



Entropy-based air quality monitoring network optimization using NINP and Bayesian maximum entropy

Ali Haddadi¹ · Mohammad Reza Nikoo² · Banafsheh Nematollahi³ · Ghazi Al-Rawas² · Malik Al-Wardy⁴ · Mehdi Toloo⁵ · Amir H. Gandomi^{6,7}

Received: 1 February 2023 / Accepted: 11 June 2023 / Published online: 24 June 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Effectual air quality monitoring network (AQMN) design plays a prominent role in environmental engineering. An optimal AQMN design should consider stations' mutual information and system uncertainties for effectiveness. This study develops a novel optimization model using a non-dominated sorting genetic algorithm II (NSGA-II). The Bayesian maximum entropy (BME) method generates potential stations as the input of a framework based on the transinformation entropy (TE) method to maximize the coverage and minimize the probability of selecting stations. Also, the fuzzy degree of membership and the nonlinear interval number programming (NINP) approaches are used to survey the uncertainty of the joint information. To obtain the best Pareto optimal solution of the AQMN characterization, a robust ranking technique, called Preference Ranking Organization METHod for Enrichment Evaluation (PROMETHEE) approach, is utilized to select the most appropriate AQMN properties. This methodology is applied to Los Angeles, Long Beach, and Anaheim in California, USA. Results suggest using 4, 4, and 5 stations to monitor CO, NO₂, and ozone, respectively; however, implementing this recommendation reduces coverage by 3.75, 3.75, and 3 times for CO, NO₂, and ozone, respectively. On the positive side, this substantially decreases TE for CO, NO₂, and ozone concentrations by 8.25, 5.86, and 4.75 times, respectively.

Keywords Air quality · Bayesian maximum entropy (BME) · Fuzzy set theory · Multi-criteria decision-making (MCDM) · Nonlinear interval number programming (NINP) · Transinformation entropy (TE)

Responsible Editor: Marcus Schulz

✉ Mohammad Reza Nikoo
m.reza@squ.edu.om

Ali Haddadi
alihaddadi1998@gmail.com

Banafsheh Nematollahi
bnema001@ucr.edu

Ghazi Al-Rawas
ghazi@squ.edu.om

Malik Al-Wardy
mwardy@squ.edu.om

Mehdi Toloo
mehdi.toloo@vsb.cz

Amir H. Gandomi
gandomi@uts.edu.au

² Department of Civil and Architectural Engineering, Sultan Qaboos University, Muscat, Oman

³ Department of Environmental Sciences, University of California, Riverside, USA

⁴ Department of Soils, Water and Agricultural Engineering, Sultan Qaboos University, Muscat, Oman

⁵ Department of Systems Engineering, Faculty of Economics, VŠB - Technical University of Ostrava, Ostrava, Czech Republic

⁶ Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia

⁷ University Research and Innovation Center (EKIK), Óbuda University, 1034 Budapest, Hungary

¹ Department of Civil and Environmental Engineering, Shiraz University, Shiraz, Iran

Abbreviations

BME	Bayesian maximum entropy
CO	Carbon monoxide
NINP	Nonlinear interval number programming
NO ₂	Nitrogen dioxide
NSGA-II	Non-dominated sorting genetic algorithm II
PROMETHEE	Preference Ranking Organization METHod for Enrichment Evaluation
TE	Transinformation entropy
TE-D	Transinformation entropy-distance

Introduction

Efficient air quality monitoring networks (AQMN) provide information that allows the source and location of air pollution to be identified and, thus, have attracted substantial attention in environmental engineering. Such information is beneficial for preventing the catastrophic damage of pollution to humans' livelihoods (Chen et al. 2006; Dong et al. 2022). Nowadays, techniques associated with effective air quality monitoring system designs have been focused in several studies based on communication and computing approaches. In more detail, assessing the air quality systems gives engineers the golden opportunity to manage this environmental issue with efficient adaptive and intelligent techniques (Hadj Sassi and Fourati 2022). Different simulation techniques have been applied for effective AQMN design, including the figure-of-merit (FOM) approach (Liu et al. 1986), cell approach (Elkamel et al. 2008), passive diffusion sampler-based model (Lozano et al. 2009), industrial source complex short-term 3 (ISCST3) model (Zoroufchi Benis et al. 2015), dispersion model (Zoroufchi Benis et al. 2016), geographic information system (GIS), cluster analysis (Li et al. 2018), interpolation method (Boubrima et al. 2019), and simple linear regression approach (Galán-Madruga 2021). In addition, the MCA method (Elkamel et al. 2008; Zoroufchi Benis et al. 2016), the ISCST3 model (Sharma and Chandra 2008; Kansal et al. 2011), GIS, and kriging (Chung et al. 2019; Galán-Madruga 2021; Wang et al. 2021) have been used to predict the amount of air pollutants around an emission source. However, although these simulation models are suitable for estimating the air pollutants' concentration around an emission source, this study focuses on designing an effective AQMN for a large area from existing monitoring stations and estimated ones. To this extent, these tools can be supplemented using robust optimization algorithms (Trujillo-Ventura and Ellis 1991; Kao and Hsieh 2006). Therefore, the optimal design of AQMNs draws substantial attention because of the air pollution severity associated with the monitoring stations' installation and maintenance.

Applied optimization algorithms for designs of AQMNs started from geospatial and statistical techniques to evolutionary and heuristic approaches (Sun et al. 2019; Verghese and Nema 2022), such as heuristic optimization algorithm (Elkamel et al. 2008), ant colony and genetic optimization model (Zoroufchi Benis et al. 2016), genetic algorithm (Hao and Xie 2018), stepwise genetic algorithm (Li et al. 2019), ant colony optimization (Rathee et al. 2019), and clustering analyses (Lu et al. 2011; Stolz et al. 2020). However, as far as we know, the literature has not considered optimizing the AQMN parameters using non-dominated sorting genetic algorithm II (NSGA-II), a robust multi-objective optimization method (Yazdandoost et al. 2022), principally in the presence of uncertainties and shared information between stations. Hence, the present study develops an optimization model using an NSGA-II optimization algorithm concerning the uncertainties and the stations' transfer information as a novel method for optimal AQMN characterization.

Employing uncertainties within the NSGA-II optimization model for the AQMN characterization is the most critical element for the optimum design of the AQMN. Mutual information, related uncertainties, and the fuzzy analytic hierarchy process (Mofarrah and Husain 2010) can serve this purpose. To this extent, while transinformation entropy (TE) is a well-known approach for calculating each monitoring station's mutual information with another monitoring station, the preliminary investigation pointed out that adopting TE within the NSGA-II optimization framework for the AQMN using a tool for generating an accurate estimation of potential air quality monitoring stations such as Bayesian maximum entropy (BME) have not been considered in the literature. This highlights the need for presenting an effectual NSGA-II optimization model based on the TE approach using the BME technique for accurate optimal AQMN characterization, which can be empowered when the numbers are assumed to have interval natures using a robust nonlinear interval number programming (NINP) tool (Jiang et al. 2008; Nematollahi et al. 2022).

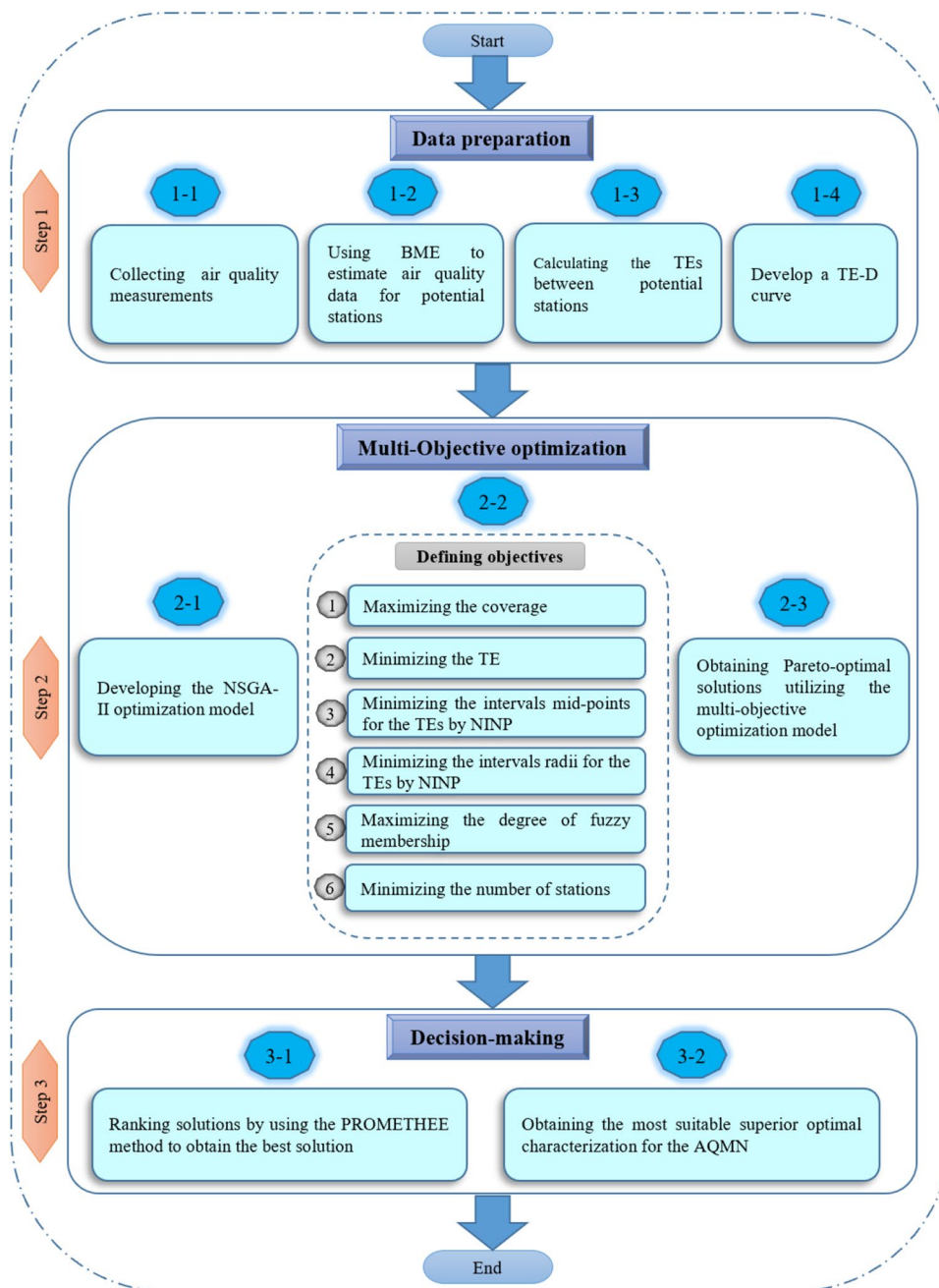
This study suggests a novel multi-objective optimization framework using NSGA-II based on the TE and NINP approaches using the BME method, considering fuzzy set theory for the accurate and optimal characterization of AQMNs. In this innovative framework, first, the BME method is applied to generate potential monitoring stations; then, the data of the existing and potential stations are collected. Second, the TE method calculates the shared information between each station pair, plotting TE-distance (TE-D) curves. Furthermore, in this stage, the desired configuration of the monitoring network is obtained. Third, NSGA-II multi-objective optimization model is operated using the objectives of minimizing the number of monitoring stations, the value of TE between

selected stations, and the average and radius of variations for the TE values and maximizing the fuzzy degree of membership for the amount of TE and the AQMN coverage to obtain a series of Pareto optimum AQMN stations. Finally, the Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) is adopted to rank the Pareto optimal solutions acquired from the multi-objective framework to select the most superior optimum AQMN characterization. To prove the usability of this methodology, it is applied to the real-world AQMNs in Los Angeles, Long Beach, and Anaheim cities as the

substantial metropolitan in California, USA. To sum it up, the presented novel multi-objective optimization model fills the gaps in the literature by.

1. Making the multi-objective optimization framework using NSGA-II for AQMN characterization
2. Adopting the BME method to generate potential air quality monitoring stations as a set of input data to the multi-objective optimization framework
3. Utilizing the TE method to calculate the stations' shared information

Fig. 1 Schematic view of the suggested framework to design a superior optimum AQMN



4. Applying the NINP technique in the optimization algorithm for the AQMN to include the interval natures of the numbers
5. Considering the TE and NINP approaches simultaneously through the multi-objective optimization framework
6. Maximizing the fuzzy degree membership for the TEs and the AQMN in the optimization model
7. Employing the PROMETHEE method to obtain the superior optimum AQMN characterization
8. Obtaining the superior optimal AQMN characterization for the significant metropolitan in California, USA, using the proposed innovative framework

Methodology

Figure 1 shows that the proposed framework in this study finds the best appropriate optimal AQMN within three primary stages: (i) preparing data, (ii) multi-objective optimization, and (iii) multi-criteria decision-making (MCDM). It is noteworthy that MATLAB® is used for data analysis, developing algorithms, and making models.

Stage 1 collects air quality concentration measurement data, generates potential monitoring stations with the help of the BME method, and calculates the TE of all potential and existing monitoring stations to develop a TE-D curve. To successfully estimate the potential stations of AQMN, the BME method uses data from all existing stations (i.e., stations' locations and pollutant concentrations) as the input.

Stage 2 develops an optimization model using NSGA-II to acquire the Pareto optimal solutions for the AQMN with

six objective functions: maximizing the AQMN coverage and the degree of fuzzy membership and minimizing (i) the TE values, (ii) the average of TE variations, (iii) the TE variation radii, and (iv) the number of stations. Finally, stage 3 ranks solutions acquired from the optimization framework using the PROMETHEE technique to acquire the most suitable optimal characterization of the AQMN.

Transinformation entropy

Transinformation entropy measures the quantity of information about a random variable by observing another variable (Vicente et al. 2011). When the transinformation entropy is high between two variables, it indicates that the value of one variable gives essential information about the value of the other variable. Conversely, if the transinformation entropy is low, it suggests that the two variables are mostly independent. Transinformation entropy has a symmetrical characteristic, meaning that the information obtained about A when B is known is equal to that obtained about B when A is known. This entropy measure is always non-negative and has a zero value only if the two variables are independent.

Multi-objective optimization model

In this research, the optimization model uses NSGA-II to (i) maximize the coverage of the AQMN (Eq. (1)), (ii) minimize the amount of the TE for the potential stations (Eq. (2)), (iii) minimize the average of TE variations (Eq. (3)), (iv) minimize the radius of TE variations (Eq. (4)), (v) maximize the fuzzy degree of membership for the stations (Eq. (5)), and (vi) minimize the number of stations (Eq. (6)).

Fig. 2 The study area

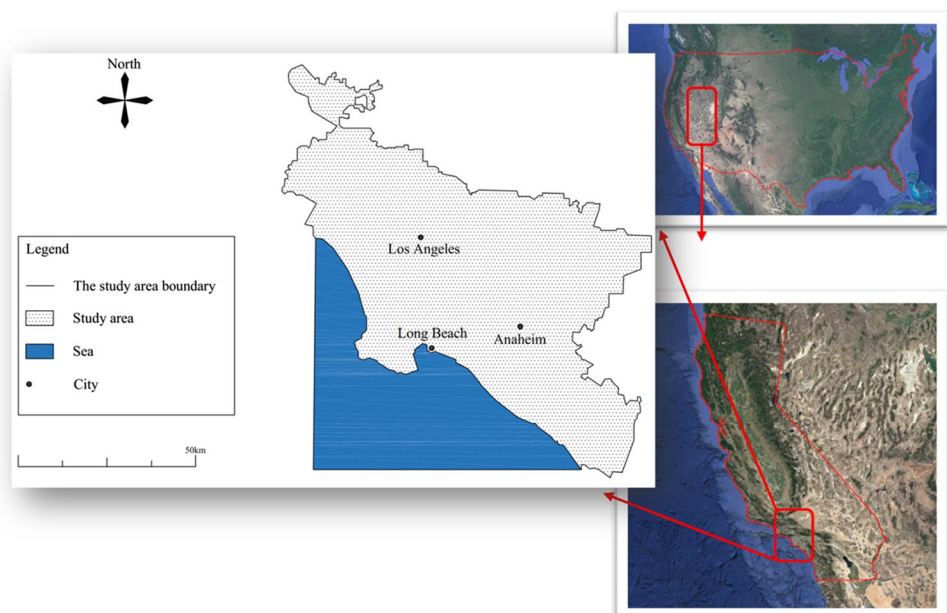
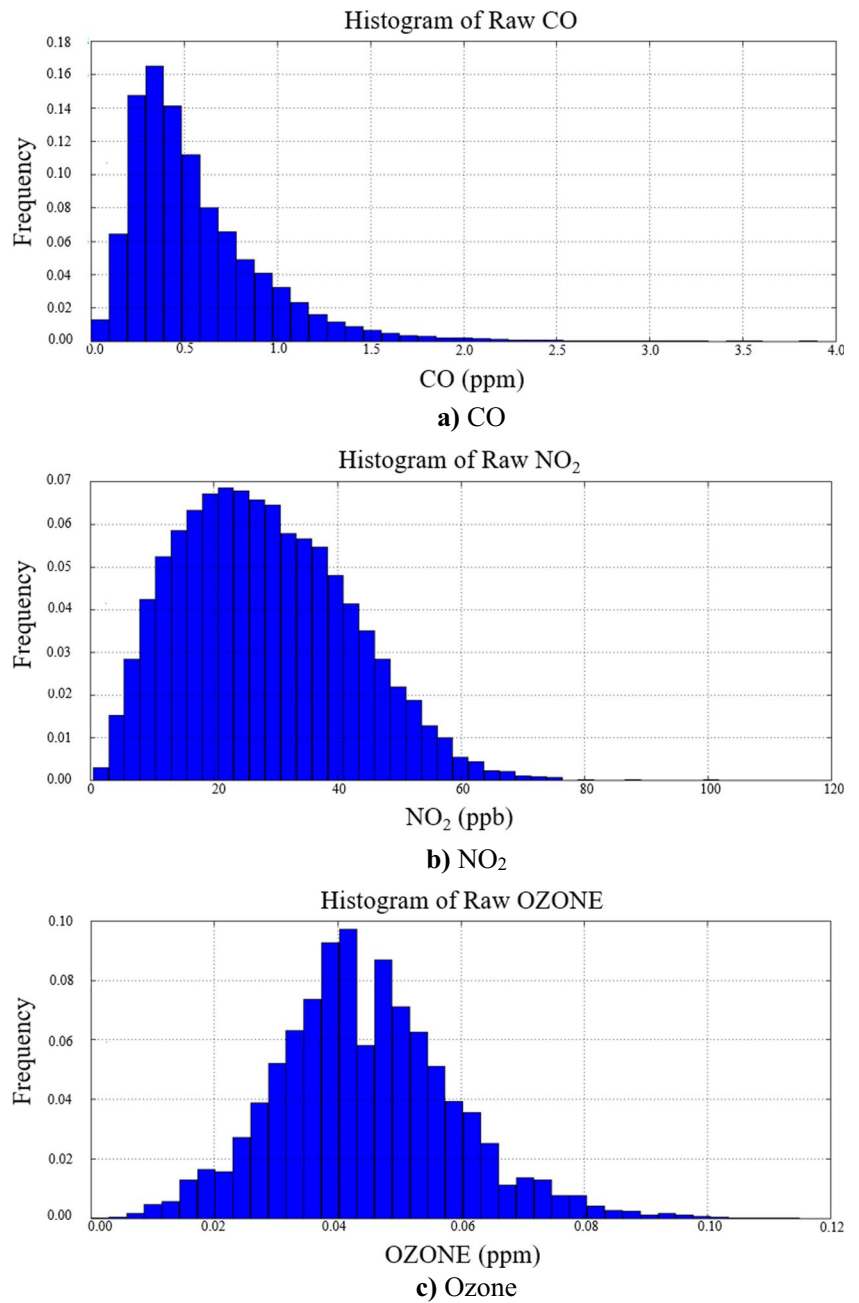


Fig. 3 Distribution of pollutants’ concentrations in the existing stations



$$\text{Maximize } Z_1 = \sum_{i=1}^{N_p} \left\{ \begin{array}{l} b_i \left(\frac{D_{\max} - D_i}{D_{\max} - D_{\text{opt}}} \right) \text{ if } D_i > D_{\text{opt}} \\ b_i \text{ if } D_i \leq D_{\text{opt}} \end{array} \right\} \quad (1)$$

$$\text{Minimize } Z_4 = \sum_{i=1}^{N_p} b_i (\max(\text{TE}_i) - \min(\text{TE}_i)) / 2 \quad (4)$$

$$\text{Minimize } Z_2 = \sum_{i=1}^{N_p} \left\{ \begin{array}{l} b_i \left(\frac{\text{TE}_{\max} - \text{TE}_i}{\text{TE}_{\max} - \text{TE}_{\min}} \right) \text{ if } D_i < D_{\text{opt}} \\ b_i \text{ if } D_i \geq D_{\text{opt}} \end{array} \right\} \quad (2)$$

$$\text{Minimize } Z_5 = \sum_{i=1}^{N_p} b_i (1 - D_{m_i}) / N_p \quad (5)$$

$$\text{Minimize } Z_3 = \sum_{i=1}^{N_p} b_i (\max(\text{TE}_i) + \min(\text{TE}_i)) / 2 \quad (3)$$

$$\text{Minimize } Z_6 = \frac{\left(\sum_{i=1}^{N_p} b_i - N_{\min} \right)}{(N_p - N_{\min})} \quad (6)$$

$$N_{\min} = (D_{\max}/D_{\text{opt}}) + 1 \tag{7}$$

where Z_i is the i th objective function; b_i is the auxiliary binary variable that is equal to 1 if there is a monitoring station; otherwise, it is zero; N_p is the number of monitoring stations; N_{\min} is the monitoring stations' minimum number; D_i is the distance between the monitoring station i and the nearest monitoring station; D_{opt} and D_{\max} are the optimum and maximum distances between stations, respectively; TE_i is the TE between the monitoring station i and the nearest monitoring station; TE_{\max} is the maximum amount of

TE between stations; TE_{\min} is the minimum amount of TE between stations; and D_{m_i} is the degree of membership for each station.

The pseudocode of the algorithm mentioned above is presented in Fig. S3, Section S2, Supplementary Material.

PROMETHEE

The PROMETHEE prefers to rank solutions according to criteria, while VIKOR and TOPSIS are two aggregating function-based multiple-criteria decision-making algorithms demonstrating the concept of “closeness to the ideal” (Opricovic and Tzeng 2004). In addition, PROMETHEE is an easy-to-use method and can be employed in authentic planning problems (Ülengin et al. 2001). More importantly, this method is suitable for both quantitative and qualitative data, making it more user-friendly than other methods. Also, it consists of two components helping understand the process of ranking solutions as simply as possible. Finally, the PROMETHEE method is a robust MCDM method to rank a group of Pareto fronts, $= \{T_1, T_2, \dots, T_n\}$, according to a set of objectives, $C = \{C_1, C_2, \dots, C_m\}$, that are different from each other (Brans and Vincke 1985; Pourshahabi et al. 2018). This study considers an identical weight to all objectives because all objectives are equally important. Although the different sets of weights may lead to different results, our approach can easily handle all possible sets of weights given by the decision maker. The obtained results illustrate the capability of the methodology to use desirable weights and criteria based on existing conditions in different case studies.

Case study

The suggested framework in this research is used to design the AQMNs for Los Angeles, Long Beach, and Anaheim in California (USA) (Fig. 2). These cities are regarded as the most industrial, populated, and polluted cities in the USA, with an area of about 5000 km², in which many people are exposed to air pollution with its health risks.

In this case study, concentrations of all three pollutants, nitrogen dioxide (NO₂), carbon monoxide (CO), and ozone, are collected from each of the 15 monitoring stations in the study area between January 2015 and December 2016 by the US Environmental Protection Agency (U.S. EPA 2022). The locations of existing stations are plotted in Fig. 5, and the information about all three pollutants is from the EPA's website.

It is notable for asserting that the main reason for selecting these 15 stations from all existing ones is significant gaps in the time series data of removed monitoring stations. Data deficiency does not estimate data for estimated monitoring stations using BME. The precise locations of the chosen stations are available in Table S1, Section S6, Supplementary Material.

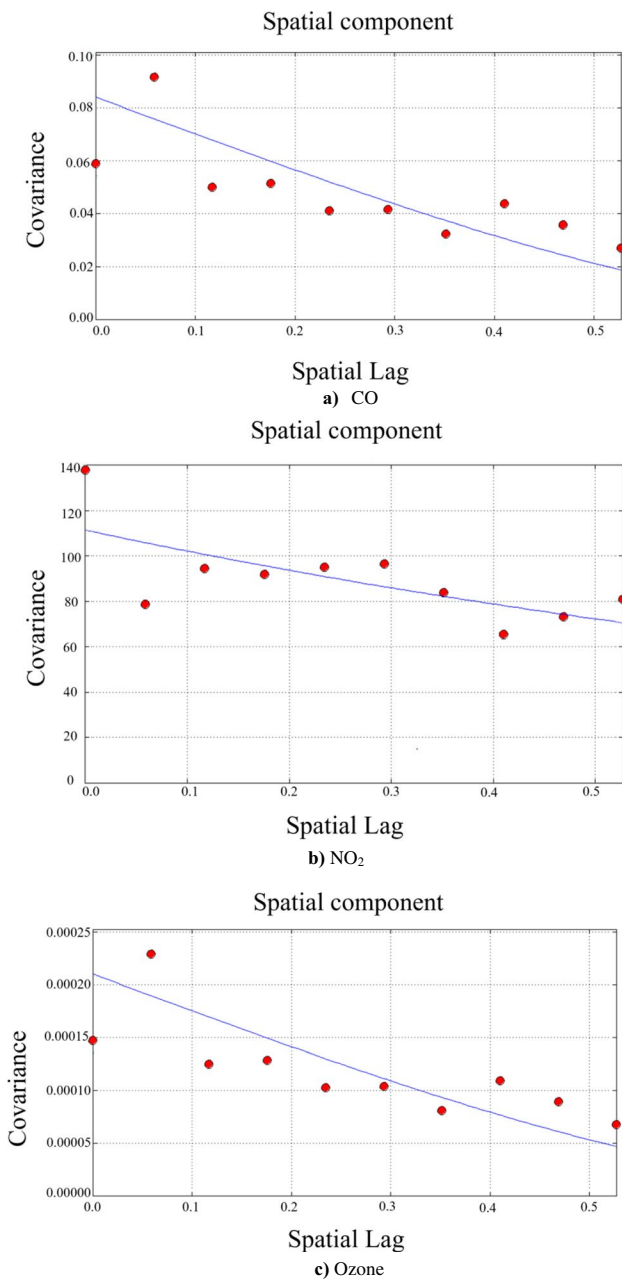


Fig. 4 Spatial covariance models of pollutants

Fig. 5 Location of the existing and BME potential stations

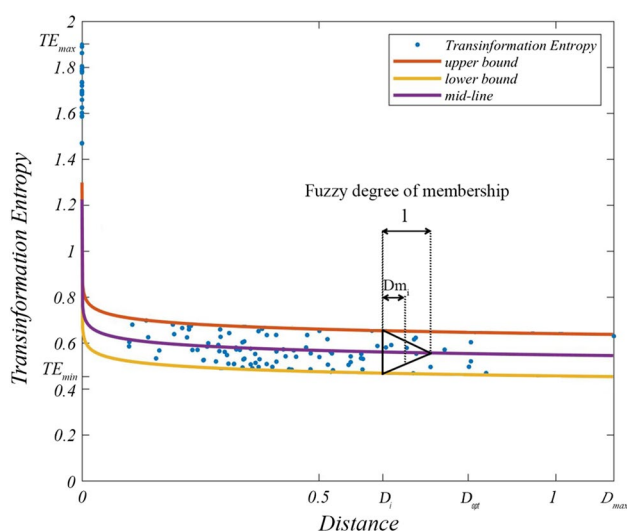
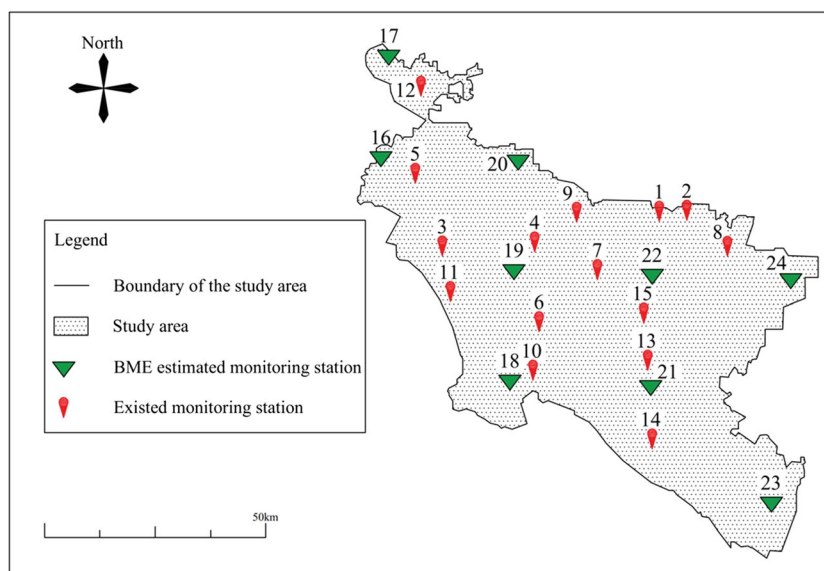


Fig. 6 A diagram of the TE-D curve and the fuzzy degree of membership (D_m)

Results and discussion

BME for potential monitoring station prediction

The BME approach predicts three air pollutants' concentrations, including CO, NO₂, and ozone in the selected locations among the existing stations to provide input data for the proposed optimization framework. Applying BME to obtain the potential stations is performed following a 3-step procedure.

First, the daily pollutants' concentrations, the stations' longitudes and latitudes, and the number of stations are collected as inputs to the BME model. After that, the distribution of pollutant concentrations is determined using histograms of the pollutants' frequencies, as illustrated in Fig. 3.

Second, the obtained data are analyzed spatially and temporally to estimate the pollutants' concentrations in considered time intervals and locations. The mean trend analysis calculates the data averages to provide a smooth time series through these analyses. Then, these spatial and temporal mean trends are plotted, as illustrated in Figs. S4 and S5, Section S7, Supplementary Material. Subsequently, the most suitable time and space covariance models are defined and fitted with their related parameters (Fig. 4).

Finally, the BME approach predicts nine potential stations, wherein their covariance structure of time and space in determining latitudes, longitudes, and periods are similar to the existing stations. The locations of these potential stations are selected to cover the study area and help find the most optimized AQMN (Fig. 5). The detailed BME potential stations' longitudes, latitudes, and station numbers are noted in Table S3, Section S8, Supplementary Material.

The mathematical models are regarded as a robust tool to select the most proper position for monitoring stations in AQMN optimization (Modak and Lohani 1985). The results show the BME method's importance in redesigning and improving the AQMN of the case study. For instance, monitoring stations 17 and 24 for CO, 18 and 21 for NO₂, and 17 and 21 for ozone are used in the most efficient AQMN. They constitute almost 50% of monitoring stations in the designed AQMN for air pollutants.

Transinformation entropy approach

The TE approach is used in this study to calculate the mutual information between each pair of stations. For this purpose, first, a matrix of order 24×24 is calculated to obtain the

Table 1 The multi-objective optimization model results

	No. of stations	Coverage	TE	TE variations		Fuzzy degree of membership
				Radii	Averages	
CO						
Minimum	0.02	0.04	1.2	0.28	1.16	0.02
Maximum	1	0.45	14.64	2.4	10.54	0.93
Average	0.42	0.13	7.11	1.15	4.99	0.42
NO₂						
Minimum	0.02	0.04	1.32	0.3	1.43	0
Maximum	1	0.44	13.89	2.49	12.9	0.9
Average	0.41	0.13	7.05	1.18	6.06	0.39
Ozone						
Minimum	0.04	0.04	1.05	0.37	2.26	0.05
Maximum	1	0.28	12.35	2.25	14.8	0.9
Average	0.44	0.11	6.76	1.15	7.47	0.45

TEs between all existing and potential stations. Each cell in this matrix presents the value of $TE(a,b)$ in which a and b are the number of stations or a th row and b th column in the TE matrix.

After that, a TE-D curve is plotted according to the TEs between each pair of stations and their corresponding distances, as displayed in Fig. 6. This figure illustrates that the maximum TE belongs to the nearest stations, and as the distance increases, the amount of the TE decreases. Besides, in this TE-D curve, the point about zero inclination is presented as the optimum value of the distance (D_{opt}) since the TE is minimum after this point.

Finally, the fuzzy degree of membership (D_m) is used to reduce uncertainty in the amount of TE between each pair of stations, which is zero in the lower and upper bounds of the TE-D curve and 1 in the curve fitted on all points of the TE-D curve.

Multi-objective optimization model

The suggested multi-objective optimization framework uses the abovementioned six objectives, resulting in a series of Pareto optimal solutions, as indicated in Table S4, Section S9, Supplementary Material. The population size and generation in this NSGA-II multi-objective optimization model are 200 and 2400, respectively. Table 1 demonstrates the maximum, minimum, and average values for the AQMN coverage, TE, average of TE variations, radii of TE variations, fuzzy degree of membership, and the number of stations for CO, NO₂, and ozone contaminants obtained from 150 acquired Pareto optimal solutions.

Relation between the number of monitoring stations, coverage, and transformation entropy

Figure 7 demonstrates the relation between the number of monitoring stations of monitoring networks, their coverage, and transformation entropy. This figure shows the relationship between the number of stations varying from 4 to 24, the spatial coverage, and redundant information for all pollutants. The amount of spatial coverage objective is to the negative power of 1 because the purpose is to maximize it in the multi-objective optimization algorithm.

PROMETHEE decision-making model

The PROMETHEE method is applied to find the optimal AQMN characterization among all Pareto optimal solutions by ranking the multi-objective optimization model results. Specifically, every solution is rated based on entering, leaving, and net flows. Table 2 presents the top ten resolutions for CO, NO₂, and ozone concentrations based on equal weights for the six objectives, from which the optimal AQMN with the most desirable solutions is determined. In other words, the selected AQMN allows the amount of information transferred between stations, related uncertainty of TE, the number of stations to be minimized, and the coverage to be maximized simultaneously. The location of these stations is depicted on the map shown in Fig. 8. As can be seen, the net flow and ranking of the solutions are directly related. Furthermore, this table illustrates that solution Nos. 30, 35, and 38 provide the best AQMN characterization, including 4, 4, and 5 stations for CO, NO₂, and ozone,

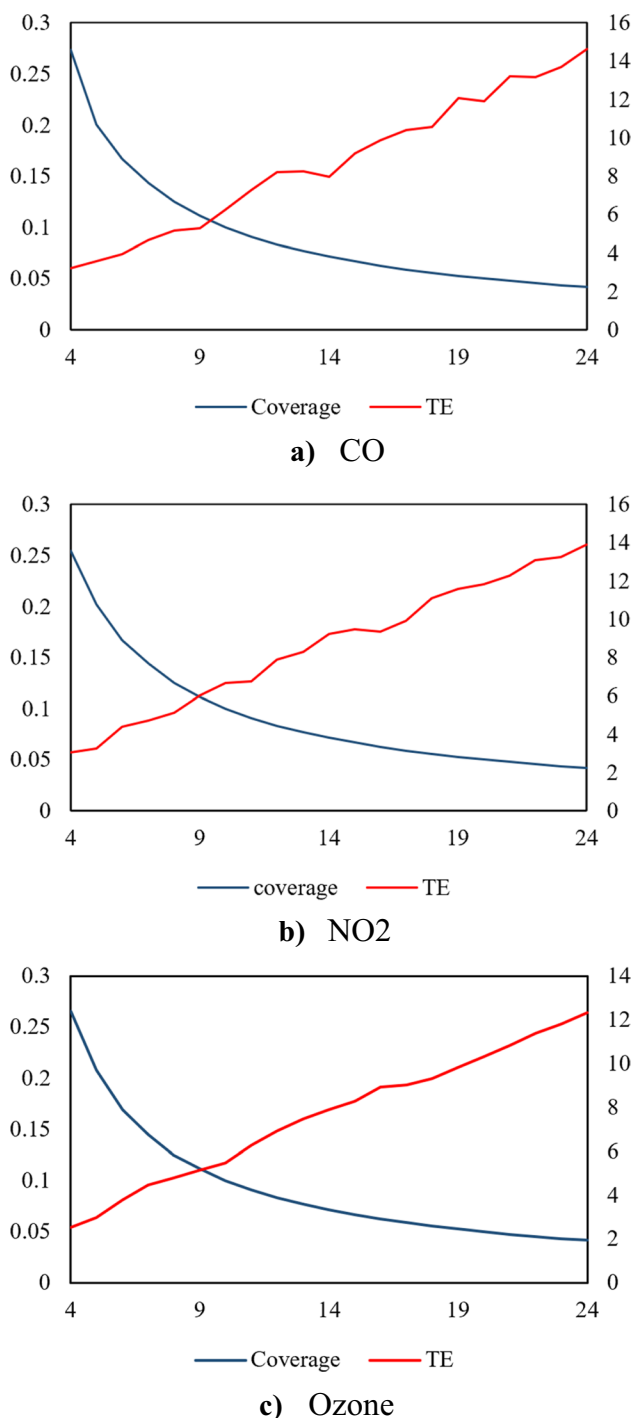


Fig. 7 The relationship between the number of monitoring stations in the monitoring networks, coverage, and TE

respectively. Detailed results for the Pareto optimal solutions in 6 scenarios are provided in Table S5, Section S10, Supplementary Material.

The optimum number of monitoring stations directly affects expenses associated with operation, maintenance, and assessment performance (Zhao et al. 2022).

According to Table 3, a simple comparison between the TE of previous AQMNs and proposed ones for air pollutants shows that observing the joint information decreases considerably using the presented methodology while the number of monitoring stations declines. In addition, the uncertainties of the observing data, including average TE variations, radii of TE variations, and fuzzy degree of membership, significantly decrease. Last but not least, this framework determines monitoring stations with an acceptable level of coverage and shared information.

The objective values of all three pollutants for different numbers of stations are compared in Fig. 9. Based on this figure, the amount of redundant information (TE) for monitoring networks with 12, 14, and 9 stations is relatively close. Therefore, having more than 9 stations does not significantly help to reduce the mutual information between monitoring stations. By comparing the current monitoring networks with 15 stations to the proposed networks with 4, 4, and 5 stations for monitoring CO, NO₂, and ozone, respectively, the results indicate a significant reduction in the number of stations and the amount of coverage by 3.75 times for both CO and NO₂ and 3 times for ozone. Despite this reduction, the proposed networks show a decrease in the amount of redundant information for CO, NO₂, and ozone concentrations by 8.25, 5.86, and 4.75 times, respectively. This outstanding reduction in the amount of shared information compared to the number of monitoring stations and coverage means a relatively desirable level of information can be achieved at a much lower cost. However, if decision-makers have a larger budget, they can have more coverage by installing more monitors.

Sensitivity analysis

A series of sensitivity analyses are performed based on the model's results in different scenarios to show that the PROMETHEE method is sensitive to varying objectives' weights. Figure 10 demonstrates the weights of the objectives in 6 scenarios. In addition, the results of sensitivity analyses are given in Fig. 11, including the best solution for each scenario obtained from the PROMETHEE method. According to the solutions obtained from different scenarios, it is clear that TE and its related uncertainties are determining factors exerting a pronounced effect on the number of stations in scenarios 4 and 5. On the other hand, the coverage raises the number of stations in scenarios 1, 2, and 3. However, in some cases, such as scenarios 1 and 2, TE and its associated uncertainties impact the number of stations. However, in scenario 3, where coverage has a significantly higher weight, the number of stations is higher than the others.

Table 2 The ranking of the best ten solutions based on the PROMETHEE method

Type of pollutant	Ranking of solution	No. of solution	Leaving flow ($\varphi +$)	Entering flow ($\varphi -$)	Net flow	No. of stations
CO	1	30	0.19865	0.055185	0.143465	4
	2	45	0.178086	0.041907	0.136179	5
	3	125	0.175781	0.042141	0.13364	5
	4	39	0.18931	0.055736	0.133574	4
	5	135	0.16795	0.035377	0.132573	6
	6	117	0.186651	0.056012	0.130639	4
	7	83	0.171772	0.042976	0.128796	5
	8	16	0.212227	0.083516	0.12871	3
	9	54	0.212227	0.083516	0.12871	3
	10	116	0.184845	0.056321	0.128523	4
NO ₂	1	35	0.193952	0.056214	0.137738	4
	2	58	0.193952	0.056214	0.137738	4
	3	49	0.176573	0.042228	0.134346	5
	4	41	0.189989	0.056326	0.133663	4
	5	52	0.175125	0.044921	0.130204	5
	6	63	0.172937	0.042772	0.130164	5
	7	121	0.171004	0.043097	0.127907	5
	8	1	0.170441	0.043208	0.127233	5
	9	10	0.212512	0.085641	0.126871	3
	10	76	0.161974	0.036225	0.125749	6
Ozone	1	38	0.198874	0.068133	0.13074	5
	2	145	0.197865	0.068188	0.129677	5
	3	29	0.223855	0.095712	0.128143	4
	4	49	0.223855	0.095712	0.128143	4
	5	129	0.171603	0.045155	0.126447	7
	6	89	0.179288	0.053634	0.125653	6
	7	27	0.217993	0.095781	0.122212	4
	8	73	0.217993	0.095781	0.122212	4
	9	35	0.191004	0.068816	0.122188	5
	10	70	0.176177	0.054092	0.122085	6

Fig. 8 The locations of the superior optimal AQMN in the study area

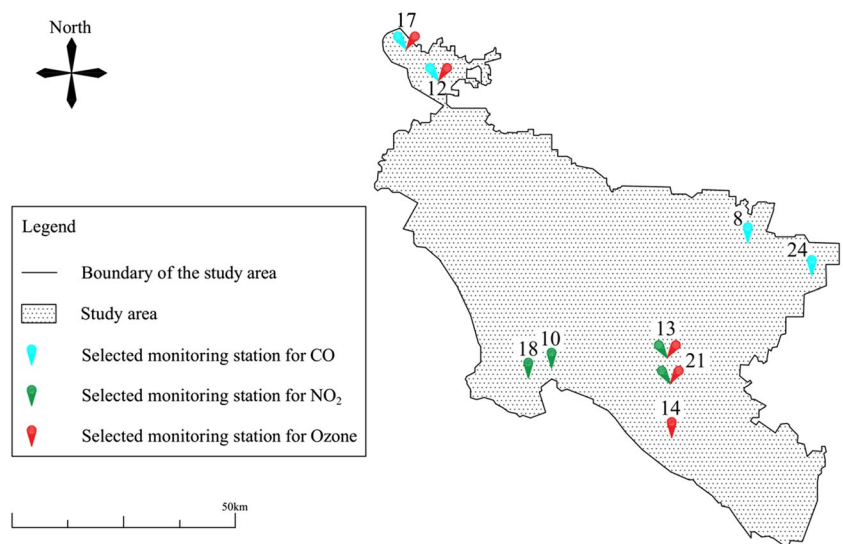


Table 3 The comparison of the optimization model objectives for current monitoring stations and the proposed ones

	No. of stations ¹	Coverage ²	TE	TE variations		Fuzzy degree of membership
				Radii	Averages	
CO						
Current stations	0.57911	0.06667	12.0332	1.49258	6.52372	0.47916
Proposed stations	0.06284	0.25	1.45807	0.3957	1.7133	0.16667
NO ₂						
Current stations	0.57934	0.06667	11.0758	1.55117	7.99504	0.50766
Proposed stations	0.06521	0.25	1.88818	0.41897	2.2251	0.16667
Ozone						
Current stations	0.56862	0.06667	9.95877	1.40749	9.1845	0.48228
Proposed stations	0.08931	0.2	2.09486	0.47023	3.10996	0.19634

¹According to Eq. (6), this column is the normalized value of the number of stations

²The objective value of the coverage is the value obtained from Eq. (1) to the negative power of 1 because the optimization algorithm aims to maximize it

Conclusions

A novel multi-objective optimization model using NSGA-II is suggested in this study to obtain the most suitable optimum air quality monitoring network (AQMN) characterizations. The presented framework adopts transinformation entropy (TE) and fuzzy nonlinear interval number programming (NINP) to evaluate the multi-objective model's mutual information and uncertainties among AQMN stations. The input data are obtained from 15 existing stations and nine potential stations generated by the Bayesian maximum entropy (BME) method for three contaminants, namely nitrogen dioxide (NO₂), carbon monoxide (CO), and ozone, allowing for the selection of the optimal AQMN with more options. The Pareto optimal results of this NSGA-II optimization model are prepared considering six optimization objectives: AQMN coverage, TE, average of TE variations, radii of TE variations, fuzzy degree of membership, and the number of stations. These Pareto optimal solutions are ranked using the PROMETHEE method, giving the same weight to all objectives to find the most appropriate AQMN characterization. Finally, to successfully show the efficiency and advantages of the methodology mentioned above, it is applied to Los Angeles, Long Beach, and Anaheim in Southern California,

USA. The results illustrate that the AQMNs determined by the PROMETHEE method consist of 4, 4, and 5 stations for CO, NO₂, and ozone concentrations from all 24 existing and potential stations, respectively. This reveals that the optimal AQMN exhibits good information even though the stations' number is reduced. This unparalleled significant achievement is gained since the BME approach generates the potential AQMN stations, and the shared information between the stations is considered using the TE and fuzzy NINP techniques within the innovative multi-objective optimization model. It is worth noticing that this methodology can be implemented in other areas with valid data within a fair distribution in those regions to acquire a good estimation using BME. More importantly, there should be no significant gaps in the time series data. In addition, other methods such as TE, NINP, and the fuzzy degree of membership are numerical methods that can be used in other case studies. Future studies can develop a framework using the non-dominated sorting genetic algorithm III (NSGA-III) to strengthen the optimization model and the value of information (VOI) theory to gain the most information. Also, the results of the future studies' framework can be compared with those of the methodology of this study.

Fig. 9 The comparison of monitoring networks with different numbers of stations

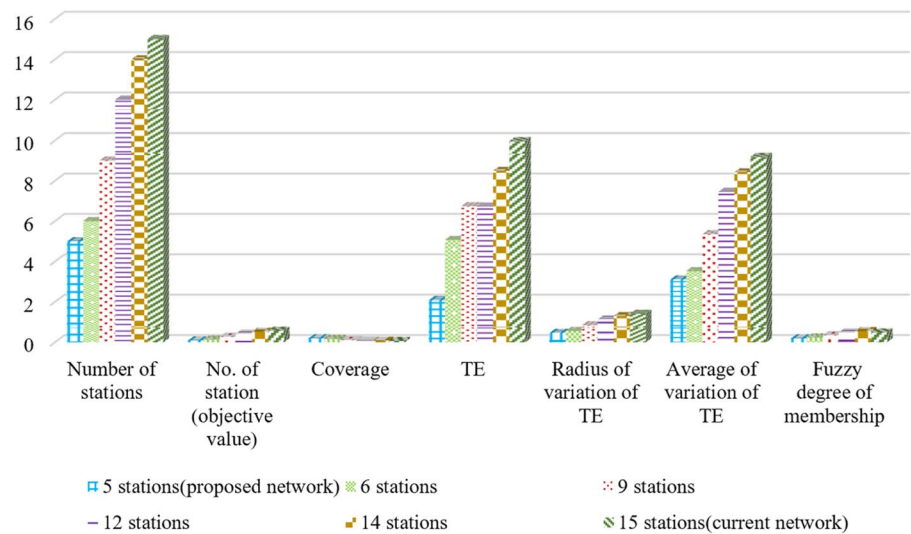
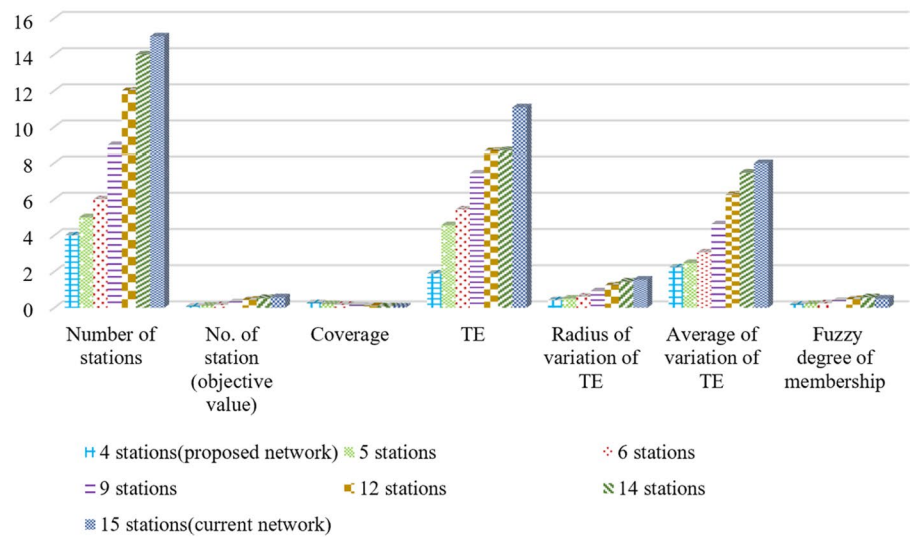
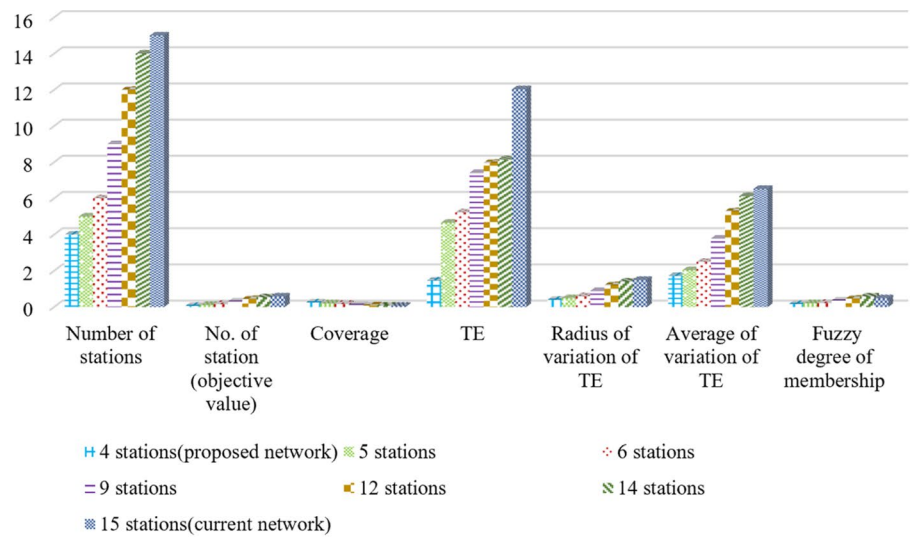


Fig. 10 The objectives weights in each scenario

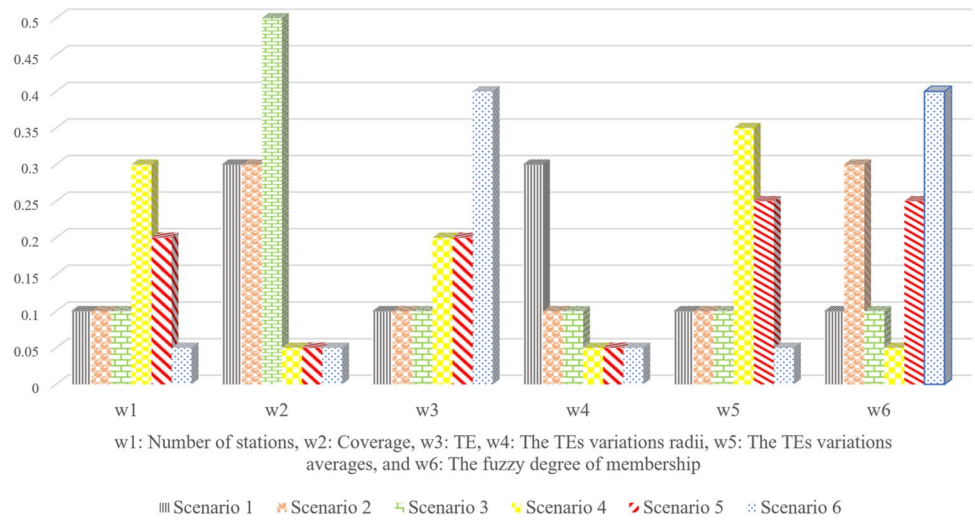
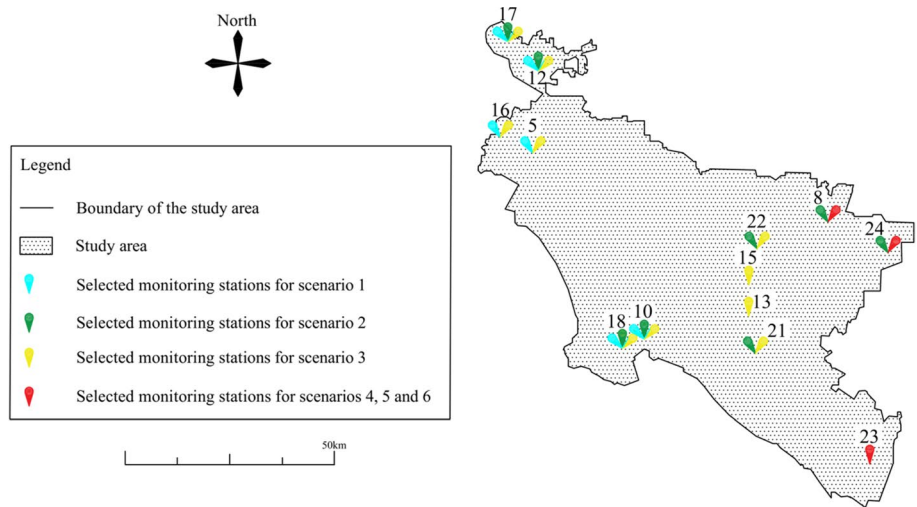
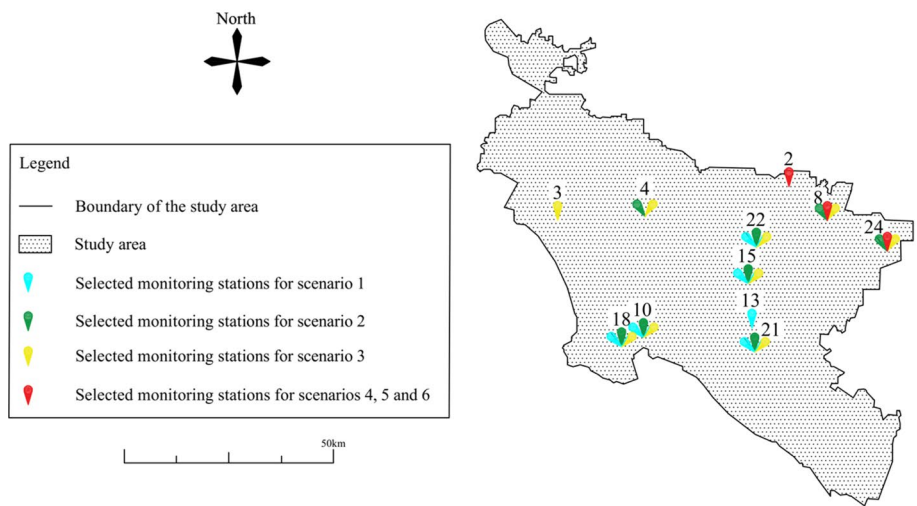


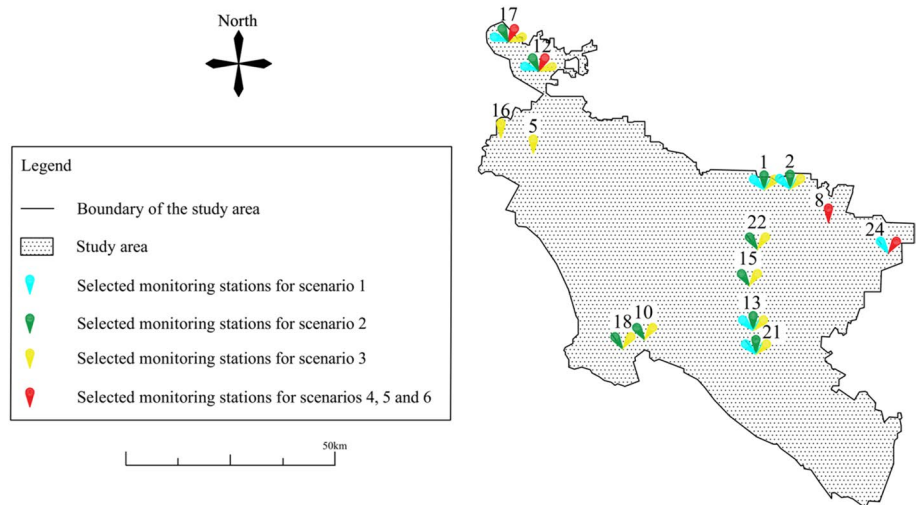
Fig. 11 Results of sensitivity analysis for different weights



a) CO



b) NO₂



c) Ozone

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11356-023-28270-w>.

Author contribution A. H.: conceptualization, methodology, coding, software, investigation, resources, data curation, writing of original draft, visualization, and review and editing; M. R. N.: supervision, conceptualization, methodology, coding, software, investigation, data curation, validation, and writing which included review and editing; B. N.: writing which included conceptualization and professional writing, as well as review and editing; G. A., M. A., and M. T.: review and editing; A. H. G.: validation and review and editing.

Data availability The corresponding author will provide data supporting the findings of this study upon reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

- Boubrima A, Bechkit W, Rivano H (2019) On the deployment of wireless sensor networks for air quality mapping: optimization models and algorithms. *IEEE/ACM Trans Netw* 27(4):1629–1642. <https://doi.org/10.1109/TNET.2019.2923737>
- Brans JP, Vincke P (1985) Note—A preference ranking organisation method: (the PROMETHEE method for multiple criteria decision-making). *Manage Sci* 31(6):647–656. <https://doi.org/10.1287/mnsc.31.6.647>
- Chen CH, Liu WL, Chen CH (2006) Development of a multiple objective planning theory and system for sustainable air quality monitoring networks. *Sci Total Environ* 354(1):1–19. <https://doi.org/10.1016/j.scitotenv.2005.08.018>
- Chung SY, Venkatraman S, Elzain HE, Selvam S, Prasanna MV (2019) Supplement of missing data in groundwater-level variations of peak type using geostatistical methods. *GIS and Geostatistical Techniques for Groundwater Science*. Elsevier 33–41. <https://doi.org/10.1016/B978-0-12-815413-7.00004-3>
- Dong J, Wang B, Wang X, Cao C, Chen S, Du W (2022) Optimization of sensor deployment sequences for hazardous gas leakage monitoring and source term estimation. *Chin J Chem Eng*. <https://doi.org/10.1016/j.cjche.2022.06.033>
- Elkamel A, Fatehifar E, Taheri M, Al-Rashidi MS, Lohi A (2008) A heuristic optimization approach for air quality monitoring network design with the simultaneous consideration of multiple pollutants. *J Environ Manag* 88(3):507–516. <https://doi.org/10.1016/j.jenvm.2007.03.029>
- 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>
- Hadj Sassi MS, Fourati LC (2022) Comprehensive survey on air quality monitoring systems based on emerging computing and communication technologies. *Comput Netw* 209:108904. <https://doi.org/10.1016/j.comnet.2022.108904>
- 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>
- Jiang C, Han X, Liu GR, Liu G (2008) A nonlinear interval number programming method for uncertain optimization problems. *European Journal of Operational Research* 188(1):1–13. <https://doi.org/10.1016/j.ejor.2007.03.031>
- Kansal A, Khare M, Sharma CS (2011) Air quality modelling study to analyse the impact of the World Bank emission guidelines for thermal power plants in Delhi. *Atmos Pollut Res* 2(1):99–105. <https://doi.org/10.5094/APR.2011.012>
- Kao JJ, Hsieh MR (2006) Utilizing multi-objective analysis to determine an air quality monitoring network in an industrial district. *Atmos Environ* 40(6):1092–1103. <https://doi.org/10.1016/j.atmosenv.2005.11.003>
- Li M, Wang WL, Wang ZY, Xue Y (2018) Prediction of PM_{2.5} concentration based on the similarity in air quality monitoring network. *Build Environ* 137:11–17. <https://doi.org/10.1016/j.buildenv.2018.03.058>
- Li J, Zhang H, Luo Y, Deng X, Grieneisen ML, Yang F, Di B, Zhan Y (2019) Stepwise genetic algorithm for adaptive management: application to air quality monitoring network optimization. *Atmos Environ* 215:116894. <https://doi.org/10.1016/j.atmosenv.2019.116894>
- Liu MK, Avrin J, Pollack RI, Behar JV, McElroy JL (1986) Methodology for designing air quality monitoring networks: I. Theoretical aspects. *Environ Monit Assess* 6(1):1–11. <https://doi.org/10.1007/BF00394284>
- Lozano A, Usero J, Vanderlinden E, Ruez J, Contreras J, Navarrete B (2009) Air quality monitoring network design to control nitrogen dioxide and ozone, applied in Malaga, Spain. *Microchem J* 93(2):164–172. <https://doi.org/10.1016/j.microc.2009.06.005>
- Lu WZ, He HD, Dong LY (2011) Performance assessment of air quality monitoring networks using principal component analysis and cluster analysis. *Build Environ* 46(3):577–583. <https://doi.org/10.1016/j.buildenv.2010.09.004>
- Modak PM, Lohani BN (1985) Optimization of ambient air quality monitoring networks. *Environ Monit Assess* 5(1):1–19. <https://doi.org/10.1007/BF00396391>
- Mofarrah A, Husain T (2010) A holistic approach for optimal design of air quality monitoring network expansion in an urban area. *Atmos Environ* 44(3):432–440. <https://doi.org/10.1016/j.atmosenv.2009.07.045>
- Nematollahi B, Hashempour Bakhtiari P, Talebbeydokhti N, Rakhshandehroo GR, Nikoo MR, Gandomi AH (2022) A stochastic conflict resolution optimization model for flood management in detention basins: application of fuzzy graph model. *Water* 14(5):774. <https://doi.org/10.3390/w14050774>
- Opricovic S, Tzeng GH (2004) Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS. *Eur J Oper Res* 156(2):445–455. [https://doi.org/10.1016/S0377-2217\(03\)00020-1](https://doi.org/10.1016/S0377-2217(03)00020-1)
- Pourshahabi S, Talebbeydokhti N, Rakhshandehroo G, Nikoo MR (2018) Spatio-temporal multi-criteria optimization of reservoir water quality monitoring network using value of information and transinformation entropy. *Water Resour Manag* 32(10):3489–3504. <https://doi.org/10.1007/s11269-018-2003-8>
- Rathee M, Kumar S, Gandomi AH, Dilip K, Balusamy B, Patan R (2019) Ant colony optimization based quality of service aware energy balancing secure routing algorithm for wireless sensor networks. *IEEE Trans Eng Manag* 68(1):170–182. <https://doi.org/10.1109/TEM.2019.2953889>
- Sharma S, Chandra A (2008) Simulation of air quality using an ISCST3 dispersion model. *CLEAN-Soil, Air, Water* 36(1):118–124. <https://doi.org/10.1002/clen.200700036>

- Stolz T, Huertas ME, Mendoza A (2020) Assessment of air quality monitoring networks using an ensemble clustering method in the three major metropolitan areas of Mexico. *Atmos Pollut Res* 11(8):1271–1280. <https://doi.org/10.1016/j.apr.2020.05.005>
- Sun C, Li VO, Lam JC, Leslie I (2019) Optimal citizen-centric sensor placement for air quality monitoring: a case study of city of Cambridge, the United Kingdom. *IEEE Access* 7:47390–47400. <https://doi.org/10.1109/ACCESS.2019.2909111>
- Trujillo-Ventura A, Ellis JH (1991) Multi-objective air pollution monitoring network design. *Atmos Environ A Gen Top* 25(2):469–479. [https://doi.org/10.1016/0960-1686\(91\)90318-2](https://doi.org/10.1016/0960-1686(91)90318-2)
- Ülengin F, Topcu YI, Şahin ŞÖ (2001) An integrated decision aid system for Bosphorus water-crossing problem. *Eur J Oper Res* 134(1):179–192. [https://doi.org/10.1016/S0377-2217\(00\)00247-2](https://doi.org/10.1016/S0377-2217(00)00247-2)
- U.S. EPA (2022). <https://www.epa.gov/outdoor-air-quality-data/download-daily-data>
- Vergheze S, Nema AK (2022) Optimal design of air quality monitoring networks: a systematic review. *Stoch Env Res Risk Assess* 36(10):1–16. <https://doi.org/10.1007/s00477-022-02187-1>
- Vicente R, Wibrál M, Lindner M, Pipa G (2011) Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci* 30:45–67. <https://doi.org/10.1007/s10827-010-0262-3>
- Wang F, Liu X, Bergquist R, Lv X, Liu Y, Gao F, Li C, Zhang Z (2021) Bayesian maximum entropy-based prediction of the spatiotemporal risk of schistosomiasis in Anhui Province, China. *BMC Infect Dis* 21(1):1–12. <https://doi.org/10.1186/s12879-021-06854-6>
- Yazdandoost F, Razavi H, Izadi A (2022) Optimization of agricultural patterns based on virtual water considerations through integrated water resources management modeling. *Int J River Basin Manag* 20(2):255–263. <https://doi.org/10.1080/15715124.2021.1879093>
- Zhao L, Zhou Y, Qian Y, Yang P, Zhou L (2022) A novel assessment framework for improving air quality monitoring network layout. *J Air Waste Manag Assoc* 72(4):346–360. <https://doi.org/10.1080/10962247.2022.2027295>
- Zoroufchi Benis K, Fatehifar E, Ahmadi J, Rouhi A (2015) Optimal design of air quality monitoring network and its application in an oil refinery plant: an approach to keep health status of workers. *Health Promot Perspect* 5(4):269. <https://doi.org/10.15171/hpp.2015.032>
- Zoroufchi Benis K, Fatehifar E, Shafiei S, Keivani Nahr F, Purfarhadi Y (2016) Design of a sensitive air quality monitoring network using an integrated optimization approach. *Stoch Env Res Risk Assess* 30(3):779–793. <https://doi.org/10.1007/s00477-015-1189-z>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.