban areas and protected vegetation areas present much lower levels of NO and higher levels of O<sub>3</sub> in summer, especially in the protected vegetation areas. It is also possible to observe an evolution in the levels of air pollution throughout the year, showing higher values in parameters such as NO during the Autumn and Winter seasons. The air pollution observed in these three air quality monitoring stations does not suffer hardly any variation between 2007 and 2008.

Regarding the application of dimensionality reduction techniques, CMLH obtains excellent performance identifying structure of data. CMLHL gives us an easy visual representation of the data, finding internal structures in the data with the naked eye.

Finally, the clustering techniques find relationships between data in a more precise way. Cluster allocation for some of the data is clearly defined. The main disadvantage is the lack of the easy graphical representation of the results that neural projection techniques output. Comparing the two clustering methods applied, can conclude that both work in a correct way for this case study, providing similar results. However fuzzy c-means, thanks to its allocation percentages, gives a more precise answer giving the possibility of a fine adjustment in the allocation of samples to clusters.

In future work, will expand the years of study and methods used to draw new conclusions about the behavior of atmospheric conditions in selected areas of study, as well as the validity of these techniques in the case study.

## References

1. San José, R., Pérez, J.L., González, R.M.: An operational real-time air quality modelling system for industrial plants. *Environ. Model Softw.* **22**, 297–307 (2007)
2. The Aporta project as a driver of the re-use of public sector information in Spain (2015). <http://datos.gob.es/content/proyecto-aporta-como-impulsor-de-reutilizacion-de-informacion-del-sector-publico-espana>
3. Corchado, E., Perez, J.C.: A three-step unsupervised neural model for visualizing high complex dimensional spectroscopic data sets. *Pattern Anal. Appl.* **14**, 207–218 (2011)
4. Corchado, E., Arroyo, A., Tricio, V.: Soft computing models to identify typical meteorological days. *Logic J. IGPL* **19**, 373–383 (2010)
5. Chattopadhyay, G., Chattopadhyay, S., Chakraborty, P.: Principal component analysis and neurocomputing-based models for total ozone concentration over different urban regions of India. *Theoret. Appl. Climatol.* **109**, 221–231 (2011)

6. Glezakos, T.J., Tsiligiridis, T.A., Iliadis, L.S., Yialouris, C.P., Maris, F.P., Ferentinos, K.P.: Feature extraction for time-series data: an artificial neural network evolutionary training model for the management of mountainous watersheds. *Neurocomputing* **73**, 49–59 (2009)
7. Abdi, H., Williams, L.J.: Principal component analysis. *Wiley Interdisciplinary Rev.: Comput. Stat.* **2**, 433–459 (2010)
8. Li, X., Lin, S., Yan, S., Xu, D.: Discriminant locally linear embedding with high-order tensor data. *IEEE Trans. Syst. Man, Cybern. Part B: Cybern.* **38**, 342–352 (2008)
9. Shao, C., Hu, H.: Extension of ISOMAP for imperfect manifolds. *J. Comput.* **7**, 1780–1785 (2012)
10. Corchado, E., Han, Y., Fyfe, C.: Structuring global responses of local filters using lateral connections. *J. Exp. Theor. Artif. Intell.* **15**, 473–487 (2003)
11. Arroyo, A., Corchado, E., Tricio, V.: Atmospheric pollution analysis by unsupervised learning. In: Corchado, E., Yin, H. (eds.) *IDEAL 2009. LNCS*, vol. 5788, pp. 767–772. Springer, Heidelberg (2009)
12. Jain, A.K., Maheswari, S.: Survey of recent clustering techniques in data mining. *J. Curr. Comput. Sci. Technol* **3**, 72–78 (2013)
13. Kassomenos, P., Vardoulakis, S., Borge, R., Lumberras, J., Papaloukas, C., Karakitsios, S.: Comparison of statistical clustering techniques for the classification of modelled atmospheric trajectories. *Theoret. Appl. Climatol.* **102**, 1–12 (2010)
14. Pires, J.C.M., Sousa, S.I.V., Pereira, M.C., Alvim-Ferraz, M.C.M., Martins, F.G.: Management of air quality monitoring using principal component and cluster analysis—Part I: SO<sub>2</sub> and PM<sub>10</sub>. *Atmos. Environ.* **42**, 1249–1260 (2008)
15. Smola, A.J., Schölkopf, B.: A tutorial on support vector regression. *Stat. Comput.* **14**, 199–222 (2004)
16. Pearson, K.: On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **2**, 559–572 (1901)
17. Oja, E.: Principal components, minor components, and linear neural networks. *Neural Netw.* **5**, 927–935 (1992)
18. Oja, E.: Neural networks, principal components, and subspaces. *Int. J. Neural Syst.* **1**, 61–68 (1989)
19. Corchado, E.F.C.: Connectionist techniques for the identification and suppression of interfering underlying factors. *Int. J. Pattern Recognit Artif Intell.* **17**, 1447–1466 (2003)
20. Corchado, E., MacDonald, D., Fyfe, C.: Maximum and minimum likelihood Hebbian learning for exploratory projection pursuit. *Data Min. Knowl. Disc.* **8**, 203–225 (2004)
21. Ding, C., He, X.: K-means clustering via principal component analysis, vol. 29 (2004)
22. Gao, X., Xie, W.: Advances in theory and applications of fuzzy clustering. *Chin. Sci. Bull.* **45**, 961–970 (2000)
23. Region of Madrid, Area Air Quality - Air Quality Network (2015). [http://gestionaria.madrid.org/azul\\_internet/html/web/ListaEstacionesAccion.icm?ESTADO\\_MENU=3\\_2](http://gestionaria.madrid.org/azul_internet/html/web/ListaEstacionesAccion.icm?ESTADO_MENU=3_2)
24. Council of Madrid, Air Quality (2015). <http://www.mambiente.munimadrid.es/opencms/opencms/calair/ContaAtmosferica/portadilla.html>
25. Snelder, T.H., Dey, K.L., Leathwick, J.R.: A procedure for making optimal selection of input variables for multivariate environmental classifications. *Conserv. Biol.* **21**, 365–375 (2007)