Atmospheric Environment 122 (2015) 22–30

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Optimization of air monitoring networks using chemical transport model and search algorithm

Shin Araki [a, b, *], Koki Iwahashi [a], Hikari Shimadera [a], Kouhei Yamamoto [c], Akira Kondo [a]

[a] *Graduate School of Engineering, Osaka University, Japan*
[b] *Otsu Public Health Center, Shiga, Japan*
[c] *Graduate School of Energy Science, Kyoto University, Japan*

HIGHLIGHTS

- Air monitoring network is optimized by minimization of the mean kriging variance.
- We propose a hybrid of a genetic algorithm and simulated annealing.
- No previous observation is needed as kriging variance is derived from simulations.
- The hybrid algorithm outperforms the two single algorithms.

ARTICLE INFO

Article history:
Received 17 March 2015
Received in revised form
6 September 2015
Accepted 8 September 2015
Available online 10 September 2015

Keywords:
$PM_{2.5}$
NO_2
O_3
Genetic algorithm
Simulated annealing
Japan

ABSTRACT

Air monitoring network design is a critical issue because monitoring stations should be allocated properly so that they adequately represent the concentrations in the domain of interest. Although the optimization methods using observations from existing monitoring networks are often applied to a network with a considerable number of stations, they are difficult to be applied to a sparse network or a network under development: there are too few observations to define an optimization criterion and the high number of potential monitor location combinations cannot be tested exhaustively. This paper develops a hybrid of genetic algorithm and simulated annealing to combine their power to search a big space and to find local optima. The hybrid algorithm as well as the two single algorithms are applied to optimize an air monitoring network of $PM_{2.5}$, NO_2 and O_3 respectively, by minimization of the mean kriging variance derived from simulated values of a chemical transport model instead of observations. The hybrid algorithm performs best among the algorithms: kriging variance is on average about 4% better than for GA and variability between trials is less than 30% compared to SA. The optimized networks for the three pollutants are similar and maps interpolated from the simulated values at these locations are close to the original simulations (RMSE below 9% relative to the range of the field). This also holds for hourly and daily values although the networks are optimized for annual values. It is demonstrated that the method using the hybrid algorithm and the model simulated values for the calculation of the mean kriging variance is of benefit to the optimization of air monitoring networks.

1. Introduction

Air monitoring networks have been developed in various areas in the world for 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 of the network. This issue is often referred to as a network design problem and has been widely discussed (e.g., Brus and Heuvelink, 2007; Wu and Bocquet, 2011). The network design problem usually aims at minimization of a design criterion that may be based on observations or other information about the field of interest: to achieve this aim, subsets of

* Corresponding author. Graduate School of Engineering, Osaka University, Yamadaoka 2-1, Suita, Osaka 565-0871, Japan.
E-mail addresses: araki@ea.see.eng.osaka-u.ac.jp (S. Araki), iwahashi@ea.see.eng.osaka-u.ac.jp (K. Iwahashi), shimadera@see.eng.osaka-u.ac.jp (H. Shimadera), yamamoto@energy.kyoto-u.ac.jp (K. Yamamoto), kondo@see.eng.osaka-u.ac.jp (A. Kondo).

http://dx.doi.org/10.1016/j.atmosenv.2015.09.030
1352-2310/

potential monitor locations are selected by an algorithm.

A design criterion is often defined with the notion of entropy where a set of locations which maximize the entropy at the monitored sites is searched for (e.g., Zidek et al., 2000; Fuentes et al., 2007). Another popular criterion is defined with a geostatistical estimation method which is called kriging, where the theoretical interpolation error averaged over the region of interest, i.e. the mean kriging variance is minimized (e.g., Baume et al., 2011; Wu and Bocquet, 2011). When the mean kriging variance is used as a criterion, observations obtained from existing monitoring network are often used to construct a variogram for the calculation of the mean kriging variance. In these cases, it is assumed that the network represents the spatial distribution sufficiently, thus the network of interest is relatively dense where the efficiency of the network is focused, i.e. reduction of stations. Therefore, this method is difficult to apply to a sparse network or a network under development that insufficiently represents the spatial distribution of the pollutant of interest.

In the field of air quality study, the chemical transport model (CTM), that simulates physical and chemical processes including emission, advection, photochemical reactions and deposition, has been extensively used at various ranges of spatial and temporal scale, not only to obtain a spatial distribution, but also to establish an effective strategy for the control of the concentrations of air pollutants (e.g., Emmons et al., 2010; Chatani et al., 2014). Thus, simulated concentrations from CTM with sufficiently high spatial resolution can be an alternative to observations to derive a variogram to compute the mean kriging variance as a design criterion. Simulations have also been used for sampling optimization by Kumral and Ozer (2013) in mine planning.

Once the criterion is defined, the network design problem can be treated as a combinatorial optimization problem. When a network is small enough, complete enumeration of all possible combinations is possible. For a large network, however, this will run into a combinatorial explosion. To deal with this difficulty, search algorithms have been applied to the optimization of large networks. For instance, Ruiz-Cardenas et al. (2010) applied genetic algorithm (GA) for an O_3 monitoring network with several hundred stations in the United States. Wu and Bocquet (2011) applied simulated annealing (SA) to the optimization of an O_3 monitoring network over France. Given that the search space is huge when using simulated fields, superior ability both for a global and a local search is required for a search algorithm to be applied. GA is able to search in a large space, but is often not able to find the local optimal solution (Ruiz-Cardenas et al., 2010). On the other hand, SA is able to find locally optimal solutions, but is often trapped in regions far from the global optimum (Ruiz-Cardenas et al., 2010). Araki et al. (2015) developed a hybrid of GA and SA (HGS) and successfully applied it to the optimization of a $PM_{2.5}$ monitoring network using simulated values obtained from CTM for the computation of the mean kriging variance. However, the performance of HGS was not compared to those of other optimization algorithms in their study, thus the advantages of HGS have not been demonstrated. In addition, HGS was applied only for a $PM_{2.5}$ monitoring network, and the possibility of application to other pollutants, that might have different spatial distribution features, was not examined.

In this paper, CTM is used to generate spatial distributions of air pollutant concentrations to derive variograms for the computation of the mean kriging variance as a design criterion, and each of the algorithm including HGS, GA and SA is applied to $PM_{2.5}$, NO_2 and O_3 in the Kinki region of Japan respectively, repeating each setting 30 times to capture random effects. The performances of the algorithms are compared against each other in terms of the quality of the solutions such as the mean and the standard deviation of the mean kriging variance of the respective trials. Fields have been interpolated by ordinary kriging using the CTM simulated values at the selected sites in the optimized networks, and the errors between the interpolated and simulated fields are computed. The difference between the algorithms for each of the pollutants is discussed and the capability and the applicability of HGS are evaluated.

2. Methodology

2.1. Chemical transport model

The chemical transport model used in this study is the Community Multiscale Air Quality model (CMAQ) (Byun and Ching, 1999) version 5.0.1 which was driven with the Weather Research and Forecasting model (WRF) (Skamarock et al., 2009) version 3.5.1. Meteorological fields were produced using WRF configured with the same physics options as those used by Shimadera et al. (2014); also emission data for the air quality simulations was produced in a similar way. The other settings involved in the simulations and CMAQ configurations are detailed in Shimadera et al. (2015).

The WRF/CMAQ model was 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 consisted 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 area where the optimization algorithms are tested, which is shown as coloured area in Fig. 1. The horizontal resolution is 4 km and the number of grids is 68×72 for domain 3. The annual and daily values used for the network optimization are computed by averaging the hourly CTM outputs over the corresponding time periods. This simulation is identical to that used in Araki et al. (2015).

The performance of the model is detailed in Shimadera et al. (2015) and summarized as follows: the statistical measures obtained from the comparison between observed and simulated daily concentrations indicate that the model simulates the temporal and spatial variation patterns of $PM_{2.5}$, NO_2 and O_3 well with the Pearson's correlation coefficient being 0.76, 0.82, and 0.77 for $PM_{2.5}$, NO_2 and O_3 respectively. The scatter plots of the simulated and observed values of annual means for $PM_{2.5}$ and NO_2, and annual means of daily maximum 8-hr mean concentrations for O_3, which are used to derive variograms for the computation of the mean kriging variance, are presented in Fig. 2 with root mean squared error (RMSE) and R^2 between the observed and simulated values. The number of observations for $PM_{2.5}$, NO_2 and O_3 is 8, 219 and 188 respectively. The reason for the limited number of observations for $PM_{2.5}$ is because the $PM_{2.5}$ network in Japan started to be developed since 2009, a year before the target year of this simulation. The concentrations of O_3 lie in a relatively narrow range, which results in low R^2 value for O_3. However, RMSE of O_3 is approximately 10% of the mean values of O_3. Therefore, these simulated concentrations have sufficient quality to derive variograms for the computation of the mean kriging variance as the optimization criterion. The spatial distributions of $PM_{2.5}$, NO_2 and O_3 are given in Fig. 1. Both $PM_{2.5}$ and NO_2 show highest concentrations in densely populated areas, where NO_2 is more concentrated to the agglomeration of Osaka. The higher and lower concentration areas of $PM_{2.5}$ are generally distributed evenly. On the other hand, higher concentrations of NO_2 are found in limited areas where megacities are located, while the lower areas are widely distributed. The spatial distribution of O_3 is generally the reverse of that of NO_2. This is because the O_3 concentrations are affected by titration with NO_x, and higher NO_x concentrations might cause lower O_3 concentrations in urban areas with relatively large anthropogenic emissions. The histograms of the simulated values in the candidate areas i.e. in the coloured area in Fig. 1 are given in Fig. 3, which reflect the characteristics of the