

# The GIS-Based Approach for Optimal Design of Air Quality Monitoring Network for Management of Chemical Clusters

Zhengqiu ZHU

College of Information System and Management, National University of Defense Technology Changsha  
China 410073  
admin@steven-zhu.me

Rongxiao WANG

College of Information System and Management, National University of Defense Technology Changsha  
China 410073  
pandaself@testbot.cn

Bin CHEN\*

College of Information System and Management, National University of Defense Technology Changsha  
China 410073  
nud tcb9732@gmail.com

Liang MA

College of Information System and Management, National University of Defense Technology Changsha  
China 410073  
15243659741@qq.com

Sihang QIU

College of Information System and Management, National University of Defense Technology Changsha  
China 410073  
qiusihang@gmail.com

Xiaogang QIU

College of Information System and Management, National University of Defense Technology Changsha  
China 410073  
Michael.qiu@139.com

## ABSTRACT

An industrial district with chemical plants producing inside poses a great threat to the surrounding atmospheric environment and human health. Therefore, designing a proper and available air quality monitoring network (AQMNs) is essential for assessing the effectiveness of deployed pollution controlling strategies and facilities in dealing with reducing pollutants in the planning stage of emergency management. Whereas monitoring facilities located at inappropriate sites would affect data validity. Thus, in this paper, a geospatial technique-Bayesian Maximum Entropy (BME) in conjunction with a multi-objective optimization model was utilized to optimize the design of an AQMN of gas sensors. Our developed atmospheric dispersion simulation system was employed to generate 'real' historical data for the above method and an experiment was implemented to illustrate the feasibility of the proposed approach. This work is expected to facilitate a decision-making process for determining an appropriate AQMN and assist the management work of environmental protection authorities.

## CCS Concepts

•Social and professional topics→ Computing industry

## Keywords

Bayesian maximum entropy, Multi-objective optimization model, Air quality monitoring network, Atmospheric dispersion simulation system

## 1 INTRODUCTION

Controlling gaseous pollutants is substantial and urgent in the

\*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.  
*EM-GIS'17*, November 7–10, 2017, Redondo Beach, CA, USA

©2017 Association for Computing Machinery.

ACM ISBN 978-1-4503-5493-6/17/11...\$15.00.

<https://doi.org/10.1145/3152465.3152479>

situation of today's atmospheric environment. Chemical industrial activities are an important factor leading to the deterioration of the atmospheric environment, and especially in developing countries (e.g. China and India), this is the case. Unfortunately, the byproducts produced during the production processes are noxious, even sometimes highly toxic, and often they are discharged to the nearby atmospheric environment without purification treatment. As a result, the atmospheric quality in these countries is extremely poor [13], leading to substantial health problems for the residents [10] and to the destruction of the ecosystem.

Faced with these problems, governments in developing countries have introduced a series of measures to abate atmospheric pollution [7; 8]. These measures include promulgation of standards, norms and emergency plans for environmental quality as well as air quality monitoring and control [31]. One of the most substantial tasks is the creation of an Air Quality Monitoring Networks (i.e. AQMN), constituted of measurement devices, to detect and monitor the disposed atmospheric pollutants in the chemical cluster for environmental protection authorities. Establishing such a proper AQMN to evaluate the spatiotemporal distribution of gaseous pollutants and the effectiveness of pollution controlling strategies is essential to ensure the health of the surrounding residents and the sustainability of ecosystem. Objectives and necessities of such monitoring networks are reported frequently in the literature [9; 18; 24; 25; 29] and can be summarized as follows: (i) Objectives related to air pollution legislation; long term land use planning and the announcement of emergency situations; (ii) The evaluation of exposure of population and other potential receptors; (iii) The controlling of emissions from significantly important sources (e.g. thermal power plant); (iv) Analysis of air pollution data to conclude emission trends in air pollution or for further research.

Moreover, the minimization of network cost covered by these objectives has been frequently reinterpreted as a constraint on the available budget [23; 31]. Previous methods in the literature fail to accomplish the task of designing a network capable of fulfilling all of the objectives above. Most of the reported methods applied to specific situations wherein one or two of the previous objectives are considered [22].

Generally, existing methods of establishing an AQMN typically account for parameters related to ambient concentrations of gaseous pollutants of interest such as atmospheric transport and

dispersion, diffusion source characteristics, secondary reactions, deposition characteristics and local topography [2]. The objective of these network design methods is usually aiming at identifying the sites of maximum contaminant concentration, maximum contaminant dosage and maximum population protection as well as covering the maximum urban areas with the minimum number of monitoring stations [9; 12; 14; 17; 20; 21; 30].

Among the previous works on AQMN design, Goldstein [11] designed an AQMN in the greater London area based on the concept of a spatial correlation analysis. A statistical measure of information content was used to evaluate the availability of a particular AQMN in Canada [26]. Moreover, interpolation techniques were taken in Netherlands when assessing the errors in the spatial analysis [32]. Afterwards, air quality simulation models and population exposure information was applied to generate representative combined patterns; and then McElroy et al. applied the concepts of ‘sphere of influence’ (SOI) and ‘figure of merit’ (FOM) to determine the minimum numbers of monitoring stations required [19]. A methodology which involves the multiple-criteria method, spatial correlation technique in conjunction with fuzzy analytic hierarchy process was also used to determine the optimum number of ambient air quality stations [23]. It can be concluded from these studies that almost all these methods focus on optimizing the number and layout of fixed monitoring stations for urban areas. However, the ambient air quality monitoring problem in our work is totally different from the domains of research mentioned above.

Our research goal is to design an appropriate AQMN to evaluate the spatial and temporal distribution of pollutants and the effectiveness of pollution control strategies for the chemical cluster incorporating fixed air quality monitoring stations with hazardous gas sensors. In the initial construction of a chemical cluster, a few fixed monitoring stations would generally be arranged based on economic, social and meteorological factors [16]. However, the price of fixed monitoring station is too high (i.e. up to a million); therefore a few fixed monitoring stations cannot effectively cover the entire industrial district. Extremely, the key areas are likely to not be covered by these monitoring stations while a leakage incident occurs. Thus the unavailability and irrationality of initial construction of an AQMN in the chemical cluster is an essential factor of ineffective supervision, which may lead to the occurrence of emergent accidents. In light of that, the proposed method in this paper designs a monitoring network of gas sensors through analyzing historical monitoring data based on existing fixed monitoring stations.

In this article, a Geographic Information System-based (GIS-based) method is introduced to design a comprehensive AQMN wherein both high-accuracy air quality monitoring stations and gas sensors are modeled. First, study area and some additional information are introduced. Then, Bayesian Maximum Entropy (BME) is applied to generate concentration distribution of gaseous pollutants based on historical monitoring data from gas sensors. Finally, a multi-objective optimization model is built on account of the concentration distribution aiming at optimizing the monitoring network of gas sensors. The proposed approach, which involves two kinds of inspection resources, is a supplement of existing monitoring approaches, and greatly improves the validity of monitoring data.

## 2 Methods

In this section, a geospatial technique based on the BME method is presented to acquire the spatiotemporal distribution of concentration of gaseous pollutants in a chemical cluster by importing multi-year monitoring data from multi-monitoring spots. With the spatiotemporal distribution of airborne contaminants, the optimal layout of gas sensors can be determined based on the optimization target of maximum contaminant concentration or maximum contaminant dosage. Together with the fixed monitoring stations, a comprehensive monitoring network is built up to achieve valid monitoring of gaseous pollutants in a chemical cluster. Moreover, decision-makers can utilize the spatiotemporal information of monitoring measurements provided by the AQMN to develop planning and mitigation strategies.

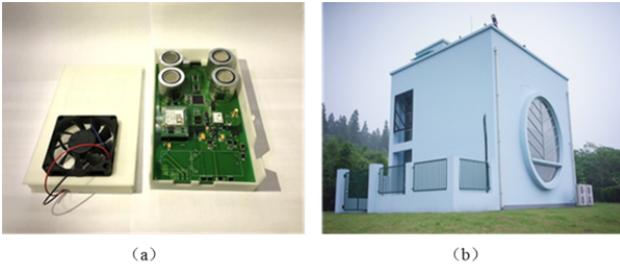
### 2.1 Study area

Basically, a chemical industrial park or a chemical cluster is composed of numerous chemical companies, an inspection agency and functional departments (e.g. hospitals, hotels, police offices and etc.). Moreover, a chemical company may possess several chemical plants in the chemical industrial park. Our study area is not an exception. Figure 1 shows a refinery GIS map of a chemical cluster in Shanghai, China which is also the study area used in our research. Through investigating the emission of  $SO_2$  and referring to the main byproducts information of chemical plants, five possible  $SO_2$  emission sources are located in this area. On the map, the 19 small circles are the complete set of discharge points for all contaminants, among which the five blue circles are the  $SO_2$  discharge spots. After projecting the WGS84 geographic coordinates into UTM Cartesian coordinates, the resulting locations of all the candidate sources are listed in Table 1 with concrete information. Moreover, the triangles indicate the fixed high-accuracy air quality monitoring stations which cannot support the ability to monitor the whole chemical cluster; the area marked by the black quadrilateral box is main working area of the site while the area below the working region is the sea. Meanwhile, a practicality picture of the two inspection resources is shown in Figure 2. These inspection resources (i.e. five monitoring stations and gas sensors) are operating to inspect 55 chemical plants with 243 releasing spots. In contrast, the measurement accuracy of high-accuracy air quality monitoring stations is a thousand times more accurate compared to that of gas sensors. Moreover, the measurements of gas sensors should be considered as hard data when utilized in BME method. The monitoring data in the past year of the five fixed monitoring stations are used as imports in source estimation methods to determine the actual frequent emission sources of  $SO_2$ . Through calculation and our investigation, the chimney of the sulfuric acid recovery (SAR) system and Waste incinerator for acrylonitrile (AR) were the major locations contributing to the emission of  $SO_2$ . Then, the source term as well as historical meteorological data is imported into an atmospheric dispersion simulation tool to generate ‘real’ historical concentration data of  $SO_2$  in the area of Shanghai chemical cluster. KD-ADSS developed by National University of Defense Technology is used as simulation tool in this study. Besides, KD-ADSS is a Gaussian-model based simulation system which has been validated by the commercial software PHAST, the Indianapolis field study and a study of the

Fukushima Dai-ichi nuclear accident. The detailed setup of study area will be presented in subsection 4.2-experimental settings.



**Figure 1:** Refinery GIS-Map of study area.



**Figure 2:** Inspection resources of inspection agency ((a) is the gas sensors and (b) is one of the high-accuracy air quality monitoring stations).

**Table 1:** Cartesian coordinates of SO<sub>2</sub> emission points and additional information.

No.	X	Y	Height	Contaminants
1	-132.575	-1317.63	50	SO <sub>2</sub> , NOx, VOC, NH <sub>3</sub>
2	-302.901	-1483.42	68	SO <sub>2</sub> , NOx, vitriol fog
3	267.1415	0.359916	27	SO <sub>2</sub> , PM2.5/PM10
4	861.3643	147.0462	27	SO <sub>2</sub> , PM2.5/PM10
5	1532.017	-142.542	30	CO, SO <sub>2</sub> , NOx, PM2.5/PM10, HF, HCl

## 2.2 Bayesian maximum entropy

In air quality studies, the concentration distribution of a particular pollutant – such as SO<sub>2</sub> – is represented in form of a spatiotemporal random field (**S/TRF**)  $X(\mathbf{p})$  which assumes values at space/time points  $\mathbf{p} = (\mathbf{s}, t)$ , where  $\mathbf{s}$  is location vector while  $t$  denotes time [1]. Generally, environmental authorities are concerned with the estimation values of the pollutant **S/TRF** at unmeasured points with available dataset and physical knowledge. The estimation process leads to a spatiotemporal map which presents the concentration distribution of the pollutant in space and time. BME, a space-time data analysis method in a modern statistical framework introduced by Christakos [1; 5], provides an effective, efficient and accurate way for the

estimation of concentration distribution. It is worth to note that BME has already been proved to perform better in spatiotemporal analysis compared to Kriging technique and interpolation approach [3].

The publicly available SEKS-GUI software library [33] is utilized in this paper to implement the space-time BME analysis. The software solves the fundamental BME equations of spatiotemporal dependence analysis and mapping as follows [15]:

$$\begin{cases} d\chi(\mathbf{g} - \bar{\mathbf{g}})e^{\mu^T g} = 0 \\ d\chi_{\xi_s} e^{\mu^T g} - Af_K(\mathbf{p}) = 0 \end{cases}. \quad (1)$$

where  $\mathbf{g}$  is a vector of  $g_a$ -functions ( $a = 1, 2, \dots$ ) and  $\bar{\mathbf{g}}$  denotes the statistical expectation,  $\mu$  is a vector of  $\mu_a$ -coefficients that depends on the space-time coordinates and is related to  $\mathbf{g}$  (i.e., the  $\mu_a$  indicates the relative significance of each  $g_a$ -functions in the composite solution sought), the  $\xi_s$  represents the site-specific knowledge bases available,  $A$  is a normalization parameter, and  $f_K$  is the attribute probability density function (PDF) at each point. The parameters of  $\mathbf{g}$  and  $\xi_s$  are inputs to the equation, whereas the unknowns are the  $\mu$  and  $f_K$  across space and time.

BME is practical because the fundamental equations make no restrictive assumptions about the underlying probability distributions (non-Gaussian laws are automatically incorporated) and the shape of the space-time predictor (non-linear predictors are allowed). Hence, the BME framework can handle with a broader scope of knowledge bases (KB) types and uncertain data [6].

General KB (G-KB) and site-specific KB (or S-KB) are integrated in the fundamental BME equations. Specifically, the G-KB includes physical laws, theoretical models of space-time dependence (covariance, semivariogram, etc.), empirical relations, and logic-based assertions that are concerned with the pollution  $X(\mathbf{p})$ . The S-KB usually consists of observed hard data and soft data (measurements with a significant amount of uncertainty). In our study, the G-KB contains theoretical covariance models while the S-KB includes airborne pollutant data.

SEKS-GUI represents the prediction grid by the space-time vectors  $\mathbf{p}_k$ , in which case the fundamental BME equations compute the complete prediction PDF  $f_K$  at each  $\mathbf{p}_k$ . After determining the objective of study and PDF  $f_K$ , predictions of  $X(\mathbf{p})$  can be derived at each spatiotemporal node  $\mathbf{p}_k$  of mapping grid.

The software thus generates informative airborne pollutant maps that completely cover the spatial and temporal continua within their respective extents, which enables establishment of optimization model in the next subsection.

## 2.3 Multi-objective optimization model

In this subsection, two single objectives are proposed to optimize the design of an AQMN of gas sensors based on the results of BME. In conjunction with the fixed monitoring stations, a comprehensive air quality monitoring network with the ability of global and valid monitoring in a chemical cluster is built up.

**2.3.1 Maximum concentration detection capability.** Maximum concentration detection capability (CDC) is defined here as the maximum grids number captured by AQMN of gas sensors wherein the pollutant levels are mostly exceeding the threshold value of standard value or average value. The model, based on this objective, is established as follows:

$$\text{Max } o_{cdc} = \sum_{i=1}^T d_i \quad (2)$$

$$\text{s.t. } d_i = \sum_{j \in M_i} y_j \quad \forall i \quad (3)$$

$$\sum_{j=1}^J y_j \leq Q \quad (4)$$

$$y_j \in \{0,1\} \quad \forall j \quad (5)$$

Where the notation of  $d_i$  is the variable indicating the number of grids where the pollutant level exceeding the threshold value of standard value is detected in the  $i^{th}$  month;  $T$  is the total number of months in a year;  $y_j$  is a binary integer that indicates whether a gas sensor is placed in grid  $j$ ;  $J$  denotes the total number of grids in study area;  $M_i$  is the set of grids in the  $i^{th}$  month with a pollutant level greater than the threshold of standard value or average value; and  $Q$  is the upper limit of the number of gas sensors in an AQMN.

**2.3.2 Maximum dosage detection capability.** Some areas may have a low incidence of high-level pollution but a large dosage due to long term exposure [16]. Therefore, using the previous mentioned the objective, CDC, alone when designing an AQMN may be inadequate. The objective and the model for maximum dosage detection capability (DDC) can be formulated as follows:

$$\text{Max } o_{DDC} = \sum_{j \in N} y_j \quad (6)$$

$$\text{s.t. } C_j = \sum_{i=1}^T C_{ij} \quad (7)$$

$$j \in N, \text{ if } C_j \geq (\sum_{j=1}^J C_j) / |J| \quad (8)$$

$$\sum_{j=1}^J y_j \leq Q \quad (9)$$

$$y_j \in \{0,1\} \quad \forall j \quad (10)$$

Where the notation of  $C_{ij}$  represents the pollutant level at grid  $j$  in the  $i^{th}$  month. The first and second constraints indicate that the notation of  $N$  is the set of grids with an accumulated dosage pollutant level greater than the threshold of average value. Of these two objective models described above, the CDC and DDC can be applied independently or combined into a multi-objective model. In this paper, the combination of the two objectives seems to have a little impact on the final result in our experiments. Therefore, we utilize the different single objective to design an AQMN of gas sensors respectively.

### 3 Dataset

The five fixed monitoring air quality stations in Shanghai chemical cluster were constructed according to the regional project for air quality conservation, established by empirical judgment and the governmental law [7]. After projecting the WGS84 geographic coordinates into UTM Cartesian coordinates, the resulting locations of these monitoring stations are listed in Table 2 with some additional information.

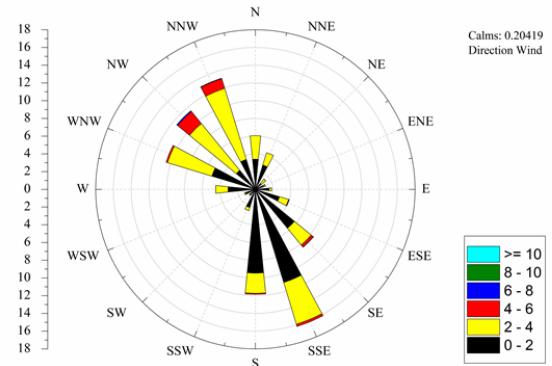
These fixed monitoring stations contain high-accuracy reactors which are able to collect up to 118 categories of airborne contaminants (e.g. VOCs, NOx, SO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> and etc.). The measurement interval of gaseous pollutants lasts no longer than a few seconds, which can be considered as continuous collections.

Therefore, these data is suitable for hourly average, daily average, monthly average and annually average analysis. The trend analysis is the focus of environmental protection authorities. The historical dataset used in this study consisted of up to 118 categories of airborne pollutants and meteorological data for the period December 2015 – November 2016 at the five fixed monitoring stations in Shanghai chemical cluster. The total amount of the dataset is 20430248 entries while the available data contains 20290164 entries. The ineffective data is associated with sites where temporary failures, warm-up of devices or stations working on a non-systematic temporal basis existed.

**Table 2: Cartesian coordinates of fixed monitoring stations and additional information.**

No.	X	Y	Explanation
A	-1022.9	223.	North-West Station
B	-796.9	-431.5	Secco Station
C	3575.3	3518.9	North-East Station
D	1620.6	1293.3	Union Road Station
E	4151.1	1292.3	Covestro Station

Pollutant distribution in the atmosphere varies under different meteorological conditions. Wind speed and directions change remarkably in different months, imposing great uncertainty on the distribution of airborne contaminants. Figure 3 shows the annual wind-rose diagram for the study area. In detail, the figure fully illustrates that Shanghai is affected by the obvious monsoon. The southeast monsoon prevails at Shanghai in summer while the northwest monsoon prevails in winter. Generally, most concentration distribution of a specific gaseous pollutant is distributed in the downside direction of the wind. Therefore, it is possible for researchers to predict the approximate distribution of the gaseous pollutants in a chemical cluster through analyzing the regular variation of wind speed and wind direction.



**Figure 3: Annual wind-rose diagram for the research area.**

### 4 Experiments

As the cooperation with Shanghai chemical cluster has just started up, it is impossible to collect long term monitoring data of gas sensors. Fortunately, measurement data in the past year of fixed monitoring stations in conjunction with historical meteorological data can be applied to estimate the parameters of diffusion source term (i.e. location and release rate) of a particular

gaseous pollutant through source estimation methods. Interested readers are referred to QIU and ZHU [27; 28], more technical and theoretical details were introduced. Then, the source term as well as historical meteorological data is imported to our developed atmospheric dispersion simulation software named as KD-ADSS. After simulation, the concentration distribution of a particular airborne contaminant in the past few years is served as ‘real’ historical concentration distribution data. Moreover, the concentration data is extracted at locations of gas sensors as the historical monitoring data. Finally, the ‘real’ historical monitoring data is imported to mixed integer linear program to generate optimal design of an AQMN.

#### 4.1 Experimental settings

**4.1.1 Setup of study area.** Further, the study area of Shanghai chemical cluster can be simplified as a  $2000m \times 3000m$  rectangle on account of experimental conditions. The refinery is shown in Figure 4. Specifically, the red circles represent the two main release spots calculated through source estimation method (1 denotes Waste incinerator for acrylonitrile (AR) and 2 denotes the chimney of the sulfuric acid recovery (SAR) system). The proper locations of these two releasing spots are (400,400) and (200,300) respectively. Moreover, the rectangle is divided into 150 quadrature grids wherein the size is  $200m \times 200m$ . The initial layout of the gas sensors is designed as follows according to the wind field analysis. Besides, the maximum number of gas sensors in an AQMN is no more than 60. All of the gas sensors are positioned at the height of 20 meters in the center of each grid.

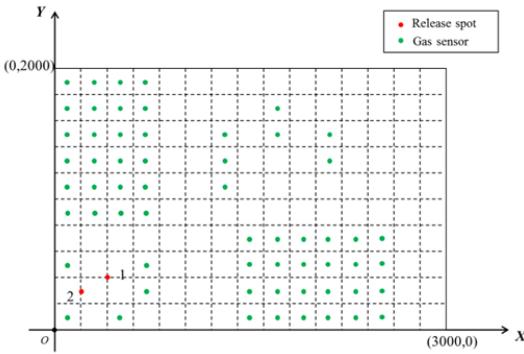


Figure 4: Simplified study area for experiment.

**4.1.2 Setup of atmospheric dispersion simulation.** To simulate a contaminant dispersion scenario, other than the information of emission source  $\theta$  and meteorological parameters (wind direction  $d$  and wind speed  $v$ ), the environmental parameters (e.g. atmospheric stability and terrain type) must also be considered. For the Gaussian diffusion coefficients  $\sigma_y$  and  $\sigma_z$ , they can be expressed as [4]:

$$\begin{cases} \sigma_y(D_x) = a_y D_x (1 + b_y D_x)^{-c_y} \\ \sigma_z(D_x) = a_z D_x (1 + b_z D_x)^{-c_z} \end{cases} \quad (11)$$

where  $D_x$  denotes the downwind distance of the interest point. The parameters  $a_y, b_y, c_y, a_z, b_z, c_z$  depend on the environmental conditions (i.e. atmospheric stability and terrain type). According to the terrain type of Shanghai chemical cluster (i.e. urban) and atmospheric stability in Shanghai (i.e. C), the parameters

$a_z, b_z, c_z, a_z, b_z, c_z$  is determined at value of (0.16, 0.0004, -0.5, 0.08, 0.0015, -0.5).

Through analyzing the daily concentration data collected by fixed monitoring stations, the daily hour-average concentration trend is discovered. In Figure 5, the X-axis is the time series of one day while the Y-axis represents the main atmospheric contaminants monitored by monitoring stations. The background color of this figure is white, which means concentration value of atmospheric pollutants is zero. Furthermore, a darker area represents higher gas concentration in this figure. It can be concluded from the color-bar that black is darker than grey and white means that the concentration of the former is greater than that of the latter. From the figure, it is obvious that discharging behavior of chemical plants clearly has temporal characteristics. The discharging amount of atmospheric pollutants in the time unit of 12-24 hours is far greater than that in the time unit of 1-12 hours. Therefore, the ‘real’ release rates  $q$  of the two release spots vary from 0 to  $5 g \cdot s^{-1}$ . Moreover, the release rate during the period of 12-24 hours is set twice of that during the period of 0-12 hours in simulation. Besides, through analyzing the monthly-average concentration data of  $SO_2$ , it is concluded in Figure 6 that the discharging amount of  $SO_2$  during the period from February to July is greater than that during the period from August to January. Thus, the release rate in summer is set greater than that in winter.

Then, all the input parameters and data including emission source, historical meteorological data, environmental parameters and Gaussian diffusion coefficients are imported to conduct daily dispersion for a decade in KD-ADSS. Further, the generated ‘real’ historical dataset is utilized to produce monthly-averaged  $SO_2$  measurements from a total of 60 monitoring locations of gas sensors. Finally, these monthly-averaged measurements as well as geospatial data would be imported to SEKS-GUI library.

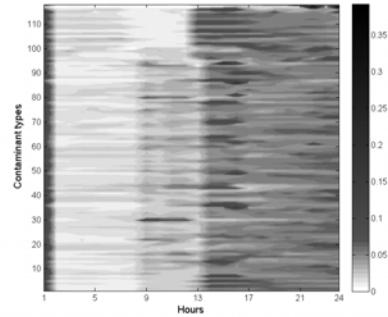
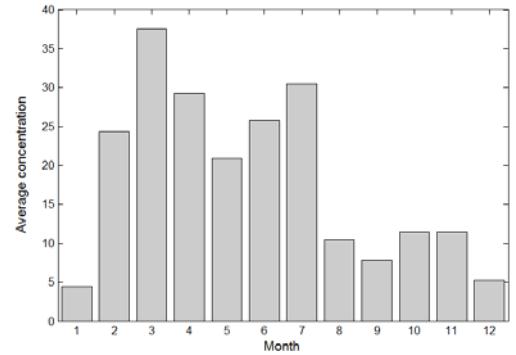


Figure 5: Daily hour-average concentration trend during the past year.



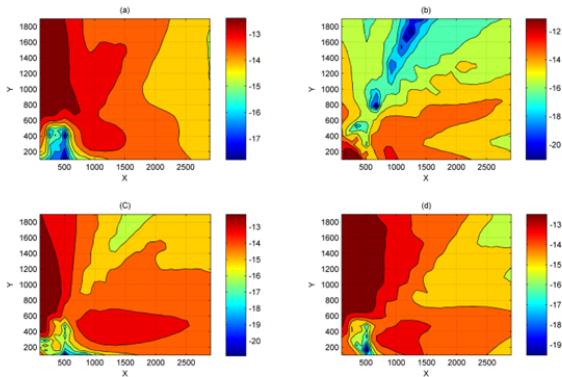
**Figure 6: Monthly-average concentration trend of SO<sub>2</sub>.**

**4.1.3 Setup of SEKS-GUI software library.** SEKS-GUI implements the BME methodology for spatiotemporal analysis. The following paragraph is the workflow of how this analysis works in SEKS-GUI.

Firstly, the hard data information with exact measurements, the soft data with certain uncertainty and output GIS grid are imported to SEKS-GUI. Then, the data is detrended and brought from raw input information into suitable processing form. For the detrending, Gaussian kernel smoothing is applied across the dataset. Moreover, a data transformation aiming at reshaping the detrended data set from the original space of values (original-space) into a space where their distribution resembles a Gaussian one (transformation-space) is implemented. Besides, a covariance analysis is conducted to investigate correlation patterns among the data in the next stage. Finally, we have to select and initialize the type of BME prediction. Different options are ranked with respect to the time and complicity required for the computations, starting with the fastest and simplest one and ending with the most time-consuming and complicated. In visualization, SEKS-GUI offers a bundle of mapping options to display the BME prediction results once the BME output MAT-file is loaded.

## 4.2 Results and discussions

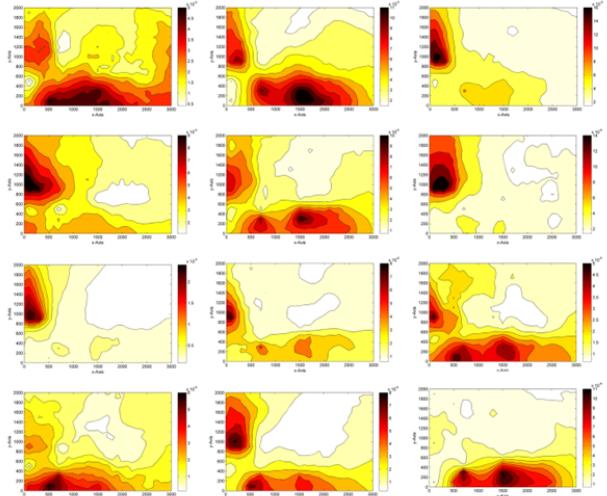
**4.2.1 Results of dispersion simulation.** To acquire the ‘real’ mean concentration distribution of SO<sub>2</sub> during the past ten years, about 86,400 dispersion scenarios (i.e. 24(hour) × 30(day) × 12(month) × 10(year) ) based on historical meteorological data were run through our developed KD-ADSS. Figure 7 indicates the simulated concentration distribution of SO<sub>2</sub> in January, June, July and December of 2016. To better exhibiting the dispersion effect, the concentration data in Figure 7 was logarithmically processed and sophisticatedly interpolated. From the figure, it is concluded that north-west part and south-east part of study area were seriously affected by the gaseous pollutant of SO<sub>2</sub> due to the influence of monsoon. After obtaining the ‘real’ dispersion data, concentration data at the monitoring spots of gas sensors are extracted from the dataset. The amount of the ‘real’ monitoring data in SEKS-GUI is 7200 entries (i.e. 60(sites) × 12(month) × 10(year) ).



**Figure 7: Mean Concentration distribution of SO<sub>2</sub> in 2016 generated by KD-ADSS ((a) denotes distribution in January;**

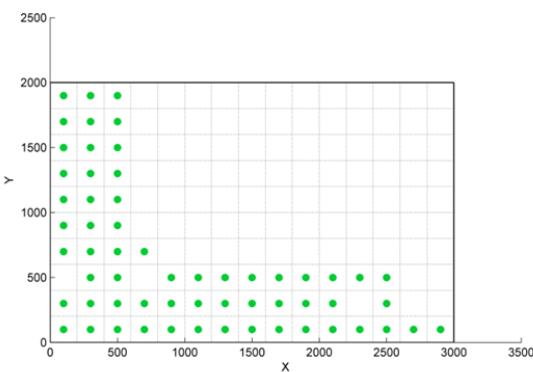
**(b) denotes distribution in June; (c) denotes distribution in July; (d) denotes distribution in December).**

**4.2.2 Results of BME analysis.** After a 10-year dataset of ‘real’ monitoring concentration as well as geospatial information of prediction grid is loaded, the detrending stage, transformation stage and covariance analysis stage were then conducted. After that, BME prediction and visualization of the results are shown in Figure 8. The figure illustrates the averaged concentration distribution trend in each month during the past ten years. It is worth noting that the concentration data utilized in this figure is the predicted raw data without interpolation and log-likelihood. Obviously, the north-west part of study area was severely affected in summer time because of the south-east monsoon while the south-east part of study area was severely influenced in winter time due to the north-west monsoon. Moreover, the concrete concentration data and geospatial data of this figure are exported to the following research of designing a valid AQMN.



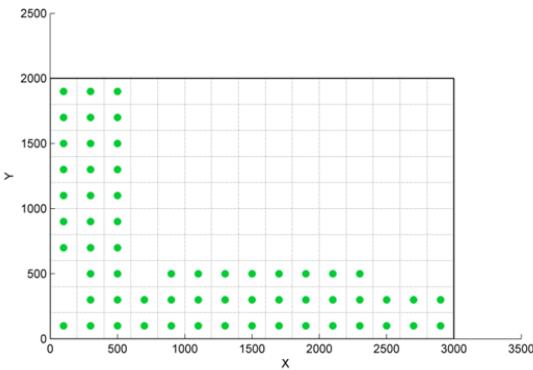
**Figure 8: Mean Concentration distribution of SO<sub>2</sub> during the past ten years by SEKS-GUI (from left to right, top to bottom, the sub-figure denotes January to December in turn).**

**4.2.3 Results of multi-objective optimization model.** With respect to the first objective—maximum concentration detection capability, the essence of the corresponding mixed integer linear program (MILP) is to detect the most frequent grids wherein high concentration value excessing the standard value or the average value often occurs. After importing the dataset acquired from the BME result to the corresponding MILP, the optimal design of an AQMN on account of maximum concentration detection capability is exhibited in Figure 9. Compared to the initial layout of gas sensor AQMN, several monitoring spots which initially are positioned at the north-east part are removed. Moreover, all of the gas sensors are located in the north-west and south-east direction of the release spot of SO<sub>2</sub>.



**Figure 9: Optimal design of an AQMN on account of maximum concentration detection capability.**

Considering the second objective—maximum dosage detection capability, the optimal design of an AQMN is exhibited in Figure 10. Compared to the result of CDC, the structure of the AQMN has barely changed. The grids wherein long term exposure occurs are also located in the north-west and south-east part of the chemical cluster. Slight variations appearing near the release sources may result from the height of monitoring spots.



**Figure 10: Optimal design of an AQMN on account of maximum dosage detection capability.**

It must be some concerns about the extreme circumstances that the dispersion of gaseous pollutants may spread to the north-east part of the study area without detection. This is the case, but there are also three high-accuracy air quality monitoring stations fixed in the north-east part of the study area. Therefore, the comprehensive AQMN constituted of gas sensors and monitoring stations is valid and available.

## 5 CONCLUSIONS

This paper applies a geospatial technique-BME in conjunction with a multi-objective optimization model to optimize the design of an AQMN of gas sensors. To deal with the problem of lacking long-term historical monitoring data, our developed atmospheric dispersion simulation system is employed to generate ‘real’ historical data based on the results of source estimation and wind analysis. Then, an experiment is implemented to illustrate the feasibility of the proposed approach. Results show that BME

prediction of mean concentration distribution not only reveals the distribution regularity of gaseous pollutants, but also provides essential data for designing an AQMN. This work has been proved to have the ability to facilitate a decision-making process for determining an appropriate AQMN and assist the management work of environmental protection authorities.

## ACKNOWLEDGMENTS

This study is supported by National Key Research & Development (R&D) Plan under Grant No. 2017YFC0803300 and the National Natural Science Foundation of China under Grant Nos. 71673292, 61503402 and Guangdong Key Laboratory for Big Data Analysis and Simulation of Public Opinion.

## REFERENCES

- [1] CHRISTAKOS, G., 2005. *Random Field Models in Earth Sciences*. Dover Publications: Mineola, NY USA. ISBN: 0486438724.
- [2] BALDAUF, R.W., LANE, D.D., and MAROTE, G.A., 2001. Ambient Air Quality Monitoring Network Design for Assessing Human Health Impacts from Exposures to Airborne Contaminants. *Environmental Monitoring and Assessment* 66, 1 (2001/01/01), 63–76. DOI: <http://dx.doi.org/10.1023/A:1026428214799>.
- [3] BOGAERT, P., CHRISTAKOS, G., JERRETT, M., and YU, H.L., 2009. Spatiotemporal modelling of ozone distribution in the State of California. *Atmospheric Environment* 43, 15 (2009/05/01/), 2471–2480. DOI: <http://dx.doi.org/http://dx.doi.org/10.1016/j.atmosenv.2009.01.049>.
- [4] CARRASCAL, M.D., PUIGCERVER, M., and PUIG, P., 1993. Sensitivity of Gaussian plume model to dispersion specifications. *Theoretical and Applied Climatology* 48, 2 (1993/06/01), 147–157. DOI: <http://dx.doi.org/10.1007/BF00864921>.
- [5] CHRISTAKOS, G., 1991. On certain classes of spatiotemporal random fields with applications to space-time data processing. *IEEE Transactions on Systems, Man, and Cybernetics* 21, 4, 861–875. DOI: <http://dx.doi.org/10.1109/21.108303>.
- [6] CHRISTAKOS, G. and HRISTOPULOS, D.T., 1998. *Spatiotemporal environmental health modelling: A tractatus stochasticus*. Kluwer Academic Publishers, Boston, MA USA.
- [7] Standing Committee of the Sixth National People's Congress, 2000. Law of the People's Republic of China on the Prevention and Control of Atmospheric Pollution. *Environmental Policy Collection*. Retrieved October 7, 2009 from <http://digital.library.unt.edu>.
- [8] DALTON, S. and TIERNEY, M., 2007. *Collections: Earth and Environmental Science Collection Development Policy*. Nyu Libraries. Retrieved May 7, 2015 from <http://digital.library.nyu.edu/collections/policies/envscien.html>.
- [9] DARBY, W.P., OSSENBRUGGEN, P.J., and GREGORY, C.J., 1974. OPTIMIZATION OF URBAN AIR MONITORING NETWORKS. *Journal of the Environmental Engineering Division* 100, 4, 577–591.
- [10] Tong, Z., Yang, B., Hopke, P.K. and Zhang, K.M., 2017. Microenvironmental air quality impact of a commercial-scale biomass heating system. *Environment Pollution* 220, (2017/01), 1112–1120. DOI: <http://dx.doi.org/10.1016/j.envpol.2016.11.025>
- [11] GOLDSTEIN, I.F., 1982. Rationalization 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* (1967) 16, 8 (1982/01/01/), 1061–1070. DOI: [http://dx.doi.org/http://dx.doi.org/10.1016/0004-6981\(82\)90482-6](http://dx.doi.org/http://dx.doi.org/10.1016/0004-6981(82)90482-6).

- [12] HARRISON, R.M. and DEACON, A.R., 1998. Spatial Correlation of Automatic Air Quality Monitoring at Urban Background Sites: Implications for Network Design. *Environmental Technology* 19, 2 (1998/02/01), 121-132. DOI= <http://dx.doi.org/10.1080/09593331908616664>.
- [13] HSU ANGEL, E.D., LEVY MA, SHERBININ ALEX, 2016 *Environmental Performance Index*. Yale University: New Haven, CT USA. Retrieved January 25, 2016 from <http://www.epi.yale.edu>.
- [14] KAINUMA, Y., SHIOZAWA, K., and OKAMOTO, S.I., 1990. Study of the optimal allocation of ambient air monitoring stations. *Atmospheric Environment. Part B. Urban Atmosphere* 24, 3 (1990/01/01), 395-406. DOI= [http://dx.doi.org/http://dx.doi.org/10.1016/0957-1272\(90\)90047-X](http://dx.doi.org/http://dx.doi.org/10.1016/0957-1272(90)90047-X).
- [15] KANEVSKI, M., 2010. *Advanced Mapping of Environmental Data: Geostatistics, Machine Learning and Bayesian Maximum Entropy*. Wiley, New York, NY USA.
- [16] KAO, J. J. and HSIEH, M. R., 2006. Utilizing multiobjective analysis to determine an air quality monitoring network in an industrial district. *Atmospheric Environment* 40, 6 (2006/02/01/), 1092-1103. DOI= <http://dx.doi.org/http://dx.doi.org/10.1016/j.atmosenv.2005.11.003>.
- [17] LIU, M.K., AVRIN, J., POLLACK, R.I., BEHAR, J.V., and MCELROY, J.L., 1986. Methodology for designing air quality monitoring networks: I. Theoretical aspects. *Environmental Monitoring and Assessment* 6, 1 (1986/01/01), 1-11. DOI= <http://dx.doi.org/10.1007/BF00394284>.
- [18] LUDWIG, F.L. and SHELAR, E., 1978. SITE SELECTION FOR THE MONITORING OF PHOTOCHEMICAL AIR POLLUTANTS. U.S. EPA Report No.EPA-450/3-78-013, Research Triangle Park, NC.
- [19] 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, 1 (1986/01), 13-34. DOI= <http://dx.doi.org/10.1007/bf00394285>.
- [20] MILLER, A.A. and SAGER, T.W., 1994. Site Redundancy in Urban Ozone Monitoring. *Air & Waste* 44, 9 (1994/09/01), 1097-1102. DOI= <http://dx.doi.org/10.1080/10473289.1994.10467305>.
- [21] MODAK, P.M. and LOHANI, B.N., 1985. Optimization of ambient air quality monitoring networks. *Environmental Monitoring and Assessment* 5, 1 (1985/03/01), 21-38. DOI= <http://dx.doi.org/10.1007/BF00396392>.
- [22] MOFARRAH, A. and HUSAIN, T., 2009. Methodology for designing air quality monitoring network. In *Second International Conference on Environmental & Computer Science*. IEEE Computer Society Press, New York, NY USA.
- [23] 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, 3 (2010/01/01/), 432-440. DOI= <http://dx.doi.org/http://dx.doi.org/10.1016/j.atmosenv.2009.07.045>.
- [24] MUNN, R.E., 1981. The Design of Air Quality Monitoring Networks. MacMillan Publishers Ltd, London UK.
- [25] OTT, W.R., 1977. Development of Criteria for Siting Air Monitoring Stations. *Journal of the Air Pollution Control Association* 27, 6 (1977/06/01), 543-547. DOI= <http://dx.doi.org/10.1080/00022470.1977.10470453>.
- [26] PICKETT, E.E. and WHITING, R.G., 1981. The design of cost-effective air quality monitoring networks. *Environmental Monitoring and Assessment* 1, 1 (1981/03), 59-74. DOI= <http://dx.doi.org/10.1007/bf00836876>.
- [27] QIU, S., CHEN, B., WANG, R., ZHU, Z., WANG, Y., and QIU, X., 2017. Estimating contaminant source in chemical industry park using UAV-based monitoring platform, artificial neural network and atmospheric dispersion simulation. *RSC Advances* 7, 63, 39726-39738. DOI= <http://dx.doi.org/10.1039/C7RA05637K>.
- [28] QIU, S., CHEN, B., ZHU, Z., WANG, Y., and QIU, X., 2016. Source term estimation using air concentration measurements during nuclear accident. *Journal of Radioanalytical and Nuclear Chemistry* 311, 1, 165-178. DOI= <http://dx.doi.org/10.1007/s10967-016-4941-z>.
- [29] SEINFELD, J.H., 1972. Optimal location of pollutant monitoring stations in an airshed. *Atmospheric Environment* 6, 11 (1972/11/01/), 847-858. DOI= [http://dx.doi.org/http://dx.doi.org/10.1016/0004-6981\(72\)90056-X](http://dx.doi.org/http://dx.doi.org/10.1016/0004-6981(72)90056-X).
- [30] SER N ARBELOA, F.J., P REZ CASEIRAS, C., and LATORRE ANDR S, P.M., 1993. Air quality monitoring: Optimization of a network around a hypothetical potash plant in open countryside. *Atmospheric Environment. Part A. General Topics* 27, 5 (1993/04/01/), 729-738. DOI= [http://dx.doi.org/http://dx.doi.org/10.1016/0960-1686\(93\)90190-A](http://dx.doi.org/http://dx.doi.org/10.1016/0960-1686(93)90190-A).
- [31] TRUJILLO-VENTURA, A. and HUGH ELLIS, J., 1991. Multiobjective air pollution monitoring network design. *Atmospheric Environment. Part A. General Topics* 25, 2 (1991/01/01/), 469-479. DOI= [http://dx.doi.org/http://dx.doi.org/10.1016/0960-1686\(91\)90318-2](http://dx.doi.org/http://dx.doi.org/10.1016/0960-1686(91)90318-2).
- [32] VAN EGMOND, N.D. and ONDERDELINDEN, D., 1981. Objective analysis of air pollution monitoring network data; spatial interpolation and network density. *Atmospheric Environment* 15, 6 (1981/01/01/), 1035-1046. DOI= [http://dx.doi.org/http://dx.doi.org/10.1016/0004-6981\(81\)90104-9](http://dx.doi.org/http://dx.doi.org/10.1016/0004-6981(81)90104-9).
- [33] YU, H.-L., KOLOVOS, A., CHRISTAKOS, G., CHEN, J.-C., WARMERDAM, S., and DEV, B., 2007. Interactive spatiotemporal modelling of health systems: the SEKS-GUI framework. *Stochastic Environmental Research and Risk Assessment* 21, 5 (2007/08/01), 555-572. DOI= <http://dx.doi.org/10.1007/s00477-007-0135-0>.