Journal of High Speed Networks 21 (2015) 141–153
DOI 10.3233/JHS-150516
IOS Press

Optimization of air pollutant monitoring stations with constraints using genetic algorithm

Kai Wang [a], Hong Zhao [a,*], Yanan Ding [a], Tao Li [a], Lujian Hou [b] and Fengjuan Sun [b]
[a] *College of Information Technical Science, NanKai University, Tianjin, China*
E-mail: zhaoh@nankai.edu.cn
[b] *Jinan Academy of Environmental Protection Science, Jinan, China*
E-mail: houlujian@vip.sina.com

Abstract. The design of air quality monitoring stations (AQMS) is very important for environmental protection department. The task, however, has been proven to be in the class of nondeterministic polynomial (NP)-hard problem. The powerful search capability of the genetic algorithm (GA) is helpful for selecting optimal monitoring sites. A mathematical programming model based method is proposed, in which environmental, social and economical factors are considered and the GA is used to optimize AQMS. Modelling results indicate that the proposed method outperforms the method of random site of AQMS. The proposed method/framework is also suitable for the design of communication network. For example, the wireless base stations can be well placed by similar method, providing sufficient signal strength for better service with lower cost.
Keywords: Air pollution monitoring sites, genetic algorithm, optimization, constraints

1. Introduction

Air pollution is a worldwide problem and considered to be a critical issue in many countries. The monitoring of pollution level in the atmosphere is of significance, especially to those residents living in the city. Siting air quality monitoring stations reasonably is an important task for environmental protection authorities and department, involving: (1) ensuring that the air quality standard is reached; (2) planning and implementing air quality protection and air pollution control strategies; and (3) preventing or responding quickly to air quality deterioration. Therefore, the environmental protection department need to site air quality monitoring stations effectively.

Different studies of sitting AQMS can be found. Tseng, Chang and Ni-Bin Chang [23] took environmental factors and social factors as target functions, and then used competitive co-evolutionary genetic algorithm to calculate their results. Rachid Abida and Marc Bocquet [1], Olivier Saunier and Marc Bocquet [21] used a simulated annealing algorithm to solve the cost function. The application adopts Boltzmann annealing and geometric cooling schedule. Kanaroglou and Michael Jerrett [12] developed a methodology to optimally site a dense network of AQMS based on the "location–allocation" approach, which offers the flexibility to integrate an assortment of variables into the "demand surface". In particular, his methodology seeks both to identify the areas with the highest spatial variability and to carry out the measurements in areas of specific sociodemographic interest. Ching-Ho Chen and Wei-Lin Liu [6] proposed a similar approach to the development of sustainable air quality monitoring networks. Their procedure, which responds to multi-objective planning, simultaneously considering environmental, social and economic objectives. Abdullah Mofarrah and Tahir Husain [18] presented an objective methodology

*Corresponding author. E-mail: zhaoh@nankai.edu.cn.

for determining the optimum number of ambient AQMS in a monitoring network. Their methodology integrates the multiple-criteria method with the spatial correlation technique. The pollutant concentration and population exposure data are used in their methodology.

The main purpose of this study is to develop a mathematical model for optimizing AQMS by using GA. The environmental, social and economic objectives should be simultaneously considered to select the optimal sites for AQMS.

2. Optimization principles and mathematical model

Monitoring objectives considered in the optimal allocation of AQMS through multi-objective analysis may include: Monitoring stations should be sited (1) in areas of highly populated region; (2) in areas where pollution concentrations are expected to be the highest; (3) in areas where the highest frequency of violation can be detected; (4) in areas where significant economic growth is expected to occur; and (5) near major industrial sources. However, different decision-makers may have different objectives of interest limited by budget. Three objectives considered in this paper consist of environmental, social and economic factors.

A normal scheme designed for the siting of AQMS is to divide a study area into a continuous grid system in which each grid represents a candidate site in AQMS. Hence, an optimization analysis, using mathematical programming models as screening tools, is designed to pick up the most appropriate subset within the proposed grid system in the sense that the design objectives can be optimized with respect to the budget, coverage effectiveness, spatial correlation and/or concentration differentiation constraints. A grid defined in the study area is thus equivalent to a candidate site waiting for the possible selection in the multi-objective trade-off process. The objective function and constraint set are to be sequentially defined in a great detail below.

If the area being studied is divided into a number of continuous grids with equal size, then the selection of AQMS can be made in the designated grid system. The objective function prepared to achieve such planning goals can be stated as below.

2.1. Environmental factor

The monitoring data of the optimal sites of AMQS should be able to reflect the maximum detection capability of the highest pollution concentrations in the study area. On one hand, by using data from optimal monitoring sites in spatial interpolation, the bias between information are related to the spatial distribution of pollutant concentration and the actual distribution should be minimized. On the other hand, utilizing the optimal sites of AMQS can detect the over-standard pollutant concentration, in other words, the optimal sites of AMQS should be sited in place with higher pollution concentration. Overall, the objective function on the environmental factor prepared to achieve such planning goals can be stated as below,

$$E_{\text{environment}} = \sum_{p \in S-V} \sum_{t \in T} w\big(i(p,t) - m(p,t)\big) \cdot \big|i(p,t) - m(p,t)\big|, \tag{1}$$

where S is the set of all candidate sites; V is the set of the optimal sites; T is the set of all time period; $m(x,t)$ is the concentration of pollutant that is simulated by the U.S. EPA's Community Multiscale Air Quality (CMAQ) [3] modeling system in grid x at t time period; $i(x,t)$ is the estimated concentration of pollutant that is obtained by spatial interpolation method in grid x at t time period. Here inverse distance interpolation is used, as shown in Eq. (2),

$$i(p,t) = \begin{cases} m(p,t) & \text{if } p \in V, \\ \dfrac{\sum_{\|s-p\|<r} \frac{m(s,t)}{\|s-p\|}}{\sum_{\|s-p\|<r} \frac{1}{\|s-p\|}} & \text{otherwise.} \end{cases} \tag{2}$$

$w(x)$ is defined as follows:

$$w(x) = \begin{cases} w_1 & \text{if } x \geqslant 0, \\ w_2 & \text{otherwise,} \end{cases} \tag{3}$$

where $w(x)$ is a weighting coefficient, which specify the scale of both environmental principles that minimize bias of concentration and the over-standard concentration.

When $w_1 = 1$, it indicates that the bias can be so negligible in the case that the predicted concentration is greater than or equal to the actual concentration. That is, the principle of the over-standard concentration has a higher priority. When $w_1 = w_2$, it indicates that the bias between the interpolated concentration and the actual concentration is treated equally in any case. That is, only using the principle of minimum bias of concentration without taking into account the principle of the over-standard concentration. When setting the value of w_1 and w_2 in practice, it should be $w_1 < w_2$. That is, the bias has a small impact on the optimal sites of AMQS in the case that the predicted concentration is greater than or equal to the actual concentration. Otherwise, the bias has a great impact on the optimal sites of AMQS.

2.2. *Social factor*

The optimal sites of AMQS should cover highest number of population. The distribution of sensitive receptors is difficult to obtain; hence, only population information is used for designing the optimal goal.

The objective function considering both the environmental and social factors can be stated as below,

$$\sum_{p \in S-V} \sum_{t \in T} \mathrm{pop}(p) \cdot w\big(i(p,t) - m(p,t)\big) \cdot \big|i(p,t) - m(p,t)\big|, \tag{4}$$

where $\mathrm{pop}(x)$ is the population in the grid x.

The optimization objective is used to focus on the densely populated area in distribution of optimal sites. More monitoring sites should be set in the region with a large population. At the same time, through introducing population information in the optimization objective, population exposure can be better appraised through optimal distribution of sites.

2.3. *Design of constraints*

According to the technical regulation for selection of ambient air quality monitoring stations (on trial) from Ministry of Environmental Protection of the People's Republic of China [17], sites should be set under following conditions.

(1) The arithmetic mean of $PM_{2.5}$ concentration in all air quality assessment sites in a certain region should represent the overall average value of the $PM_{2.5}$ concentration in this region. The regional total average value can be calculated with arithmetic mean of simulation in the region as its estimate, and the relative margin of error between the average pollutant concentrations at all select air quality appraisement sites during the same period of time and the estimate should be no more than 10%. The mathematical form of the constraint is shown in Eq. (5),

$$-10\% \leqslant \frac{\sum_{v \in V} \sum_{t \in T} \frac{m(v,t)}{|T| \cdot |V|} - \sum_{s \in S} \sum_{t \in T} \frac{m(s,t)}{|T| \cdot |S|}}{\sum_{s \in S} \sum_{t \in T} \frac{m(s,t)}{|T| \cdot |S|}} \leqslant 10\%, \tag{5}$$

where $|T|$, $|V|$, $|S|$ respectively refers to the number of elements in the set T, V, S.

(2) The arithmetic mean of regional simulation is used as the regional overall mean to calculate the estimates at 30th, 50th, 80th and 90th percentiles which are compared with the 30th, 50th, 70th and 90th percentiles calculated with average pollutant concentration at all select air quality assessment sites during the same period of time. The relative margin of error between these two should be no more than 15%. The mathematical form of the constraint is shown in Eq. (6),

$$-15\% \leqslant \frac{\text{Percent}(V, per) - \text{Percent}(S, per)}{\text{Percent}(S, per)} \leqslant 15\%, \tag{6}$$

where Percent(X, per) denotes per percentiles in the set X, which takes values at 30th, 50th, 80th, 90th.

(3) Each city or district should have at least one monitoring site. The mathematical form of the constraint is shown in Eq. (7),

$$\sum_{s \in S_i} \text{index}(s) \geqslant 1, \quad i = 1, 2, \ldots, M, \tag{7}$$

where M is the number of cities; S_i is the gird set in the ith city, and if one gird consists of regions from different cities, it will be considered to belong to the city of the biggest region; index(s) is an indicator function, of which calculation method is shown in Eq. (8),

$$\text{index}(s) = \begin{cases} 1 & \text{if } s \in V, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

(4) According to the economic factors, total optimization cost should be within the reach of budget. The mathematical form of the constraint is shown in Eq. (9),

$$\sum_{s \in V} c(s) \leqslant C, \tag{9}$$

where $c(s)$ is the cost for building and maintaining a monitoring site in grid s, and C refers to the total budget. Each monitoring site is laid out under similar conditions, so the cost of every monitoring site can be likely denoted as *cost*. In this sense, the constraint can be rewritten as shown in Eq. (10),

$$\sum_{s \in S} \text{index}(s) \leqslant \frac{C}{\cos t}. \tag{10}$$

According to the above analysis, the optimization model for AQMS is developed in this study. If the area being studied is divided into N grids in CMAQ model, the length of time considered for the optimization of AQMS is T, the total budget for the installation and maintenance of monitoring sites is C. The mathematical programming model designed in this study is shown below:

obj.

$$\min \sum_{p=1}^{N} \sum_{t=1}^{L} \text{pop}(p) \cdot w\big(i(p,t) - m(p,t)\big) \cdot \big|i(p,t) - m(p,t)\big| \cdot \big(1 - x(p)\big) \tag{11}$$

s.t.

$$\begin{cases} -10\% \leqslant \dfrac{\sum_{v\in V}\sum_{t\in T}\frac{m(v,t)}{|T|\cdot|V|} - \sum_{s\in S}\sum_{t\in T}\frac{m(s,t)}{|T|\cdot|S|}}{\sum_{s\in S}\sum_{t\in T}\frac{m(s,t)}{|T|\cdot|S|}} \leqslant 10\%, \\ -15\% \leqslant \dfrac{\text{Percent}(V, per) - \text{Percent}(S, per)}{\text{Percent}(S, per)} \leqslant 15\%, \\ \sum\limits_{s\in S_i} \text{index}(s) \geqslant 1, \quad i = 1, 2, \ldots, M, \\ \sum\limits_{s\in S} \text{index}(s) \leqslant \dfrac{C}{\cos t}. \end{cases} \tag{12}$$

3. GA based siting AQMS scheme

GA is an adaptive and robust optimization and search technique which borrows the ideas of natural selection and "survival of the fittest" from natural evolution [8]. In GA, the potential solution of a problem is encoded into a form that is analogue to the chromosomes of biological systems. The simulation process begins with an initial population of random chromosomes, which iteratively evolves over multiple reproductive cycles (called generations or iterations) through the application of three probabilistic genetic operators: selection, crossover and mutation. The standard GA introduced by Holland [10], which employs binary chromosomes and binary genetic operators, was chosen for the studied cases in this paper.

3.1. Chromosome representation

The encoding scheme of chromosomes has a major impact on the performance because it can severely limit the search space observed by the system. For the proposed GA siting AQMS scheme where the variable components are in real space, a binary encoding scheme is used in order to move the representation closer to the problem domain. A chromosome in this study corresponds to an optimization siting scheme, the chromosomal structure is depicted in Table 1.

The study denotes a 2-dimensional $h = [h^1, h^2]$ genetic vector as a vector of an optimal site, where each component corresponds to X or Y coordinate of the grid with binary form. The genetic structure is depicted in Table 2.

3.2. Initialization

The population of binary-coded chromosomes $\{h_q = [h_q^1, h_q^2], q = 1, \ldots, Q\}$ is initialized by employing a random approaches, where Q is known as the "population size". The purpose of using a random generation is to distribute the initial trial solutions intelligently and maintains the diversity. To achieve this, the initialization procedure produces the random genes as each variable h_q^1 and h_q^2 equals to one integer value within the range of [1, 100].

Table 1

Chromosomal structure

The number of Gene	Gene 1	Gene 2	…	Gene N

Table 2

Genetic structure

X coordinates	Y coordinates

3.3. Fitness evaluation

By convention, the fitness function should be a positive value. Equation (4) provides the mechanism for evaluating the fitness of each chromosome and, therefore, serves as the fitness function of the proposed GA approach. As the aim is to optimizing siting AQMS, then the lowest value in (4) corresponds to the best chromosome.

3.4. Genetic operators

Based on the above define fitness function, three basic types of genetic operators are required to modify the population: selection, crossover and mutation. Selection is a process used for choosing parent chromosomes to participate in reproduction for the next generation, and among many selection schemes available, the roulette wheel sampling scheme [9] is used.

Crossover is a crucial operator that combines two or more parent chromosomes to produce new offspring chromosomes. A suitably designed crossover can significantly accelerate the search process [14]. The proposed GA approach adopts a single point crossover.

A partially mapped crossover operators are explained below, where parent-1 and parent-2 denote parent chromosomes while child-1 and child-2 denote the created ones.

Referring to Fig. 1, each parent is randomly divided into three sections (head, middle and tail). A new offspring (e.g., child-1) is created by the following procedure. The middle of child-1 is created by referring to that of parent-2, which is a string (1, 8, 2, 7). Both the head and tail of child-1 are created by referring to parent-1. If the gene values in the head/tail sections of parent-1 do not appear in child-1, we copy them in the exact positions of child-1 (e.g., gene values 6 and 3). Finally, for those vacant genes in child-1, we place their values by sequentially referring to the unassigned genes in parent-1. This yields child $-1 = (9, 6, 3, 1, 8, 2, 7, 4, 5)$.

The mutation operator randomly alters some values in a chromosome with a probability determined by the mutation rate p_m. This can result in entirely new offspring chromosomes. Mutation is used very sparingly in most GAs. Typically, the mutation rate p_m is generally less than 0.1 [9]. We consider multi nonuniform mutation [14], which is a dynamic (population dependent) mutation operator aimed at improving single-element tuning and reducing the drawback of random mutation.

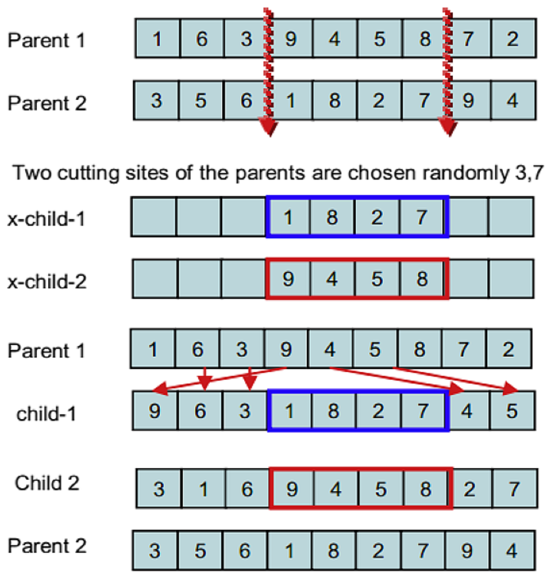

Fig. 1. A partially mapped crossover operator. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/JHS-150516.)

3.5. Replacement

After a predefined number of offspring has been produced through the above genetic operators, a replacement strategy is required in order to modify the old population with the new generation [14]. An elitist strategy is also used to improve algorithm performance [10], which appends the best performing chromosome of a previous generation to the current population and thus ensures that the chromosome with the best fitness value always survives to the next generation.

4. Modelling results

4.1. Area and period of study

This study was conducted by using the data of Shandong province, China. The inner grid of computational modeling domain covers all of Shandong province with 88 × 88 horizontal grid cells at 9 km × 9 km spacing, using the Lambert conformal map projection centered at (34.9°N, 113.1°E), including 25 vertical layers with the top at 74 hPa (∼20.1 km), and the first layer is centered at ∼13 m. The outer grid covers all of China with 70 × 58 horizontal grid cells at 27 km × 27 km spacing, which is used to provide the pollutions concentration initial and boundary conditions for inner grid. For all simulations, the meteorological initial and boundary conditions were constructed from the National Centers for Environmental Prediction Final Global Analyses data, available every 6 h at 1° × 1° spatial resolution. The $PM_{2.5}$ concentration in January, April, July and October of 2012 were simulated respectively.

4.2. Data generation

Distribution optimization is based on the study of regional air quality. Generally, data needed for analysis for regional air quality comes from three sources: regional historical monitoring data, data simulated by model, combination of historical monitoring and simulated data. Simulated data, gained from air quality model, is spatially and temporally more accurate and economical. Common air quality models include: ISC3 [2,6,13,18,23], ADMS [4], AERMOD [4] and Model-3/CAMQ [5,7] etc. The spatiotemporal range, resolution, list of pollution sources and meteorological data of the simulated region must be considered in selection of air quality model.

The simulation of air quality must be able to identify the features of complex atmospheric pollution, changing trend and the spatiotemporal distribution of pollutants so that it can reflect the whole process of regional air pollution.

The U.S. EPA's Community Multi-scale Air Quality (CMAQ) modeling system [3] has been widely used in scientific research fields and business forecast fields around the world. The CMAQ was adopted in this study to simulate the temporal-spatial distributions of $PM_{2.5}$ concentration, and the environmental monitoring data of monitoring sites which were used to verify the simulation results. The Weather Research and Forecasting (WRF) model-Advanced Research WRF (WRF-ARW) were used to simulate meteorological inputs for CMAQ [22]. The Sparse Matrix Operator Kernel Emission System (SMOKE) [11] model was used to compute the emissions for CMAQ from the annual emissions inventory based on empirical formula, and area-, mobile-, point-, and biogenic-source emissions were processed by SMOKE in this study. The processing flow is shown in Fig. 2: firstly, data on orographic features and meteorological grids is input to WRF model to get meteorological data needed by CAMQ model; secondly, the list of pollutants emission is processed by SMOKE model to gain source intensity data needed by CAMQ model; finally data on spatiotemporal distribution of $PM_{2.5}$ concentration is received from CAMQ.

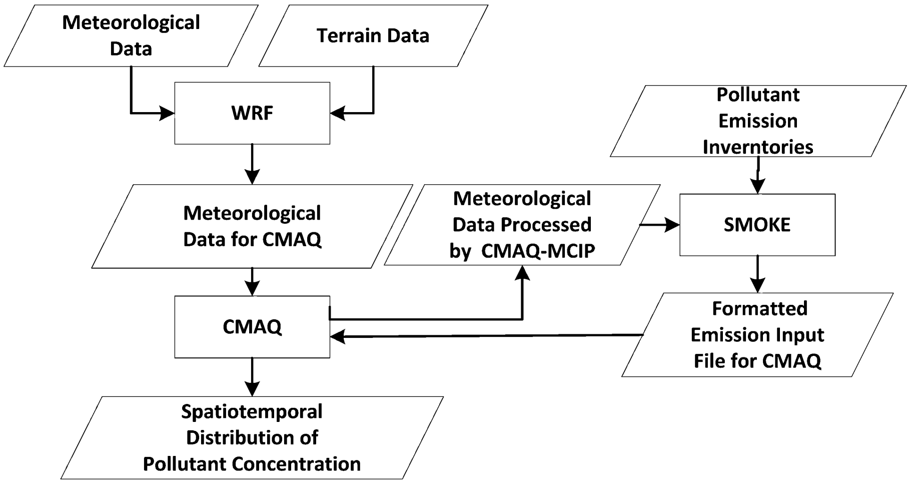

Fig. 2. Models-3/CMAQ processing flow.

4.3. Appraisement of Models-3/CMAQ simulation results

The appraisement of air quality model's simulation results is conducted mainly through quantitative and comparative analysis of simulation and observation data. Quantitative comparison referring to statistical analysis of the magnitude of bias between simulated and observed values is to assess reliability of simulation results. Nowadays, the most common appraisement indexes of model simulation results include Mean Bias (MB), Normalized Mean Bias (NMB), Normalized Mean Error (NME) and Root Mean Square Error (RMSE), which are shown in Eqs (13)–(16),

$$MB = \frac{1}{|T|}\sum_{t\in T}(C_m - C_o), \tag{13}$$

$$NMB = \frac{\sum_{t\in T}(C_m - C_o)}{\sum_{t\in T} C_o}, \tag{14}$$

$$NME = \frac{\sum_{t\in T}|C_m - C_o|}{\sum_{t\in T} C_o}, \tag{15}$$

$$RMSE = \sqrt{\frac{1}{|T|}\sum_{t\in T}(C_m - C_o)^2}, \tag{16}$$

where C_m and C_o denote respectively simulated and observed values, T refers to the set of all time, and $|T|$ indicates the number of elements in the set T.

The *MB* and *RMSE* reflect the magnitude of bias and error between simulated and observed values, and the *NMB* as well as *RMSE* indicate the magnitude of relative bias and error between simulation and observed values. If values of both *NMB* and *NME* are less than 50%, simulation results will be considered acceptable [24].

Since January 2012, $PM_{2.5}$ automatic monitoring devices using β-ray method have been set in Quancheng Square of Jinan (inland capital city); $PM_{2.5}$ automatic monitoring data from monitoring site in Quancheng Square can be used to calibrate the simulation results of Models-3/CAMQ model.

The procedure is: compare $PM_{2.5}$ monitoring data from monitoring site in Quancheng Square in January, April, July, 2012 with $PM_{2.5}$ concentrations in their corresponding grids in Models-3/CAMQ model simulation; then calculate bias and error according to equations from (13) to (16). The calculation results are shown in Table 3.

Table 3
Verification of simulated data from monitoring site in Quancheng Square

Month	*MB*	*NMB*	*NME*	*RMSE*
January	−0.005	−0.04	0.35	0.205
April	0.031	0.421	0.552	0.202
July	−0.223	0.275	0.174	0.034

Table 4
Individual Air Quality Index standards of $PM_{2.5}$

24 hour average $PM_{2.5}$ concentrations range (mg/m^3)	Individual Air Quality Index (AQI) value	AQI levels	Individual Air Quality Index (AQI) values	Health implications
0–35	0–50	Level 1	Good	Air quality is considered satisfactory, and air pollution poses little or no risk.
35–75	51–100	Level 2	Moderate	Air quality is acceptable, but for some pollutants, there may be a moderate health concern for a very small number of people. People who are unusually sensitive to ozone may experience respiratory symptoms.
75–115	101–150	Level 3	Unhealthy for sensitive groups	Although the general public is unlikely to be affected in this AQI range, people with lung disease, older adults and children are at a greater risk from exposure to ozone; those with heart and lung disease, older adults and children are at greater risk from airborne particles.
115–150	151–200	Level 4	Unhealthy	Everyone may begin to experience adverse health effects, and members of the aforesaid sensitive groups may experience more serious effects.
150–250	201–300	Level 5	Very unhealthy	Everyone may begin to experience adverse health effects, and members of the aforesaid sensitive groups may experience more serious effects.
Above 250	>300	Level 6	Hazardous	This would trigger a health warning of emergency conditions, and the entire population is more likely to be affected.

According to the fact that in error analysis of simulation results at this monitoring site, both *NMB* and *NME* values were below 50%, so that the Models-3/CAMQ simulation data reflected the concentration of particles in a reasonably accurate way. Therefore, a conclusion can be made that distribution optimization of monitoring site can be feasible conducted on basis of the output results of Models-3/CAMQ model.

4.4. Method and results

According to the "Ambient Air Quality Standards" (GB3095-2012) [15] and Ambient Air Quality Index (AQI) Technical Requirements (HJ633-2012, on trial) [16], $PM_{2.5}$ concentration can be divided into 6 levels, which is shown in Table 4. The experiments were designed as the following steps.

Firstly, the CMAQ modeling system produces temporal-spatial distribution of $PM_{2.5}$ concentration. The spatial distribution of $PM_{2.5}$ concentration at 10 o'clock on October 25th, 2011 is shown in Fig. 3.

Secondly, according to optimization goals, the optimization model is solved by using genetic algorithm, the red dots in Figs 4 and 5 indicate the optimal sites that are selected in the case of $w_1 = 1$ and $w_1 = 0$. Finally, the data of $PM_{2.5}$ concentration on the optimal sites of AMQS is verified with spatial interpolation method.

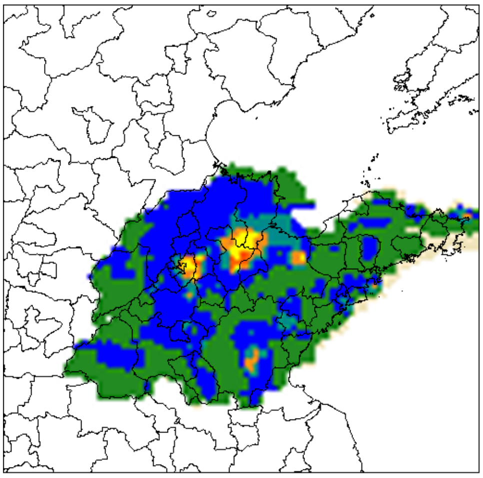

Fig. 3. Outputs of CMAQ model. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/JHS-150516.)

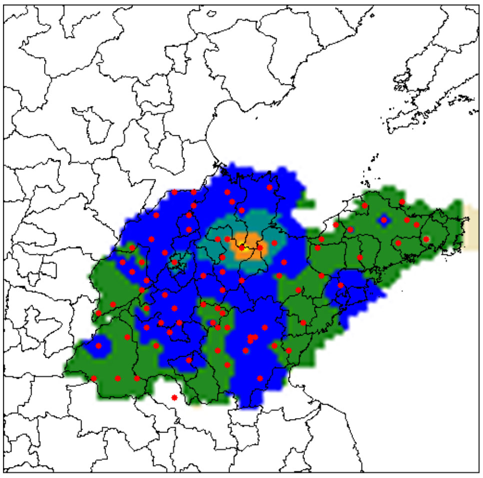

Fig. 4. Optimal sites of AQMS ($w_1 = 1$). (The colors are visible in the online version of the article; http://dx.doi.org/10.3233/JHS-150516.)

The comparison of results of random siting of AQMS and the optimal siting of AQMS are presented in Table 5. Figure 6 is the graphical representation of the data in Table 5. Obviously, with regard to meet the requirement of the same accuracy rate, using the method of optimal siting of AQMS can reduce the number of AQMS significantly. If the over-standard detection accuracy rate is request to achieve over 90%, using the method of optimal sites of AQMS will need to site a hundred AQMS, however, using the method of random sites of AQMS will only need to site eighty AQMS.

5. Conclusions

The AQMS represents an essential tool to monitor and control atmospheric pollution. The use of some specific criteria in conjunction with the mathematical models provides a general approach to select the optimal sites of AQMS. In this study, a new GA-based siting AQMS scheme was proposed. The scheme simultaneously considered the environmental, social and economic objectives. The optimal site task has been proved to be in the class of

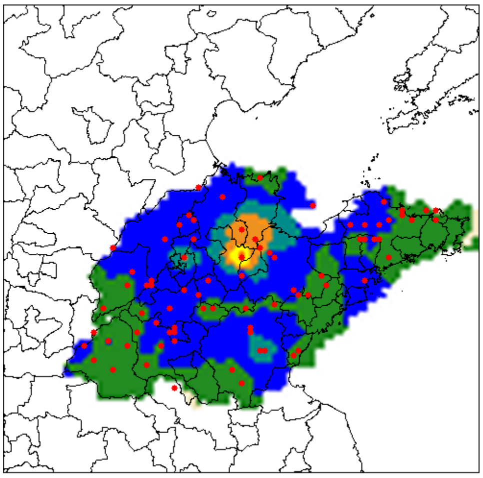

Fig. 5. Optimal sites of AQMS ($w_1 = 0$). (The colors are visible in the online version of the article; http://dx.doi.org/10.3233/JHS-150516.)

Table 5
Random sites of AQMS vs. optimal sites of AQMS

The number of AQMS	The accuracy rate whether over-standard detection		The accuracy rate of the grade detection		The accuracy rate of the grade (plus or minus 1) detection	
	Random siting of AQMS	Optimal siting of AQMS	Random siting of AQMS	Optimal siting of AQMS	Random siting of AQMS	Optimal siting of AQMS
10	80.01	82.92	46.54	50.78	82.57	86.47
20	83.63	85.66	53.42	55.53	87.52	90.75
30	85.70	87.69	57.00	59.86	90.15	92.57
40	86.66	88.42	58.70	62.14	91.42	93.33
60	88.48	89.50	62.14	64.62	93.39	94.30
80	89.29	90.34	64.56	66.49	94.02	95.21
100	89.81	90.85	65.77	67.78	94.77	95.75
150	91.19	91.57	68.85	69.93	95.89	96.47

NP-hard problem. The proposed GA approach carried out a global search by manipulating and maintaining a population of candidate solutions to find the optimal site. Simulation results demonstrate that the GA-based siting scheme outperforms random siting of AQMS. In further studies, more attention will be paid on a combination strategy of GA and other local search algorithm.

Furthermore, the proposed method/framework is also suitable for the design of communication network. For example, the wireless base stations can be well placed by similar method, providing sufficient signal strength for better service with lower cost [19,20]. Or, more specifically, the following aspects of our work can be referred for the task of communication network design:

(1) Three problems should be considered for the study of station optimization: (a) What data are needed and how to obtain them? (b) What principles and constraints should be considered and how to convert them to mathematical model? (c) What algorithm can calculate approximate optimum solution within an acceptable time?
(2) The method used in this paper, converting the principles and constraints to the mathematical programming model, can be referred during the modelling of communication network optimization.
(3) GA method used in this paper, including the encoding scheme, genetic operators, etc., can be directly applied to solve the mathematical model for the application of communication network design.

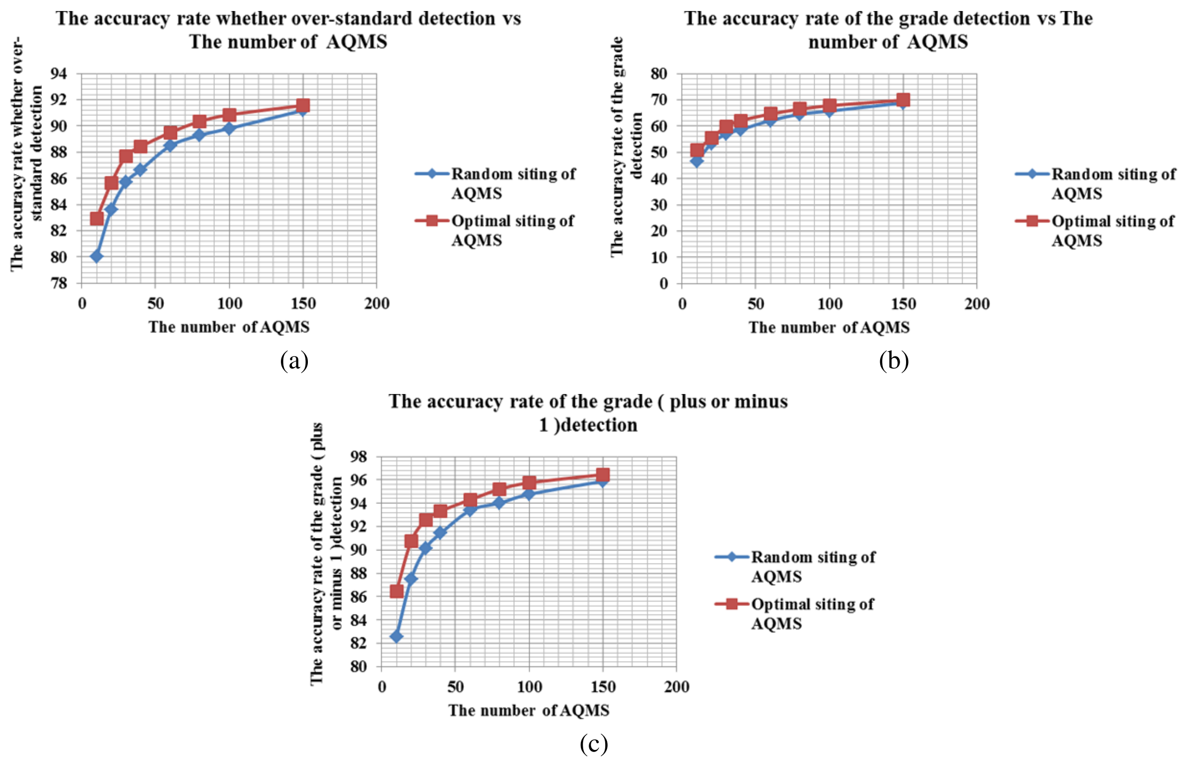

Fig. 6. Graphical representation of the data in Table 5: (a) The accuracy rate whether over-standard detection vs. The number of AQMS. (b) The accuracy rate of the grade detection vs. The number of AQMS. (c) The accuracy rate of the grade (plus or minus 1) detection vs. The number of AQMS. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/JHS-150516.)

Acknowledgements

This work was partly supported by the Public Welfare Scientific Research Project sponsored by Ministry of Environmental Protection of PRC under Grant No. 2010467007; the National Natural Science Foundation of China under Grant No. 61201424 and 61212005; the Natural Science Foundation of Tianjin, China under Grant No. 12JCYBJC10100.

References

[1] R. Abida and M. Bocquet, Design of a monitoring network over France in case of a radiological accidental release, *Atmospheric Environment* **42** (2008), 5205–5219.

[2] M.S. Al-Rashidi, V. Nassehi and R.J. Wakeman, Investigation of the efficiency of existing air pollution monitoring sites in the state of Kuwait, *Environmental Pollution* **138** (2005), 219–229.

[3] D.W. Byun and K.L. Schere, Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system, *Appl. Mech. Rev.* **59** (2006), 51–77.

[4] D.J. Carruthers, M.D. Seaton, C.A. McHugh, X. Sheng, E. Solazzo and E. Vanvyve, Comparison of the complex terrain algorithms incorporated into two commonly used local-scale air pollution dispersion models (ADMS and AERMOD) using a hybrid model, *Journal of the Air & Waste Management Association* **61**(11) (2011), 1227–1235.

[5] W. Chea, J. Zheng, S. Wang, L. Zhong and A. Lau, Assessment of motor vehicle emission control policies using Model-3/CMAQ model for the Pearl River Delta region, China, *Atmospheric Environment* **45**(9) (2011), 1740–1751.

[6] C.-H. Chen and W.-L. Liu, Development of a multiple objective planning theory and system for sustainable air quality monitoring networks, *Science of the Total Environment* **354** (2006), 1–19.

[7] L. Chen, H. Zhao, B. Han and Z. Bai, Combined use of WEPS and Models-3/CMAQ for simulating wind erosion source emission and its environmental impact, *Science of the Total Environment* **466**(467) (2014), 762–769.
[8] L. Davis, *Handbook of Genetic Algorithms*, Van Nostrand, New York, 1991.
[9] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
[10] J.H. Holland, *Adaptation in Natural and Artificial Systems*, 2nd edn, Univ. Michigan Press, Ann Arbor, MI, 1992.
[11] IE (UNC Institute for the Environment), SMOKE version 2.7 user's manual, 2010, available at: https://www.cmascenter.org/help/documentation.cfm?MODEL=smoke&VERSION=2.7.
[12] P.S. Kanaroglou and M. Jerrett, Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location–allocation approach, *Atmospheric Environment* **39** (2005), 2399–2409.
[13] N.A. Mazzeo and L.E. Venegas, Design of an air-quality surveillance system for Buenos Aires city integrated by a NO_x monitoring network and atmospheric dispersion models, *Environ. Model. Assess.* **13** (2008), 349–356.
[14] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlin, Germany, 1992.
[15] Ministry of Environmental Protection of the People's Republic of China, Ambient air quality standards, February 2012.
[16] Ministry of Environmental Protection of the People's Republic of China, Technical regulation on ambient air quality index (on trial), February 2012.
[17] Ministry of Environmental Protection of the People's Republic of China, Technical regulation for selection of ambient air quality monitoring stations (on trial), September 2013.
[18] A. Mofarrah and T. Husain, A holistic approach for optimal design of air quality monitoring network expansion in an urban area, *Atmospheric Environment* **44** (2010), 432–440.
[19] T. Oda, A. Barolli, E. Spaho, L. Barolli, F. Xhafa and M. Younas, Effects of population size for location-aware node placement in WMNs: Evaluation by a genetic algorithm-based approach, *Personal and Ubiquitous Computing* **18**(2) (2014), 261–269.
[20] S. Sakamoto, E. Kulla, T. Oda, M. Ikeda, L. Barolli and F. Xhafa, A comparison study of hill climbing, simulated annealing and genetic algorithm for node placement problem in WMNs, *Journal of High Speed Networks* **20**(1) (2014), 55–66.
[21] O. Saunier and M. Bocquet, Model reduction via principal component truncation for the optimal design of atmospheric monitoring networks, *Atmospheric Environment* **43** (2009), 4940–4950.
[22] W.C. Skamarock, J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, W. Wang and J.G. Powers, A description of the advanced research WRF version 2, NCAR Technical Note NCAR/TN 468+STR, 2005.
[23] C.C. Tseng and N.-B. Chang, *Assessing Relocation Strategies of Urban Air Quality Monitoring Stations by GA-Based Compromise Programming*, Vol. 26, Environmental International, 2001, pp. 523–541.
[24] J. Zhang, A performance evaluation study of Models-3/CMAQ model for simulating PRD regional air quality, Master's degree thesis, South China University of Technology, 2010.