16th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
8-11 September 2014, Varna, Bulgaria

AIR MONITORING NETWORK OPTIMIZATION METHOD USING CHEMICAL TRANSPORT MODEL AND METAHEURISTICS

Shin Araki[1, 2], Hikari Shimadera[3], Kouhei Yamamoto[4], Akira Kondo[2]

[1]Otsu City Public Health Center, Otsu, Japan
[2]Graduate School of Engineering, Osaka University, Suita, Japan
[3]Center for Environmental Innovation Design for Sustainability, Osaka University, Suita, Japan
[4]Graduate School of Energy Science, Kyoto University, Kyoto, Japan

Abstract: The air monitoring network design is a critical issue because the monitoring stations should be allocated properly so that they adequately represent the concentrations in the domain of interest. Previous studies have developed methods based on observations obtained from existing networks with a considerable number of stations. These methods, however, are difficult to apply to a sparse network or a network under development. In this paper, a unique method is presented using metaheuristics integrated with a chemical transport model. The WRF/CMAQ simulated concentrations are provided to metaheuristics including genetic algorithm (GA) and simulated annealing (SA). In addition, the hybrid algorithm that combines the GA and SA is developed to make best use of the two algorithms. The performances of these algorithms are compared to each other by the application to the Kinki region in Japan and the best combination with WRF/CMAQ simulation is examined. The HGS best performs among others in terms of minimum, mean and standard deviation of the quality measure. The weighing factor based on concentrations is successfully implemented so that more stations are placed in higher concentration areas. These results demonstrate the effectiveness of the proposed method for the air monitoring network optimization.

Key words: *$PM_{2.5}$; WRF/CMAQ; Genetic algorithm; Simulated annealing; Japan*

INTRODUCTION

Air monitoring networks have been developed in various areas in the world for the environmental, epidemiological, policy evaluation and/or emission surveillance purposes by national or local governments. These networks should be allocated properly so that they adequately represent the concentrations in the domain of interest to accomplish the purposes. This issue is often referred to as a network design problem and has been widely discussed. The network design problem generally consists of design criterion, concentration estimation method and an algorithm for the selection of a subset from potential locations (Wu et al., 2011). A criterion is often defined with geostatistical estimation method which is called kriging, where the theoretical extrapolation error (kriging variance) is minimized (e.g., Wu et al., 2011). Once the criterion is defined, the network design problem is a combinatorial optimization problem. When a network is small enough, complete enumeration of all combinations could be possible. For a large network, however, this will run into a combinatorial explosion. To deal with this difficulty, metaheuristics have been applied to the optimization of large networks (e.g., Ruis-Cardenas et al., 2010; Wu et al., 2011). Previous studies including above mentioned ones apply metaheuristics to observations obtained from existing monitoring networks with a considerable number of stations. These methods assume that networks represent the spatial distribution sufficiently, thus are often applied to relatively dense networks where the efficiency of the network i.e., reduction of stations is focused. Therefore, these methods are difficult to apply to a sparse network or a network under development that insufficiently represents the spatial distribution of the pollutant of interest. Although the optimization of such networks is necessary as well, no procedure that can be applied to such networks has been proposed so far to our best knowledge.

Chemical transport models simulate physical and chemical processes including emission, advection, photochemical reactions and deposition. These models have been extensively used at various ranges of

spatial and temporal scale, not only to obtain the spatial distribution, but also to establish an effective strategy for the control of the concentrations of air pollutants. Thus, simulated concentrations by chemical transport models with sufficiently high spatial resolution and long term calculation can be an alternative to observations for the network optimization. Given that the search space is huge and continuous when using simulated fields, the superior ability both for a global and local search is required for the search algorithm. The GA is able to search in a large space, but often not able to find the local optimal solution (Ruiz et. al., 2010). On the other hand, the SA algorithm is able to find locally optimal solutions, but is often trapped in regions far from the global optimum (Ruiz et. al., 2010). Therefore, for the optimization using the simulated concentrations, the hybrid algorithm that combines the GA and SA should be considered to make best use for the GA and SA and to make up for the limitation of the two algorithms.

In this paper, a unique method is proposed using metaheuristics integrated with a chemical transport model that can be applied to the optimization of a network with limited number of stations or a network under development. As many previous studies have optimized air monitoring networks by minimizing mean kriging variance, it is also adopted to the algorithms in this paper. In addition to the GA and SA, the hybrid GA and SA (HGS) algorithm is developed and applied to the monitoring network of $PM_{2.5}$ in the Kinki region in Japan. The performances of these three algorithms are compared to each other in terms of the quality of the solutions in order to reveal the best combination with model simulated values.

CHEMICAL TRANSPORT MODEL

The Community Multiscale Air Quality model (CMAQ) (Byun and Ching, 1999) version 5.0.1 is driven with the Weather Research and Forecasting model (WRF) (Skamarock et al., 2009) version 3.5.1 with the same physics options as those used by Shimadera et al. (2014). Emission data and other settings involved in the simulations and CMAQ configurations are detailed in Shimadera et al. (2014).

OPTIMIZATION ALGORIGHMS

The concentrations simulated by WRF/CMAQ are mean concentrations in grid cell. These simulated values are treated as point values at the grid cell centroids. The points that are on land areas are the candidate points for the optimized network. The network configuration is expressed as a binary data, a vector consisting of zeros and ones with the length of the candidate points, where ones represent gauged sites and zeroes represent ungauged sites. The position of the binary string corresponds to that of the candidate points. Empirical variogram is computed from all the candidate points and theoretical variogram is fitted to it. Mean kriging variance, which is the fitness value of the selected combination of points, is then calculated with this fitted variogram and the simulated concentrations at the selected points. When calculating mean kriging variance, sequentially selected 15 points that on the outer side of the study area are included in order not to overestimate the variance at the boundary area.

Genetic algorithm

The basic scheme of the GA algorithm in this paper is similar to the standard GA which consists of population, selection, crossover and mutation. In the crossover, gauged sites in both parents will be gauged and ungauged sites in both parents will be ungauged in the offspring. Half of the sites that are gauged in one of the parents and not gauged in the other will be randomly conversed. The probability of crossover occurrence is set to 0.9. In the mutation, 5% of gauged sites are randomly selected and swapped with the same number of randomly selected ungauged sites with a probability of 0.2. The solutions with the best 5% fitness remain in the next generation. These operations stop when the total number of generations reaches 400, or when the number of generations without any improvement reaches 100.

Simulated annealing

The SA algorithm is an iterative search algorithm which starts with a randomly selected initial solution. A new potential solution is created by exchanging a randomly selected gauged site for a randomly selected ungauged site in the pre-defined search window of the selected gauged site. The fitness is then computed for the new potential solution and compared to that of the current solution. The new potential solution is accepted when the fitness has improved. Even when the fitness has degraded, the new potential solution is accepted with a certain probability. The size of the search window is fixed in the first 500 iterations at 7×7 cells of the original grid. Then it is decreased after 500 iterations down to a 5×5 window, and a 3×3

from 1 000 iterations and onward. The total number of iterations for each run is set to 10 000. The algorithm also stops when the number of iterations with no progress reaches 500.

Hybrid algorithm

The basic scheme of the HGS is similar to the GA described above. The SA algorithm with 30 iterations is applied to the solutions in the population after 50 generations. After 50 generations, the crossover and mutation probability is set to 0.3 and 0 respectively. The SA is only applied to the solutions with the best 15% fitness. The size of the search window is decreased according to the number of generations; 7 × 7 during the first 25 generations with the SA, then 5 × 5 during the next 25 generations, and finally 3 × 3 until the end. The total number of iterations is set to 300, but the algorithm also stops when the number of generations with no progress reaches 10 after the SA is applied.

APPLICATION TO $PM_{2.5}$ MONITORING NETWORK

Study Area and setups

The proposed method is applied to the $PM_{2.5}$ monitoring network in Kinki region in Japan (134.2°E – 136.5°E, 33.4°N – 35.8°N) where megacities such as Osaka, Kyoto and Kobe are located. The air quality standard of $PM_{2.5}$ in Japan was set in the year 2009 and the number of monitoring stations started to increase rapidly, for instance, from 57 stations in the study area in the year 2011 to 116 in the year 2013, but still remains much fewer than those of the other air pollutants such as NO_2 with 216 stations. In addition, many of the $PM_{2.5}$ monitoring stations are placed in populated areas, while a few stations in less populated or mountain areas.

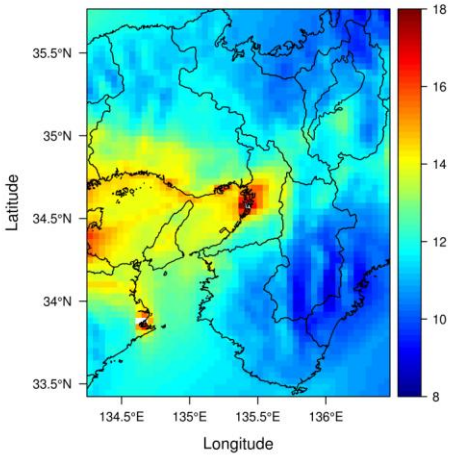

Figure 1. The spatial distribution of the PM2.5 annual mean obtained by WRF/CMAQ simulation for the Japanese fiscal year 2010. Unit is μg m^{-3}.

WRF/CMAQ model is run from April 2010 to March 2011 (Japanese fiscal year 2010) with an initial spin-up period of 22-31 in March 2011. The horizontal domains consist of three domains: domain 1 covering a wide area of Northeast Asia, domain 2 covering the main land of Japan, and domain 3 covering the study area. The horizontal resolution is 4 km and the number of grid is 68 × 72 for domain 3.

The GA, SA and HGS algorithms are applied to three different network sizes; 1) 57 stations, which is consistent with the number of stations in the year 2011, 2) 116 stations, consistent with that in the year 2013, 3) 150 stations. It should be noted that the locations of the existing stations are not considered here because the application of the observation-free method is focused. The performance is compared in terms of minimum, mean and standard deviation (SD) of mean kriging variance.

Model performance

The performance of the model is detailed in Shimadera et al. (2014) and summarized as follows. The model well simulates the temporal and spatial variation patterns daily concentrations, and approximately reproduces the monthly means except for the underestimates in summer, which results in the underestimation of the annual means. Given that the underestimation is spatially uniform in general, the simulated annual means have sufficient quality to be provided to the network optimization algorithms. The spatial distribution of the simulated concentrations of annual mean is given in Figure 1.

Optimization results

The indicators of the performance calculated from 20 realizations of each algorithm for the three cases are provided in Table 1. Regarding the comparison between the performance of the GA and SA, the SA shows smaller minimum and mean, while the GA consistently shows smaller SD. However, the HGS outperforms the GA and SA with smaller minimum and mean as well as much smaller SD. The best

network design obtained by the HGS for the three network sizes are presented in Figure 2 together with the spatial distribution of the concentrations predicted by ordinary kriging with the simulated values at the selected sites. The selected sites in the best network designs are placed homogeneously in general for all the three algorithms (not shown here for those by the GA and SA). The spatial distributions sufficiently capture the feature of the simulated fields for all the cases. Although the difference in the pattern of the spatial distribution between the network sizes is not significant, the larger the network size, the better the predicted field reproduces the higher and lower concentration areas.

Table 1. The statistical measures of the mean kriging variance ($\times 10^{-6}$) obtained from 20 realizations of the GA, SA and HGS for each network size *n*

	GA			SA			HGS		
n	**Min.**	**Mean**	**SD**	**Min.**	**Mean**	**SD**	**Min.**	**Mean**	**SD**
57	1692	1704	8.6	1650	1662	15.6	1645	1650	2.6
116	1199	1204	2.9	1159	1171	9.5	1155	1158	1.7
150	1049	1055	2.8	1017	1027	7.9	1013	1014	1.0

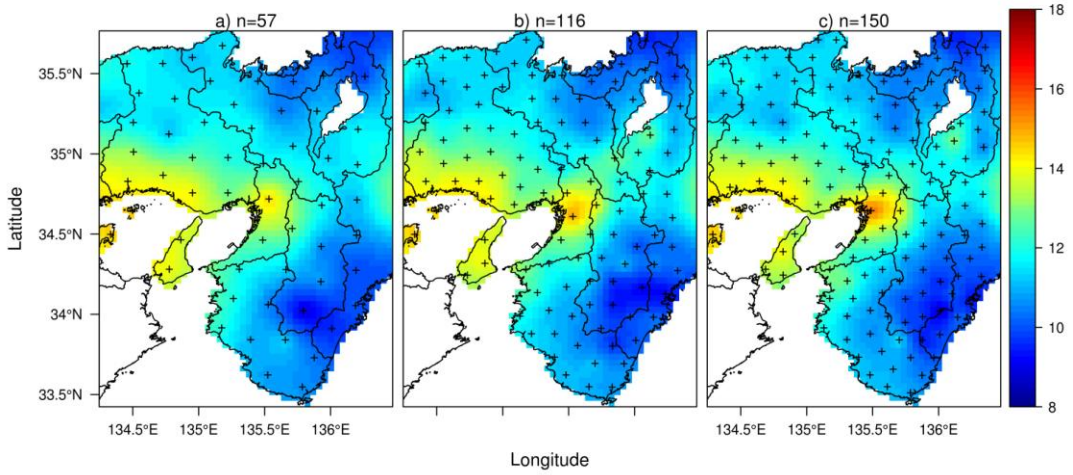

Figure 2. The optimized network by the HGS for the network size a) n=57, b) n=116 and c) n=150. The concentration fields are generated by ordinary kriging with the simulated values at the selected sites. The crosses represent the selected sites. Unit is μg m^{-3}.

Weighing factor

One might consider the basic strategy where more monitoring stations should be placed in the higher concentration areas. To achieve this, a weighing factor can be implemented in the algorithm. Kriging variance is multiplied by the factor $w(s) = \{(c(s) - c_{min})/c_{sd}\}^2$, where $c(s)$ is a concentration at s, c_{min} and c_{sd} are the minimum and standard deviation of the simulated concentrations respectively. The network is optimized by the HGS incorporated with $w(s)$ for the three network sizes and the results are compared to those obtained without the factor. The mean kriging variance as well as R^2 and root mean squared error (RMSE) between the simulated values and the predicted fields by ordinary kriging with the simulated values at the selected sites are presented in Table 2. The quality measures for the results obtained without the factor are identical to the mean values from 20 realizations by the HGS in the previous section. The optimized networks obtained with the weighing factor are given in Figure 3 for the three network sizes.

By the weighing factor, some sites shift from lower concentration areas to higher concentrations areas. As a consequence, the distributions predicted by ordinary kriging (Figure 3) show better correspondence with the higher concentration areas in the simulated field (Figure 1) than those obtained without the weighing factor (Figure 2). The distributions also show smaller areas of lower concentrations where less sites are selected. Although the difference in R^2 is not significant, mean kriging variance and RMSE of the network obtained with the weighing factor are larger than those without the factor, which is due to the effect of the factor that realizes less homogeneous placement of stations.

Table 2. The comparison of the quality measures from the optimized network with and without the weighing factor by the HGS for the three network sizes. MKV represents mean kriging variance ($\times 10^{-6}$). Unit of RMSE is μg m^{-3}.

	With the weighing factor			Without the weighing factor		
n	**MKV**	R^2	**RMSE**	**MKV**	R^2	**RMSE**
57	1796	0.84	0.59	1649	0.84	0.49
116	1298	0.86	0.53	1158	0.89	0.40
150	1140	0.89	0.48	1014	0.90	0.38

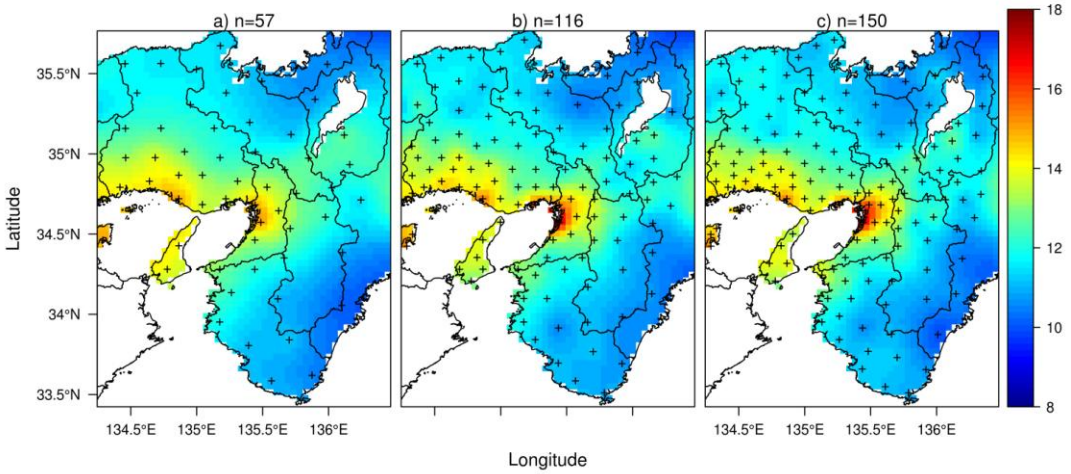

Figure 3. The optimized network by the HGS with the concentration based weighing factor for the network size a) n=57, b) n=116 and c) n=150. The concentration fields are generated by ordinary kriging with the simulated values at the selected sites. The crosses represent the selected sites. Unit is μg m^{-3}.

CONCLUSION

In this paper, a unique method integrating metaheuristics with WRF/CMAQ model is presented that can be applied to the optimization of an air monitoring network with limited number of stations or a network under development. In addition to the metaheuristics such as the GA and SA, the HGS algorithm is developed and compared to the two algorithms with the results from the application to the monitoring network of $PM_{2.5}$ in the Kinki region in Japan for three different network sizes. The HGS consistently outperforms the rest of the two algorithms over the three cases. This demonstrates that the HGS successfully combine the advantages of the GA and SA, and works best with WRF/CMAQ simulated values. The concentration based weighing factor is successfully implemented in the HGS resulting in the optimized network with more stations in the relatively higher concentration areas. When the basic strategy where more stations should be placed in populated areas is considered, the weighing factor based on population or population density can be readily implemented into the HGS in a similar way. The proposed method that integrates the HGS algorithm with WRF/CMAQ model has broad utility and can provide information of vital use on the air monitoring network optimization.

REFERENCES

Byun, D.W. and Ching, J.K.S., 1999: Science algorithms of the EPA models-3 Community Multi-scale Air Quality (CMAQ) modelling system. NERL, Research Triangle Park, NC.

Ruiz-Cardenas, R., Ferreira, M.A.R. and Schmidt, A.M., 2010: Stochastic search algorithms for design of monitoring networks. *Environmetrics*, **21**, 102-112.

Shimadera H., Kojima T., Kondo A. and Inoue Y., 2014: Evaluation of air quality model performance for long-term $PM_{2.5}$ simulation in Japan. Extended abstract on HARMO16, H16-104.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Baker,D.M., Duda, M.G., Huang, X.-Y., Wang, W. and Powers, J.G., 2009: A description of the advanced research WRF version 3. NCAR/TN-475+STR.

Wu, L. and Bocquet M., 2011: Optimal redistribution of the background ozone monitoring stations over France. *Atmos. Environ.*, **45**, 772-783.