

Optimization of Air Quality Monitoring Network Using GIS Based Interpolation Techniques

Mohammed Mujtaba Shareef^{1*}, Tahir Husain¹, Badr Alharbi²

¹Faculty of Engineering and Applied Science, Memorial University, St. John's, Canada

²National Center for Environmental Technology, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia

Email: mujtaba.shareef@gmail.com

Received 5 February 2016; accepted 20 May 2016; published 23 May 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper proposes a simple method of optimizing Air Quality Monitoring Network (AQMN) using Geographical Information System (GIS), interpolation techniques and historical data. Existing air quality stations are systematically eliminated and the missing data are filled in using the most appropriate interpolation technique. The interpolated data are then compared with the observed data. Pre-defined performance measures root mean square error (RMSE), mean absolute percentage error (MAPE) and correlation coefficient (r) were used to check the accuracy of the interpolated data. An algorithm was developed in GIS environment and the process was simulated for several sets of measurements conducted in different locations in Riyadh, Saudi Arabia. This methodology proves to be useful to the decision makers to find optimal numbers of stations that are needed without compromising the coverage of the concentrations across the study area.

Keywords

Air Quality Monitoring Network, GIS, Interpolation, Kriging, IDW, Riyadh

1. Introduction

It is well known that the air pollution causes adverse effects on human health in addition to the impact on environment. Due to rapid urbanization and industrialization, air pollution assumes high significance particularly in large cities. Continuous monitoring of the air pollution with a well-designed air quality monitoring network

*Corresponding author.

How to cite this paper: Shareef, M.M., Husain, T. and Alharbi, B. (2016) Optimization of Air Quality Monitoring Network Using GIS Based Interpolation Techniques. *Journal of Environmental Protection*, 7, 895-911.
<http://dx.doi.org/10.4236/jep.2016.76080>

(AQMN) is the first step in addressing this issue. Obtaining the continuously monitored data to ensure the safe levels of air quality is one of the primary objectives of AQMN, in addition to evaluating exposure hazards and implementing effective control strategies. Environmental protection agencies would be looking for an optimal design of AQMN meeting these objectives with an obvious focus on minimizing cost.

The methodology to design a new AQMN or evaluate an existing AQMN attracted the attention of several researchers. Maximum sensitivity of the collected data [1] [2] and maximum coverage factors such as intensity of emissions, source distance and meteorology [3] were one of the first techniques to design an AQMN. Statistical measure of information content [4] and Fisher's information measure [5] was used to determine the optimum number and location of monitors in a network. The design of an AQMN in the greater London area was conducted by Handscombe and Elson [6] based on the concept of a spatial correlation analysis. The same concept combining with potential of violation was used by Arbeloa *et al.* [7] to design an optimal air quality monitoring networks. Noll and Mitsutome [8] developed a method to design AQMN based on expected ambient pollutant dosage. This method ranks potential locations by calculating the ratio of a station's expected dosage over the study area's total dosage. Another methodology developed by Nakamori and Sawaragi [9] determines the representative areas of monitor stations in urban areas. A different perspective, based on the use of Shannon information, was initiated with the results of Caselton and Zidek [10], applied in a univariate setup by Sampson and Guttorp [11] and Guttorp *et al.* [12] and later extended to a multivariate context by Perez-Abreu and Rodriguez [13]. The concept of sphere of influence and figure of merit was applied by McElroy *et al.* [14] to calculate the minimum number of air quality monitoring sites. A simple methodology for siting ambient air monitors at the neighborhood scale was developed by Richard *et al.* [15] and applied as a case study in Hudson County.

Linear programming approach was also used by many researchers to site optimum AQMN. A multi-attribute utility function method was used for siting the air quality network by Kainuma *et al.* [16]. Trujillo-Ventura and Ellis [17] applied multiple objectives, including spatial coverage, violation detection, data validity, and a weighting method to determine the most suitable network for Tarragona, Spain. Chen *et al.* [18] developed a multiple objective optimization model for air quality monitoring networks in Taoyuan County, Taiwan. A multiple objective optimization model and procedure for sustainable air quality monitoring networks planning were developed. A holistic approach was adapted for optimal design of air quality monitoring network expansion in an urban area by Mofarrah and Husain [19]. In this approach multiple-criterion method with the spatial correlation technique was implemented to design an expanded air quality monitoring network of Riyadh city in Saudi Arabia. Tseng and Ni-Bin [20] proposed a Genetic Algorithm (GA) based compromise programming technique for assessing the relocation of strategy of urban air quality monitoring network with respect to the multi-objective and multi-pollutant design criteria. Silva and Quiroz [21] used Shannon information index to optimize atmospheric monitoring network by excluding the least informative stations with respect to different air pollutants. An optimal design of AQMN was done around a refinery using a mathematical model based on the multiple cell approach with simultaneous multiple pollutants by Elkamel *et al.* [22]. Lu *et al.* [23] used the principal component analysis (PCA) and cluster analysis to optimize the air quality monitoring network in Hong Kong. PCA and fuzzy c-means clustering was applied for assessment of air quality monitoring in Turkey by Dogruparmak *et al.* [24]. The authors showed that a number of monitoring stations can be decreased using the methodology. Ferradas *et al.* [25] developed a methodology based on self-organizing map (SOM) artificial neural networks for integrating data about multiple measured pollutants to group monitoring stations according to their similar air quality. The proposed method considered the subsequent geographical mapping of the clusters of stations observed with the SOM, which made it possible to detect geographically different areas that share similar air pollution problems.

The strides that the field Geographical Information System (GIS) and its components (such as interpolation methods) are making as an application in almost every field are incredible. GIS and spatial interpolation techniques were also used in AQMN. Bayraktar *et al.* [26] used a Kriging-based approach to locate sampling site for assessing the air quality. Long years of the smoke data measurements were used in the determination of the region for the representation by drawing contours through the Kriging method. Then, the selection of the sampling site in this region was done based on the EPA criteria. External drift Kriging of NO_x concentrations with dispersion model output was used by Kassteele *et al.* [27] to reduce the uncertainties in parameter estimation due to the reduction of air quality network. GIS ancillary variables were used to predict volatile organic compound and nitrogen dioxide levels at unmonitored locations [28]. The predictive equations were developed by regressing the passive monitor measurements at the 22 monitored schools on land-use variables derived from GIS. Univer-

sal Kriging interpolation method was found to perform better than other methods, when a comparison was made between various interpolation methods by developing EU-wide high resolution air pollution maps [29]. The best method to model concentrations was selected on the basis of predefined performance measures; correlation coefficient and RMSE. A generalized monotonic regression based on B-splines with an application to air pollution data was proposed by Leitenstorfer and Tutz [30] and spline method was used to estimate the varying health risks from air pollution across Scotland by Duncan [31].

The methods summarized above are very useful, well established and has been implemented widely; however it appears a simple GIS based methodology would further reduce the complexities of AQMN design. The basic advantage of using GIS is that it organizes geographic data in such a way that the decision making process becomes easy. In addition to this, it provides several advanced functionalities to manage statistical and spatial data, interpolate the data to create smooth surface, extract data from the interpolated surface, and create algorithms to automate the process. Furthermore, it creates the results that can be visualized in interactive maps which will further simplify the decision making process. Taking the cue on these advantages, this paper proposes a simple and innovative process to optimize AQMN by using GIS, interpolation methods and the historical data. The existing stations are systematically eliminated by creating several interpolated maps and comparing it with the observed values. The number of stations that can be eliminated is governed by the pre-defined performance measures criteria. In recent times an increasing trend of air pollution has been observed in Riyadh city of Saudi Arabia and there is an emphasis on frequent air pollution measurements [32], hence Riyadh is used as a case study to test the proposed methodology.

2. Methodology

2.1. Interpolation Methods

Interpolation predicts values for cells in a raster using a limited number of sample data points, which helps in predicting unknown values for any geographic station. Five interpolation methods were selected to estimate the concentrations of air pollution at the unknown stations. The selected methods were 1) Inverse Distance Weighted (IDW); 2) Spline (SPL); 3) Ordinary Kriging (OK); 4) Universal Kriging (UK) and 5) Natural Neighbor (NN). These methods have been widely used in estimating the air pollution concentrations.

The IDW uses a method of interpolation that estimates cell values by averaging the values of sample data points in the neighborhood of each processing cell [33]. The closer a point is to the center of the cell being estimated, the more influence or weight it has in the averaging process and has been used in several instances to interpolate the air pollutant concentrations [34]-[36].

The SPL uses an interpolation method that estimates values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points. This method was found superior in varying health risk from air pollution [31] and also was recommended by Leitenstorfer and Tutz [30].

Kriging is an advanced geo-statistical procedure that generates an estimated surface from a scattered set of points with z-values. While Kriging is a weighted combination of monitor values, this method also uses spatial auto correlation among data to determine the weights. Generally Kriging has two different forms *i.e.* ordinary and universal Kriging [37]. In the ordinary Kriging the mean value is assumed constant and determined during interpolation and universal Kriging assumes that data follows a known trend. Ordinary and universal Kriging have previously been used with success to model ozone [38] and particles [39] at the local scale, and to model broad scale variations in background air pollution [40]. Kriging has also been implemented in several other studies [29] [41] [42].

Natural Neighbor interpolation finds the closest subset of input samples to a query point and applies weights to them based on proportionate areas to interpolate a value. It is also known as Sibson or “area-stealing” interpolation.

2.2. Performance Measures

A statistical error is the amount by which an observation differs from its expected value. The statistical indices selected to measure performance are Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Nash-Sutcliffe equation (NSE) [43], and Accuracy Factor (ACFT) [44] given by Equations (1) to (4) respectively. Pearson correlation coefficient (r) was also measured to check the strength and direction of linear

relationship between the observed and interpolated values.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Intr}_i - \text{Obs}_i)^2} \quad (1)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left(\left| \frac{\text{Intr}_i - \text{Obs}_i}{\text{Intr}_i} \right| \right) \quad (2)$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (\text{Intr}_i - \text{Obs}_i)^2}{\sum_{i=1}^n (\text{Intr}_i - \text{Obs}_i^*)^2} \quad (3)$$

$$\text{ACFT} = 10^{\left(\frac{\log \left(\frac{\text{Intr}_i}{\text{Obs}_i} \right)}{n} \right)} \quad (4)$$

where Intr_i = Interpolated value; Obs_i = Observed value; n = number of observations.

RMSE is the frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. It basically represents the sample standard deviation of the differences between predicted and observed values. RMSE gives important information in predicting the magnitude of a pollutant concentration, a measure close to zero represents good predictions. The absolute mean percentage error denoted by MAPE is calculated by dividing sum of percentage error by number of observations, and a value equal or close to zero is considered ideal. Coefficient of correlation is a measure of linear dependence between two variables, and it was chosen to get an indication of the correspondence of timing and evolution between observed and interpolated concentration values. The coefficient of efficiency (NSE) indicates the normalized fit of the model, the value ranges from $-\infty$ to 1. It compares the mean square error generated by a particular model simulation to the variance of the output sequence; a value of 1 indicates a perfect fit [43]. Ross [44] proposed a simple multiplicative factor called accuracy factor (ACFT) representing the spread of the results of the modelled data. A value of one indicates that there is perfect agreement between all the predicted and the modelled values and values larger than one indicates the less accurate average estimate.

Several studies have used error statistics in comparing observed and predicted meteorology and air quality data. RMSE was used by Shad *et al.* [42] in comparing the observed data with the predicted air pollution using fuzzy genetic linear Kriging. Monteiro *et al.* [45] used it in bias correction techniques to improve air quality ensemble predictions and Son *et al.* [46] used in the study of individual exposure to air pollution and lung function in Korea. EU-wide maps were prepared based on the predictor variables and the modelled air pollution concentrations were selected on the predefined performance measures r^2 and RMSE. A value of $\text{RMSE} < 10$ and $r^2 > 0.5$ were considered a good performance measure [29]. Singh *et al.* [47] used RMSE, ACFT and NSE as performance measures of different linear and nonlinear modeling approaches for predicting urban air quality. Solar radiation and air pollution index were estimated based on linear, exponential and logarithmic models and the similar indices were used to check performance index in China [48]. In the current study, RMSE, MAPE and r^2 were primarily used as performance measure; $\text{RMSE} < 8$ $\text{MAPE} < 25\%$ and $r^2 > 0.5$ were considered a good measure.

2.3. Station Elimination Process

The main objective of this optimization process is to eliminate as many stations as possible and filling in the missing information through the interpolated values. The steps involved in this process are illustrated in **Figure 1** and elaborated as follows.

Step 1: Selection of stations

As a first step, a single station (P) or set of stations (P_1, P_2, \dots) are selected to be eliminated from the vector dataset used for creating the raster. A particular station or set of stations can be chosen or set of all possible stations tested in a loop.

Step 2: Storing the observed values

The observed concentrations at the selected station (P ()) or stations ($P_1 ()$, $P_2 ()$...) are stored as arrays. These values will later be compared with the interpolated values.

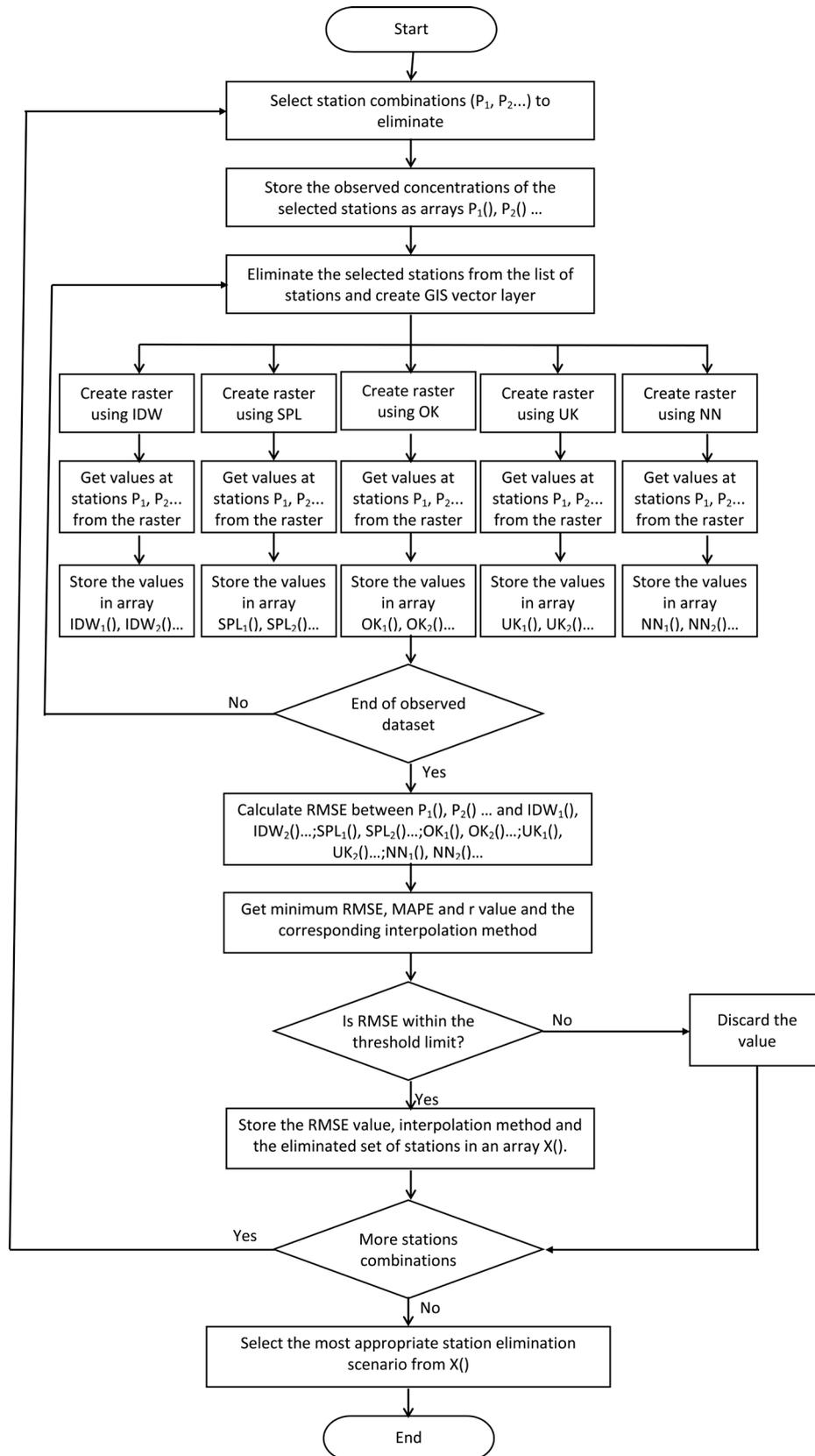


Figure 1. Station elimination process.

Step 3: Creating vector layer

A vector layer is a coordinate-based data model that represents geographic features, such as points. Each point is the station represented by the geographical coordinates. The layer has a column of “z” values used for interpolation. Several columns of “z” values *i.e.* the concentration of the pollutants are added to the vector for simulation. In this step the vector layer is created without the selected stations.

Step 4: Creating raster from the vector

Raster is defined as a spatial data model that defines space as an array of equally sized cells arranged in rows and columns, and composed of single or multiple bands. Each cell contains an attribute value and location coordinates. Rasters are created using the vector data and applying appropriate interpolation techniques. In this step five rasters are created using IDW, Spline, Ordinary Kriging, Universal Kriging and Natural neighbor methods.

Step 5: Extracting the interpolated value

This step calculates the predicted concentrations. The values at the eliminated stations are extracted from the created rasters. The values for each interpolation method (IDW, SPL, OK, UK, and NN) are stored in separate arrays (IDW₁() , IDW₂()...; SPL₁() , SPL₂()...; OK₁() , OK₂()...; UK₁() , UK₂()...; and NN₁() , NN₂()...). This process of interpolation and value extraction are repeated for all the observed datasets.

Step 6: Performance measure

The process of interpolation and value extraction generates arrays of the interpolated values. In this step, the performance measures are applied to observed and interpolated values. RMSE, Bias, and correlation coefficient are calculated for the selected stations and the five interpolation methods. The interpolation method that generates the minimum performance measure is then chosen and the others are discarded. These measures are compared with the pre-defined threshold limits and if the measures are within the limit, they are stored in the possible station combinations array C ().

Step 7: Repeat for another station combination

The process is repeated for another station combination until all the combinations are exhausted.

Step 8: Finding the best possible station combination

Best possible station combinations can be chosen from the list of possible station combination array C (). The decision maker can then choose from the list of possible station combinations which can be eliminated from the AQMN.

2.4. Study Location and Field Measurements

The proposed methodology is applied to the city of Riyadh, Saudi Arabia. The city is divided into sixteen cells that are identical in area and each cell is 12 km × 12 km. The measurements were carried out intermittently from September 2011 to September 2012. Most of the measurements have been conducted approximately in the center of each cell (**Figure 2**) with two equipped mobile air quality monitoring stations capable of monitoring meteorological variables, as well as CO, O₃, NO_x, CH₄, OC, EC and PM_{2.5}. However, the methodology suggested in this study is implemented for the following four criteria pollutants *i.e.* SO₂, NO_x, O₃ and CO. The type of sensors used with their respective method of monitoring for the pollutants utilized in this study are: NO, NO₂ and NO_x-Chemiluminescence; CO-Dual Beam NDIR; O₃-UV Photometer; and SO₂-UV Fluorescence.

As the measurements are staggered, in order to get a continuous dataset, 24 datasets were prepared from the available measurements by averaging the hourly measurements for the entire study period for all the 16 stations. These 24 datasets were used for simulation to create the raster with different interpolation techniques and compared with the observed values. ESRI's ArcGIS (ESRI, Redlands) exposes several functions to run the interpolations and extract the necessary data. These functions were customized to run the simulation process through Microsoft Visual Basic for application (VBA) environment in ArcGIS.

3. Results and Discussion

The pollutant concentration data were collected from 16 stations as shown in **Figure 2**. It is assumed that a minimum of 8 stations are needed to produce reliable interpolated maps, therefore up to 8 possible elimination scenarios were simulated. **Table 1** outlines the simulation details such as the number of stations to be eliminated, possible number of combinations, number of simulations and the approximate time of simulation (on a high end PC). As seen in the table, the station combinations and the computing power required increase with the increase in number of stations. The methodology proposed in this study is tested on 4 pollutants namely O₃, NO_x and

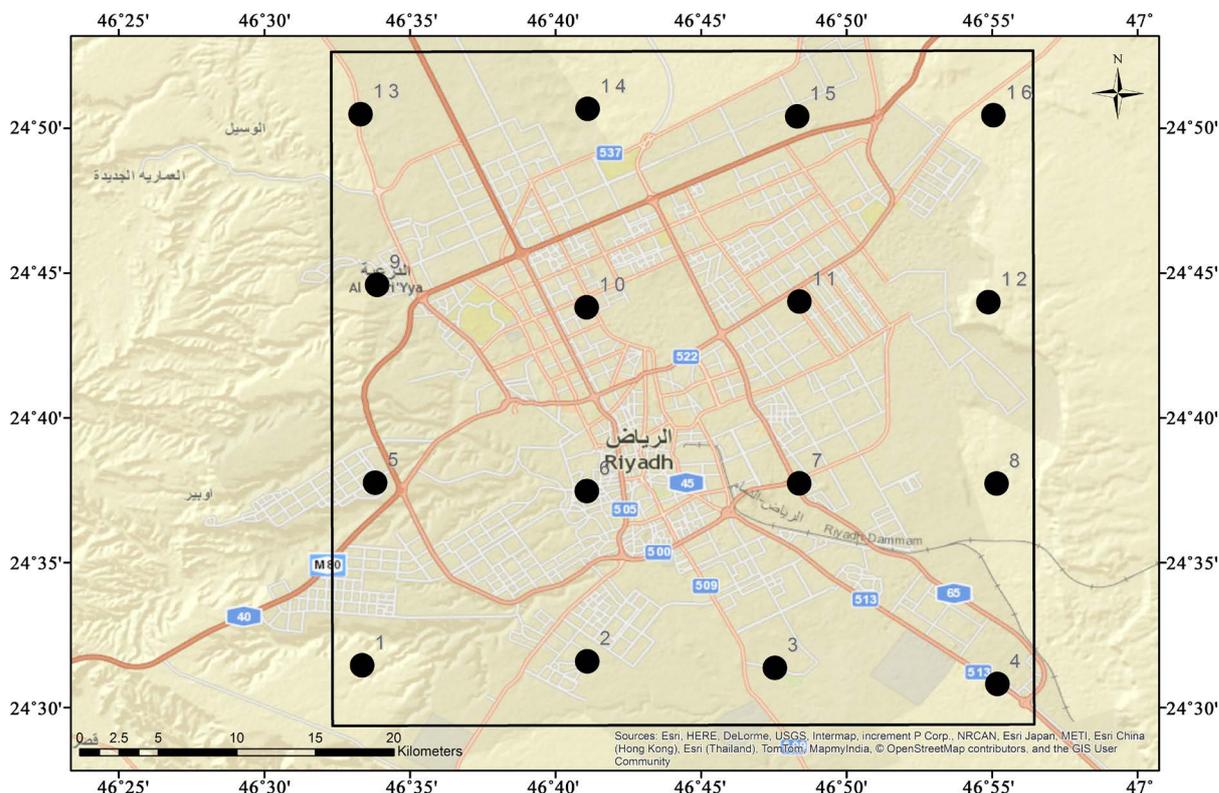


Figure 2. Distribution of AQMN in Riyadh.

Table 1. Simulation parameters.

No. of stations to be eliminated	Possible combinations	No. of simulations	Approximate time of simulation (hours)
1	16	1920	0.3
2	120	14,400	2.25
3	560	67,200	10.5
4	1820	218,400	34
5	4368	524,160	81
6	8008	960,960	150
7	11,440	1,372,800	214
8	12,870	1,544,400	241

SO₂ and CO, and the results of the simulations are discussed as follows. The interpolation was performed using five methods *i.e.* IDW, Spline, OK, UK and NN. The IDW and UK outperformed other methods particularly in terms of producing lower RMSE, MAPE values and higher value of r^2 .

3.1. O₃

Table 2 presents the results of the simulation for O₃. The values of RMSE, r^2 , MAPE, NSE and ACFT are shown along with the number of stations to be eliminated. The variations of RMSE, MAPE and r^2 with respect to stations are illustrated in Figures 3-5 respectively. The station 11 was the best single station that could be eliminated and the stations 4, 5, 6, 8, 11, 12, 15, and 16 were the best eight stations. RMSE and MAPE increased from 3.224 to 7.018 and 4.041 to 49.61 from one station elimination to eight respectively. Similarly NSE and ACFT decreased from 0.944 to 0.775 and 0.960 to 1.138, respectively. In order to limit MAPE within 25%, maximum number of station that could be eliminated were 6 (4, 8, 11, 12, 13, and 15), this also generated a satisfactory RMSE (6.113) and r^2 (0.831). Accuracy factor *i.e.* ACFT was 1.015 which is considered to be very

good performance. The remaining 10 stations that are required to produce satisfactory maps of ozone are shown in **Figure 6**; these stations could essentially be considered as hotspots of O₃ emissions.

The primary sources of ground level O₃ are automobiles, cement and power plants, construction activities and biogenic or natural sources. Small industries such as paint shops, dry cleaners and bakeries are also known to contribute O₃. These 10 stations fall within these areas. The stations 10 and 7 are located in dense residential area, and the stations 2 and 3 are in industrial area. The agricultural areas are found near station 9, and construction activities are reported near station 14. Station 1 and 16 are the city outskirts where small scale industries are located.

3.2. NO_x

The simulation results for NO_x concentrations are shown in **Table 3**. For the one station elimination scenario, it

Table 2. Performance measures for O₃ simulations.

No. of station(s) to be eliminated	Optimal station(s) combination that can be eliminated	Interpol. method	RMSE	r ²	MAPE	NSE	ACFT
1	11	IDW	3.224	0.953	4.041	0.944	0.960
2	11, 15	IDW	3.811	0.925	5.153	0.919	0.973
3	4, 11, 15	UK	3.964	0.913	7.991	0.901	1.003
4	4, 11, 12, 15	UK	4.588	0.911	15.699	0.884	1.099
5	4, 5, 11, 15, 16	IDW	5.495	0.862	22.805	0.844	1.095
6	4, 8, 11, 12, 13, 15	UK	6.113	0.831	25.322	0.821	1.015
7	4, 5, 8, 11, 12, 15, 16	IDW	6.488	0.832	36.469	0.809	1.081
8	4, 5, 6, 8, 11, 12, 15, 16	IDW	7.018	0.817	49.610	0.775	1.138

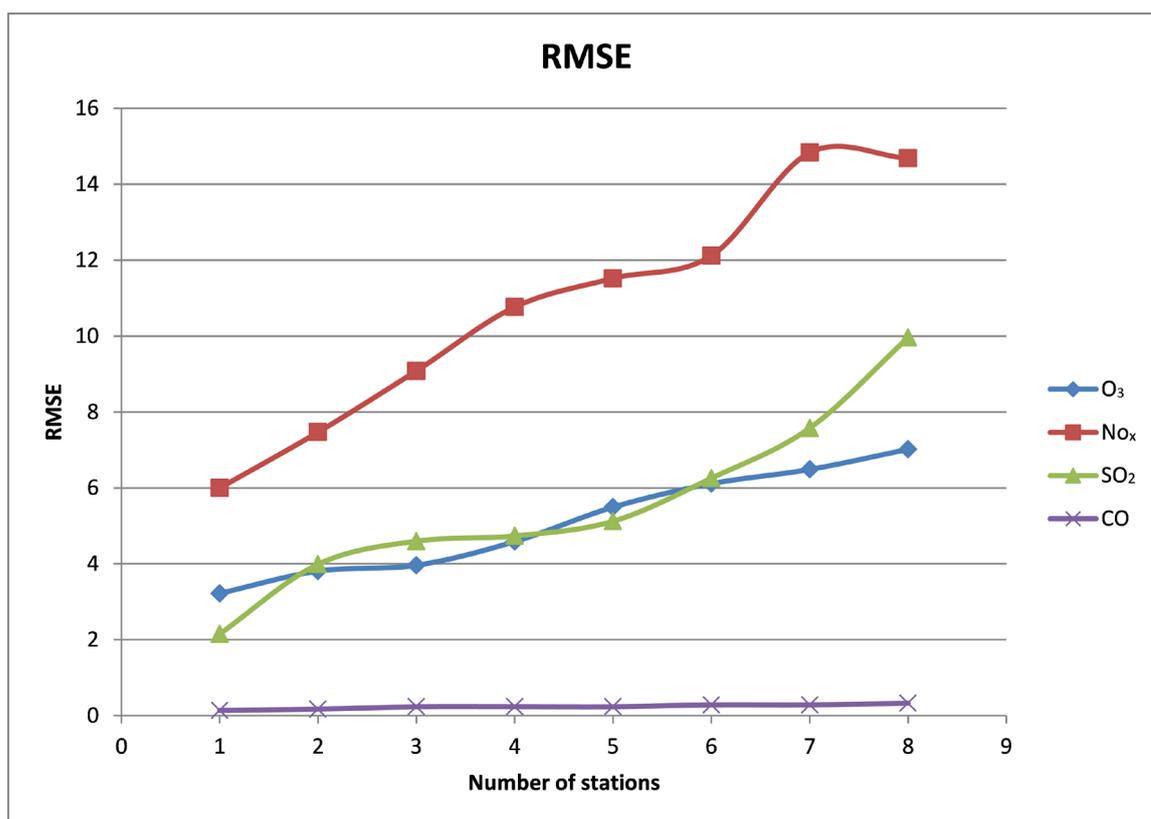


Figure 3. Variation of RMSE with respect to combination of stations that are eliminated.

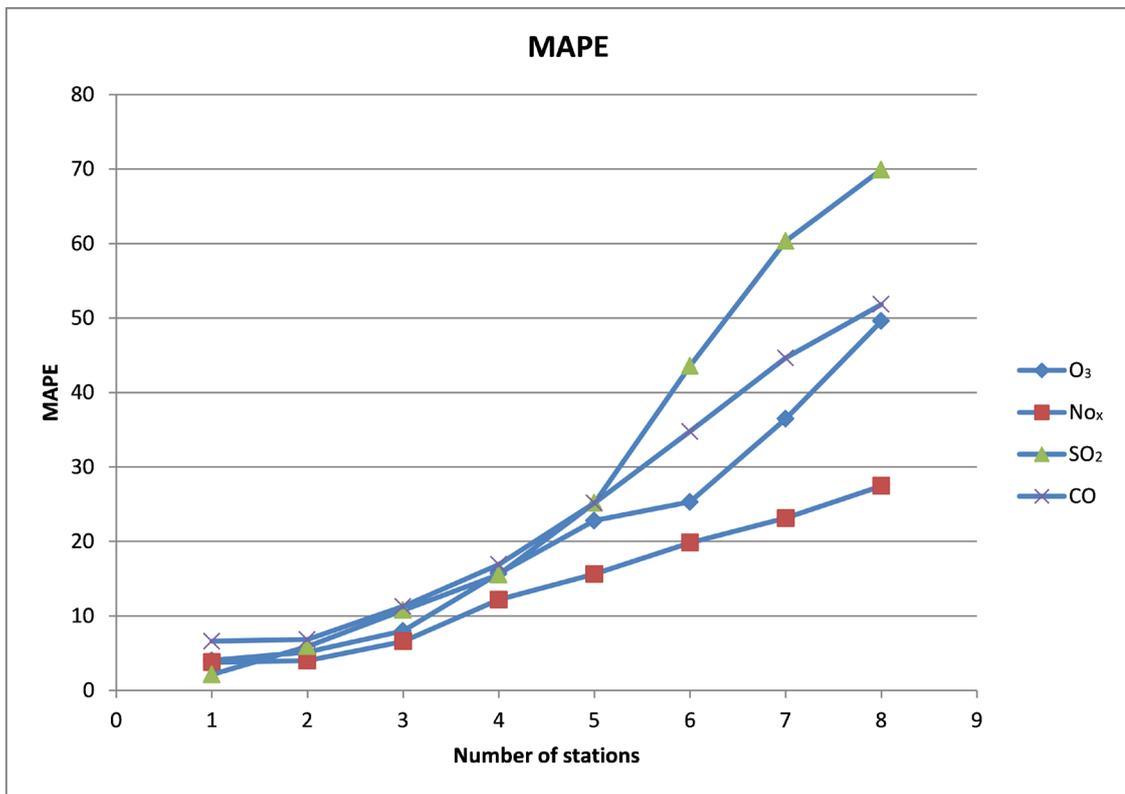


Figure 4. Variation of MAPE with respect to combination of stations that are eliminated.

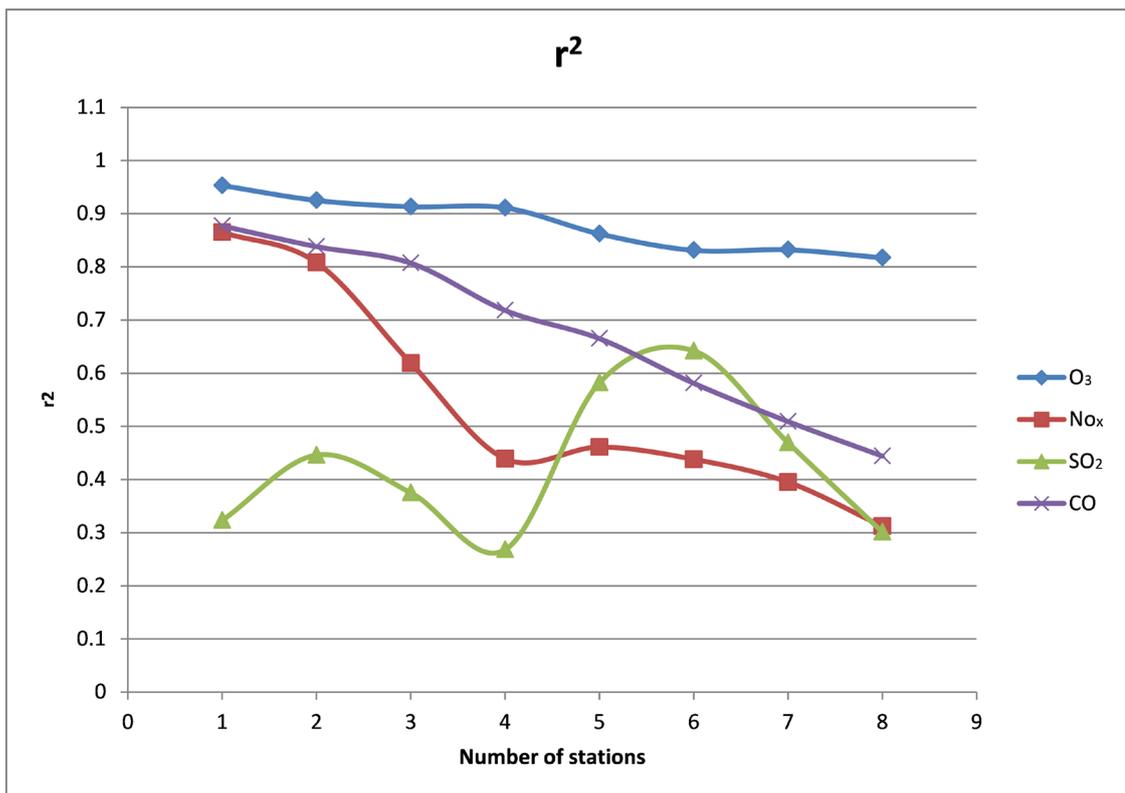


Figure 5. Variation of r² with respect to combination of stations that are eliminated.

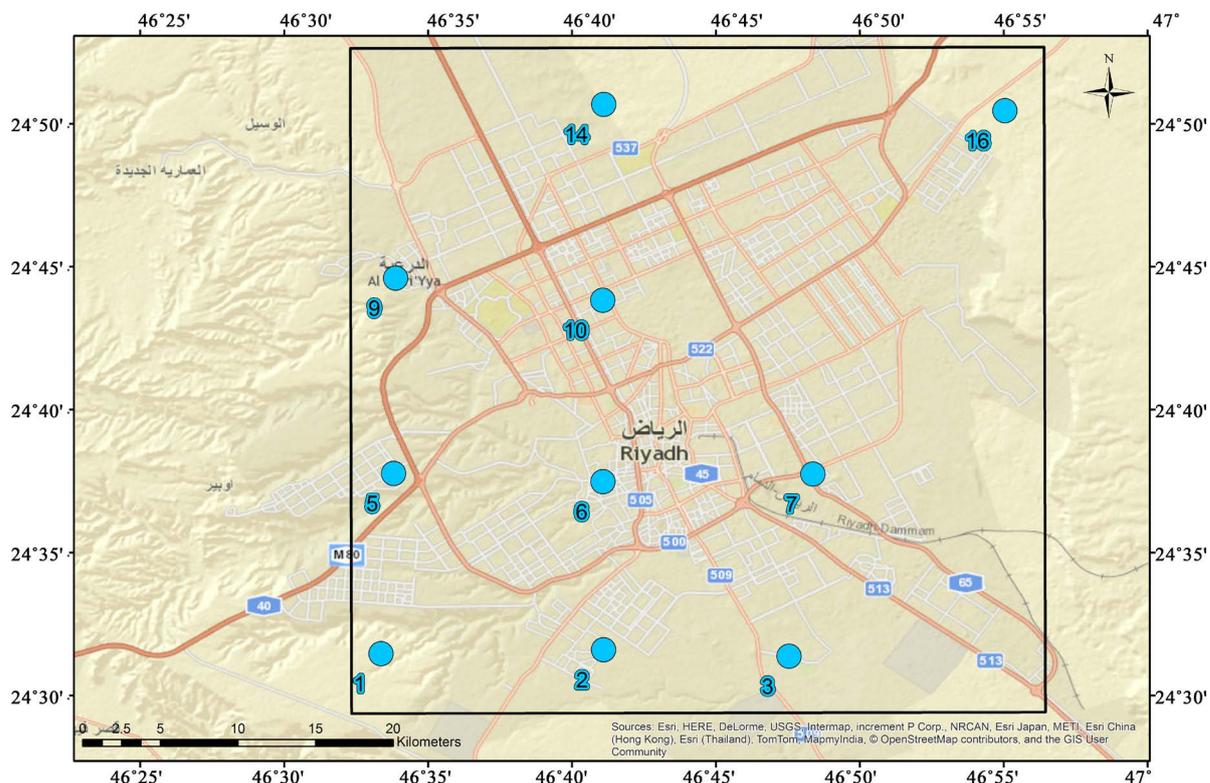


Figure 6. Stations required to generate O₃ concentration maps (RMSE < 7 and MAPE < 25).

Table 3. Performance measures for NO_x simulations.

No. of station(s) to be eliminated	Optimal station(s) combination that can be eliminated	Interpol. method	RMSE	r ²	MAPE	NSE	ACFT
1	11	IDW	5.999	0.865	3.792	0.848	1.020
2	3, 11	IDW	7.470	0.808	4.006	0.731	0.961
3	3, 5, 11	UK	9.084	0.619	6.591	0.599	0.984
4	3, 5, 7, 11	UK	10.766	0.439	12.184	0.430	1.025
5	3, 5, 7, 11, 14	IDW	11.521	0.461	15.621	0.459	1.022
6	3, 5, 7, 11, 14, 16	UK	12.119	0.438	19.839	0.438	1.019
7	3, 5, 6, 7, 11, 14, 16	IDW	14.834	0.395	23.148	0.341	0.966
8	1, 3, 5, 6, 7, 11, 14, 16	IDW	14.684	0.312	27.468	0.287	0.977

was found that station 11 was the best one with a RMSE = 5.999; MAPE = 3.792 and r² = 0.864. These values increased to 14.684, 27.468 and 0.312, respectively, for 8 stations eliminations (**Figures 3-5**). For MAPE scenario of ≤25, 7 a maximum of 7 stations could be eliminated (3, 5, 6, 7, 11, 14, and 16), however RMSE and r² values were not within the predefined limits. Taking the MAPE as priority, a minimum of 9 stations is needed to produce satisfactory NO_x concentration maps as shown in **Figure 7**. In addition to the industrial emissions, the primary source of NO_x is from the automobile emissions, these are reasonably evenly distributed over the study area, and hence less number of stations is required to produce the concentration maps. Stations 4 and 8 signify the presence of industries and the other stations are spread over the densely populated areas in the city where the automobiles move predominantly. The ACFT = 0.966 and NSE = 0.341 supports the elimination of the 7 stations.

3.3. SO₂

Station 1, which is located outskirts of the City, was the best one station to be eliminated with RMSE = 2.155,

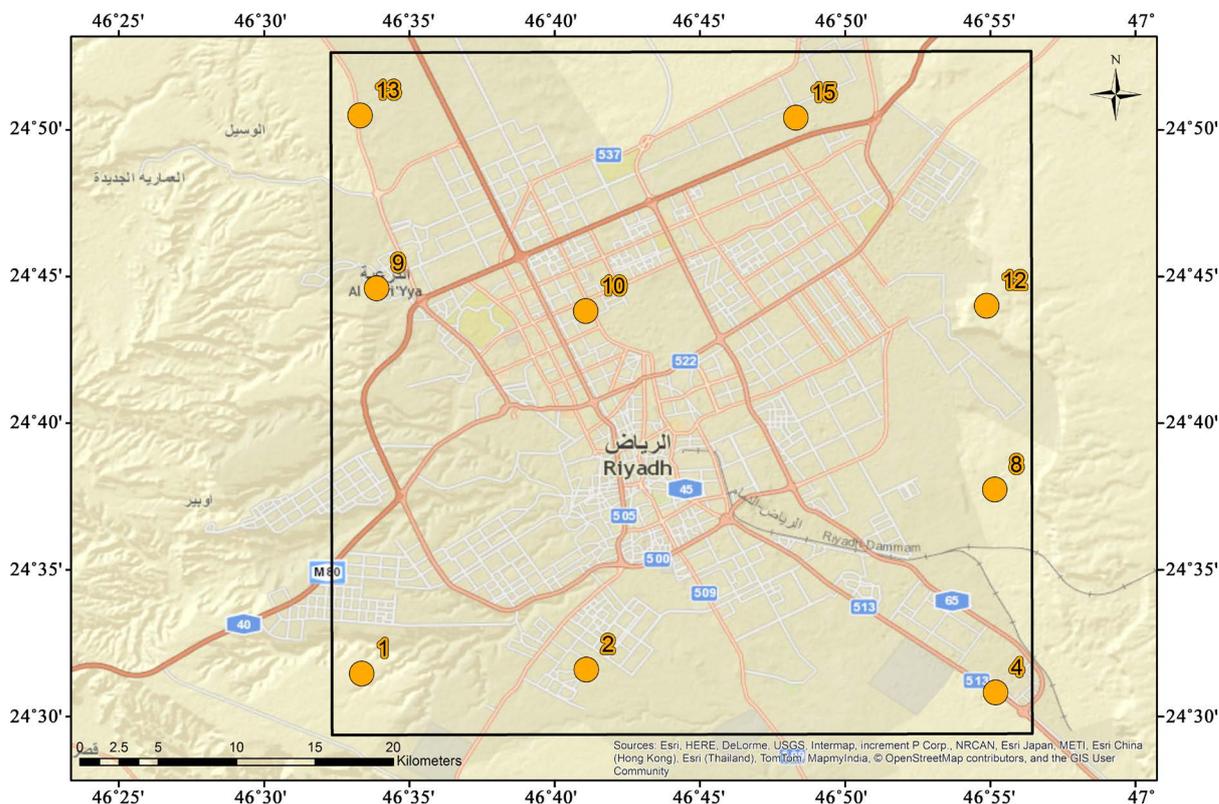


Figure 7. Stations required to generate NO_x concentration maps (RMSE < 15 and MAPE < 25).

MAPE = 2.155 and $r^2 = 0.944$. For eight stations, these parameters increased to 9.962, 69.931 and 0.301 respectively as shown in **Table 4**. To limit the MAPE value to <25, a maximum of 5 stations could be eliminated. **Figure 8** shows the rest of 11 stations required to produce the SO_2 concentration maps. The higher number of stations is needed for this, as the sources of SO_2 are distributed with a wide variation in the concentration. The stations in the south *i.e.* 2, 3, 7, 8 are predominantly located in the industrial area where power plants and refineries are located. Other stations are in the residential area, where exhaust from automobile could be the source of SO_2 . Station 9 is in the agriculture area and 14 and 16 are located in the automobile workshops and small scale industries zones.

3.4. CO

The results of simulations run on CO data are shown in **Table 5**. Since the concentrations of CO were very small, RMSE values were also very small ranging from 0.144 to 0.329 (**Figure 3**). MAPE values ranged from 6.608 to 51.586 for one station to eight station elimination as shown in **Figure 4**. The station 10 was the best single station that could be eliminated and the stations 2, 4, 7, 9, 10, 14, 15, and 16 were the best eight stations. However, in order for the MAPE to be within 25, a maximum of only 5 stations could be eliminated as shown in **Table 5**. The NSE, ACFT and r^2 values were satisfactory for this 5 station elimination. **Figure 9** shows the rest of 12 stations that are needed to make good CO concentration maps. As shown in the figure, most of the stations are located in industrial zones or the highly densely populated areas.

3.5. Overall

As observed from **Tables 2-5**, there is no single station which is common among the list of possible elimination stations. **Table 6** outlines the stations required against each pollutant to achieve MAPE level of 25. This indicates that the sources of the pollutants are highly varied with respect to the location. It will be up to the decision maker to prioritize the pollutant and select the stations. **Tables 7-10** illustrate the statistical parameters taking the priority stations for O_3 , NO_x , SO_2 and CO, respectively. For the case of O_3 as priority, MAPE for NO_x

Table 4. Performance measures for SO₂ simulations.

No. of station(s) to be eliminated	Optimal station(s) combination that can be eliminated	Interpol. method	RMSE	r ²	MAPE	NSE	ACFT
1	1	UK	2.155	0.323	2.155	-1.507	0.983
2	1, 4	IDW	3.991	0.446	5.925	0.421	1.029
3	1, 4, 13	UK	4.599	0.375	10.792	0.355	1.042
4	1, 4, 6, 13	IDW	4.737	0.268	15.562	0.167	1.063
5	1, 4, 6, 13, 15	IDW	5.124	0.582	25.187	0.428	1.140
6	1, 4, 6, 11, 13, 15	IDW	6.258	0.642	43.558	0.284	1.248
7	1, 4, 6, 7, 11, 13, 15	UK	7.579	0.469	60.337	0.038	1.278
8	1, 4, 6, 7, 9, 11, 13, 15	UK	9.962	0.301	69.931	0.100	1.279

Table 5. Performance measures for CO simulations.

No. of station(s) to be eliminated	Optimal station(s) combination that can be eliminated	Interpol. method	RMSE	r ²	MAPE	NSE	ACFT
1	10	IDW	0.144	0.877	6.608	0.841	0.914
2	10, 16	UK	0.174	0.838	6.842	0.775	0.901
3	2, 10, 16	UK	0.237	0.807	11.283	0.671	0.863
4	2, 10, 14, 16	UK	0.239	0.718	16.928	0.659	0.915
5	2, 9, 10, 14, 16	IDW	0.236	0.665	25.182	0.647	0.953
6	2, 4, 9, 10, 14, 16	IDW	0.287	0.581	34.782	0.476	0.870
7	2, 4, 7, 9, 10, 14, 16	IDW	0.288	0.509	44.646	0.467	0.951
8	2, 4, 7, 9, 10, 14, 15, 16	UK	0.329	0.444	51.856	0.374	0.926

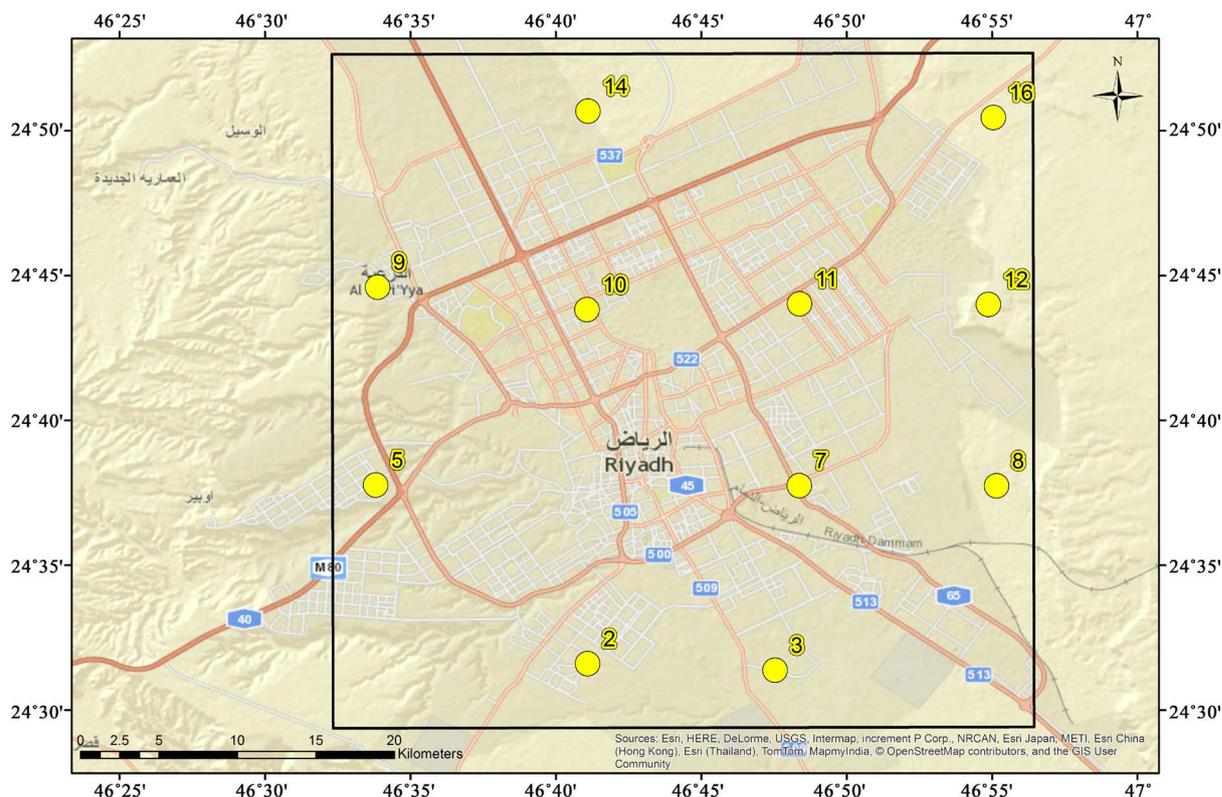


Figure 8. Stations required to generate SO₂ concentration maps (RMSE < 5 and MAPE < 25).

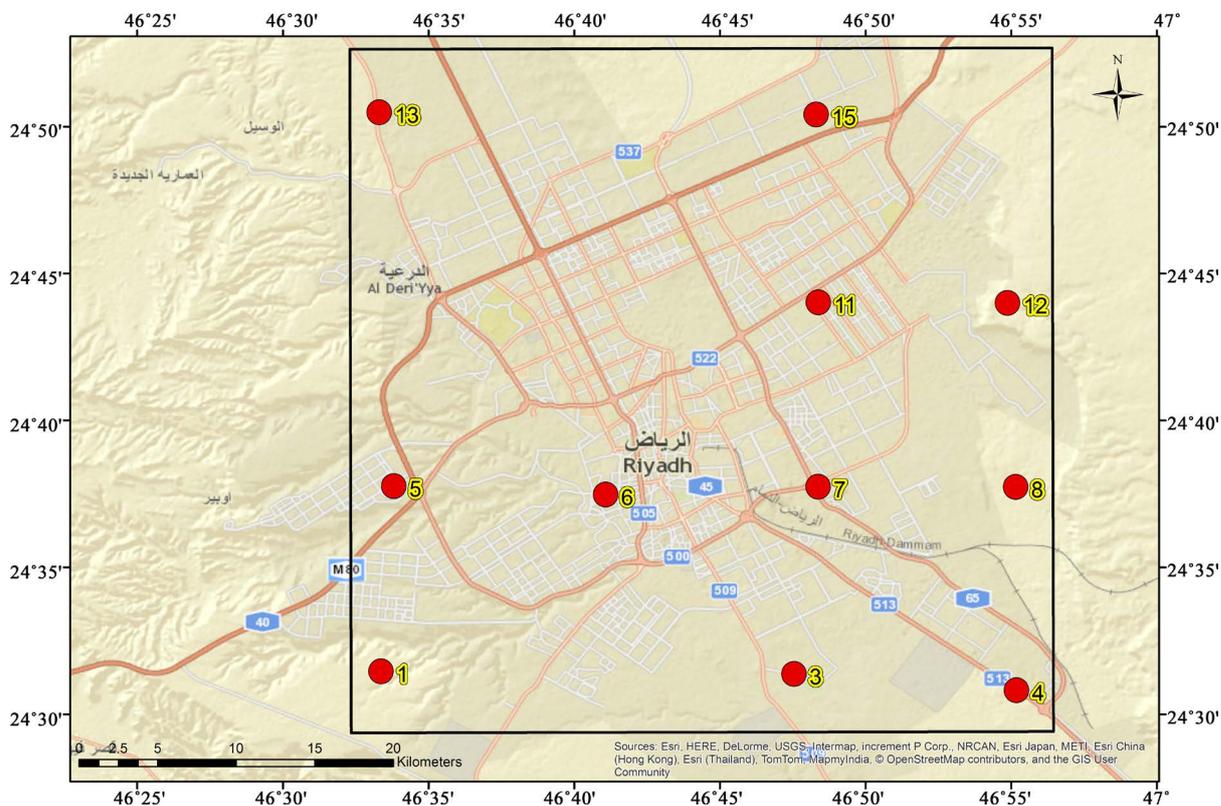


Figure 9. Stations required to generate CO concentration maps (RMSE < 0.24 and MAPE < 25).

Table 6. Pollutant and the stations needed to achieve MAPE < 25.

Pollutant	Stations needed to achieve MAPE < 25
O ₃	1, 2, 3, 5, 6, 7, 9, 10, 14, 16
NO _x	1, 2, 4, 8, 9, 10, 12, 13, 15
SO ₂	2, 3, 5, 7, 8, 9, 10, 11, 12, 14, 16
CO	1, 3, 4, 5, 6, 7, 8, 11, 12, 13, 15

Table 7. Parameters with priority as O₃.

Pollutant	RMSE	r ²	MAPE	NSE	ACFT
O ₃	6.114	0.831	25.322	0.821	1.016
NO _x	24.173	0.208	57.374	0.189	1.122
CO	0.490	0.234	82.670	0.221	1.112
SO ₂	17.579	0.056	240.894	-4.570	2.302

Table 8. Parameters with priority as NO_x.

Pollutant	RMSE	r ²	MAPE	NSE	ACFT
NO _x	14.834	0.396	23.148	0.341	0.966
CO	0.565	0.285	73.538	0.275	1.183
O ₃	10.736	0.815	85.331	0.432	1.438
SO ₂	37.826	0.109	198.364	-0.194	0.900

Table 9. Parameters with priority as SO₂.

Pollutant	RMSE	r ²	MAPE	NSE	ACFT
SO ₂	5.125	0.582	25.128	0.431	1.139
O ₃	11.834	0.550	28.723	0.496	0.929
NO _x	26.512	0.063	51.687	0.045	1.101
CO	0.740	0.182	80.081	0.141	1.038

Table 10. Parameters with priority as CO.

Pollutant	RMSE	r ²	MAPE	NSE	ACFT
CO	0.236	0.665	25.181	0.647	0.953
NO _x	17.480	0.287	31.760	0.283	1.047
O ₃	15.769	0.343	60.332	0.310	1.004
SO ₂	20.571	0.149	327.356	-1.631	3.067

and CO is about 57 and 82 while SO₂ has a very high MAPE value of over 240. Considering NO_x as priority, MAPE value for CO and O₃ were about 73 and 85. In this case also SO₂ has very high MAPE (over 198) as shown in **Table 8**. The MAPE value of O₃ and NO_x were less than 50 in the case of SO₂ priority stations, while CO was little over 80. Taking CO as priority produced a MAPE value of 31 and 60 for NO_x and O₃ while SO₂ produced exorbitantly high value of 327 (**Table 10**).

4. Conclusions

A simple method of optimizing the AQMN is proposed using GIS, interpolation techniques and historical data. Existing air quality stations are systematically eliminated and the missing data are filled in using the most appropriate interpolation technique. The interpolated data are then compared with the observed data. Pre-defined performance measures RMSE, MAPE and r² were used to check the accuracy of the interpolated data. NSE and ACFT supported the validity of the interpolated data. The process was simulated for several sets of observed data using an algorithm developed in GIS environment. In order to achieve a MAPE value of 25 or less, no combination of station could be eliminated for all the pollutants. The pollutants could be prioritized to achieve the most optimal scenario. The results of the prioritization showed that the most optimal scenario was for the SO₂ stations, which achieved MAPE for O₃, NO_x and CO about 28, 51 and 80, respectively.

This methodology proves to be useful to the decision makers to find optimal numbers of stations that are needed without compromising the coverage of the concentrations across the study area. Although it is a simple procedure, it does have few limitations. A continuous set of data is required to get the reliable simulation results, owing to the unavailability of such continuous dataset; the staggered dataset is averaged as hourly data for a day and simulated in present case study. Secondly, the process is computing intensive and hence requires large computing resources, though not very expensive these days. Lastly, more parameters can be included in the performance measures to get the most appropriate results.

Acknowledgements

We gratefully acknowledge the financial support of King Abdulaziz City for Science and Technology (KACST) under grant number 32-594.

References

- [1] Koda, M. and Seinfeld, J.H. (1978) Air Monitoring Siting by Objective. EPA-600/4-7-036. US Environmental Protection Agency, Las Vegas.
- [2] Caselton, W.F. and Husain, T. (1980) Hydrologic Networks: Information Transmission, *Journal of the Water Resources Planning and Management Division, ASCE*, **106**, 503-520.
- [3] Hougland, E.S., Oades, T.W. and Shank, K.E. (1980) Design of the SO₂ and Particulate Air Monitoring Network for the Oak Ridge National Laboratory Fossil Steam Plant. *Proceedings of the UCC-ND and GAT Waste Management Se-*

- minar, Friendship, 22-23 April 1980, 255-264.
- [4] Pickett, E.E. and Whiting, R.G. (1981) The Design of Cost-Effective Air Quality Monitoring Networks. *Environmental Monitoring and Assessment*, **1**, 59. <http://dx.doi.org/10.1007/BF00836876>
 - [5] Husain, T. and Khan, S.M. (1983) Air Monitoring Network Design Using Fisher's Information Measures—A Case Study. *Atmospheric Environment*, **17**, 2591-2598. [http://dx.doi.org/10.1016/0004-6981\(83\)90087-2](http://dx.doi.org/10.1016/0004-6981(83)90087-2)
 - [6] Handscombe, C.M. and Elson, D.M. (1982) Rationalisation of the National Survey of Air Pollution Monitoring Network of the United Kingdom Using Spatial Correlation Analysis: A Case Study of the Greater London Area. *Atmospheric Environment*, **4B**, 395. [http://dx.doi.org/10.1016/0004-6981\(82\)90195-0](http://dx.doi.org/10.1016/0004-6981(82)90195-0)
 - [7] Arbeloa, S.F.J., Caseiras, C.P. and Andres, P.M.L. (1993) Air Quality Monitoring: Optimization of a Network around a Hypothetical Potash Plant in Open Countryside. *Atmospheric Environment*, **27A**, 729-738. [http://dx.doi.org/10.1016/0960-1686\(93\)90190-A](http://dx.doi.org/10.1016/0960-1686(93)90190-A)
 - [8] Noll, K.E. and Mitsutomi, S. (1983) Design Methodology for Optimum Dosage Air Monitoring Site Selection. *Atmospheric Environment*, **17**, 2583-2590. [http://dx.doi.org/10.1016/0004-6981\(83\)90086-0](http://dx.doi.org/10.1016/0004-6981(83)90086-0)
 - [9] Nakamori, Y. and Sawarangi, Y. (1984) Interactive Design of Urban Level Air Quality Monitoring Network. *Atmospheric Environment*, **8**, 793-799. [http://dx.doi.org/10.1016/0004-6981\(84\)90263-4](http://dx.doi.org/10.1016/0004-6981(84)90263-4)
 - [10] Caselton, W.F. and Zidek, J.V. (1984) Optimal Monitoring Networks Designs. *Statistics and Probability Letters*, **2**, 223-227. [http://dx.doi.org/10.1016/0167-7152\(84\)90020-8](http://dx.doi.org/10.1016/0167-7152(84)90020-8)
 - [11] Sampson, P.D. and Guttorp, P. (1992) Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *Journal of the American Statistics Association*, **87**, 108-126. <http://dx.doi.org/10.1080/01621459.1992.10475181>
 - [12] Guttorp, P., Le, N.D., Sampson, P.D. and Zidek, J.V. (1993) Using Entropy in the Redesign of an Environmental Monitoring Network. Technical Report, Department of Statistics, University of British, Columbia, 116.
 - [13] Perez-Abreu, V. and Rodriguez, J.E. (1996) Index of Effectiveness of a Multivariate Environmental Monitoring Network. *Environmetrics*, **7**, 489-501. [http://dx.doi.org/10.1002/\(sici\)1099-095x\(199609\)7:5<489::aid-env224>3.0.co;2-0](http://dx.doi.org/10.1002/(sici)1099-095x(199609)7:5<489::aid-env224>3.0.co;2-0)
 - [14] McElroy, J.L., Behar, J.V., Meyers, T.C. and Liu, M.K. (1986) Methodology for Designing Air Quality Monitoring Networks: II. Application to Las Vegas, Nevada, for Carbon Monoxide. *Environmental Monitoring and Assessment*, **6**, 13-34. <http://dx.doi.org/10.1007/BF00394285>
 - [15] Richard, W.B., Russell, W.W. and David, K.H. (2002) Methodology for Siting Ambient Air Monitors at the Neighborhood Scale. *Journal of the Air & Waste Management Association*, **52**, 1433-1442. <http://dx.doi.org/10.1080/10473289.2002.10470870>
 - [16] Kainuma, Y., Shiozawa, K. and Okamoto, S. (1990) Study of the Optimal Allocation of Ambient Air Monitoring Stations. *Atmospheric Environment. Part B. Urban Atmosphere*, **24**, 395-406. [http://dx.doi.org/10.1016/0957-1272\(90\)90047-x](http://dx.doi.org/10.1016/0957-1272(90)90047-x)
 - [17] Trujillo-Ventura, A. and Ellis, J.H. (1991) Multiobjective Air Pollution Monitoring Network Design. *Atmospheric Environment. Part A. General Topics*, **25**, 469-479. [http://dx.doi.org/10.1016/0960-1686\(91\)90318-2](http://dx.doi.org/10.1016/0960-1686(91)90318-2)
 - [18] Chen, C., Liu, W. and Chen, C. (2006) Development of Multiple Objective Planning Theory and System for Sustainable Air Quality Monitoring Networks. *Science of the Total Environment*, **354**, 1-19. <http://dx.doi.org/10.1016/j.scitotenv.2005.08.018>
 - [19] Mofarrah, A. and Husain, T. (2010) A Holistic Approach for Optimal Design of Air Quality Monitoring Network Expansion in an Urban Area. *Atmospheric Environment*, **44**, 432-440. <http://dx.doi.org/10.1016/j.atmosenv.2009.07.045>
 - [20] Tseng, C.C. and Ni-Bin, C. (2001) Assessing Relocation Strategies of Urban Air Quality Monitoring Stations by GA-Based Compromise Programming. *Environment International*, **26**, 523-541. [http://dx.doi.org/10.1016/S0160-4120\(01\)00036-8](http://dx.doi.org/10.1016/S0160-4120(01)00036-8)
 - [21] Silva, C. and Quiroz, A. (2003) Optimization of the Atmospheric Pollution Monitoring Network at Santiago de Chile. *Atmospheric Environment*, **37**, 2337-2345. [http://dx.doi.org/10.1016/S1352-2310\(03\)00152-3](http://dx.doi.org/10.1016/S1352-2310(03)00152-3)
 - [22] Elkamel, A., Fatehifar, E., Taheri, M., Al-Rashidi, M.S. and Lohi, A. (2008) A Heuristic Optimization Approach for Air Quality Monitoring Network Design with the Simultaneous Consideration of Multiple Pollutants. *Journal of Environmental Management*, **88**, 507-516. <http://dx.doi.org/10.1016/j.jenvman.2007.03.029>
 - [23] Lu, W., He, H. and Dong, L. (2011) Performance Assessment of Air Quality Monitoring Networks Using Principal Component Analysis and Cluster Analysis. *Building and Environment*, **46**, 577-583. <http://dx.doi.org/10.1016/j.buildenv.2010.09.004>
 - [24] Dogruparmak, S.C., Keskin, G.A., Yaman, S. and Alkan, A. (2014) Using Principal Component Analysis and Fuzzy C-Means Clustering for the Assessment of Air Quality Monitoring. *Atmospheric Pollution Research*, **5**, 656-663. <http://dx.doi.org/10.5094/APR.2014.075>
 - [25] Ferradas, G.F., Minarro, M.M., Terres, I.M.M. and Martinez, F.J.M. (2010) An Approach for Determining Air Pollu-

- tion Monitoring Sites. *Atmospheric Environment*, **44**, 2640-2645. <http://dx.doi.org/10.1016/j.atmosenv.2010.03.044>
- [26] Bayraktar, H. and Turalioglu, F.S. (2005) A Kriging-Based Approach for Locating a Sampling Site—In the Assessment of Air Quality. *Stochastic Environmental Research and Risk Assessment*, **19**, 301-305. <http://dx.doi.org/10.1007/s00477-005-0234-8>
- [27] Kasstele, J., Stein, A., Dekkers, A.L.M. and Velder, G.J.M. (2009) External Drift Kriging of NO_x Concentrations with Dispersion Model Output in a Reduced Air Quality Monitoring Network. *Environmental and Ecological Statistics*, **16**, 321-339. <http://dx.doi.org/10.1007/s10651-007-0052-x>
- [28] Smith, L., Mukerjee, S., Gonzales, M., Stallings, C., Neas, L., Norris, G. and Ozkayanak, H. (2006) Use of GIS and Ancillary Variables to Predict Volatile Organic Compound and Nitrogen Dioxide Levels at Unmonitored Locations. *Atmospheric Environment*, **40**, 3773-3787. <http://dx.doi.org/10.1016/j.atmosenv.2006.02.036>
- [29] Beelan, R., Hoek, G., Pebesma, E., Vienneau, D., de Hoogh, K. and Briggs, D.J. (2009) Mapping of Background Air Pollution at a Fine Spatial Scale across the European Union. *Science of the Total Environment*, **407**, 1852-1867. <http://dx.doi.org/10.1016/j.scitotenv.2008.11.048>
- [30] Leitenstorfer, F. and Tutz, G. (2006) Generalized Monotonic Regression Based on B-Splines with an Application to Air Pollution Data. *Biostatistics*, **8**, 654-673. <http://dx.doi.org/10.1093/biostatistics/kxl036>
- [31] Duncan, L. (2012) Using Spline Models to Estimate the Varying Health Risks from Air Pollution across Scotland. *Statistics in Medicine*, **31**, 3366-3378. <http://dx.doi.org/10.1002/sim.5420>
- [32] Alharbi, B., Shareef, M.M. and Husain, T. (2015) Study of Chemical Characteristics of Particulate Matter Concentrations in Riyadh, Saudi Arabia. *Atmospheric Pollution Research*, **6**, 88-98. <http://dx.doi.org/10.5094/APR.2015.011>
- [33] ESRI. Redlands, USA. <http://www.esri.com>
- [34] Wong, D.W., Yuan, L. and Perlin, S.A. (2004) Comparison of Spatial Interpolation Methods for the Estimation of Air Quality Data. *Journal of Exposure Analysis and Environmental Epidemiology*, **14**, 404-415. <http://dx.doi.org/10.1038/sj.jea.7500338>
- [35] Hoek, G., Brunekreef, B., Goldbohm, S., Fischer, P. and van den Brandt, P.A. (2002). Association between Mortality and Indicators of Traffic-Related Air Pollution in the Netherlands: A Cohort Study. *Lancet*, **360**, 1203-1209. [http://dx.doi.org/10.1016/S0140-6736\(02\)11280-3](http://dx.doi.org/10.1016/S0140-6736(02)11280-3)
- [36] Jerrett, M., Shi, Y.L., Gapstur, S.M., Thun, M.J., Pope, C.A., Burnett, R.T., Beckerman, B.S., Turner, M.C., Krewski, D., Thurston, G., Martin, R.V., van Donkelaar, A., Hughes, E., *et al.* (2013) Spatial Analysis of Air Pollution and Mortality in California. *American Journal of Respiratory and Critical Care Medicine*, **188**, 593-599. <http://dx.doi.org/10.1164/rccm.201303-0609OC>
- [37] Burrough, P. and McDonnell, R. (1998) Principles of Geographic Information Systems. Oxford University Press, New York.
- [38] Liu, L.J.S. and Rossini, A.J. (1996) Use of Kriging Models to Predict 12-Hour Mean Ozone Concentrations in Metropolitan Toronto—A Pilot Study. *Environment International*, **22**, 677-692. [http://dx.doi.org/10.1016/S0160-4120\(96\)00059-1](http://dx.doi.org/10.1016/S0160-4120(96)00059-1)
- [39] Cressie, N. (2006) Block Kriging and Lognormal Spatial Processes. *Mathematical Geology*, **38**, 413-443. <http://dx.doi.org/10.1007/s11004-005-9022-8>
- [40] Lefohn, A.S. and Pinkerton, J.E. (1988) High Resolution Characterization of Ozone Data for Sites Located in Forested Areas of the United States. *Journal of the Air Pollution Control Association*, **38**, 1504-1511. <http://dx.doi.org/10.1080/08940630.1988.10466489>
- [41] Zou, B., Zhan, F.B., Zeng, Y., Yorke, C. and Liu, X. (2011) Performance of Kriging and EWPM for Relative Air Pollution Exposure Risk Assessment. *International Journal of Environmental Research*, **5**, 769-778.
- [42] Shad, R., Mesgari, M.S., Abkar, A. and Shad, A. (2009) Predicting Air Pollution Using Fuzzy Linear Membership Kriging in GIS. *Computers, Environment and Urban Systems*, **33**, 472-481. <http://dx.doi.org/10.1016/j.compenvurbsys.2009.10.004>
- [43] Nash, J.E. and Sutcliffe, I.V. (1970) River Flow Forecasting through Conceptual Models Part I—A Discussion of Principles. *Journal of Hydrology*, **10**, 282-290. [http://dx.doi.org/10.1016/0022-1694\(70\)90255-6](http://dx.doi.org/10.1016/0022-1694(70)90255-6)
- [44] Ross, T. (1996) Indices for Performance Evaluation of Predictive Models in Food Microbiology. *Journal of Applied Microbiology*, **81**, 501-508.
- [45] Monteiro, A., Ribero, I., Tchepel, O., Sa, E., Ferreira, J., Carvalha, A., Martins, V., Strunk, A., Galmarini, S., Elbern, H., Schaap, M., Builtjes, P., Miranda, A.I. and Borrego, C. (2013) Bias Correction Techniques to Improve Air Quality Ensemble Predictions: Focus on O₃ and PM over Portugal. *Environmental Modeling & Assessment*, **18**, 533-546. <http://dx.doi.org/10.1007/s10666-013-9358-2>
- [46] Son, J.-Y., Bell, M.L. and Lee, J.-T. (2010) Individual Exposure to Air Pollution and Lung Function in Korea: Spatial

Analysis Using Multiple Exposure Approaches. *Environmental Research*, **110**, 739-749.
<http://dx.doi.org/10.1016/j.envres.2010.08.003>

- [47] Singh, K.P., Gupta, S., Kumar, A. and Shukla, S.P. (2012) Linear and Nonlinear Modeling Approaches for Urban Air Quality Prediction. *Science of the Total Environment*, **426**, 244-255. <http://dx.doi.org/10.1016/j.scitotenv.2012.03.076>
- [48] Zhao, N., Zeng, X. and Han, S. (2013) Solar Radiation Estimation Using Sunshine Hour and Air Pollution Index in China. *Energy Conversion and Management*, **76**, 846-851. <http://dx.doi.org/10.1016/j.enconman.2013.08.037>