Air Quality, Atmosphere & Health
https://doi.org/10.1007/s11869-022-01254-4

Performance assessment of air quality monitoring networks. A specific case study and methodological approach

D. Galán-Madruga[1] · J. Cárdenas-Escudero[2] · P. Broomandi[3] · R. Oleniacz[4] · J. O. Cáceres[2]

Received: 8 July 2022 / Accepted: 24 September 2022

Abstract
This work presents a new methodological approach to evaluating the long-term performance of an existing air quality monitoring network (AQMN). The AQMN is essential in controlling human beings' exposure to air pollutants, and the performance should be assessed over time. Still, there is not a harmonised method at the legislative level. In this work, 2008–2016 NO_2 data recorded by the Community of Madrid's AQMN were used for developing the suggested methodology, and 2007, 2017, and 2020 NO_2 data were involved in testing the aptitude of the proposed methodology to check the performance along the time. Chemometric techniques were employed to suggest the most representative non-redundant fixed stations within the target AQMN, reducing up to ~ 80% of the original number of fixed monitoring stations (from 23 to 5 fixed stations). The influence of the temporal frame used in developing the exposed methodology showed a variability lower than 5%. The spatial NO_2 distribution pictured by the current versus recommended fixed stations showed a higher than 95% similarity. This recommended approach can also be applied to short-time data. The exhibited methodology is a valuable tool for supporting AQMN managers in decision-making concerning AQMN management and complementing European Legislation guidelines concerning air pollutants monitoring using AQMN.

Keywords Air quality · Ambient air measurement network · Optimization · Non-redundant stations · Environmental applications

Introduction

In the last decades, numerous scientific studies established links between human health and harmful impacts of air pollution exposure (Zhu et al. 2019; Ghaffarpasand et al. 2020; Lamphar et al. 2022), which point to atmospheric pollution as the principal environmental risk to the human being at a global level (Kolasa-Więcek and Suszanowicz 2019; Madruga et al. 2019). For this reason, the European Union develops Air Quality Directives (Directive 2004/107/EC; Directive 2008/50/EC) for laying down air quality objectives to reduce toxic effects on human health and the environment. In this context, the European Union urge the Member States to monitor the air quality within their territories to control exposure to air pollutants. In this sense, the Member States establish air quality monitoring networks (AQMNs) in their regions to verify compliance with those air quality objectives. AQMN performs a principal function in the improvement of monitoring strategies (Mofarrah and Husain 2010; Galán-Madruga 2021a, b) and for assisting authorities in decision making, clarifying local/regional specific sources and sinks of air pollutants, studying the dynamic behaviour of air pollutants, verifying dispersion models, and checking compliance of statistical models (Benis and Fatehifar 2015). Therefore, they are an essential part of environmental management, particularly regarding air quality and controlling air pollutant exposure (Chen et al. 2006). An adequate

✉ D. Galán-Madruga
david.galan@isciii.es

[1] Department of Atmospheric Pollution, National Center for Environment Health, Health Institute Carlos III. Ctra. Majadahonda a Pozuelo, km 2.2, 28220 Madrid, Spain

[2] Laser Chemistry Research Group, Department of Analytical Chemistry, Faculty of Chemistry, Complutense University of Madrid, Plaza de Ciencias 1, 28040 Madrid, Spain

[3] Department of Civil and Environmental Engineering, School of Engineering and Digital Sciences, Nazarbayev University, Nursultan (Astana), Kazakhstan

[4] Department of Environmental Management and Protection, Faculty of Geo-Data Science, Geodesy and Environmental Engineering, AGH University of Science and Technology, 30 Mickiewicza Av, 30-059 Krakow, Poland

installation would lead to support the high efficiency of the network, which consists of fixed monitoring stations for measuring air pollutants, such as nitrogen oxides (NOx), carbon monoxide (CO), sulphur dioxide (SO_2), benzene (C_6H_6), and particulate matter (PM), including PM_{10} and $PM_{2.5}$, and provide relevant spatial information regarding a target area (Gómez-Losada et al. 2014).

In this frame, primordial features in the AQMNs design are the (i) distribution of the fixed monitoring stations within the target area and (ii) determination of a sufficient and reliable number of sampling points for air quality measurements. Relative to point (i), the passive methodology has been demonstrated to be a helpful tool for selecting the most representative locations within a determined area for measuring air pollutants, given that it allows tracking a vast target territory (Chakraborty et al. 2017). However, as a limitation, passive samplers only apply to gaseous pollutants. Regarding point (ii), air quality directives have criteria, in terms of the number of inhabitants, to help the Member States set a minimum number of sampling points to measure target air pollutants in their territories. Nevertheless, the air quality networks usually have more locations of measure than the minimum number assigned by the directives due to the efforts of governments to control air pollutant exposure. Broadly, the networks could include redundant fixed monitoring stations within their setup, which would be translated into a reduction in AQMN efficiency. On the other hand, the emergence of new emission sources could change the gradient of the spatial distribution of air pollutants within a target territory. Consequently, the non-redundant stations could become redundant ones and vice versa, thereby modifying the possible effectiveness of each fixed monitoring station.

Based on the previously mentioned arguments, AQMN performance should be assessed along the time regarding the representativeness of fixed stations within the AQMN to optimise the network layout. It is relevant to highlight that there is currently no harmonised methodology for estimating the representativeness of the fixed stations belonging to an AQMN (Martin et al. 2015), confirming the development of those studies as a priority subject.

Within the described frame, scientific literature exhibits a highly limited number of studies. Some studies proposed approaches for optimising AQMNs layout (Rizzo et al. 2004; Ibarra-Berastegi et al. 2009; Dincer et al. 2016; Baca-López et al. 2021; Zeydan and Pekkaya 2021). Karppinen et al. (2000), at the European level, evaluated the AQMN performance by comparing predicted NOx and NO_2 levels vs those concentrations measured by the fixed stations included in the Helsinki AQMN. Sarigiannis et al. (2007) designed a new method based on satellite remote sensing of the troposphere for multiobjective optimising air quality monitoring networks. They applied this approach in Brescia (Italy), showing significant potential for enhancing the cost-effectiveness of air quality networks at the urban and regional levels. In Spain, Ibarra-Berastegi et al. (2010) combined self-organising maps and cluster analysis to evaluate the capability of a concrete network. This methodology was applied to the Bilbao network. In the same context, pollution data and social and economic indicators were used by Munir et al. (2019) to propose an integrated air quality monitoring network in Sheffield. At the Asian level, prestigious research groups provided valuable solutions to this issue. In this sense, a Shanghai University research group in China exposed a performance assessment and adjustment program for Shanghai's AQMN using historical data for several air pollutants from January 1 to August 22, 2014, to identify redundant fixed stations (Zhao et al. 2015). They used a combination of principal components analysis and assignment method, as clustering analysis to verify the previous two techniques' outcomes. Wang et al. (2018) extended the previous work to optimise the AQMN layout, employing correlation analysis, principal component analysis, assignment method, clustering analysis, and correspondence analysis. This approach was implemented in Xi'an city's AQMN using historical data series from January 1 to December 31, 2016. Huang et al. (2019) used a combined technique based on a Gaussian model and source area to assess long-term AQMN performance using the surveillance efficiency for hydrogen sulphide (H_2S) in a chemical industrial park in Shanghai, China. In America, Silva et al. (2003) applied a multivariate effectiveness index to an existing AQMN at Santiago de Chile, based on the Shannon information index, to exclude the least informative fixed stations within the target AQMN. Austin et al. (2013) engaged clustering analysis to identify spatial patterns based on the composition of $PM_{2.5}$ particles. Within the same objective, Soares et al. (2018) employed hierarchical clustering techniques based on associativity analysis for optimising AQMN design.

The valuable research previously cited works supposed a remarkable advance in assessing AQMN design to identify redundant stations. Nevertheless, they were not addressed a validation process concerning pragmatic enforcement of the developed approaches using ambient air pollution data non-incorporated in creating these. In this context, studies were not conducted to assess the time impact and loss of spatial information derived from removing redundant stations within the target AQMN.

This research aims to provide a stepwise-developed methodological framework to assess the long-time performance of existing air quality networks and improve their design. This novel approach could serve as a benchmark framework to elaborate a harmonised methodology, given that it is currently unavailable. Then, a case study is addressed to offer a practical application in its implementation once defining the methodological framework. For that, the performance of a specific AQMN is appraised for a determined air pollutant, although the leading objective of this work is not to improve the design of the attended network. For achieving that objective, air quality data recorded by the Community of Madrid's AQMN between 2008 and 2016 were examined, and ambient air pollution data measured in 2007, 2017, and 2020 were used to test the practical application of the proposed methodology. Among all polluting compounds measured by the target AQMN, nitrogen dioxide (NO_2) data were attended to avoid repeating the approach with the rest of the pollutants. NO_2 is highly important because it is mandatory in the European Member States. Directive 2008/50/EC sets a limit value for ambient air of 40 µg /m^3 on average in a calendar year, a concentration also supported by the World Health Organization (WHO 2006). For this reason, we attended annual average NO_2 concentrations to develop the proposed methodology, which aligns with other scientific studies on the covered subject (Galán-Madruga and García-Cambero 2022). The annual approach was tested using a minor temporal scale (daily mean NO_2 levels). NO_2 pollutant is responsible for about 92% of asthma cases and plays a crucial role in forming other major pollutants, such as ozone and particulate matter (Bhardwaj et al. 2020).

Materials and methods

Area of study and reference data package

This work was conducted in the Community of Madrid (Spain). It has over 6,751,251 inhabitants in 2021, according to the Spanish National Institute of Statistics and a surface of 802,180 hectares (ha) (INE 2021). It is located in the centre of the Iberian Peninsula; therefore, it is an interior region of Southern Europe with 179 districts. It has a population density of ~ 800 people per km^2 of land area, one of the most densely populated South European regions (Eurostat 2022). The Community of Madrid is characterised by stable atmospheric conditions at the meteorological level, with a warm temperature as the main climate, summer hot and dry at the precipitation level, according to the World Map of Köppen – Geiger Climate Classification (Kottek et al. 2006). The selected study time for achieving the proposed objective encompassed 2008 to 2016. Over this period, meteorological values of 15.0 °C (minimum and maximum: 14.0 and 16.0 °C, respectively) of average temperature, 184.8 W/m^2 (173–191 W/m^2) of solar radiation, 57.9% (54–61%) relative humidity, 932.4 mbar (928–939 mbar) of barometric pressure and 2.70 m/s (2.0–3.1 m/s) wind speed.

The original air quality dataset was recorded by the Community of Madrid's AQMN between 2008 and 2016 and was acquired from the Community of Madrid's open data portal (Comunidad de Madrid 2022). For the picked period, the AQMN was conformed of 23 fixed measurement stations for measuring ambient air NO_2 levels, distributed in six homogeneous zones within its territory; three are agglomerations, and three are rural areas.

The agglomeration zones include Henares Corridor 915 km^2, South Urban Area 1413 km^2, and Northeast Urban Area 1016 km^2, while Tajuña basin 942 km^2, Alberche basin 1,181 km^2, and North Mountain 1,951 km^2 correspond to rural zones. Two criteria supported the organisation of the fixed stations included in the network: location and primary pollution focus. In the site's function, urban, suburban and rural stations were accounted for, while that traffic, industrial, or background stations correspond to the primary pollution focus (see Fig. 1). The location of all fixed stations complies with the macro- and micro-implantation requirements in current European Legislation. The regional government control and manages the AQMN by securing its maintenance and validating monitored air quality data. The NO_2 levels were measured in all fixed stations with automatic analysers. The chemiluminescence was the measuring method used in all the analysers (it is the normalised reference method according to EN 14,211:2012 Standard). In order to offer an overview of the stability of the values obtained by the automatic NO_2 analysers included in the target network, the measurement uncertainty in all analysers was lower than 15%, defined as an expanded uncertainty. The ambient air mass concentrations were expressed in µg/m^3.

Description of the methodological framework

As background information, an AQMN is constituted by fixed measurement stations furnishing pollution information at the sampling points and surroundings. To evaluate the spatial representativeness of fixed stations in terms of monitored pollution information, Santiago et al. (2013) developed an approach based on computational fluid dynamics simulations applied to two urban areas (Pamplona and Madrid, Spain) as a case study. Based on their results, the spatial representative area of a concrete fixed station corresponds to the domain where concentrations fall within an interval of ± 20% of measured levels at that monitoring station. Similarly, Righini et al. (2014) supported the same conclusion. Within this research area, a highlight investigation group conducted a study to assess the representativeness of a specific fixed station using a methodology relying on statistical/geostatistical analyses in the city of Varese (Italy). More information in Yatkin et al. (2022).

Furthermore, given that each fixed station provides air pollution information in a concrete area, the spatial distribution gradient across a target region may be known by combining the pollution information generated by all stations. In this sense, the AQMN performance depends on

Fig. 1 Location, type of fixed monitoring station and agglomeration, and rural zones within the Community of Madrid's AQMN (2008–2016). Note: The white surface corresponds to Madrid City (It has its own air quality monitoring network)

the efficiency of fixed stations (the greater the number of redundant stations, the lower the network's performance).

The analysis techniques' sequence used in developing the suggested methodology is listed below:

1. A correlation analysis (CA) is addressed, using the original database, to assess the connection degree between pairs of fixed monitoring stations within the target AQMN to prove whether they report similar air pollution information, resulting in redundant fixed stations.
2. A clustering analysis was used with the original database to identify non-redundant fixed measurement stations within the target AQMN.
3. Position measures and population data are applied to outcomes reached by the clustering analysis to recommend the most representative non-redundant fixed stations within the AQMN to measure the target air pollutant in the studied domain.
4. Geostatistical estimation techniques were employed to test the practical application of the suggested methodological framework using a database not included in its development. For that, a geographic information system (GIS) was used. In this sense, and as innovative scientific contributions, this work quantifies the (i) impact of the temporal range covering the proposed approach and (ii) loss of spatial information reported by AQMN by removing all redundant stations and those non-proposed non-redundant stations.

Based on the previous sequence, Fig. 2 shows a flow diagram representing the step succession addressed for assessing and improving the AQMN performance.

Investigating the presence of redundant fixed stations within the target AQMN

As a first step for assessing existing AQMN performance, the possible presence of fixed stations measuring similar pollution information within the network context should be studied. In this regard, a CA between pairs of fixed stations would provide knowledge on the association degree between them, given that the correlation defines the existing relationship between study variables that change, associate, or take place in a not awaited way (Merriam-Webster Dictionary). Therefore, this analysis type assists in considering the complex relation among diverse datasets, supplying information concerning the weight of each variable (Mikheev and Kazakov 2017). The correlation grade relied on Pearson's

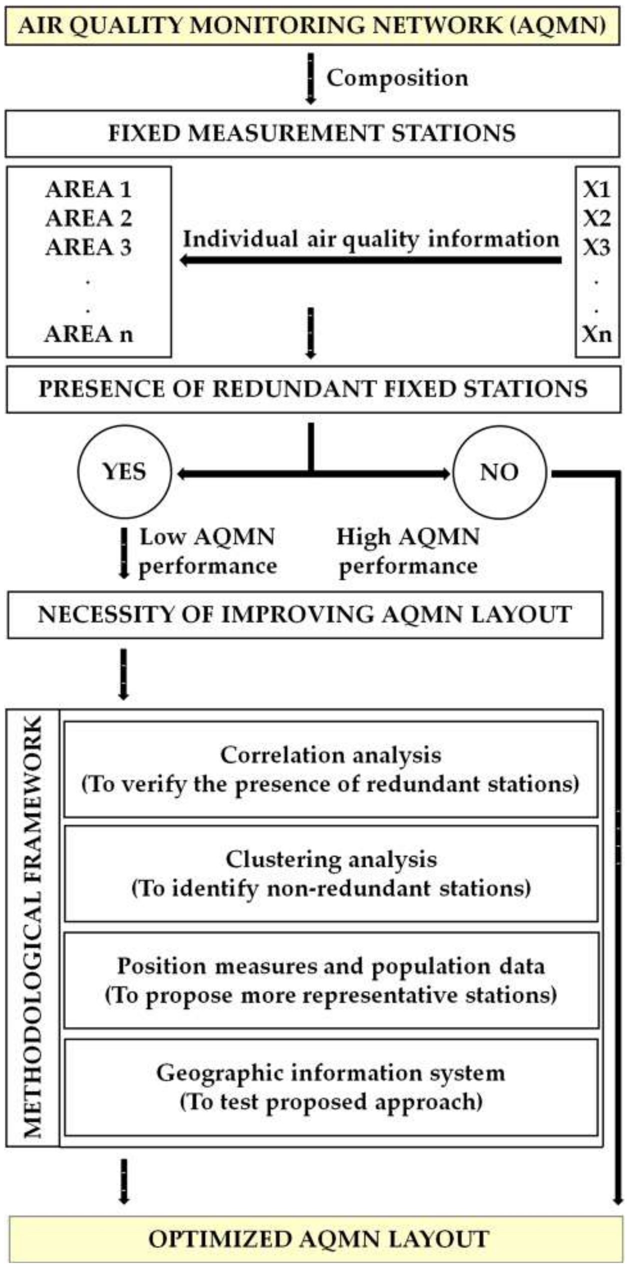

Fig. 2 Flowchart picturing stepwise suggested methodological framework

coefficient of correlation (Galán-Madruga et al. 2022), which establishes the linear dependence between two target variables, and its calculus is sustained by Eq. 1.

$$r = \frac{\sum (x - m_x) \bullet (y - m_y)}{\sqrt{\sum (x - m_x)^2 \bullet (y - m_y)^2}} \quad (1)$$

where r is Pearson's correlation coefficient, $\boldsymbol{m_x}$ and $\boldsymbol{m_y}$ are equivalent to the mean value of x and y variables. Here, $\boldsymbol{m_x}$ and $\boldsymbol{m_y}$ correspond to annual average NO_2 concentrations monitored in the x and y fixed measurement stations.

Identifying the non-redundant fixed monitoring stations

Once redundant fixed stations were confirmed within the target AQMN, a study was conducted to identify the non-redundant fixed stations. A clustering analysis based on k-means algorithms was performed in this study (k-means clustering with ten maximum iterations). It is widely used in different science areas due to its efficiency and simplicity (Jain 2010), and it is catalogued as an exclusive partitional clustering algorithm (Salem et al. 2018). The execution of the k-means clustering analysis was widely described by Govender and Sivakumar (2020). Briefly, the clustering k-means algorithm minimises the total sum of squared distances between each component and its nearest cluster centre (Maione et al. 2019). It aims to group diverse elements with similar intrinsic features. In this work, the fixed stations included in the AQMN constitute the components of the clusters, and the annual average NO_2 concentration monitored in each station sustains the standard feature.

Before running clustering analysis, indicative variables need to be set. In this regard, those reported by Galán Madruga et al. (2018) were employed in this work. In addition, the cluster standard deviation was used as a cluster membership identifier, and the Euclidean distance was selected as a spatial indicator.

Once running the clustering analysis, several homogeneous clusters are generated. Each cluster is represented by a mean NO_2 value averaged from the annual average NO_2 concentration recorded by each fixed station included within this cluster. Each station of the cluster assumes this mean NO_2 value. Applying the clustering technique to this work aims to identify the non-redundant fixed stations.

Proposal of the most representative non-redundant fixed monitoring stations for measuring ambient NO_2 levels in the target AQMN.

The outcomes reported by applying the clustering analysis to the original pollution database allowed recognising the set of non-redundant fixed stations within the target AQMN. Among all non-redundant stations, those most representative ones displayed the lowest Euclidean distance value. Then, it is summed along the covered years the number of times each station has achieved this scale criterion (lowest Euclidean distance value), obtaining the total representativeness for each fixed station for the studied period.

An analysis of position measures was applied to the results of total representativeness, classifying it in the function of the quartiles scale. Then, using quartile outcomes and population data, the most representative non-redundant fixed stations were proposed as a function of station type. The population data were obtained from the 2019 Census (INE 2021). Note that this research does not aim to propose new measuring NO_2 sites within the target AQMN but to evaluate the long-time performance of the existing target AQMN.

Testing the proposed methodological approach

As a primordial aspect within research studies aiming to provide methodologies for improving any science scope, developing a validation process supposes a highly relevant exercise, which will offer certainty concerning the practical application of the proposed methodological tool.

The methodological framework exhibited in this study was submitted to a rigorous validation once proposing the most representative non-redundant fixed monitoring stations for measuring NO_2 in the Community of Madrid's AQMN. Therefore, it quantified the (i) temporal influence of the original dataset (air quality data from 2008 to 2016) on the suggested methodology and (ii) loss of information of the spatial pollution gradient reported by the new AQMN design regard with the original.

The methodological development was tested using pollution NO_2 levels recorded by AQMN in 2007, 2017, and 2020. Note that the elaboration of the recommended methodology did not include these years.

Data treatment and elaboration of isolines concentration maps

Original dataset treatment was performed with the software IBM SPSS Statistics v28.0 (IBM Corp., Armonk, NY, USA). On the other hand, Surfer for Windows (Win32): Surface Mapping System, v.6.04 (Golden Software, Inc., Golden, CO, USA) was used as a geographic information system for building NO_2 iso-concentrations maps. According to previous environmental studies, the kriging method was employed as a geostatistical estimation tool for the spatial interpolating NO_2 levels in those points not measured within the target domain (Park et al. 2020; Araki et al. 2021). This interpolation method is also known as the spatial self-covariance optimal interpolation method. It is based on the relationship of the (i) measured sample points and their spatial locations, (ii) the non-monitored points needing to be evaluated, and (iii) the structural information supplied by the variating function (Meng 2021). Therefore, the application of the kriging method requires the semivariogram calculus, which was defined as:

$$\gamma_{(A_i,B_j)} = \frac{1}{2}\boldsymbol{Var}\left[Z_{A_i} - Z_{B_j}\right] \quad (2)$$

where $\boldsymbol{Var}$ is the variance, $\boldsymbol{A_i}$ and $\boldsymbol{B_j}$ correspond to two points where the target variable has been measured and associated with spatial data (Wackernagel 2003).

In order to execute the kriging method, next parameters were set: variogram model: lineal type, variogram anisotropy: ratio = 1 and angle = 0, nugget effect: error variance = 0 and drift type: No drift.

Results and discussion

Air quality status in the target territory over the studied period

Given that sources of pollution and meteorological conditions may vary along the annuities involved in this work, a general investigation describing air quality over the studied period will provide relevant information. Regarding health, the AQMN plays a primordial role in evaluating human beings' exposure to air pollutants. Within this frame, the annual NO_2 level averaged for each studied year was lower than the limit value of 40 μg/m^3 established in the current Legislation (see Fig. S1). Broadly, it exhibited a downward trend in NO_2 levels, reaching a decrease of 3.5 μg/m^3 (13%, expressed as a relative value) from 2008 to 2016, mainly sustained by the successive environmental policies implanted by the regional government to reduce or limit the polluting emissions into the atmosphere. In this regard, the Community of Madrid has conducted an outstanding effort, setting up several air quality plans, such as Air Quality and Climate Change Strategies, from 2006 to 2012 and from 2013 to 2020.

In the function of the type of environment, a more detailed analysis reported that the highest annual average NO_2 levels over the studied period were observed in urban stations, followed by suburban and rural stations, respectively (Fig. S2). Nonetheless, the urban environment evidenced the highest pollution declines (13.8, 5.0, and 2.5 μg/m^3 at the urban background, urban traffic, and urban industrial sites, respectively). In contrast, the suburban and rural environments practically kept polluting levels. This fact may be explained by the emission sources being more numerous in urban environments than at suburban and rural sites, which translated into the highest NO_2 urban levels (Domínguez-López et al. 2014; Núñez-Alonso et al. 2019; Oleniacz and Gorzelnik 2021). The leading polluting sources in urban locations drive road traffic, public and private transport networks, industrial activities, and domestic heating systems. Over time, implementing environmental plans to reduce the pollution released into the atmosphere generate higher NO_2 levels decrease in urban environments. They are shown the outcomes reached when applying development exposed in the "Materials and methods" section on the original air-quality dataset.

Investigating the presence of redundant fixed stations within the target AQMN

Table S1 shows Pearson's coefficients of correlation obtained when applying CA to the original air quality database. To interpret the outcomes exhibited in Table S1, the approach proposed by Dancey and Reidy (2007) was addressed. They set a rating for Pearson's coefficient of correlation in the function of the relationship value between study-correlated variables. Based on that, they established five categories, named: zero (when the value of the coefficient of correlation is 0), weak (when the scaling value falls into interval ± 0.1¸0.3), moderate (interval ± 0.4¸0.6), strong (interval ± 0.7¸0.9), and perfect (scaling value equal to ± 1.0). In this context, to further ease the understanding of correlation study outcomes, Fig. 3 pictures the distribution, in relative scale, of the correlation coefficient in the function of the type of fixed station and for all stations.

Considering a cut-off level for Pearson's coefficient of correlation of 0.5, a deeper study evidenced that Pearson's coefficients of correlation > 0.5 explained more than 50% of the distribution in urban traffic and background stations (57.14 and 100.00%, respectively). This finding is highly relevant because the fixed stations catalogued as urban traffic and background stations represent almost 50% of the total number of measuring NO_2 sites within the target AQMN. Pearson's coefficients > 0.5 accounted for 20 and 30% distribution for suburban and rural background stations. Finally, the lower distribution levels for Pearson's coefficients > 0.5 pointed toward the urban industrial stations.

Considering all stations, the application of CA revealed a moderate correlation between the pairs of fixed stations for measuring ambient air NO_2, with a distribution close to 40% for the selected cut-off level. These findings suggest that several fixed monitoring stations within the AQMN may be redundant, providing reasonably similar polluting information concerning NO_2 levels. Therefore, several zones in the covered territory might have a similar pollution behaviour.

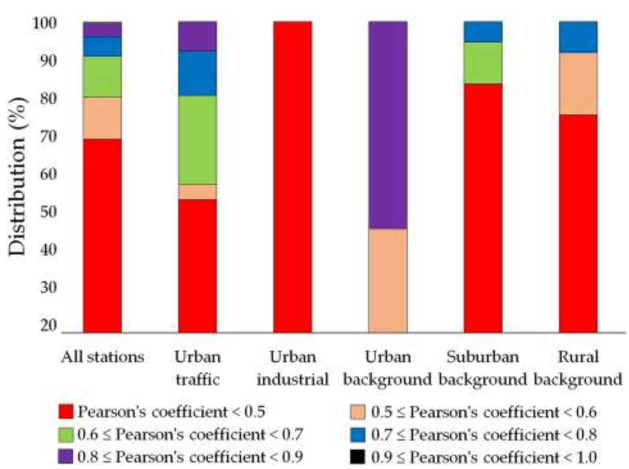

Fig. 3 Distribution of Pearson's coefficient of correlation classified by fixed station category (2008–2016). Note: The code ES1811A station (suburban traffic, $N = 1$) has been included in the suburban background category

Identifying the non-redundant fixed monitoring stations

The CA enforcement allows laying down the connectivity degree between two study variables potentially correlated, in this case, pollution data measured by pairs of fixed monitoring stations. Nevertheless, this analysis technique does not allow identifying the non-redundant fixed stations within AQMN to upgrade the network layout and enhance its performance and efficiency. Therefore, a study in this sense is fully justified.

In this regard, a k-means clustering technique was applied to annual average NO_2 concentrations measured at the twenty-three fixed monitoring stations belonging to the target AQMN to point to the non-redundant fixed stations. When executing a *k*-means clustering analysis, the number of clusters (value of *k*) is essential. In this sense, the individual *k* values used to implement the clustering technique were $k = 20$, 15, 10, and 5, respectively.

Once the previously listed four clusters were executed, the coefficient of determination for each cluster was computed when correlating current annual average NO_2 concentrations versus those estimated by clustering analysis. Among all outcomes, the cluster selected for identifying the non-redundant fixed stations (hereafter conceptualised as the nominated cluster, NC) presented a coefficient of determination equal to or higher than 0.95. As shown in Fig. S3, cluster number 5 was the first group meeting this premise for all researched years, thereby appointing cluster 5 as NC. The average value of the coefficient of determination (for cluster 5) along the investigated period was 0.967 ± 0.076 (range: 0.958–0.977). Within cluster number 5, Euclidean distance and standard deviation attained maximum values of 4.330 ± 0.725 and 1.594 ± 0.305 μg/m^3, respectively.

Within this frame, Table S2 reported Euclidean distance values corresponding to the NC. The non-redundant fixed stations by studied year are linked to the lowest Euclidean distances. This parameter is a distinctive variable among components included in the same group (Penkova 2017). According to the outcomes, the non-redundant fixed stations are displayed in Table 1. The stations encompassed within the same cluster are highly dissimilar from stations associated with other clusters (Maione et al. 2019).

Note the wide variability of obtained outcomes regarding the number of non-redundant fixed stations. More than 15 fixed stations were designated as non-redundant at least once

Table 1 Non-redundant fixed monitoring stations in the function of the studied year (Note: They are assigned by station code. No common part is included to reduce the station code)

Station code	2008	2009	2010	2011	2012	2013	2014	2015	2016
1563				X				X	
1564			X		X	X			
1565			X		X				
1567		X	X	X	X	X	X		
1568		X		X				X	
1611		X		X	X	X		X	X
1612	X			X			X		
1613	X								
1801	X			X	X	X			
1803				X					X
1806							X		X
1807	X			X		X			
1808	X			X		X	X	X	
1810		X	X		X	X		X	
1811			X						
1838		X		X			X	X	
1869	X	X	X		X	X	X	X	X
1890			X		X		X		
2028	X			X				X	X

during the studied time, which brings out the complexity of air quality network management.

By counting the number of times that urban, suburban, and rural fixed stations were labelled as non-redundant between 2008 and 2016, the urban stations displayed a higher representativeness degree than suburban and rural ones (Fig. S4). In concrete, the urban representativeness ranged between 43 (2014) and 86% (2012), while stations placed at suburban sites oscillated between 0 (2012 and 2016) and 33% (2009), and finally between 9 (2011) and 29% (2014) in the case of fixed stations located at rural sites (see Fig. S5).

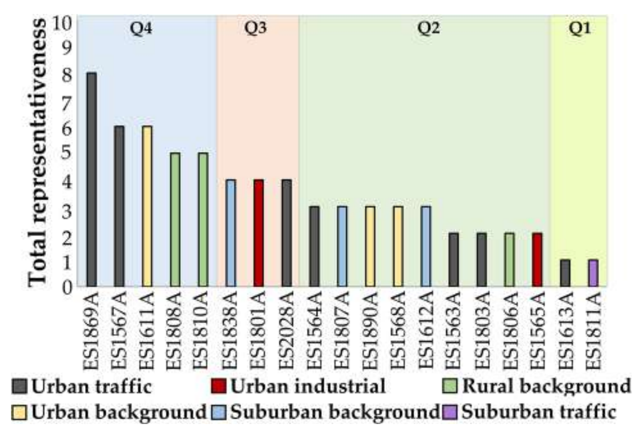

Fig. 4 Total representativeness of fixed monitoring stations of the Community of Madrid for measuring ambient air NO_2 concentrations from 2008 to 2016

Proposing the most representative non-redundant fixed monitoring stations for measuring ambient NO_2 levels in the target AQMN

According to the approach reported in the "Materials and methods" section, the total representativeness for each station was calculated along the covered period. Then, position measures and population data were applied to propose the most representative non-redundant fixed stations within the set of non-redundant stations. Figure 4 represents the total representativeness for each station along the respected study period grouped by quartiles (hereafter Q, results for Q4, 3, 2, and 1: 6.0, 4.5, 3.0, and 2.0, respectively) and type of station.

As previously mentioned, the following requirements were regarded to propose one non-redundant fixed station in the function of the station type: All quartiles should be represented.

1. The station with the larger representative scale in each quartile should be picked.
2. The station covering a higher inhabitant number should be selected whether several stations within the same quartile reach a similar representativeness value.

Based on those criteria, the first selected relevant station leads toward ES1869A station (Coslada, urban traffic station). Due to that, this one sustains a higher representativeness level (Q4) during the covered years. However, several stations share the highest representativeness level for the rest of the quartiles; therefore, the number of inhabitants criteria will decant the suggested measuring site (Table S3). In this frame, within the Q3, the ES2028A station (Getafe) has a higher number of inhabitants than ES1801A station (Arganda del Rey) and ES1838A station (Algete), respectively. Nonetheless, the ES2028A station (Getafe) was not recommended because an urban traffic station had already been selected in a quartile of a larger scale (ES1869A station, Coslada). Instead, the ES1801A station (Arganda del Rey, urban industrial station) was recognised as a primordial location for measuring NO_2 (population higher than ES1838A station, Algete).

Within Q2, five fixed stations accomplished the same number of representativeness. Nevertheless, the ES1568A station (Móstoles, urban background station) is a main measuring NO_2 point because it has a higher number of inhabitants than the rest of the stations with similar representativeness in this quartile. Finally, the fixed station with higher representativeness and the number of inhabitants within the Q1 drives to an urban traffic station (ES1613A station, Colmenar Viejo). Despite that, this site was not regarded as a relevant monitoring point because the ES1869A station (Coslada) was already recognised within the category of urban traffic station. Therefore, the ES1811A site (Villarejo de Salvanés) was selected in the category of suburban traffic station.

Up to here, the non-redundant stations proposed for measuring NO_2 in the target atmosphere sum up a total of four stations, encompassing the following types: urban traffic (ES1869A, Coslada), urban-industrial (ES1801A, Arganda del Rey), urban background (ES1568A, Móstoles), and suburban traffic (ES1811A, Villarejo de Salvanés). However, given that the initial objective was to propose a non-redundant measuring site per station type, the category of rural background station should be advanced. In this sense, within the rural background station's category, the monitoring station with a higher representativeness degree and population is the ES1808A station (San Martín de Valdeiglesias), covering the five stations classes.

Since each atmospheric pollutant has its peculiarity in terms of emission sources, atmospheric chemical reactions, dispersion rate, and lifetime (Tuck 2021), the suggested methodology should be employed in a particular way. To offer evidence regarding this consideration, research carried out in the same AQMN used in this study for optimising the monitoring of PM_{10} particles identified eight fixed stations as the more representative non-redundant fixed stations for measuring PM_{10} within the target region (Galán-Madruga 2021a, b). However, comparing the fixed stations selected for measuring PM_{10} with those proposed ones for NO_2 in this study, only two stations coincided (ES1568A, Móstoles, and ES1811A, Villarejo de Salvanés).

Based on the previous development, it is relevant to highlight that fixed monitoring NO_2 sites in the target territory declined from twenty-three to five fixed stations, with a percentage decrease of nearly 80%. Other authors also reported outcomes in this sense. For example, in 2008, a relevant Portuguese research group evaluated the Oporto Metropolitan Area's AQMN layout using principal component analysis and clustering techniques. They used ambient air pollution data (SO_2 and PM_{10}) monitored at fourteen fixed stations from January 2003 to December 2005. The outcomes displayed that six and two stations were enough to measure SO_2 and PM_{10} (decrease of 57 and 85%, respectively) (Pires et al. 2008). Similarly, Hao et al. (2018) developed a method to optimise the network size and the measuring points' locations for Shijiazhuang city (China). The results showed an improvement of 60.99% (SO_2) and 76.06% (NO_2) within the target urban area.

Other research studies developed approaches based on distinct analysis techniques to identify redundant fixed monitoring stations within AQMN (see Table S4). All studies provided a significant advance in the AQMN optimisation and design. Many of these studies identified redundant fixed stations (Pires et al. 2009), while others reported alternative measuring sites (Munir et al. 2019) or identified appropriate sites in a target territory to locate fixed stations in order to set up an AQMN (Hacıoğlu et al. 2016; Kollati and Debnath 2021). Gorzelnik and Oleniacz (2019) assessed the suitability of new air quality monitoring stations in Krakow's AQMN regarding spatial and temporal variability of PM_{10} concentrations. In this frame, Verghese and Nema (2022) conducted a comprehensive literature review study covering the relevant conventional and contemporary approaches in the field of AQMN optimization over the last four decades, highlighting major drawbacks and advantages.

The recommended strategy in this work may offer significant applications within the environmental management field, particularly in air quality management. This strategy may be combined with other techniques of air quality assessment, such as atmospheric pollution forecasting studies (Videnova et al. 2006), to easy identification the most appropriate sites for locating fixed monitoring stations for measuring air pollutants. Although the proposed method has been applied to a specific AQMN as a case study, this one may be implemented in any air quality network at the global level to optimise its design by restructuring the network. The enhancement in the layout of an existing network would drive the most accurate measurements, given that redundant fixed stations would influence data validity (Kao and Hsieh 2006). The AQMN optimisation would carry beneficial consequences assuming different sceneries concerning the removed fixed monitoring stations:

Scenery 1 (They are not relocated). It would be obtained a significant reduction in expenses derived from the network maintenance, in terms of investment for acquiring automatic observation analytical instrumentation used in the AQMN, data quality assurance activities, such as calibration operations of measurement instruments and acquisition of reference materials. This fact results in highly pertinent within a frame of limited economic resources.

Scenery 2 (They are relocated). The target surface covered by AQMN would be increased by relocating the removed fixed stations to new sampling points to measure the same pollutant or new compounds not previously measured in the AQMN.

In any case, reducing the number of fixed stations would not carry a decline in AQMN performance, being sustained by the body of findings reported in this work.

Given that the European Legislation is a pertinent and necessary instrument to control atmospheric pollution and improve air quality, the methodology proposed in this work also has a potential application at a legislative level. The current air quality European legislation lays down considerations at the macro-and microscale level about the location of fixed stations and the minimum number of measurement sites within an AQMN. This last consideration is established in function of the zone's population needing to be controlled (Directive 2008/50/EC). Nevertheless, this Legislation does not include considerations or methods allowing (i) identifying non-redundant fixed stations within an AQMN and (ii) proposing the most representative non-redundant fixed stations. Therefore, implementing the methodological framework recommended in this work would quickly solve this environmental issue, complementing the guidelines set by European Legislation on air pollutants monitoring using AQMN. Similarly, the developed approach may serve as a benchmark frame to devise a harmonised methodology to evaluate the AQMN performance, given that it is currently unavailable.

On the other hand, the suggested methodology may serve competent authorities as a helpful instrument for assessing the impact of air quality plans implemented in a concrete region for a long-time period. A particular emphasis deserves urban and suburban environments accounting for its potential growth in population, transport networks, town planning, and industries, which would be translated into alterations of the spatial air pollutants pattern.

Testing the proposed methodological approach

Since the proposed methodology severely reduces the number of measuring NO_2 points in the target AQMN, a validation study must be conducted to sustain the validity and reliability of the reported methodological development. In this sense, temporal and spatial variables were tested, using a GIS for building maps of NO_2 iso-concentration. To execute the validation process, two groups of data packages were regarded: pack 1 (formed by the current 23 fixed stations) and pack 2 (constituted by the proposed non-redundant 5 fixed stations). To equal the number of stations in both packs, those included in pack one but not in pack two were incorporated into pack 2. The fixed stations included in pack 1 sustained the current annual average NO_2 concentrations, and those stations of pack 2 assumed the estimated annual average NO_2 concentrations derived by clustering analysis.

At the temporal level, the impact of the period used in developing the suggested methodology was addressed using NO_2 pollution data averaged between 2008 and 2016 (see Fig. 5).

The pollution information pictured by the NO_2 spatial distribution gradient in both environmental sceneries displays a high connection grade, achieving 96.7%. Therefore, it can be concluded that the influence of the time range employed in developing the proposed methodological framework exhibits a highly acceptable value. At the annual level, a more detailed study confirmed the finding found in the global study (see Fig. S6), showing slightly higher similitudes (between 97.9 and 98.9%). However, they did not find significant differences at the global and annual levels (t-test).

At the spatial level, the loss of information regarding the spatial distribution pictured by current (pack 1) versus proposed fixed stations (pack 2) is researched using ambient

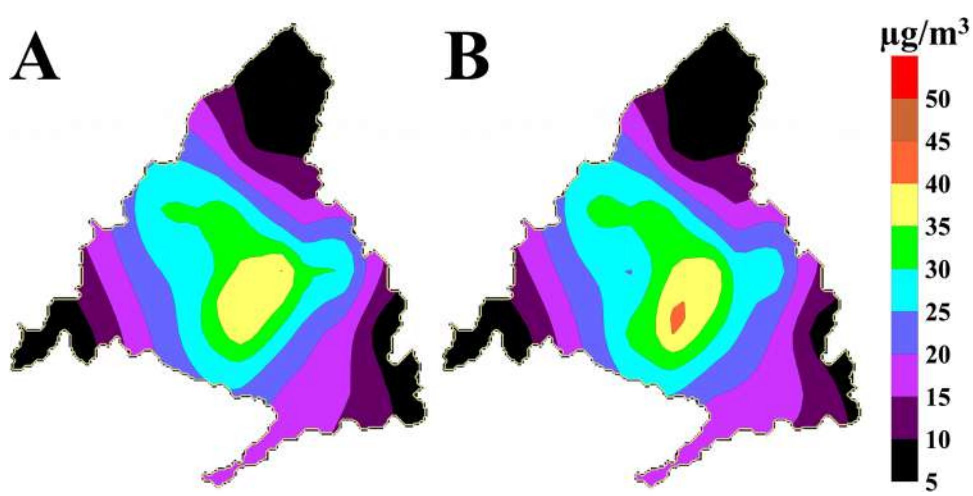

Fig. 5 Iso-concentration map for average NO_2 levels (2008–2016). **A** Currently fixed stations (pack 1, 23 fixed stations) and **B** Proposed fixed stations (pack 2, 5 fixed stations)

air NO_2 data recorded by target AQMN in 2007, 2017, and 2020. The outcomes will allow the investigation of potential discrepancies concerning spatial gradients obtained from both postulates for the three studied years. Figure 6 represents the spatial annual NO_2 distribution for those years.

The concentration gradient drawn by fixed stations included in pack one versus pack 2 provides practically the same pollution information (95.4, 96.9%, and 96.5% for 2007, 2017, and 2020, respectively), thereby exhibiting a spatial discrepancy lower than 5%. Therefore, significant differences were not observed at the temporal and spatial levels, which sustains the advised methodological framework as a helpful tool for assessing AQMN performance and improving its design.

To examine the influence of resolution of pollution data used in developing the suggested methodological framework (annual average levels), short-time data (daily average) were tested. The outcomes reached by the approach developed using annual average NO_2 concentrations were respected (the nominated cluster, $K=5$ and the non-redundant five fixed stations proposed). Given that an existing AQMN should be evaluated in a concrete time, December 12 was considered to examine the impact of the data resolution. In this sense, Tables from S5 to S12 display the results of Euclidean distance observed when running cluster number 5 using daily average NO_2 data from December 12, 2008–2016. In addition, the spatial distribution gradient reported by current vs proposed fixed stations for the selected target day was compared, according to the "Results and discussion" section. As shown (Fig. 7), the spatial information reported by both sceneries reaches a likeness level higher than 95% (minimum and maximum value: 95.1 and 99.4% for 2013 and 2015, respectively). Therefore, the methodological framework proposed using long-time pollution data is also valid by employing short-time data.

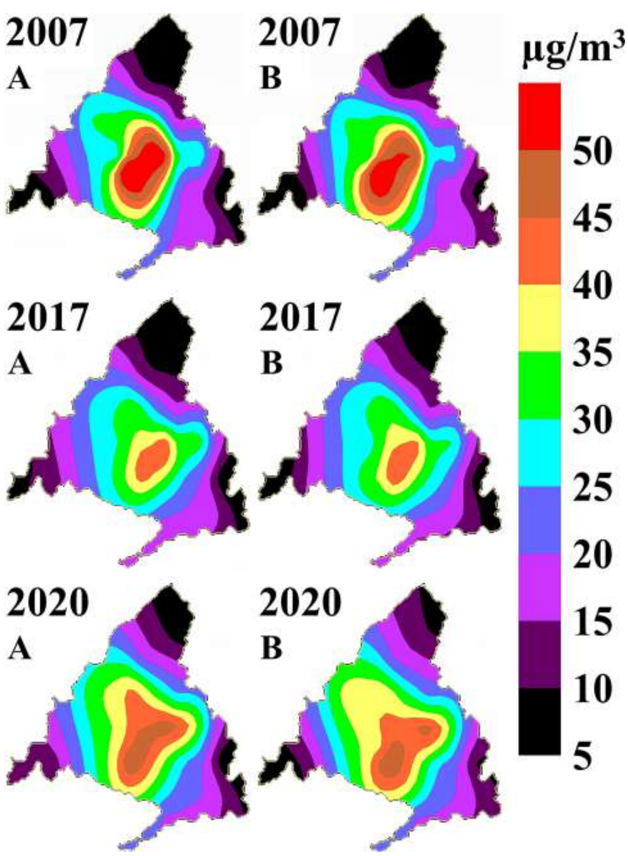

Fig. 6 Iso-concentration map for annual average NO_2 levels (2007, 2017, and 2020). **A** Currently fixed stations (pack 1) and **B** Proposed fixed stations (pack 2)

Conclusions

Since air quality networks' efficiency should be evaluated over time because of the variation influencing agents in polluting compounds in the air matrix, this research aims to furnish a stepwise methodological framework for assessing the performance of air quality networks. Currently, a harmonised method is unavailable at the legislative level.

A combined study involving chemometric analysis, position measures, and population data was conducted to achieve the proposed previously objective. In this sense, the body of outcomes exhibited in this work supports the validity of the suggested methodological framework for optimising AQMN layout by developing the following sequence: (i) evaluation of the existence of redundant fixed stations, (ii) detection of the non-redundant fixed stations, and (iii) proposal of the most representative non-redundant fixed stations within the AQMN. This approach is also adequate for assessing fixed monitoring sites' representativeness within an AQMN.

As an innovative aspect, this work is the first to conduct a testing process using pollution NO_2 data not included in the development of the proposed methodology, which is fundamental for securing the validity of the methodology. For that, a consistent validation of the suggested approach was addressed using GIS, including analysing the spatial distribution gradient of air pollutant concentrations. It is quantified the impact of (i) the temporal frame employed in developing the suggested methodology (variability $< 5\%$) and (ii) removing redundant stations and non-redundant stations non-proposed on the spatial distribution gradient regarding original AQMN capacity (similitude $> 95\%$), thereby sustaining the practical application of the recommended methodological sequence. The proposed methodological framework (based on annual average NO_2 concentrations) also applies to lower resolution NO_2 data.

The methodology reported in this study can serve as a helpful tool, within the air quality field, for managing AQMN, especially in the case of a dense network of measuring stations working with the use of expensive reference

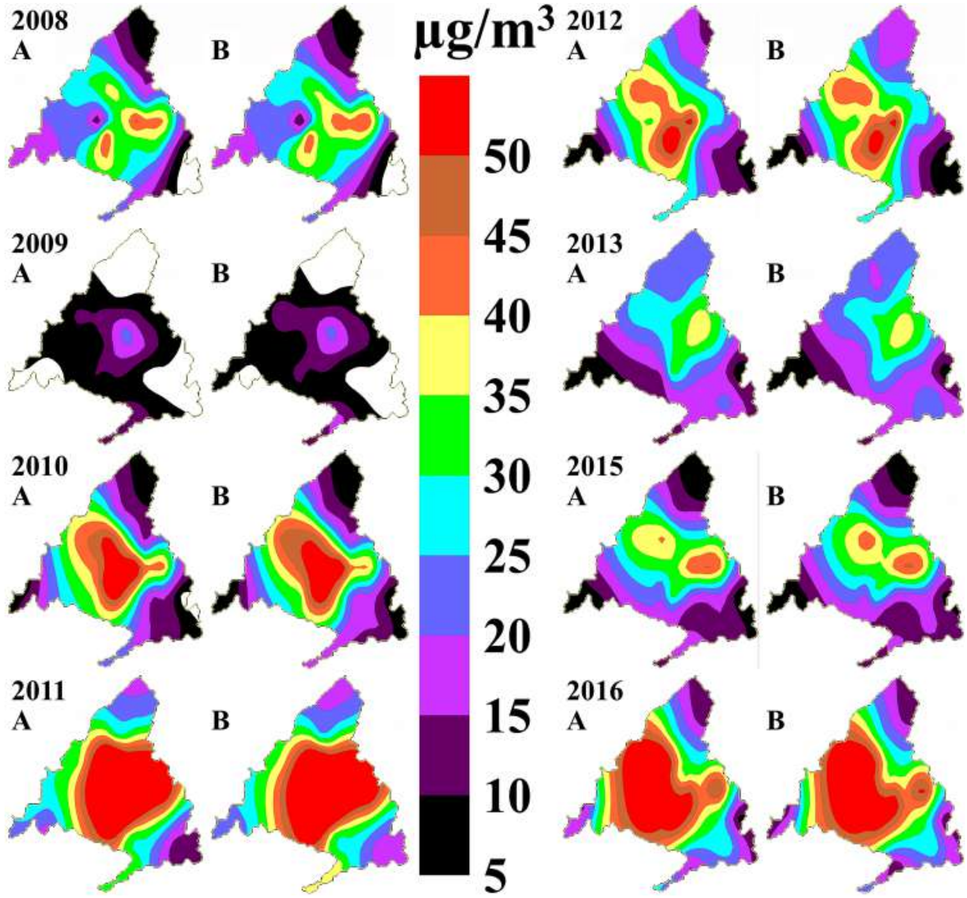

Fig. 7 Iso-concentration map for daily average NO_2 levels (December 12, 2008–2016). **A** Currently fixed stations (pack 1) and **B** Proposed fixed stations (pack 2). Note: Daily 2014 data are unavailable in the Community of Madrid's open data portal

methods. Therefore, potential applications are highly relevant for existing AQMN and implementing new AQMN in a concrete territory. In the first case, this methodological procedure would nominate the most representative non-redundant fixed stations within an existing AQMN. In the second case, it would identify the most representative monitoring sites for measuring air pollutants using previous air pollution information, reducing the required investment. On the other hand, this methodology would support the evaluation of long-term air quality management plans and help decision-making by the competent authorities.

Finally, the methodological framework would complement the guidelines set by European Legislation on air pollutants monitoring at fixed stations, acting as a benchmark for developing harmonised methodology.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11869-022-01254-4.

Author contribution All authors contributed to developing this research work. Data collection was performed by Galán-Madruga D and Cárdenas-Escudero J. Data treatment was conducted by Galán-Madruga, D. Discussion of results was elaborated by Galán-Madruga D, Broomandi P, Oleniacz R, and Cáceres JO. The first draft of the manuscript was written by Galán-Madruga, D, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate The present work does not involve human or animal subjects; therefore, ethical approval is unnecessary. Similarly, informed consent to participate is not required.

Consent for publication This work does not contain any individual person's data in any form, so consent to publish is not required.

Competing interests The authors declare no competing interests.

References

Araki S, Hasunuma H, Yamamoto K et al (2021) Estimating monthly concentrations of ambient key air pollutants in Japan during 2010–2015 for a national-scale birth cohort. Environ Pollut 284:117483. https://doi.org/10.1016/j.envpol.2021.117483

Austin E, Coull BA, Zanobetti A, Koutrakis P (2013) A framework to spatially cluster air pollution monitoring sites in US based on the

$PM_{2.5}$ composition. Environ Int 49:244–254. https://doi.org/10.1016/j.envint.2013.06.003
Baca-López K, Fresno C, Espinal-Enríquez J, Martínez-García M, Camacho-López M.A, Flores-Merino MV, Hernández-Lemus E (2021) Spatio-temporal representativeness of air quality monitoring stations in Mexico City: implications for public health. Front Public Health 12:1–15. https://doi.org/10.3389/fpubh.2020.536174
Benis KhZ, Fatehifar E (2015) Optimal design of air quality monitoring network around an oil refinery plant: a holistic approach. Int J Environ Sci Technol 12(4):1331–1342. https://doi.org/10.1007/s13762-014-0723-8
Bhardwaj R, Pruthi D (2020) Development of model for sustainable nitrogen dioxide prediction using neuronal networks. Int J Environ Sci Technol 17:2783–2792. https://doi.org/10.1007/s13762-019-02620-z
Chakraborty P, Zhang G, Cheng H, Balasubramanian P, Li J, Jones KC (2017) Passive air sampling of polybrominated diphenyl ethers in New Delhi, Kolkata, Mumbai and Chennai: Levels, homologous profiling and source apportionment. Environ Pollut 231:1181–1187. https://doi.org/10.1016/j.envpol.2017.08.044
Chen CH, Lin WL, Chen CH (2006) Development of a multiple objective planning theory and system for sustainable air quality monitoring networks. Sci Total Environ 354:1–19. https://doi.org/10.1016/j.scitotenv.2005.08.018
Comunidad de Madrid (2022) Available at http://gestiona.madrid.org/azul_internet/run/j/InformEvaluacionAccion.icm?ESTADO_MENU=7. Accessed 2 Sept 2022
Dancey CP, Reidy J (2007) Statistics without Maths for Psychology. Pearson Education
Dincer NG, Yalçin MO (2016) Revealing information and equipment redundancies in air pollution monitoring networks in Turkey. Int J Environ Sci Technol 13:2927–2938. https://doi.org/10.1007/s13762-016-1118-9
Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air. OJ L 23, 26.1.2005. 3–16
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. OJ L 152, 11.6.2008, 1–44
Domínguez-López D, Adame JA, Hernández-Ceballos MA, Vaca F, De la Morena BA, Bolívar JP (2014) Spatial and temporal variation of surface ozone, NO and NO_2 at urban, suburban, rural and industrial sites in the southwest of the Iberian Peninsula. Environ Monit Assess 186:5337–5351. https://doi.org/10.1007/s10661-014-3783-9
EN 14211:2012 Standard: Ambient Air. Standard Method for the Measurement of the Concentration of Nitrogen Dioxide and Nitrogen Monoxide by Chemiluminescence. https://standards.iteh.ai/catalog/standards/cen/46e97f97-9007-4b4c-8040-3ce5b0187ed5/en-14211-2012. Accessed 5 Sept 2022
Eurostart (2022) Available at https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20200430-1. Accessed 28 Mar 2022
Galán Madruga D, Fernández Patier R, Sintes Puertas MA, Romero García MD, Cristóbal López A (2018) Characterization and local emission sources for ammonia in an urban environment. Bull Environ Contam Toxicol 100:593–599. https://doi.org/10.1007/s00128-018-2296-6
Galán-Madruga D (2021) Importance of Air Quality Networks in Controlling Exposure to Air Pollution. In Environmental Emissions (EBook ISBN 978–1–83968–512–5). Editor: Richard Viskup. Published in London. United Kingdom. IntechOpen. Chapter 3:37–51. https://doi.org/10.5772/Intechopen.92335
Galán-Madruga D (2021) A methodological framework for improving air quality monitoring network layout. Applications to environment management. J Environ Sci 102:138–147. https://doi.org/10.1016/j.jes.2020.09.009
Galán-Madruga D, García-Cambero JP (2022) An optimized approach for estimating benzene in ambient air within an air quality monitoring network. J Environ Sci 111:164–174. https://doi.org/10.1016/j.jes.2021.03.005
Galán-Madruga D, Ubeda RM, Terroba JM et al (2022) Influence of the products of biomass combustion processes on air quality and cancer risk assessment in rural environmental (Spain). Environ Geochem Health 44:2595–2613. https://doi.org/10.1007/s10653-021-01052-4
Ghaffarpasand O, Nadi S, Shalamzari ZD (2020) Short-term effects of anthropogenic/natural activities on the Tehran criteria air pollutants: source apportionment and spatiotemporal variation. Build Environ 186:107298. https://doi.org/10.1016/j.buildenv.2020.107298
Gómez-Losada A, Lozano-García A, Pino-Mejías R, Contreras-González J (2014) Finite mixture models to characterize and refine air quality monitoring networks. Sci Total Environ 485–486:292–299. https://doi.org/10.1016/j.scitotenv.2014.03.091
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. https://doi.org/10.1016/j.apr.2019.09.009
Gorzelnik T, Oleniacz R (2019) Suitability analysis of new air quality monitoring stations in Krakow as related to assessment of spatial and temporal variability of PM_{10} concentrations. Geomat Environ Eng 13:31–45. https://doi.org/10.7494/geom.2019.13.2.31
Hacıoğlu HI, Arı A, Özkan A, Elbir T, Tuncel G, Yay OD, Gaga EO (2016) A new approach for site selection of air quality monitoring stations: multi-criteria decision-making. Aerosol Qual Res 16:1390–1402. https://doi.org/10.4209/aaqr.2014.11.0273
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. https://doi.org/10.1016/j.atmosenv.2018.01.011
Huang Z, Yu Q, Liu Y, Ma W, Chen L (2019) Optimal design of air quality monitoring network for pollution detection and source identification in industrial parks. Atmosphere 10:318–834. https://doi.org/10.3390/atmos10060318
Ibarra-Berastegi G, Sáenzm J, Ezcurra A et al (2009) Assessing spatial variability of SO_2 field as detected by an air quality network using self-organizing maps, cluster, and principal component analysis. Atmos Environ 43:3829–3836. https://doi.org/10.1016/j.atmosenv.2009.05.010
Ibarra-Berastegi G, Sáenz J, Ezcurra A, Ganzedo U, Elías A, et al (2010) Identification of redundant sensors in an air pollution network using cluster analysis and SOM. WIT Trans Ecol Environ 136. https://doi.org/10.2495/AIR100321
INE (2021) Instituto Nacional de Estadística. Available at http://www.ine.es. Accessed 28 Mar 2022
Jain AK (2010) Data clustering: 50 years beyond K-means. Pattern Recognit Lett 31:651–666. https://doi.org/10.1016/j.patrec.2009.09.011
Kao JJ, Hsieh MR (2006) Utilizing multiobjective analysis to determine an air quality monitoring network in an industrial district. Atmos Environ 40:1092–1103. https://doi.org/10.1016/j.atmosenv.2005.11.003
Karppinen A, Kukkonen J, Elolähde T, Konttinena M, Koskentalo T (2000) A modelling system for predicting urban air pollution: comparison of model predictions with the data of an urban measurement network in Helsinki. Atmos Environ 34:3735–3743. https://doi.org/10.1016/S1352-2310(00)00073-X
Kolasa-Więcek A, Suszanowicz D (2019) Air pollution in European countries and life expectancy—modelling with the use of neural network. Air Qual Atmos Health 12:1335–1345. https://doi.org/10.1007/s11869-019-00748-y

Kollati H., Debnath A (2021) A short review on different techniques used for site selection of air quality monitoring stations. JIEDM 2:27–30. https://doi.org/10.31181/jiedm200201026h

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World Map of the Köppen-Geiger climate classification updated. Meteorol Z 15:259–263. https://doi.org/10.1127/0941-2948/2006/0130

Lamphar H, Kocifaj M, Limón-Romero J, Paredes-Tavares J, Chakameh SD et al (2022) Light pollution as a factor in breast and prostate cancer. Sci Total Environ 806:150918. https://doi.org/10.1016/j.scitotenv.2021.150918

Madruga D, Úbeda R, Terroba J, dos Santos S, García-Cambero JP (2019) Particle-associated polycyclic aromatic hydrocarbons in a representative urban location (indoor-outdoor) from South Europe: Assessment of potential sources and cancer risk to humans. Indoor Air 29:817–827. https://doi.org/10.1111/ina.12581

Maione C, Nelson DR, Barbosa RM (2019) Research on social data by means of cluster analysis. Applied Comput Inform 15:153–162. https://doi.org/10.1016/j.aci.2018.02.003

Martin F, Santiago JL, Kracht O, Garcia L, Gerboles M (2015) FAIRMODE Spatial Representativeness Feasibility Study. JRC Technical Report ISBN 978–92–79–50322–1 https://doi.org/10.2788/49487. European Comission, Joint Reasearch Centre, Institute for Environment and Sustainability. Publications Office of European Union, Luxembourg

Meng J (2021) Raster data projection transformation based-on Kriging interpolation approximate grid algorithm. Alex Eng J 60:2013–2019. https://doi.org/10.1016/j.aej.2020.12.006

Merriam-Webster Dictionary. Available at https://www.merriam-webster.com/dictionary/correlation. Accessed 6 Dec 2021

Mikheev AV, Kazakov BN (2017) Correlation analysis of spectroscopic data. J Lumin 184:117–122. https://doi.org/10.1016/j.jlumin.2016.12.019

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. https://doi.org/10.1016/j.atmosenv.2009.07.045

Munir S, Mayfield M, Coca D, Jubb SA (2019) Structuring an integrated air quality monitoring network in large urban areas – discussing the purpose, criteria and deployment strategy. Atmos Environ X 100027. https://doi.org/10.1016/j.aeaoa.2019.100027

Núñez-Alonso D, Pérez-Arribas LV, Manzoor S, Cáceres JO (2019) Statistical tools for air pollution assessment: multivariate and spatial analysis studies in the Madrid region. J Anal Methods Chem 9753927. https://doi.org/10.1155/2019/9753927

Oleniacz R, Gorzelnik T (2021) Assessment of the variability of air pollutant concentrations at industrial, traffic and urban background stations in Krakow (Poland) using statistical methods. Sustainability 13(10):5623. https://doi.org/10.3390/su13105623

Park J, Ryu H, Kim E et al (2020) Assessment of $PM_{2.5}$ population exposure of a community using sensor-based air monitoring instruments and similar time-activity groups. Atmos Pollut Res 11:1971–1981. https://doi.org/10.1016/j.apr.2020.08.010

Penkova TG (2017) Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety. Procedia Comput Sci 112:99–108. https://doi.org/10.1016/j.procs.2017.08.179

Pires JCM, Martins FG, Alvim-Ferraz MCM, Pereira MC (2009) Evaluation of redundant measurements on the air quality monitoring network of Lisbon and Tagus Valley. Chem Prod Process Model 4. https://doi.org/10.2202/1934-2659.1280

Pires JCM, Sousa SIV, Pereira MC, Alvim-Ferraz MCM, Martins FG (2008) Management of air quality monitoring using principal component and cluster analysis-Part I: SO_2 and PM_{10}. Atmos Environ 42:1249–1260. https://doi.org/10.1016/j.atmosenv.2007.10.044

Righini G, Cappelletti A, Ciucci A et al (2014) GIS based assessment of the spatial representativeness of air quality monitoring stations using pollutant emissions data. Atmos Environ 97:121–129. https://doi.org/10.1016/j.atmosenv.2014.08.015

Rizzo MJ, Scheff PA (2004) Assessing ozone networks using positive matrix factorization. Environ Prog 23:110–119. https://doi.org/10.1002/ep.10018

Salem SB, Naouali S, Chtourou Z (2018) A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach. Comput Electr Eng 68:463–483. https://doi.org/10.1016/j.compeleceng.2018.04.023

Santiago JL, Martín F, Martilli A (2013) A computatioinal fluid dinamic modelling approach to assess the representativeness of urban monitoring stations. Sci Total Environ 454–455:61–72. https://doi.org/10.1016/j.scitotenv.2013.02.068

Sarigiannis DA, Saisana M (2007) Multi-objective optimization of air quality monitoring. Environ Monit Assess 136:87–99. https://doi.org/10.1007/s10661-007-9725-z

Silva C, Quiroz A (2003) Optimization of the atmospheric pollution monitoring networkat Santiago de Chile. Atmos Environ 37:2337–2345. https://doi.org/10.1016/S1352-2310(03)00152-3

Soares J, Makar PA, Aklilu Y, Akingunola A (2018) The use of hierarchical clustering for the design of optimized monitoring networks. Atmos Chem Phys 18:6543–6566. https://doi.org/10.5194/acp-18-6543-2018

Tuck AF (2021) theoretical chemistry and the calculation of the atmospheric state. Atmosphere 12:727. https://doi.org/10.3390/atmos12060727

Verghese S, Nema AK (2022) Optimal design of air quality monitoring networks: a review systematic. Stoch Environ Res Risk Assess. https://doi.org/10.1007/s00477-022-02187-1

Videnova I, Nedialkov D, Dimitrova M, Popova S (2006) Neural networks for air pollution nowcasting. Appl Artif Intell 20:493–506. https://doi.org/10.1080/08839510600753741

Wackernagel H (2003) Multivariate Geostatistics. Springer, New York

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. https://doi.org/10.1016/j.atmosenv.2018.07.040

WHO. World Health Organization (2006) Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. Global Update 2005. Summary of Risk Assessment. WHO/SDE/PHE/OEH/06.02. Available at https://apps.who.int/iris/handle/10665/69477. Accessed 9 Apr 2021

Yatkin S, Gerboles M, Belis CA, Karagulian F, Lagler F, Barbiere M, Borowiak A (2022) Representativeness of an air quality monitoring station for $PM_{2.5}$ and source apportionment over a small urban domain. Atmos Pollut Res 11:225–233. https://doi.org/10.1016/j.apr.2019.10.004

Zeydan Ö, Pekkaya M (2021) Evaluating air quality monitoring stations in Turkey by using multi criteria decision making. Atmos Pollut Res 12:101046. https://doi.org/10.1016/j.apr.2021.03.009

Zhao L, Xie Y, Wang J, Xu X (2015) A performance assessment and adjustment program for air quality monitoring networks in Shanghai. Atmos Environ 122:382–392. https://doi.org/10.1016/j.atmosenv.2015.09.069

Zhu F, Ding R, Lei R et al (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. https://doi.org/10.1016/j.rmed.2018.11.019

Zoroufchi BK, Fatehifar E (2015) Optimal design of air quality monitoring network around an oil refinery plant: a holistic approach. Int J Environ Sci Technol 12:1331–1342. https://doi.org/10.1007/s13762-014-0723-8