Time Analysis of Air Pollution in a Spanish Region through *k*-means

Ángel Arroyo[1], Verónica Tricio[2], Álvaro Herrero[1], Emilio Corchado[3]

[1]Department of Civil Engineering, University of Burgos, Burgos, Spain.
{aarroyop, ahcosio}@ubu.es
[2]Department of Physics, University of Burgos, Burgos, Spain.
vtricio@ubu.es
[3]Departamento de Informática y Automática, University of Salamanca, Salamanca, Spain.
escorchado@usal.es

Abstract. This study presents the application of clustering techniques to a real-life problem of studying the air quality of the Castilla y León region in Spain. The goal of this work is to analyze the level of air pollution in eight points of this Spanish region between years 2008 and 2015. The analyzed data were provided by eight acquisition stations from the regional network of air quality. The main pollutants recorded at these stations are analyzed in order to study the characterization of such stations, according to a zoning process, and their time evolution. Four cluster evaluation and a clustering technique, with the main distance measures, have been applied to the dataset under analysis.

Keywords. Clustering, *k*-means, air quality, time evolution.

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. Systematic measurements in Spain, are fundamental due to the health risks caused by high levels of atmospheric pollution. The measurement stations acquire data continuously. Thanks to the open data policy promulgated by the public institutions [1] these data are available for further study and analysis.

Clustering can be defined as the unsupervised classification of patterns into groups [2]. Hence, clustering (or grouping) techniques divide a given dataset into groups of similar objects, according to several different "similarity" measures. These sets of techniques have been previously applied to air pollution data [3, 4]. A clustering method for the study of multidimensional non-stationary meteorological time series was presented in [3]. Principal Components Analysis (PCA) and Cluster Analysis (CA), were applied in [4] over a 3-year period to analyze the mass concentrations of Sulfur Dioxide (SO_2) and Particulate Matter (PM10) in Oporto.

The main idea of present study is the analysis of the time evolution of the most important pollutant variables between the years 2008 and 2015. The data were recorded at eight data acquisition stations from four provinces of the region of Castilla y León, considering the zoning process stated by the European Union in [5]. Four clustering evaluation techniques [6] are applied in a first step to determine the optimal number of clusters existing in the data set. After this, *k*-means [7], combined with the most widely-used distance measures is applied to each one of the years in order to analyze the evolution of air pollution by taking into account the clustering results of the year-by-year analysis.

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

2 Clustering Techniques and Methods

Clustering is one of the most important unsupervised learning problems [8]. It can be defined as the process of organizing objects into groups whose members are similar in some way. A cluster is a collection of objects which are similar to those in the cluster and are dissimilar to those belonging to other clusters.

Those methods and measure distances are described in this section.

2.1 Cluster Evaluation Measures

Clustering validation evaluates the goodness of clustering results [6]. The two main categories of clustering validation are external and internal. The main difference is whether external information (for which *a priori* knowledge of the dataset is required) is used for clustering validation. Internal validation measures can be used to choose the best clustering algorithm, as can the optimal numbers of clusters, with no further information needed. The following four internal validation measures were all applied in the present work: Calinski-Harabasz Index [9], Silhouette Index [10], Davies-Bouldin Index [11] and Gap Index [12].

2.2 *k*-means Clustering Technique

The well-known *k*-means [13] is a partitional clustering technique for grouping data 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 there are no further changes in the centroids. 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 Sums of Squared Error (SSE) that

it attempts to minimize, Where *p()* is the proximity function, *k* is the number of the groups, c_j are the centroids, and *n* the number of rows:

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

In the case of Euclidean distance [14], the expression is equivalent to the global mean square error.

K-means technique takes distance into account to cluster the data. Different distance criteria were defined and the distance measures applied in the study are described in this subsection.

An *mx*-by-*n* data matrix *X*, which is treated as *mx* (1-by-*n*) row vectors x_1, x_2, ...,x_{mx}, and *my*-by-*n* data matrix *Y*, which is treated as *my* (1-by-*n*) row vectors y_1, y_2, ...,y_{my}.. are given. Various distances between the vector x_s and y_t are defined as follows:

Seuclidean distance

In Standardized Euclidean metrics (Seuclidean), each coordinate difference between rows in *X* is scaled, by dividing it by the corresponding element of the standard deviation:

$$d_{st}^2 = (x_s - y_t)\, V^{-1} (x_s - y_t)' \tag{2}$$

Where *V* is the *n*-by-*n* diagonal matrix the *j*th diagonal element of which is $S(j)^2$, where *S* is the vector of standard deviations.

Cityblock distance

In this case, each centroid is the component-wise median of the points in that cluster.

$$d_{st} = \sum_{j=1}^{n} \left| x_{sj} - y_{tj} \right| \tag{3}$$

Cosine Distance

This distance is defined as one minus the cosine of the included angle between points (treated as vectors). Each centroid is the mean of the points in that cluster, after normalizing those points to unitary Euclidean lengths:

$$d_{st} = 1 - \frac{x_s y_t'}{\sqrt{(x_s x_s')(y_t y_t')}} \tag{4}$$

Correlation Distance

In this case, each centroid is the component-wise mean of the points in that cluster, after centering and normalizing those points to a zero mean and a unit standard deviation.

$$d_{st} = 1 - \frac{\left(x_s - \bar{x}_s\right)\left(y_t - \bar{y}_t\right)'}{\sqrt{\left(x_s - \bar{x}_s\right)\left(x_s - \bar{x}_s\right)'}\sqrt{\left(y_t - \bar{y}_t\right)\left(y_t - \bar{y}_t\right)'}} \tag{5}$$

3 Real-life Case Study

In present study, pollutant data recorded in eight different places in the region of Castilla y León are analyzed. This region is full of vegetation varieties and large natural areas to be protected; another chance is the compensation ratio among the number of urban stations and urban background traffic stations, of which virtually lacked Castilla y León. Some representative data acquisition stations for the air quality monitoring have been selected from four provinces of the region, being these four provinces which own more available data for the study. The main reason that determines the selection of the stations listed below is the characterization of the stations: four of them are assigned to the zone division oriented to the health protection, and the other four stations are assigned to the ozone protection, according to the zoning process in Castilla y León for the assessment of air quality [15].

A compendium of European legislation on air quality is the Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe [16]. This Directive established that air quality plans should be developed for zones and agglomerations within which concentrations of pollutants in ambient air exceed the relevant air quality target values or limit values, plus any temporary margins of tolerance. Two of these zones are: the ozone protection stations and the stations for the human health protection. The eight stations selected for this study, according to the information in [17] are:

1. Burgos 4. Fuentes Blancas, Burgos. Geographical coordinates: 03º38’10’’W; 42º20’10’’N; 929 meters above sea level (masl). Data acquisition station oriented to the health protection.
2. Salamanca 6. Aldehuela park, Salamanca. Geographical coordinates: 05º38’23’’W; 40º57’39’’N; 743 masl. Data acquisition station oriented to the health protection.
3. León 4. Escolar preserve, León. Geographical coordinates: 05º33’59’’W; 42º34’31’’N; 814 masl. Data acquisition station oriented to the health protection.
4. Medina del Campo. Bus station, Valladolid province. Geographical coordinates: 04º54’33’’W; 41º18’59’’N; 721 masl. Data acquisition station oriented to the health protection.
5. Burgos 5. Teresa de Cartagena Saravia St., Burgos. Geographical coordinates: 03º43’16’’W; 42º20’44’’N; 929 (masl). Data acquisition station oriented to the study of the ozone.

6. Salamanca 5. La Bañeza St., Salamanca. Geographical coordinates: 05º39'55''W, 40º58'45''N; 797 masl. Data acquisition station oriented to the study of the ozone.
7. León 1. The Pinilla neighborhood, León. Geographical coordinates: 05º35'14''W; 42º36'14''N; 838 masl. Data acquisition station oriented to the study of the ozone.
8. Valladolid 14. Regueral bridge, Valladolid. Geographical coordinates: 04º44'02''W; 41º39'22''N; 691 masl. Data acquisition station oriented to the study of the ozone.

From the timeline point of view, data are selected between years 2008 and 2015. There are a total of 715 samples containing monthly averages. These samples are distributed as described in Table 1 (corrupted or missing data are omitted):

Table 1. Number of samples by year and for each type of protection zone.

Zone	**Year**							
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Health Protection	48	47	39	36	45	48	48	48
Ozone Protection	56	36	48	48	48	48	46	46

For each one of the station and monthly sample, the following parameters (four air quality variables) were gathered and are considered in present study:

1. Nitric Oxide (NO) - µg/m³, primary pollutant. NO is a colorless gas which reacts with ozone undergoing rapid oxidation to NO_2, which is the predominant in the atmosphere [18].
2. Nitrogen Dioxide (NO_2) - µg/m³, primary pollutant. From the standpoint of health protection, nitrogen dioxide has set exposure limits for long and short duration [18].
3. Particulate Matter (PM10) - µg/m³, 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 over long distances. Defined by the ISO as follows: *"particles which pass through a size-selective inlet with a 50% efficiency cut-off at 10 µm aerodynamic diameter. PM10 corresponds to the 'thoracic convention' as defined in ISO 7708:1995, Clause 6"* [19].
4. Sulphur Dioxide (SO_2) - µg/m³, primary pollutant. It is a gas. It smells like burnt matches. It also smells suffocating. Sulfur dioxide is produced by volcanoes and in various industrial processes. In the food industry, it is also used to protect wine from oxygen and bacteria [18].

4 Results and Discussion

The techniques described in Section 2 were applied to the case study presented in Section 3 and the results are discussed below. Table 2 shows the information on the cluster evaluation for the whole dataset (years from 2008 to 2015) performed by applying the different cluster evaluation measures. In this table, column '*k*' represents

the optimum number of clusters estimated by each one of the measures from the 'InspectedK' parameter (taking values from 2 to 6), 'Time' is the execution time (in seconds) and 'Criterion Values' corresponds to each proposed number of clusters in 'InspectedK', stored as a vector of numerical values. Each value of this vector is calculated according to the evaluation measure on cluster centroids, the number of points in each cluster, the sum of Squared Euclidean and the number of clusters.

Table 2. Cluster evaluation for the whole dataset

Cluster Evaluation Measure	***K***	**Time (s)**	**Parameters**
Calinski-Harabasz	5	1.18	Criterion Values: [209.31 252.57 133.26 288.45 239.96]
Davies-Bouldin	2	1.31	Criterion Values: [0.63 0.84 1.12 1.17 1.05]
Gap	2	98.62	Criterion Values: [1.40 1.39 1.50 1.50 1.21]
Silhouette	2	1.51	Criterion Values: [0.77 0.61 0.52 0.48 0.42]

The output of the four measures applied is k=2 in all cases, except for the Calinski-Harabasz measure. This suggested value of k=2 in three of four cases points to the usefulness of the k parameter, required as an input for the k-means subsequent clustering technique. This value of k provides information about the internal structure of the data. In this case study is equivalent to the two main subsets of data existing in the data set (health and ozone protection stations). The Gap evaluation measure was the slowest in terms of computing time.

Table 3 shows the information on the cluster evaluation distributed by years, one data set for each year.

Table 3.A Cluster evaluation distributed by years (years 2008 to 2010)

Year	**Cluster Evaluation Measure**	***K***	**Time (s)**	**Parameters**
2008	Calinski-Harabasz	4	0.65	Criterion Values: [42.52 35.93 52.81 41.83 29.29]
2008	Davies-Bouldin	2	0.67	Criterion Values: [0.64 0.91 2.08 1.80 1.43]
2008	Gap	2	48.95	Criterion Values: [0.84 1.05 0.52 1.02 0.87]
2008	Silhouette	2	0.69	Criterion Values: [0.79 0.78 0.33 0.17 0.47]
2009	Calinski-Harabasz	6	0.48	Criterion Values: [29.90 36.14 38.74 42.53 46.21]
2009	Davies-Bouldin	2	0.59	Criterion Values: [0.62 1.80 0.79 1.03 1.22]
2009	Gap	4	46.76	Criterion Values: [0.32 0.49 0.72 0.82 0.84]
2009	Silhouette	2	0.54	Criterion Values: [0.60 0.18 0.34 0.40 0.32]
2010	Calinski-Harabasz	4	0.45	Criterion Values: [39.25 33.68 42.28 34.04 38.86]
2010	Davies-Bouldin	2	0.46	Criterion Values: [0.48 0.82 0.97 0.87 1.06]
2010	Gap	3	49.35	Criterion Values: [0.62 0.88 0.70 0.62 1.06]
2010	Silhouette	2	0.58	Criterion Values: [0.77 0.39 0.37 0.32 0.28]

Applying the four cluster evaluation techniques to a subset of data for each year, the value of *k* equals 2 is selected in 65% cases and in all the years of the case study. All the values in the range of *k* (2, 6) are selected at least one time.

Table 3.B Cluster evaluation distributed by years (years 2011 to 2015)

Year	Cluster Evaluation Measure	*K*	Time (s)	Parameters
2011	Calinski-Harabasz	2	0.42	Criterion Values: [31.07 25.34 24.40 28.83 29.83]
2011	Davies-Bouldin	2	0.41	Criterion Values: [0.48 0.82 0.97 0.87 1.06]
2011	Gap	2	48.62	Criterion Values: [0.87 0.52 0.49 0.62 0.65]
2011	Silhouette	2	0.51	Criterion Values: [0.54 0.37 0.35 0.32 0.27]
2012	Calinski-Harabasz	5	0.46	Criterion Values: [25.75 37.09 20.49 41.48 26.03]
2012	Davies-Bouldin	2	0.56	Criterion Values: [0.46 0.71 0.88 0.89 1.50]
2012	Gap	2	51.53	Criterion Values: [0.89 0.84 0.56 0.75 0.45]
2012	Silhouette	2	0.54	Criterion Values [0.81 0.59 0.43 0.23 0.49]
2013	Calinski-Harabasz	3	0.43	Criterion Values: [45.92 34.85 15.16 52.23 44.89]
2013	Davies-Bouldin	2	0.46	Criterion Values: [0.65 0.99 1.13 0.83 0.78]
2013	Gap	3	50.73	Criterion Values: [0.69 0.86 0.69 1.13 0.97]
2013	Silhouette	2	0.50	Criterion Values [0.76 0.36 0.46 0.24 0.37]
2014	Calinski-Harabasz	4	0.43	Criterion Values: [45.92 34.85 15.16 52.23 44.89]
2014	Davies-Bouldin	2	0.43	Criterion Values: [0.52 0.80 1.65 1.02 1.30]
2014	Gap	2	49.84	Criterion Values: [0.85 0.92 0.87 1.05 0.98]
2014	Silhouette	2	0.45	Criterion Values: [0.72 0.39 0.35 0.25 0.23]
2015	Calinski-Harabasz	2	0.56	Criterion Values: [65.89 63.80 28.56 42.26 21.86]
2015	Davies-Bouldin	4	0.63	Criterion Values: [2.80 0.98 0.76 1.40 1.27]
2015	Gap	3	50.46	Criterion Values: [0.84 1.09 0.63 0.89 1.21]
2015	Silhouette	2	0.62	Criterion Values: [0.78 0.64 0.54 0.19 0.37]

Table 4 shows the results obtained for the *k*-means, distributed for each of the years between 2008 and 2015, with different distance criteria and a value of *k* equals 2 (value of *k* mostly selected in Table 2). In this table, 'Distance' is the distance criterion applied (see Section 2) and 'SumD' is the within-cluster sums of point-to-centroid distances in the *k*-by-1 vector. The Cluster Samples Allocation columns represent the percentage of samples from each one of the zones (Heath and Ozono) that are allocated to each one the clusters; e. g. [85 15] represents 2 clusters and 85% of samples allocated to the first cluster and 15% to the second one.

Some issues from the results in Table 4 are worth mentioning: for all the years under study, the best (minimum) value for parameter SumD is obtained when applying 'Seuclidean' distance, followed by 'Cosine'. Regarding with the sample process allocation, 'Seuclidean' distance allocates most of the samples in the same cluster in four of the eight years, despite the characterization (zoning) of its station. Clustering with 'Cosine' and 'Correlation' distances let us separate most of the samples in different clusters for all the years, according to the station characterization (zoning).

Table 4. *k*-means clustering results on yearly subsets of data (2008-2015).

Year	Distance	SumD	Cluster Samples Allocation (%)	
			HealthProtection	Ozone Protection
2008	Seuclidean	[0.08 0.02]	[85 15]	[97 3]
2008	Cityblock	[3.91 0.53]	[85 15]	[100 0]
2008	Cosine	[0.98 0.62]	[77 23]	[39 61]
2008	Correlation	[2.40 3.55]	[33 67]	[64 36]
2009	Seuclidean	[0.05 0.04]	[49 51]	[14 86]
2009	Cityblock	[1.31 2.59]	[26 74]	[72 28]
2009	Cosine	[0.96 0.48]	[81 19]	[39 61]
2009	Correlation	[2.96 1.21]	[79 21]	[36 64]
2010	Seuclidean	[0.03 0.06]	[49 51]	[4 96]
2010	Cityblock	[2.70 1.49]	[46 54]	[90 10]
2010	Cosine	[0.81 0.57]	[74 26]	[29 71]
2010	Correlation	[2.03 2.98]	[33 67]	[77 23]
2011	Seuclidean	[0.05 0.04]	[44 56]	[88 13]
2011	Cityblock	[2.45 1.61]	[44 56]	[88 13]
2011	Cosine	[0.75 1.32]	[69 31]	[29 71]
2011	Correlation	[3.93 3.10]	[31 69]	[75 25]
2012	Seuclidean	[0.07 0.06]	[51 49]	[88 13]
2012	Cityblock	[2.87 1.99]	[44 56]	[85 15]
2012	Cosine	[1.47 1.30]	[71 29]	[25 75]
2012	Correlation	[4.54 5.15]	[33 67]	[73 27]
2013	Seuclidean	[0.04 0.10]	[35 65]	[2 98]
2013	Cityblock	[1.81 3.45]	[52 48]	[6 94]
2013	Cosine	[1.59 1.66]	[79 21]	[31 69]
2013	Correlation	[5.02 6.77]	[79 21]	[31 69]
2014	Seuclidean	[0.05 0.06]	[52 48]	[4 96]
2014	Cityblock	[2.13 2.30]	[31 69]	[87 13]
2014	Cosine	[1.39 0.82]	[25 75]	[80 20]
2014	Correlation	[4.02 2.66]	[29 71]	[85 15]
2015	Seuclidean	[0.07 0.04]	[69 31]	[94 6]
2015	Cityblock	[3.00 1.62]	[65 35]	[27 73]
2015	Cosine	[1.02 0.79]	[65 35]	[31 69]
2015	Correlation	[3.69 2.65]	[58 42]	[27 73]

Fig. 1 shows the evolution between the years 2008 to 2015 of the parameter SumD (Sums of point-to-centroid distance), when applying *k*-means (*k*=2) and the different distance measures applied. It can be seen that the lowest value of SumD for all the years is obtained when applying 'Seuclidean' distance. This means a high level of compactness in the samples of data when applying this distance measure. Another important aspect to be highlighted is that the highest values for SumD are obtained in

years 2012, 2013 and 2014. The 'Correlation' distance measure performs in a different way from the other three distance, presenting the highest value in year 2013, the lowest in 2014 and increasing in the last year when the other three distances decrease. This is because 'Correlation' depends of the typical deviation. Although the years from 2012 to 2015 present lower levels of air pollution in the pollutants analyzed, the typical deviation, especially in NO and NO_2, is bigger than in previous years, major pollution peaks exist in these years of low pollution in the region of Castilla y León.

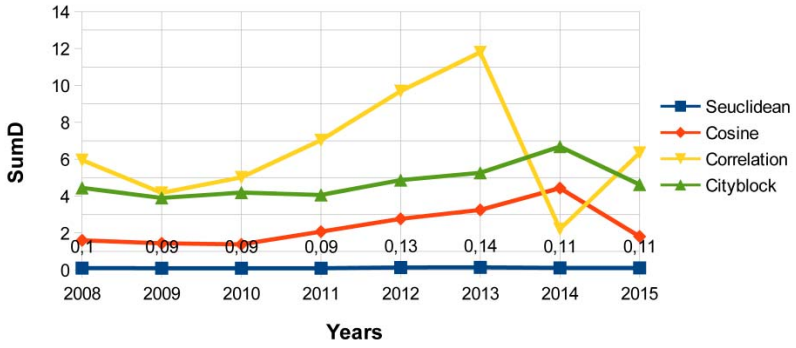

Fig. 1. Year evolution of the SumD parameter.

5 Conclusions and Future Work

Main conclusions derived from obtained results (see Section 4) can be divided into two groups; at first, those regarding the analysis of air quality conditions in the case study considered. Secondly, those related to the behaviour of the two clustering techniques applied in the case study.

Talking about the air quality conditions in the eight selected places, grouped by the data acquisition station type, the average monthly levels of air pollution in the stations oriented to the ozone protection are lower than those recorded in the health oriented stations, especially in NO and NO_2. The evolution in the period of time analyzed (2008-2015) shows higher levels of air pollution between 2008 and 2011, when compared with the subsequent years. By working with monthly data average, the pollutant concentration levels are smoothed in both areas.

Regarding the applied clustering techniques, clustering measure techniques are a very useful set of techniques to determine the optimal value for parameter *k* (number of clusters). The four techniques applied obtained similar results, not being very appropriate the use of Gap Index with large datasets due to high elapsed time. When applying *k*-means with the different measure distance explained in Section 2, 'Seuclidean' distance is the best in terms of creating compact clusters of data, as parameter SumD takes the lowest values, but is not the best technique in the sample process allocation, where tends to keep samples from stations of different zones in the same cluster of data. 'Cosine' distance measure offers the best balance between a good sample allocation process and a not very high value for parameter 'SumD'.

Future work will consist of extending proposed analysis to a wider time period, data from different locations and some other clustering techniques.

References

1. Government of Spain - Aporta Project, http://administracionelectronica.gob.es
2. Jain, A. K., Murty, M. N., Flynn, P. J.: Data Clustering: A Review. ACM computing surveys (CSUR) 31(3), 264-323 (1999).
3. Kassomenos, P., Vardoulakis, S., Borge, R., Lumbreras, J., Papaloukas, C., Karakitsios, S.: Comparison of statistical clustering techniques for the classification of modelled atmospheric trajectories. Theor. Appl. Clim. (102) 1–12 (2010).
4. 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: SO2 and PM10. Atmospheric Environment, 42 (6): 1249–1260 (2008).
5. European Commission - Air Quality Standards, http://ec.europa.eu/environment/air/quality/standards.htm
6. Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J.: Understanding of internal clustering validation measures. IEEE International Conference on Data Mining. 911-916 (2010).
7. Anil K, J.: Data clustering: 50 years beyond K-means. Pattern Recognition Letters 31, 651-666 (2010).
8. Barlow, H.: Unsupervised learning. Neural computation 1, 295-311 (1989).
9. Caliński, T., Harabasz, J.: A dendrite method for cluster analysis. Communications in Statistics-theory and Methods 3, 1-27 (1974).
10. Rousseeuw, P.J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20, 53-65 (1987).
11. Davies, D.L., Bouldin, D.W.: A cluster separation measure. Pattern Analysis and Machine Intelligence, IEEE Transactions on 224-227 (1979).
12. Tibshirani, R., Walther, G., Hastie, T.: Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63, 411-423 (2001).
13. Ding, C., He, X.: K-means clustering via principal component analysis. Proceedings of the twenty-first international conference on Machine learning, 29 (2004).
14. Danielsson, P.E.: Euclidean distance mapping. Computer Graphics and Image Processing 14, 227-248 (1980).
15. Government of Castilla y León - Zoning of the territory in Castilla y León, http://www.jcyl.es/
16. European Union Law - Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, http://eur-lex.europa.eu/
17. Government of Castilla y León - Annual reports of the Air Quality, http://www.medioambiente.jcyl.es/
18. PubChem - PubChem Compounds, https://pubchem.ncbi.nlm.nih.gov/compound
19. ISO - International Organization for Standardization. - PM10/PM2.5, https://www.iso.org/