Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jes

JES JOURNAL OF ENVIRONMENTAL SCIENCES

www.jesc.ac.cn

A methodological framework for improving air quality monitoring network layout. Applications to environment management

David Galán-Madruga

Department of Atmospheric Pollution, National Center for Environment Health, Health Institute Carlos III, Ctra. Majadahonda a Pozuelo km 2,2 28220 Madrid, Spain

ARTICLE INFO

Article history:
Received 2 March 2020
Revised 7 September 2020
Accepted 7 September 2020
Available online 30 September 2020

Keywords:
Air quality
Ambient air monitoring network
PM_{10} particles
Redundant stations
Optimization
Environment management

ABSTRACT

This work aims to provide a methodology framework which allows to improve the performance and efficiency of an air quality monitoring network (AQMN). It requires to be constituted by a minimum and reliable number of measurement sites. Nevertheless, the AQMN efficiency should be assessed over time, as a consequence of the possible emergence of new emission sources of air pollutants, which could lead to variations on their spatial distribution within the target area. PM_{10} particles data monitored by the Community of Madrid's (Spain) AQMN between 2008 and 2017 were used to develop a methodology to optimize the AQMN performance. The annual spatial distribution of average PM_{10} levels over the studied period monitored by all current stations vs those more representative was provided by a geographic information system (GIS), and the percentage of similarity between both postulates was quantified using simple linear regression (> 95%). As one innovative tool of this study, the practical application of the proposed methodology was validated using PM_{10} particles data measured by AQMN during 2007 and 2018, reaching a similitude degree higher than 95%. The influence of temporal variation on the proposed methodological framework was around 20%.

The proposed methodology sets criteria for identifying non-redundant stations within AQMN, it is also able to appropriately assess the representativeness of fixed monitoring sites within an AQMN and it complements the guidelines set by European legislation on air pollutants monitoring at fixed stations, which could help to tackle efforts to improve the air quality management.

Introduction

Atmospheric pollution is globally recognized as a major environmental risk to human health. World Health Organization reported 4.2 million deaths every year, approximately, as a result of exposure to ambient outdoor air pollution (http://www.who.int/airpollution/en/). It comprises a wide variety of pollutants, both in gaseous and particulate phases (Landis et al., 2019). Atmospheric particulate matter covers a complex mixture of solid and liquid particles of organic and inorganic substances suspended in the ambient air (Chernyshev et al., 2019; Tsiflikiotou et al., 2019) and it is linked to a growing concern for public health as a consequence of their injurious impacts (Kim et al., 2018; Yoo et al., 2018; Zhu et al., 2019). For this reason, the European Union develops Air Quality Directives (Directive 2008/50/EC) for setting air quality objectives in or-

E-mail: david.galan@isciii.es

https://doi.org/10.1016/j.jes.2020.09.009
1001-0742/© 2020 The Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences. Published by Elsevier B.V.

der to reduce harmful effects on human health and environment and it lays down measurement methods for assessing ambient air quality in the Member States, such as fixed measurements (by automatic observation instruments), modelling techniques and indicative measurements, so as their combination (Pisoni et al., 2019). Member States establish air quality monitoring networks (AQMN) in their territories for verifying compliance with those air quality objectives, therefore, the air quality management relative to assess the atmospheric pollutant concentrations should be suitably managed in any AQMN.

AQMN plays a remarkable role in the development of monitoring strategies (Mofarrah and Husain, 2010) and for assisting authorities in decision making. Consequently, the AQMN design is a transcendent factor for assessing human being exposure to air pollutants (Pope and Wu, 2014). AQMN consists of fixed measurement stations. These should be attributed conveniently in the domain of interest for picturing air pollutant information appropriately. Thus, one of the keys in the AQMN layout is the distribution of monitoring stations as well as the determination of a sufficient and confident number of sampling points for carrying out those air quality measurements. These features are associated with the air quality network management, which should focus on monitoring the largest possible target study area, covering the maximum information on the spatial gradient of air pollutants, and reducing the fixed stations within the AQMN to a reliable and non-redundant number. Thus, the network must not duplicate information on air pollutants. Nonetheless, given that new emission sources of air pollutants can emerge in a target territory, a process of AQMN optimization should be conducted to guarantee the presence of non-redundant fixed monitoring stations within the AQMN, which it would be translated in an improvement of its performance.

In order to solve the air quality management issues regarding the identification of possible redundant fixed measurement stations within an AQMN, the use of combined statistical tools, such as chemometric and artificial neural network techniques could be addressed, which contribution would lead to improve the target region or city's air quality assessment program. These statistical methods could be applied to both air quality data sampled by passive methodology and those measured by fixed methods in AQMN. Although the acquisition of the passive samplers supposes a low economical investment and allows to sample at a wide study area using a large number of sampling points (Bozkurt et al., 2018), this method is not recognized among the reference methods designated by European Directive. Nevertheless, whereas fixed techniques (automatic observation instruments) used in AQMN for monitoring air quality would lead to a less number of sampling points than the passive methods, those fixed measurement methods would comply with those proposed by European Directive.

In recent decades, some studies exploited analysis methods to improve the air quality management programs using historical data series of atmospheric pollutants measured by AQMN. Zhao et al. (2015a, 2015b) exposed a performance assessment and adjustment program for Shanghai's AQMN and considered historical data for different atmospheric pollutants. The research was conducted on a short-period study (from 1 January to 22 August 2014). They chose a combination of principal components analysis and assignment method for identifying redundant fixed stations. On the other hand, they employed a clustering analysis to verify the results obtained from the aforementioned analysis techniques. Briefly, they addressed a proposal to take full advantage location of fixed measurement stations.

Wang et al. (2018) expanded the Zhao et al. (2015a, 2015b) approach in order to optimize the AQMN layout. They executed a correlation analysis, principal component analysis, assignment method, clustering analysis and correspondence analysis using historical data series of air pollutants from Xi'an city's AQMN from 1 January to 31 December 2016. They carried out a second verification of the proposed fixed stations.

The aforementioned studies led to the identification of non-redundant fixed monitoring stations within an AQMN, nevertheless, the practical application of the proposed approaches was not validated by using air pollutant data non included in the development of these approaches. Similarly, they did not assess spatial information percentage that was lost when redesigning the AQMN due to the removal of redundant fixed measurement stations.

The guiding thread of the present work is to propose a methodology for improving the AQMN layout. For this goal, PM_{10} particles (less than 10 µm in diameter) data recorded by the Community of Madrid's AQMN between 2008 and 2017 were examined. The evaluation of PM_{10} particles is of remarkable importance because it is mandatory in European member states. European Directive set a limit value for ambient air of PM_{10} 40 µg/m^3 on average in a calendar year and PM_{10} 20 µg/m^3 according to WHO (Franzin et al., 2020).

As one innovative tool of this study, the proposed methodology was validated using PM_{10} particles data measured by AQMN during 2007 and 2018.Target pollutant data registered by all currently fixed stations vs those proposed by this methodology were compared and discussed. Another key aspect in this work pointed to the lost spatial information once that the non-proposed fixed stations by this methodology were removed within AQMN. For that, spatial distribution maps for PM_{10} particles were developed by using geographical information systems (GIS) in order to assess the similarity degree between both postulates (current vs proposed fixed stations) using a simple linear regression technique.

The present work aims to develop a combined methodology to: (i) evaluate the presence of possible redundant fixed measurement stations within AQMN and preliminary identify those non-redundant ones, (ii) propose those more representative fixed stations within AQMN, (iii) validate the new proposed design, (iv) assess the loss of spatial information due to the removal of redundant fixed monitoring stations within AQMN, and (v) address possible applications to the air quality management.

1. Materials and methods

1.1. *Area of study*

The present study was conducted in the Community of Madrid (Spain), with a population over 6,500,000 inhabitants, approximately, and a surface slightly greater than 8000 km^2 (http://www.ine.es). It is consisted of 179 municipalities and located in the center of the Iberian Peninsula, therefore, it can be identified as a representative interior region of South Europe.

The researched period covered from 2008 to 2017 and focused on average annual PM_{10} particle concentrations monitored at fixed measurement stations of the Community of Madrid's AQMN.

During the researched period, Community of Madrid's AQMN was constituted by 23 fixed measurement stations distributed in six homogeneous areas of the territory, which are divided into three agglomerations (Corredor del Henares, Urbana Sur, Urbana Noroeste) and three rural areas (Cuenca del Tajuña, Cuenca del Alberche and Sierra Norte). They were managed by the regional government, which assures the measured data validation and network maintenance (Fig. S1). The station organization is based on two criteria: (i) location (Urban, Suburban or Rural) and (ii) main pollution source (Traffic, Industrial or Background). Information about all fixed stations is shown in Table S1. Ambient air PM_{10} particles were measured at all fixed stations except for the stations located in Algete, Alcorcón, Valdemoro and Collado Villalba (between 2015 and 2017), San Martín de Valdeiglesias (2015 and 2016) and Villarejo de Salvanés (2008 and between 2015 and 2017). Data on target air pollutant was acquired from the Community of Madrid's open data portal (http://www.comunidad.madrid).

1.2. Evaluation of the presence of possible redundant fixed measurement stations within AQMN and preliminary identification of non-redundant fixed stations

As an action preceding the preliminary identification of non-redundant fixed stations within the investigated AQMN, a study to verify whether the air pollution information provided by each of the 23 fixed stations for measuring PM_{10} particles could be linked among several stations was carried out, which would confirm the presence of possible redundant fixed stations. This objective was examined by correlation analysis (COA) and was applied on average annual PM_{10} concentrations between 2008 and 2017.

Correlation is defined as the existing relation between phenomena or things or between mathematical or statistical variables which tend to vary, be associated, or occur together in a way not expected by chance alone, according to the Merriam-Webster dictionary (2019). Therefore, this method measures the weight of the reciprocal relationship between two variables (Mikheev and Kazakov, 2017) and it is an efficient technique to reveal the complex relationship existing among diverse datasets (Zhang et al., 2017), which supports to understand the connectivity degree between two fixed stations regarding recorded PM_{10} data. High COA values would explain a strong degree of connection between the PM_{10} dataset measured by two fixed stations. This means that the presence of emission sources for target pollutant at both stations has a strong probability of similarity. Although the COA does not identify redundant fixed stations, it is a useful tool as a preliminary study to confirm the existence of no-efficient stations.

Once confirmed that the AQMN displays redundant fixed stations, a pre-identification study of non-redundant fixed monitoring stations was conducted. For this aim, a principal component analysis (PCA) was tested. PCA was created before the Second World War, regardless, the wider application of this method only occurred in the 1960s, during the "Quantitative Revolution" in the Natural and Social Sciences. (Maćkiewicz and Ratajczak, 1993). Liu et al. (2018) widely clarified the mathematical development which explains the PCA. Briefly, PCA consists of a classic chemometric technique and it applies an rotational algorithm to classify a dataset of possibly correlated variables into a smaller set of values of linearly uncorrelated variables (Chen et al., 2019). The new set of values preserve most of the information of the original dataset (Li et al., 2019). In this process, it was used the varimax method and adopted an eigenvalues > 1 (the Kaiser Criterion), as criteria in order to select the principal components (PCs) which explained most of the cumulative variance. Subsequent PCs were accounted until reaching up cumulative variance $> 90\%$.

Then, a multiple linear regression technique (MLR) was applied to the new dataset extracted by the PCA process in order to quantify the weight of each fixed monitoring station within the PCs. A fixed station with an elevated weight suggests that this one is relevant for monitoring target pollutant and, therefore, its removal within AQMN should not be taken into account.

1.3. Proposal of more representative fixed measurement stations

The results reached by the combined PCA-MLR technique were tested by an artificial neural network (ANN) analysis (Li et al., 2014). This application is based on a method of partitive clustering (k-means clustering with 10 maximum iterations) and, for addressing its implementation, it is required to previously establish indicative variables. In this case, those variables reported by Galán Madruga et al. (2018) were used. Briefly, Euclidean distance was employed as the spatial indicator, and cluster standard deviation, as the cluster membership identifier.

The ANN analysis identifies clusters constituted by fixed monitoring stations with common features in any variable (in this case, the average annual PM_{10} concentration), which appoints empirical homogeneous groups, and as dissimilar as possible from stations associated with other clusters (Maione et al., 2019). Govender and Sivakumar (2020) widely described the execution of the clustering analysis. Finally, simple statistical analysis techniques were addressed to evaluate the annual representativeness of the proposed fixed stations.

1.4. Validation of the proposed fixed stations within assessed AQMN

Once more representative fixed monitoring stations within the Community of Madrid's AQMN were proposed, a study for evaluating confidence degree of them was led. While the reported methodological framework encompassed from 2008 to 2017, the validation process covered 2007 and 2018.

1.5. Assessment the lost spatial information due to the removal of non-proposed fixed monitoring stations within investigated AQMN

The maps of spatial annual PM_{10} particle distribution in the Community of Madrid surface over the studied period were built using data coming from (i) pre-identified and non-redundant fixed stations included in the AQMN and (ii) those stations proposed by combined COA/PCA-MLR/ANN technique. According to other research studies (Liu et al., 2017), in order to picture those points where concentrations of the target air pollutant were not measured, those levels were extrapolated by using a Geographic Information System (GIS), and the Kriging method as a geostatistical estimation tool (Beauchamp et al., 2018).

The PM_{10} iso-concentration maps were generated using Surfer for Windows (Win32): Surface Mapping System, v.6.04. (Golden Software, Inc., Golden, CO, USA). Simple linear regression analysis was used to quantify the percentage of similarity of the interpolated spatial information between pre-identified vs proposed fixed monitoring stations.

1.6. Statistical analysis

Analysis of the dataset was performed with the software IBM SPSS Statistics v22.0 (IBM Corp., Armonk, NY, USA). Statistical significance between paired samples (non-redundant vs proposed fixed monitoring stations) was evaluated by a paired t-test (Yang et al., 2018).

2. Results and discussion

2.1. Evaluation of the presence of possible redundant fixed measurement stations

Table S2 shows the resulting Pearson's coefficients once applying the COA on average annual PM10 concentrations measured at all fixed stations of Community of Madrid's AQMN between 2008 and 2017. COA disclosed values of Pearson's coefficient of correlation between 0.7 and 1.0 in most cases, which evidenced a strong correlation among many pairs of stations.

One common interpretation of the Pearson's coefficient values was proposed by Dancey and Reidy (2007). They classified as zero (value 0), weak (±0.1˜ ±0.3), moderate (±0.4–0.6), strong (±0.7–0.9) and perfect (±1) the association degree between two correlated variables. Applying 0.7 as a recommended cut-off value for Pearson's coefficient of correlation, a more detailed analysis regarding the representativeness of coefficient of correlation showed that in urban traffic stations, Pearson's coefficients < 0.7 explained 71.43% of the total representativeness, while suburban and rural background stations, Pearson's coefficients > 0.7 explained 60.00% and 70.00%, respectively, reaching up 100.00% at urban industrial and urban background stations, as it is shown in Fig. 1.

In most cases, the COA results revealed that there was a clear and strong correlation between the fixed monitoring sites for measuring PM_{10}, which could evidence the presence of redundant fixed stations within researched AQMN. In order to solve this air quality management issue, the application of a preliminary study to identify potential redundant fixed monitoring sites is justified, recognizing the need for effective monitoring sites (Tseng and Chang, 2001).

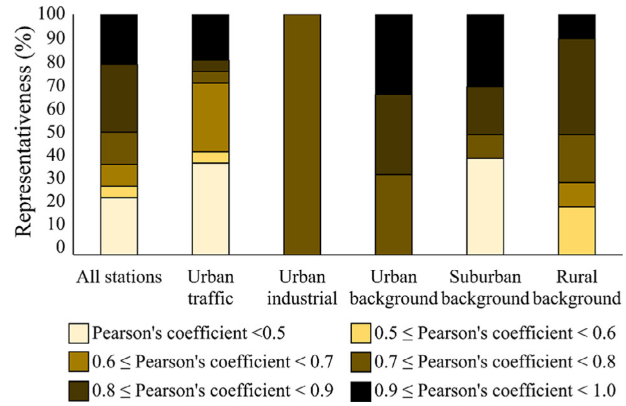

Fig. 1 – Intervals of Pearson's coefficient classified by station category (2008–2017).

Table 1 – Results of the PCA technique for annual PM_{10} data monitored at the 23 fixed stations in Community of Madrid's AQMN (between 2008 and 2017).

Location	PC1	PC2	PC3
ALH	**0.948**	0.167	0.219
ALB	0.528	0.410	0.730
ALG	**0.887**	-0.181	0.414
AGR	0.643	0.632	0.377
COS	**0.872**	0.342	0.050
RVM	**0.878**	-0.007	0.395
TDA	**0.941**	0.287	0.137
ALC	0.679	0.701	0.094
ARJ	**0.936**	0.290	0.060
FUE	**0.973**	0.203	0.066
GET	**0.839**	0.238	0.142
LEG	**0.917**	0.335	0.147
MOS	**0.930**	0.362	0.005
VAL	**0.879**	0.209	0.350
CLV	0.043	-0.167	**-0.958**
COV	0.460	-0.394	**0.788**
MAJ	0.006	**-0.989**	0.046
EAT	**0.943**	0.161	0.090
GDS	**0.909**	0.152	0.315
SMV	**0.875**	0.141	0.168
VDP	0.417	**0.860**	0.122
ODT	**0.863**	0.322	0.274
VDS	**0.942**	0.062	0.305
Eigenvalue	17.23	2.69	1.82
Variance (%)	74.93%	11.68%	7.91%
Cumulative variance (%)		86.61%	94.52%

Note: Higher factor loadings are marked in bold.

2.2. Preliminary identification of non-redundant fixed measurement stations

The rotation factor matrix revealed by the PCA technique for PM_{10} particles at the 23 fixed stations of the Community of Madrid between 2008 and 2017 is shown in Table 1. A total of three PCs were defined, reaching a cumulative variance higher

Table 2 – Results of combined PCA-MLR technique for annual PM_{10} data monitored at the 23 stations at Community of Madrid (from 2008 to 2017).

Location	PC1	PC2	PC3	Total
ALH	4.57%	0.08%	0.12%	4.77%
ALB	1.41%	0.51%	1.35%	3.28%
ALG	4.00%	0.10%	0.43%	4.53%
AGR	2.10%	1.21%	0.36%	3.67%
COS	3.86%	0.35%	0.01%	4.22%
RVM	3.91%	0.00%	0.40%	4.31%
TDA	4.49%	0.25%	0.05%	4.79%
ALC	2.34%	1.48%	0.02%	3.84%
ARJ	4.45%	0.25%	0.01%	4.72%
FUE	4.80%	0.12%	0.01%	4.94%
GET	3.57%	0.17%	0.05%	3.79%
LEG	4.27%	0.34%	0.05%	4.66%
MOS	4.39%	0.40%	0.00%	4.79%
VAL	3.92%	0.13%	0.31%	4.36%
CLV	0.01%	0.08%	2.33%	2.42%
COV	1.07%	0.47%	1.58%	3.12%
MAJ	0.00%	2.95%	0.01%	2.96%
EAT	4.51%	0.08%	0.02%	4.61%
GDS	4.19%	0.07%	0.25%	4.52%
SMV	3.89%	0.06%	0.07%	4.02%
VDP	0.88%	2.24%	0.04%	3.16%
ODT	3.78%	0.31%	0.19%	4.28%
VDS	4.51%	0.01%	0.24%	4.75%
Variance (%)	74.93%	11.68%	7.91%	94.52%

than 90% (94.52%). The most dominant PC pointed to the PC1, given that this one explained most of the original information of the primary dataset. So, PC1 explained 74.93% of the original variance, while PC2 and PC3 reached up 11.68% and 7.91%, respectively.

In order to clarify the new dataset obtained by the PCA technique, the criteria reported by Abdul Halim et al. (2018) in terms of factor loading (strong > 0.75) was considered. So, the PC1 component mainly comprised urban traffic locations, suburban and rural background stations (25% for each category) while urban background sites showed 12.50% and urban industrial and suburban traffic stations 6.25%, respectively. PC2 consisted of suburban and rural background stations and urban traffic sites was included in PC3. Note that similar values of factor loadings were found for Arganda del Rey and Alcorcón (PC1 and PC2).

It is relevant highlighted that the results of PCA technique are qualitative, as they can only distinguish variables that tend to appear linked from those ones that do not (Comero et al., 2009), therefore, and in order to sort out that limitation, a combined PCA-MLR analysis was run, using the PCA results. The goal of this combined technique was to quantify the weight of each fixed station within each principal component (see Table 2). To select those more representative fixed monitoring stations within AQMN, the total weight of each fixed station within the network was considered for sustaining 94.52% of the dataset original information. So, any fixed station should reach a contribution higher than the mean value (4.11%) in order to be included within the list of non-redundant stations.

As result, the next fixed stations were selected: ALH, ALG, COS, RVM, TDA, ARJ, FUE, LEG, MOS, VAL, EAT, GDS, ODT and VDS, providing a decrease of 39% relative to the total number of fixed stations. Herein, this set of stations is referred to as ST1. Representativeness higher than 10% was observed at urban traffic stations (13%), suburban (17%) and rural (13%) background stations, while urban background, urban industrial, suburban traffic stations reached lowest values (9%, 4% and 4%, respectively). In more detail, those stations selected within the traffic, industrial and background category represented 50, 50 and 69%, respectively, while those urban, suburban and rural types depicted 50, 83 and 60%, respectively.

The assessment of the spatial representativeness of fixed monitoring stations within an AQMN is a prominent subject, due to that it is linked to health risk assessment, population exposure to air pollution, the design of AQMN, air quality modelling and data assimilation (Martin et al., 2015).

The fixed stations framed in the ST1 list are linked to areas characterized by a specific air pollution behavior, therefore, each area should only have one monitoring site (Pires et al., 2009).

2.3. *Proposal of more representative fixed measurement stations within researched AQMN*

The final result of the combined PCA-MLR technique allowed to simplify the total number of stations of the AQMN (14 instead of 23 stations). In order to further improve the AQMN performance, a clustering analysis was applied on average annual PM_{10} concentrations monitored along the study time at those stations included inside the ST1 list, for identifying those more relevant ones, assuring to keep most of the original information from ST1 package. The clustering analysis was tested for seven groups (number of cluster: 2, 3, 4, 6, 8, 10 and 12, respectively). Then, a simple linear regression analysis was conducted between the current average annual concentrations of PM_{10} vs those estimated by clustering analysis at the stations tied to the ST1 (Fig. S2), appointing the group 6 as the more representative cluster, due to that it was the first group with a coefficient of determination higher than 0.990 (0.994 ± 0.003, expressed as an average value, see Table S3). The maximum value for standard deviation and Euclidean distance in this cluster was 0.260 ± 0.227 and 0.908 ± 1.250 $\mu g/m^3$, respectively.

Given that Euclidean distance can be used as a distinctive variable among elements of the same group, according to Penkova (2017), those stations included within cluster 6 with the lowest Euclidean distances were considered as the more representative stations within the AQMN (Table S4). For these stations, and in order to provide a more plausible interpretation of the results of cluster 6, the whole representativeness for each station was calculated in terms of the sum of annual partial representativeness (Fig. 2). The total representativeness of each station expresses the number of times that each one is represented in the studied period. Finally, the representativeness was studied separately in terms of quartile (6.3, 9.0, 9.0, and 10.0 for quartile 1, 2, 3 and 4, respectively).

Analyzing Fig. 3, in order to recognize the more representative fixed stations within Comunidad de Madrid's AQMN for measuring PM_{10} particles, it was considered: (1) broadly, each quartile must have some fixed station selected, (2) those stations pictured in the different quartiles showing higher repre-

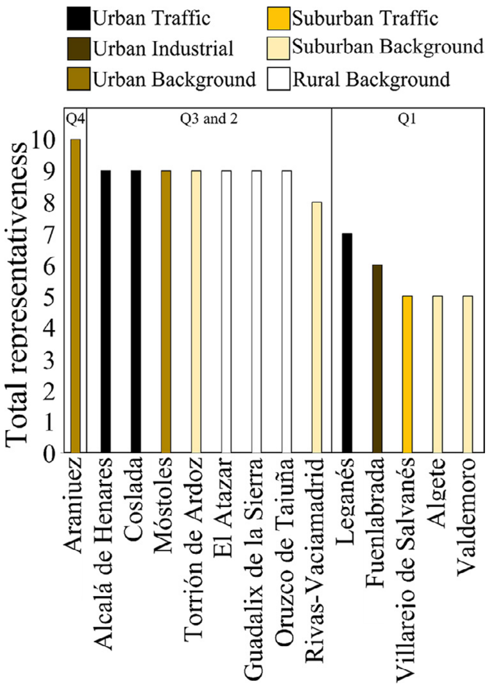

Fig. 2 – Overall representativeness of fixed stations of the Community of Madrid for monitoring air PM_{10} levels between 2008 and 2017.

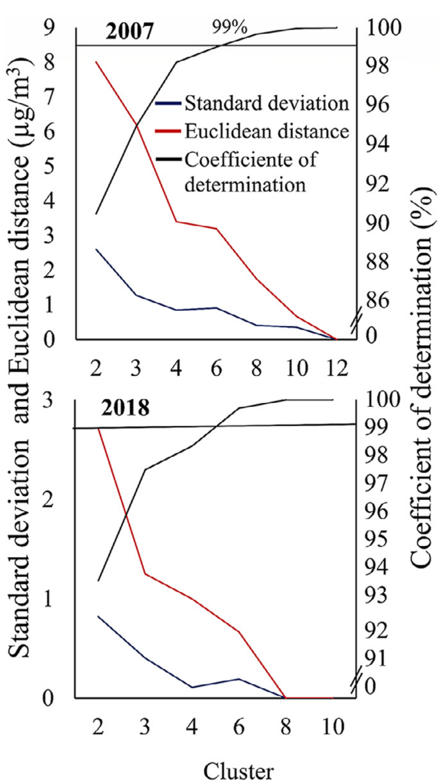

Fig. 3 – Cluster vs. selection criteria during 2007 and 2018.

sentativeness should be selected with the next priority Q4 > Q3,2 > Q1, without reiterating its category in terms of location and main pollution source, (3) if various stations display the same representativeness within the same quartile, those Municipalities with the biggest inhabitant number should be selected in order to protect the biggest population density sites of potentially harmful effects coming from PM_{10} particles.

Therefore, the first relevant station that should be recognized is Aranjuez (quartile 4), given that this one pictures the highest occurrence among all non-redundant stations (ST1 list) within Comunidad de Madrid's AQMN. In the quartile 3 and 2, while most of the stations showed the same representativeness, Alcalá de Henares and Móstoles were selected given that they hold the biggest inhabitant number (see Table S5). In the quartile 1, on the one hand, whereas Leganés should be identified as a paramount station, Fuenlabrada was selected, given that Leganes would repeat the category of the station located at Alcalá de Henares (urban traffic station) and, on the other hand, Villarejo de Salvanés was also selected because it is the only one categorized as suburban traffic station.

Up to here, the identification study has covered the next categories: urban traffic sites, urban industrial and background station and suburban traffic. Therefore, in order to encompass all station classes, a suburban and rural background station must be incorporated. Among all suburban background sites, Algete was picked due to that the rest of the stations were nearby to other already selected. In the case of rural background stations, the three locations fell into quartiles 3 and 2. El Atazar was not considered due to the lower population, while Guadalix de la Sierra and Oruzco de Tajuña were considered as relevant, given that the first one would cover the Madrid North zone (Sierra Norte Area) and the second one would maintain the only rural background site in the Cuenca de Tajuña Area.

According to the developed work using the sum of annual partial representativeness between 2008 and 2017, the more relevant PM_{10} particles monitoring stations within the Community of Madrid's AQMN were: ALH (urban traffic station), ALG (suburban background), ARJ (urban background), FUE (urban industrial), MOS (urban background), GDS (rural background), ODT (rural background) and VDS (suburban traffic). Herein, this set of stations is referred to as ST2. Note that the number of fixed stations reached a decrease up to 39% between currently fixed stations of the Community of Madrid's AQMN (23) and those no-redundant stations (14), and a slightly higher decrease than 65% regards to those more representative sites (8). Given that the concentration gradient of air pollutants in urban environments is relatively heterogeneous, the urban stations provision up to 50% of the total number of relevant stations into the ST2 pack.

As a consequence of the wide time range employed in the approach (one decade), a study on temporal influence was addressed. For that, the average PM_{10} concentrations obtained between 2008 and 2017 at each non-redundant station (ST1 list) were used, conducting a partitive clustering analysis similar to the reported in this Section. Thus, Fig. S3 pictures the calculated maximum values for the standard deviation, Euclidean distance, as well as values of coefficient of determination for each examined cluster. The interpretation of this Figure frames cluster 6 as the first cluster with a coefficient of determination > 0.990 (0.996). The determined maximum value for standard deviation and Euclidean distance in this cluster (0.317 and 0.967 µg/m^3, respectively) fell within the interval calculated by annual analysis. Therefore, encompassing the whole studied period, cluster 6 was selected for appointing the most paramount fixed stations (Table S6). ALH, ARJ, FUE, MOS, GDS, ODT, VDS, EAT and TDA were identified as primordial stations. Herein, this set of stations is referred to as ST3.

While ST3 package reached a slightly higher number of stations than ST2 pack, the similarity degree between both postulates attained up to 78%, therefore, the stations proposed in ST2 pack were affected by a temporal variability around 20%, thereby exhibiting an acceptable temporal influence.

The adequate identification of non-redundant and more relevant fixed monitoring sites would improve the AQMN management without losing environmental information, as a consequence of the lack of measurements at those not considered stations (Christakos et al., 2017), given that those could be covered by spatial extrapolation methods and/or statistical techniques (Aguilera et al., 2016).

2.4. *Assessment of the lost spatial information due to the removal of non-proposed fixed monitoring stations within investigated AQMN*

The pollution information for PM_{10} particles recorded by those stations included in ST1 and ST2 pack, respectively, was employed for picturing their annual spatial distribution in the ambient air of the Community Madrid by using a GIS. The relationship between both postulates was found by simple linear regression analysis. For that, those remaining fixed stations included in ST1 list and not in ST2 pack were taken into account for reaching a total of 14 fixed stations. Real concentrations were used at those stations incorporated within ST1 and levels estimated by clustering analysis at those linked to ST2 list.

In this sense, the maps picturing average annual concentration of PM_{10} particles between 2008 and 2017, at those stations included in ST1 and ST2 pack, are illustrated in Fig. S4.

It is important to note that the major objective of the maps is not to represent overall pictures of the spatial distribution of PM_{10} levels in the Community of Madrid ambient air, due to that pollution data is lacking from several municipalities. Rather, the goal is to compare the spatial information obtained between both postulates and to confirm the validity of the proposed fixed stations (ST2 set) as representative locations for measuring PM_{10} particles in the air of the Comunidad de Madrid.

Annual spatial information between 2008 and 2017 for both postulates is highly correlated, reaching a minimum and maximum similarity value up to 97.94% and 100.00%, respectively (average percentage 98.91%±0.78%). In terms of overall spatial distribution, average PM_{10} levels over the studied period at ST2 vs ST1 showed a similarity percentage of 98.38%, following the annual spatial information (Fig. S5). Therefore, the reported methodology for identifying the minimum number and reliability of stations non-redundant fixed within the Community of Madrid's AQMN can be developed using annual or global data monitored along a decade.

Table 3 – Current and estimated PM_{10} concentration at the fixed stations for 2007 and 2018.

Location	2017 Current	Estimated	2018 Current	Estimated
Alcalá de Henares	44	46	19	19
Algete	22	21	–	–
Coslada	–	50	–	21
Rivas-Vaciamadrid	45	46	20	20
Torrejón de Ardoz	47	46	—	21
Aranjuez	29	28	14	14
Fuenlabrada	36	36	20	20
Leganés	47	46	20	20
Móstoles	30	28	17	17
Valdemoro	–	39	–	–
El Atazar	19	21	–	12
Guadalix de la Sierra	27	28	13	14
Orusco de Tajuña	25	28	14	14
Villarejo de Salvanés	30	28	–	–

2.5. *Validation of the proposed methodological framework*

An essential issue to value in the modelling studies would be the assessment regards the confidence of the proposed approach. For that objective, a clustering analysis was conducted, according to Section 2, on actual PM_{10} particle data measured during 2007 and 2018. As it is observed in Fig. 3, the first group with a coefficient of determination higher than 0.990 was cluster 6.

Then, in order to assess the reliability of the proposed methodology, actual concentrations monitored at the stations incorporated into the ST1 pack and those estimated by clustering analysis at the stations included within the ST2 package were compared by simple linear regression, considering the remain stations within ST1 in the ST2 pack. Both postulates were highly correlated ($r^2 = 0.973$ for 2007 and 0.990 for 2018) and, therefore, non-significant differences were found between paired samples of actual and estimated PM_{10} particles concentration during 2007 and 2018 (Table 3). A pragmatic decision for establishing the relationship between two datasets (actual vs estimated) is dependent on bias determination (Eurachem, 2012). As a consequence, a mean bias value of 4.8%± 3.4% (2007) and 1.2%±1.9% (2018) was obtained between both packages. Therefore, it can be concluded that the proposed fixed monitoring stations (ST2 package) would provide reliable measurements of PM_{10} particles in the air of the Community of Madrid and comparable spatial information respect to that one supported by actually fixed stations.

Based on the aforementioned results, the combined COA, PCA and clustering technique is statistically effective for multivariate segmentation studies (Remesan et al., 2018). Besides, extrapolation results are a good predictive tool for procuring relevant spatial information on the AQMN stations. While the identification of suitable fixed monitoring sites and the removal of redundant stations can improve the management of resources within an AQMN, without losing spatial information due to the lack of measurements at the removed stations, the lost data can be covered by adequate spatial extrapolation methods (Aguilera et al., 2016).

As a consequence of the existing different types of air pollution sources for PM_{10} particles, the AQMN's performance should be regularly evaluated to infer the efficiency of the monitoring system (Castro and Pires, 2019).

2.6. Applications to the air quality management

In order to support possible applications of the proposed methodological framework toward environmental management, in special with air quality management, the next points are exhibited.

2.6.1. Assuming the AQMN is already established

Point 1: Application in order to re-layout the AQMN.

Since the major role of an AQMN is to evaluate with accuracy the national, regional, urban or rural air quality, in order to publish in real-time information about the air pollutants that affect public health, a relevant application of this work would focus on the improvement of existing AQMN layout.

On the one hand, taking into account that one of the greatest handicaps in the design of an AQMN, beside the economic investment, would point on the identification of the best potential site in order to locate the fixed measurement stations within the area of interest (Wu and Bocquet, 2011), the presented methodology would help to select those more representative fixed stations within the AQMN. So, those stations found to be inefficient or redundant should be removed of the network due to their location at inappropriate sites would likely affect data validity (Kao and Hsieh, 2006), which would lead to executing the AQMN's management program solely to those efficient stations.

This enforcement would not only be limited to atmospheric particles but it could be applied to any air pollutant measured by the AQMN.

On the other hand, as a consequence of the evolution of economic development and urbanization, among other factors (Li et al., 2017), mainly in an urban and suburban environment, the introduction of new emission sources of atmospheric pollutants could lead to the modification of the spatial patterns of those pollutants in the ambient air of a determined region, due to the increase of vehicles and industrial activities (Athira et al., 2018). Given that, countries and regional authorities have the legal obligation of designing and assessing the impacts of air quality plans whenever exceedances occur and they generally lack the proper tools to do so (Thunis et al., 2016), the implementation of this methodology would re-assess efficiency of AQMN after a wide period of time in order to review its performance in accordance with a new emission map.

Point 2: Application as regards the guidelines set by European legislation on air pollutants monitoring at fixed stations.

Environmental regulation is an important instrument to control air pollution (Wang et al., 2019) and it intends to reduce pollution emissions and improve environmental quality (Chong et al., 2017). In this frame, Directive 2008/50/EC, on ambient air quality and cleaner air for Europe, lays down considerations about the location of sampling points for measuring atmospheric pollutants, at both macro- and microscale levels, as well as criteria for determining the minimum number of sampling points for fixed measurement in order to assess compliance with limit values for the protection of human health. Both criteria are used for establishing AQMN within the Member States territory.

The minimum number of sampling points, within a territory, is set in function of next criteria: (i) the population of the agglomeration or zone which is needed to control, and (ii) whether maximum concentration exceeds the upper assessment threshold or it is between the upper and lower assessment thresholds for target air pollutant. Nevertheless, the European legislation does not set criteria or methods for evaluating the representativeness of the fixed stations within an AQMN. The execution of the aforementioned methodology would help to solve that management issue relative to the AQMN design.

2.6.2. Assuming the AQMN is not yet established

As it can be deduced from aforecited, an AQMN needs to be designed efficiently, with a minimum number of sites that guarantee the identification and the spatiotemporal patterns of air pollution and the information supplied relative to human being exposition to air pollutants (Hao and Xie, 2018). Therefore, the selection of the location of the sampling points for establishing an AQMN is an extraordinary relevant issue. To establish an AQMN, the air quality control strategies require previous convenient information about the dispersion of the air pollutants within the target study surface, as well as a thorough understanding of major emission sources. (Li et al., 2017).

Automatic observation instruments used in AQMN present geographic limitations due to the strong economic investment required for their implantation and maintenance (Schneider et al., 2017). For this reason, in order to obtain wider information about air pollutant dispersion and emission sources, many authors have used the passive methodology, given that a lower investment is needed, in order to draw the spatial behavior of air pollutants using an elevated number of sampling points. The designed methodology in the present work would conduct to identify the more representative points for appointing them as possible fixed measurement points within the AQMN, although this approach would only be useful for gaseous air pollutants and not for particulate matter due to those compounds do not diffuse in the air mass. Therefore, another application of this methodology would conduct to establish reliable and fixed sampling points within an AQMN using passive samplers instead of reference methods.

Overall, the implementation of the proposed methodology on AQMN would reduce significantly expenses associated with the maintenance of the fixed stations, and the establishment of quality assurance activities, such as calibration operations of sampling and measurement equipment and acquisition of reference materials, among other, due to the improvement in the re-utilization of the stations, removing those inefficient ones within the AQMN, while securing an adequate representation the atmospheric pollutant concentrations in the domain of interest, which accomplishes the purposes of the network (Araki et al., 2015) and further achieving health benefits through comprehensive and rigorous air quality management practice (Wu et al., 2019). Nevertheless, this decrease of expenses would not entail a decline of the original information provided by the AQMN and would guarantee the adequate characterization of the air quality in the monitored region. Besides, the broken down approach in this study could constitute a useful appliance for developing intercomparison exercises of methodologies estimating the spatial representativeness of fixed stations within AQMN.

3. Conclusions

The results reached in this work sustain the effectiveness of the proposed methodology in order to optimize the AQMN's layout, by assessing the possible presence of redundant fixed monitoring stations, identifying those non-redundant and selecting the most representative ones for improving the AQMN design. It is also able to appropriately assess the representativeness of fixed monitoring sites within an AQMN and help to tackle efforts to improve the environmental performance of air quality management.

The proposed methodology has relied on the combination of chemometric techniques and neural artificial analysis applied on air pollutant data monitored at fixed stations belonging to an AQMN. An innovative aspect is the validation of the practical application of the proposed methodology, which was conducted on air pollutant data measured over periods not included in the development of the enunciated methodology.

As relevant applications on air quality management, the implementation of the enunciated methodology is particularly paramount for AQMN with an elevated number of fixed monitoring stations, given that the economic cost derived from acquisition and maintenance of the measurement instruments could reach significate expenses. Similarly, the proposed methodology would be effective in the establishment of new networks, by limiting the required investment in regard to the implementation in the number of fixed stations. So, the AQMN layout could be adjusted according to the designed program in order to reduce expenses, investing remain cost in the monitoring of other air pollutants within the area of interest, which would assure the main role of AQMNs, given that it would provide valuable and additional data on human being's exposure to atmospheric pollution and spatial and temporal patterns of air pollutants. Therefore, the proposed methodology is posed as a useful tool to support the decision making about more-effective strategies for addressing environmental issues in a context of limited resources.

Appendix A. Supplementary data

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jes.2020.09.009.

REFERENCES

Abdul Halim, N.D., Latif, M.T., Ahamad, F., Dominick, D., Chung, J.X., Juneng, L., et al., 2018. The long-term assessment of air quality on an island in Malaysia. Heliyon 4, e01054.

Aguilera, R., Sabater, S., Marcé, R., 2016. A methodological framework for characterizing the spatiotemporal variability of river water-quality patterns using dynamic factor analysis. J. Environ. Inf. 31 (2), 97–110.

Araki, S., Iwahashi, K., Shimadera, H., Yamamoto, K., Kondo, A., 2015. Optimization of air monitoring networks using chemical transport model and search algorithm. Atmos. Environ. 122, 22–30.

Athira, V., Geetha, P., Vinayakumar, R., Soman, K.P., 2018. DeepAirNet: applying recurrent networks for air quality prediction. Proc. Comput. Sci. 132, 1394–1403.

Beauchamp, M., Malherbe, L., de Fouquet, C., Létinois, L., Tognet, F., 2018. A polynomial approximation of the traffic contributions for kriging-based interpolation of urban air quality model. Environ. Model. Softw. 105, 132–152.

Bozkurt, Z., Üzmez, Ö.Ö., Döğeroğlu, T., Artun, G., Gaga, E.O., 2018. Atmospheric concentrations of SO2, NO2, ozone and VOCs in Düzce, Turkey using passive air samplers: Sources, spatial and seasonal variations and health risk estimation. Atmos. Pollut. Res. 9, 1146–1156.

Castro, M., Pires, J.C.M., 2019. Decision support tool to improve the spatial distribution of air quality monitoring sites. Atmos. Pollut. Res. 10, 827–834.

Chen, Y., Zhang, L., Zhao, B., 2019. Identification of the anomaly component using BEMD combined with PCA from element concentrations in the Tengchong tin belt, SW China. Geosci. Front. 10, 1561–1576.

Chernyshev, V.V., Zakharenko, A.M., Ugay, S.M., Hien, T.T., Hai, L.H., Olesik, S.M., et al., 2019. Morphological and chemical composition of particulate matter in buses exhaust. Toxicol. Rep. 6, 120–125.

Chong, Z., Qin, C., Ye, X., 2017. Environmental regulation and industrial structure change in China: integrating spatial and social network analysis. Sustainability 9, 1465.

Christakos, G., Angulo, J.M., Yu, H.L., Wu, J., 2017. Space-time metric determination in environmental modeling. J. Environ. Inf. 30, 29–40.

Comero, S., Capitani, L., Gawlik, M., 2009. Positive Matrix Factorisation (PMF). An introduction to the chemometric evaluation of environmental monitoring data using PMF. EUR 23946 EN. European Commission. JRC. Sci. Tech. Rep.. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/positive-matrix-factorisation-pmf-introduction-chemometric-evaluation-environmental. (Accessed July 24, 2019) .

Dancey, C.P., Reidy, J., 2007. Statistics without Maths for Psychology. Pearson Education.

Directive 2008/50/EC of the European Parliament and of the Council on 21 May 2008 on ambient air quality and cleaner air for Europe

Eurachem, 2012. Quantifying uncertainty in analytical measurement. EURACHEM / CITAC Guide CG 4. 3rd ed.

Franzin, B.T., Guizellini, F.C., de Babos, D.V., Hojo, O., Pastre, I.A., Marchi, M.R.R., et al., 2020. . Characterization of atmospheric aerosol (PM_{10} and $PM_{2.5}$) from a medium sized city in São Paulo state, Brazil. J. Environ. Sci. 89, 238–251.

Galán Madruga, D., Fernández Patier, R., Sintes Puertas, M.A., Romero García, M.D., Cristóbal López, A., 2018. Characterization and local emission sources for ammonia in an urban environment. B. Environ. Contam. Tox. 100, 593–599.

Govender, P., Sivakumar, V., 2020. Application of k-means and hierarchical clustering techniques for analysis of air pollution: a review (1980–2019). Atmos. Pollut. Res. 11, 40–56.

Hao, Y., Xie, S., 2018. Optimal redistribution of an urban air quality monitoring network using atmospheric dispersion model and genetic algorithm. Atmos. Environ. 177, 222–233.

Kao, J.-J., Hsieh, M.-R., 2006. Utilizing multiobjective analysis to determine an air quality monitoring network in an industrial district. Atmos. Environ. 40, 1092–1103.

Kim, D., Chen, Z., Zhou, L.-F., Huang, S.-X., 2018. Air pollutants and early origins of respiratory diseases. Chronic Dis. Trans. Med. 4, 75–94.

Landis, M.S., Studabaker, W.B., Pancras, J.P., Graney, J.R., White, E.M., Edgerton, E.S., 2019. Source apportionment of ambient fine and coarse particulate matter polycyclic aromatic hydrocarbons at the Bertha Ganter-Fort McKay community site in the Oil Sands Region of Alberta, Canada. Sci. Total Environ. 666, 540–558.
Li, Jinfeng, J.F., Lu, K., Lv, W., Li, J., Zhong, L., Ou, Y., et al., 2014. . Fast increasing of surface ozone concentrations in Pearl River Delta characterized by a regional air quality monitoring network during 2006–2011. J. Environ. Sci. 26, 23–36.
Li, W., Peng, M., Wang, Q., 2019. Improved PCA method for sensor fault detection and isolation in a nuclear power plant. Nucl. Eng. Technol. 51, 146–154.
Li, Y., Chang, M., Ding, S., Wang, S., Ni, D., Hu, H., 2017. Monitoring and source apportionment of trace elements in PM 2.5: implications for local air quality management. J. Environ. Manage. 196, 16–25.
Liu, Y.J., Tran, T., Postma, G., Buydens, L.M.C., Jansen, J., 2018. Estimating the number of components and detecting outliers using angle distribution of loading subspaces (ADLS) in PCA analysis. Anal. Chim. Acta 1020, 17–29.
Liu, Z., Xie, M., Tian, K., Gao, P., 2017. GIS-based analysis of population exposure to PM 2.5 air pollution—a case study of Beijing. J. Environ. Sci. 59, 48–53.
Maćkiewicz, A., Ratajczak, W., 1993. Principal components analysis (PCA). Comput. Geosci. 19, 303–342.
Maione, C., Nelson, D.R., Barbosa, R.M., 2019. Research on social data by means of cluster analysis. Appl. Comput. Inf. 15, 153–162.
Martin, F., Santiago, J.L., Kracht, O., Garcia, L., Gerboles, M., 2015. FAIRMODE Spatial Representativeness Feasibility Study. European Comission, Joint Reasearch Centre, Institute for Environment and Sustainability. Publications Office of European Union, Luxembourg JRC Technical Report ISBN 978-92-79-50322-1 (Doi: 10.2788/49487)Available at: https://ec.europa.eu/jrc/en/publication/fairmodespatial-representativeness-feasibility-study .
Merriam-Webster dictionary. Available from: https://www.merriam-webster.com/dictionary/correlation (accessed Aug 12, 2019)
Mikheev, A.V., Kazakov, B.N., 2017. Correlation analysis of spectroscopic data. J. Lumin. 184, 117–122.
Mofarrah, A., Husain, T., 2010. A holistic approach for optimal design of air quality monitoring network expansion in an urban area. Atmos. Environ. 44, 432–440.
Penkova, T.G., 2017. Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety. Proc. Comput. Sci. 112, 99–108.
Pires, J.C.M., Pereira, M.C., Alvim-Ferraz, M.C.M., Martins, F.G., 2009. Identification of redundant air quality measurements through the use of principal component analysis. Atmos. Environ. 43, 3837–3842.
Pisoni, E., Guerreiro, C., Lopez-Aparicio, S., Guevara, M., Tarrason, L., Janssen, S., et al., 2019. . Supporting the improvement of air quality management practices: The “FAIRMODE pilot” activity. J. Environ. Manag. 245, 122–130.
Pope, R., Wu, J., 2014. A multi-objective assessment of an air quality monitoring network using environmental, economic, and social indicators and GIS-based models. J. Air Waste Manag. 64, 721–737.
Remesan, R., Bray, M., Mathew, J., 2018. Application of PCA and clustering methods in input selection of hybrid runoff models. J. Environ. Inf. 31, 137–152.
Schneider, P., Castell, N., Vogt, M., Dauge, F.R., Lahoz, W.A., Bartonova, A., 2017. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. Environ. Int. 106, 234–247.
Thunis, P., Degraeuwe, B., Pisoni, E., Ferrari, F., Clappier, A., 2016. On the design and assessment of regional air quality plans: the SHERPA approach. J. Environ. Manage. 183, 952–958.
Tseng, C.C., Chang, N.B., 2001. Assessing relocation of urban air quality monitoring stations by GA-based compromise programming. Environ. Int. 26, 523–541.
Tsiflikiotou, M.A., Kostenidou, E., Papanastasiou, D.K., Patoulias, D., Zarmpas, P., Paraskevopoulou, D., et al., 2019. Summertime particulate matter and its composition in Greece. Atmos. Environ. S1352231019303966.
Wang, C., Zhao, L., Sun, W., Xue, J., Xie, Y., 2018. Identifying redundant monitoring stations in an air quality monitoring network. Atmos. Environ. 190, 256–268.
Wang, K., Yin, H., Chen, Y., 2019. The effect of environmental regulation on air quality: a study of new ambient air quality standards in China. J. Clean. Prod. 215, 268–279.
Wu, L., Bocquet, M., 2011. Optimal redistribution of the background ozone monitoring stations over France. Atmos. Environ. 45, 772–783.
Wu, R., Song, X., Chen, D., Zhong, L., Huang, X., Bai, Y., et al., 2019. Health benefit of air quality improvement in Guangzhou, China: Results from a long time-series analysis (2006–2016). Environ. Int. 126, 552–559.
Yang, Y., Sun, L., Guo, C., 2018. Aero-material consumption prediction based on linear regression model. Procedia Comput. Sci. 131, 825–831.
Yoo, E.-H., Brown, P., Eum, Y., 2018. Ambient air quality and spatio-temporal patterns of cardiovascular emergency department visits. Int. J. Health Geogr. 17, 18.
Zhang, Y., Liu, T., Li, K., Zhang, J., 2017. Improved visual correlation analysis for multidimensional data. J. Visual Lang. Comput. 41, 121–132.
Zhao, L., Xie, Y., Wang, J., Xu, X., 2015a. A performance assessment and adjustment program for air quality monitoring networks in Shanghai. Atmos. Environ. 122, 382–392.
Zhao, L., Xie, Y., Wang, J., Xu, X., 2015b. A performance assessment and adjustment program for air quality monitoring networks in Shanghai. Atmos. Environ. 122, 382–392.
Zhu, F., Ding, R., Lei, R., Cheng, H., Liu, J., Shen, C., Zhang, C., Xu, Y., Xiao, C., Li, X., Zhang, J., Cao, J., 2019. The short-term effects of air pollution on respiratory diseases and lung cancer mortality in Hefei: a time-series analysis. Respir. Med. 146, 57–65.