Title Page

Manuscript title:

Optimization of Air Monitoring Networks Using Chemical Transport Model and Search Algorithm

Name of the authours:
Shin Araki[a, b, *], araki@ea.see.eng.osaka-u.ac.jp
Koki Iwahashi[a], iwahashi@ea.see.eng.osaka-u.ac.jp
Hikari Shimadera[a], shimadera@see.eng.osaka-u.ac.jp
Kouhei Yamamoto[c], yamamoto@energy.kyoto-u.ac.jp
Akira Kondo[a], kondo@see.eng.osaka-u.ac.jp

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

*Corresponding author:
Name: Shin Araki
Address: Graduate School of Engineering, Osaka University
Yamadaoka 2-1, Suita, Osaka 565-0871, Japan
Tel.: +81 06 6879 7670
Fax: +81 06 6879 7670
E-mail address: araki@ea.see.eng.osaka-u.ac.jp

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.

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

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 et al., 2007; Wu et al., 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 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 et al., 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., 2011; 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, Ruis-Cardenas et al. (2010) applied genetic algorithm (GA) for an O_3 monitoring network with several hundred stations in the United States. Wu et al. (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 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 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 colored area in Figure 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

124 computed by averaging the hourly CTM outputs over the corresponding time periods. This
125 simulation is identical to that used in Araki et al. (2015).
126 The performance of the model is detailed in Shimadera et al. (2015) and summarized as follows:
127 the statistical measures obtained from the comparison between observed and simulated daily
128 concentrations indicate that the model simulates the temporal and spatial variation patterns of
129 $PM_{2.5}$, NO_2 and O_3 well with the Pearson's correlation coefficient being 0.76, 0.82, and 0.77 for
130 $PM_{2.5}$, NO_2 and O_3 respectively. The scatter plots of the simulated and observed values of annual
131 means for $PM_{2.5}$ and NO_2, and annual means of daily maximum 8-hr mean concentrations for O_3,
132 which are used to derive variograms for the computation of the mean kriging variance, are
133 presented in Figure 2 with root mean squared error (RMSE) and R^2 between the observed and
134 simulated values. The number of observations for $PM_{2.5}$, NO_2 and O_3 is 8, 219 and 188
135 respectively. The reason for the limited number of observations for $PM_{2.5}$ is because the $PM_{2.5}$
136 network in Japan started to be developed since 2009, a year before the target year of this
137 simulation. The concentrations of O_3 lie in a relatively narrow range, which results in low R^2
138 value for O_3. However, RMSE of O_3 is approximately 10% of the mean values of O_3. Therefore,
139 these simulated concentrations have sufficient quality to derive variograms for the computation of
140 the mean kriging variance as the optimization criterion. The spatial distributions of $PM_{2.5}$, NO_2
141 and O_3 are given in Figure 1. Both $PM_{2.5}$ and NO_2 show highest concentrations in densely

142 populated areas, where NO_2 is more concentrated to the agglomeration of Osaka. The higher and
143 lower concentration areas of $PM_{2.5}$ are generally distributed evenly. On the other hand, higher
144 concentrations of NO_2 are found in limited areas where megacities are located, while the lower
145 areas are widely distributed. The spatial distribution of O_3 is generally the reverse of that of NO_2.
146 This is because the O_3 concentrations are affected by titration with NO_x, and higher NO_x
147 concentrations might cause lower O_3 concentrations in urban areas with relatively large
148 anthropogenic emissions. The histograms of the simulated values in the candidate areas i.e. in the
149 colored area in Figure 1 are given in Figure 3, which reflect the characteristics of the distribution
150 of each pollutant: $PM_{2.5}$ concentrations are approximately normally-distributed, the distribution
151 of NO_2 concentrations is skewed to zero and that of O_3 is relatively skewed to higher values.

152 **2.2 Design criterion**

153 The optimization criterion is minimizing the mean kriging variance, which has been used as a
154 design criterion in many previous studies (e.g., Van Groenigen, 2000; Brus et al., 2007; Wu et al.,
155 2010; Wu et al., 2011; Baume et al., 2011). A definition of a design criterion is an expression of a
156 purpose of a monitoring network. When minimization of the mean kriging variance is used as a
157 criterion, the purpose of the network can be seen to grasp the spatial distribution of a pollutant in
158 the whole region of interest well.

159 The simulated concentrations are transformed to a natural logarithmic scale before analysis, and
160 back transformed after analysis so that the predicted concentrations are positive, which is not
161 always the case without transformation (Beelen et al., 2009). The concentrations simulated by
162 CTM are treated as point values at the grid cell centroids and an empirical variogram is computed
163 from all the candidate points. A theoretical variogram is fitted to it with spherical model by
164 weighted least-squares fitting. The derived variograms for the three pollutants are given in Figure
165 4. Although the ranges of the three variograms are slightly different, the variograms are generally
166 similar, with nugget very close to zero.

167 The fitness value for a selected set of points is the mean kriging variance: kriging variance is
168 computed in all points of the grid in the region of interest from the variograms and the selected
169 set of points and averaged over the area; to avoid overestimation of the kriging variance at the
170 boundaries, fifteen equally-spaced points at the margins of the region of interest (see Figure 1)
171 are in addition regarded as selected.

172 **2.3 Optimization algorithms**

173 The optimal combination of points is searched for by the optimization algorithms. Each
174 network configuration is given by a binary string, where each element represents a candidate
175 location, with ones at gauged sites and zeros at ungauged sites. A subset of locations is selected
176 from the candidate locations according to this binary string.

2.3.1 Genetic algorithms

The basic scheme of GA algorithm in this paper is similar to the standard GA which consists of population, selection, crossover and mutation. The initial population is randomly generated and the population size is 100. In the crossover operation, gauged sites in both parents are gauged in the offspring, and ungauged sites in both parents are ungauged in the offspring. Half of the sites that are gauged in one of the parents and not gauged in the other are randomly conversed. The probability of crossover occurrence is set to 0.9. In the mutation operation, 5% of gauged sites are randomly selected and swapped with the same number of randomly selected ungauged sites with a mutation probability of 0.2. The solutions with the best 5% fitness, called elite, remain in the next generation, while others are rejected in order to keep the population size unchanged. These operations stop when the total number of generations reaches 500. The algorithm also stops when the number of generations without any improvement in the fitness reaches 40.

2.3.2 Simulated annealing

SA is an iterative search algorithm which starts with a randomly selected initial solution and the fitness value is computed. 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 value 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

value improves. Even when the fitness value degrades, the new potential solution is accepted with a certain probability, which is given by

$$p = \exp(\frac{-\Delta f}{T}), \qquad (1)$$

where Δf is the amount of degradation of the fitness value and T is a control parameter (Brus et al., 2007). T is gradually decreased with iterations by

$$T_{k+1} = \alpha T_k, \qquad (2)$$

where α is 0.9 and k is a number of iterations. The initial T (i.e., T_1) is set so that the average p in the equation (1) equals to 0.8 (Brus et al., 2007).

The size of the search window is fixed in the first 500 iterations at 18 km (approximately equals to 7 × 7 cells window). Then it is decreased after 500 iterations down to 12 km (5 × 5 window), and 6 km (3 × 3 window) from 1,000 iterations and onward. The total number of iterations for each run is set to 50,000. The algorithm also stops when the number of iterations with no progress in the fitness value reaches 4000.

2.3.3 Hybrid algorithm

HGS algorithm is identical to that applied in Araki et al. (2015) and summarized as follows. The basic scheme of HGS is similar to GA. The initial population of the size of 100 is randomly generated and GA is applied for 75 iterations. After 75th generation, SA is applied to each of the 15% best solutions for 30 iterations with the fixed search window size of 7 × 7. The rest 85%

213 solutions are kept unchanged and combined with the solutions obtained from SA. GA is then
214 applied to this set of solutions i.e. population for 1 iteration, and the next generation is obtained.
215 Again, SA is applied to the best 15% solutions for 30 iterations in the generation and the
216 solutions are combined to the rest 85% solutions. The next generation is obtained by the
217 application of 1 iteration of GA to the set of solutions. The parameters of HGS in the first 75
218 generations are identical to GA algorithm described above. After 75 generations, the crossover
219 and mutation probability is set to 0.3 and 0 respectively. After the 100th generation, the search
220 window size is reduced to 5 × 5. This window size is again decreased to 3 × 3 after 125th
221 generation and is keep unchanged. The total number of iterations is set to 155, but the algorithm
222 also stops when the number of generations with no progress in the fitness value reaches 8 after
223 75th generation. The flowchart of HGS is presented in Figure 5.

224 **2.4 Application of the algorithms**

225 The algorithms are tested in the area that includes the Kinki region in Japan (134.2°E –
226 136.5°E, 33.4°N – 35.8°N) that consists of six prefectures, and its contiguous areas, which equals
227 to the colored area in Figure 1. Only land areas are considered for the potential locations of the
228 monitoring stations. The number of the candidate locations is 1984. Megacities such as Osaka,
229 Kyoto and Kobe are located in this region. The number of monitoring stations for NO_2 and O_3 in
230 the study area is 221 and 187 respectively in the year 2011, which has remained almost

231 unchanged in recent years, while that of $PM_{2.5}$ is increasing rapidly from 57 stations in the year
232 2011 to 117 in the year 2013. This is because the national air quality standard for $PM_{2.5}$ in Japan
233 was set in the year 2009 and the monitoring network was initially developed by placing
234 continuous monitors at existing stations.

235 GA, SA and HGS algorithms were applied to three different network sizes; 1) 57 stations,
236 which is consistent with the number of stations of $PM_{2.5}$ in the year 2011, 2) 117 stations,
237 consistent with that in the year 2013, 3) 150 stations, which is approximately 3 times as much as
238 the first case. It should be noted that the locations of the existing stations were not considered
239 because the application of the observation-free method is focused. Each of the algorithms are
240 applied to each pollutant combined with each network size respectively, repeating each of these
241 27 settings 30 times to capture random effects. The performances are compared in terms of the
242 mean and the standard deviation (SD) of the fitness values of the respective 30 trials. The
243 stopping parameters of iterations for GA, SA and HGS are set so that the total effort of all the
244 three algorithms is similar for a fair comparison.

245 Data analysis was carried out using R statistical software 3.1.2 (R Core Team, 2014) with the
246 package GA (Scrucca, 2013) for the basic scheme of GA algorithm, the package gstat (Pebesma,
247 2004) for kriging.

248

3. Results and discussion

3.1 Performance of the algorithms

The mean and standard deviation (SD) of the fitness values obtained from 30 random trials of each algorithm for the three network sizes and the three pollutants are presented in Figure 6. Regarding the comparison between the performance of GA and SA, SA gives smaller means in all the cases. However, SD of GA is consistently smaller than that of SA. In the case of the network size of n=117 for $PM_{2.5}$, for instance, SD of GA is as small as 36 % of that of SA. These results indicate that SA is able to find a better solution but is not as stable as GA.

The mean of HGS is between 95 and 96% of the means of GA in all the cases and SD is 26 - 47% of that of GA. When compared to the results of SA, HGS gives smaller means than those of SA and the SD is as small as 13 - 29% of that of SA. This indicates that HGS outperforms GA and SA with a better search capability and stability.

The best network placements out of the 30 trials for each network size and pollutant obtained by HGS are presented in Figure 7. The selected sites in the best network designs are generally placed homogeneously. Although the distributions of the simulated values for the three pollutants have unique characteristics, the difference in the optimized network between the pollutants is small.

3.2 Spatial characteristics of the resulting best networks

As shown in Figure 7, there seems to be few differences between the optimal networks for the different pollutants. For a quantitative comparison, the fitness value of each pollutant is calculated using the optimized network obtained by HGS for all the pollutants. For instance, fitness values of $PM_{2.5}$ are calculated using optimized network for NO_2 and O_3, which in turn are compared to that obtained using $PM_{2.5}$-optimized network. The fitness value is computed for each solution of 30 trials. The obtained results are presented in Figure 8.

The fitness values for each pollutant generally lie in the same range, no matter which pollutant is used in the optimization. Thus, the network optimized for one pollutant also result in good fitness for the other pollutants, and the relative goodness of a network compared to others does not change between the pollutants.

As shown in Figure 4, the variograms of all said three pollutants used for the optimization are generally similar, spherical model with nugget very close to zero. Van Groenigen (2000) showed that the optimized networks obtained for linear, exponential and spherical variograms with a relatively low nugget were similar. The results in this study agree with this previous study.

3.3 Interpolation error

The monitoring network can be evaluated by how well it represents the spatial distribution of the concentrations in the domain of interest. More specifically, the concentration field is generated by ordinary kriging using the CTM simulated values at the selected sites in the

285 optimized network. The variogram is constructed using simulated values at the selected sites.
286 Fifteen equally-spaced points on the outer side of the study area are included in the same way as
287 described in the Section 2.2. The resulting map is compared to the CTM simulated field.

288 RMSE and R^2 between the simulated values and kriged values for each set of trials are
289 computed. The obtained RMSE is given in Figure 9. RMSE is less than 0.6 ($\mu g\ m^{-3}$) for all the
290 three algorithms for $PM_{2.5}$, less than 1.7 (ppb) for NO_2 and less than 1.8 (ppb) for O_3. These
291 RMSE values are 6, 8 and 9% relative to the ranges of the simulated values of $PM_{2.5}$, NO_2 and O_3
292 respectively. The difference in RMSE between the three algorithms is minute for each network
293 size and pollutant. R^2 is generally higher than 0.62 for all the cases and the difference in R^2
294 between the algorithms is also marginal (not shown here).

295 In addition, the kriged fields generated by the optimization result of the three algorithms for the
296 same size and pollutant are similar to each other and generally capture the feature of the
297 corresponding simulated fields. As an example, the error maps between the simulated and kriged
298 fields obtained from the result of HGS for the size of n=57 and for the three pollutants are given
299 in Figure 10, where positive values represent overestimates and negative values represent
300 underestimates. The errors for $PM_{2.5}$ are close to zero, between -3.6 and 1.6 ($\mu g\ m^{-3}$), over the
301 region of interest. On the other hand, underestimates for NO_2 and overestimates for O_3 are found
302 in limited areas such as agglomeration of Osaka. These areas correspond to the areas of the

highest concentrations of NO_2 or the lowest concentrations of O_3. These under/overestimates are due to the lack of the selected site in such limited areas of high/low concentrations. However, under/overestimates are not significant for most of the region except for the limited areas. These results indicate that the optimized networks obtained by the three algorithms represent the spatial distribution of the concentrations in the region of interest sufficiently.

3.4 Evaluation of the optimized network for daily and hourly values

The monitoring networks should map not only annual mean concentrations but also daily or hourly mean concentrations well. In order to evaluate the optimized network versus daily and hourly means, interpolation errors (RMSE) between the simulated and kriged fields for each day and hour of the whole year are computed in the same way as described in the section 3.3. Variograms are derived in an automated way using package automap (Hiemstra et al., 2009), because manually constructing variograms for every single day or hour of the entire year would be too time consuming.

The resulted RMSE are given in Table 1. The finer the temporal scale is, the larger RMSE becomes, which may be due to less homogeneous spatial distribution of pollutants at a finer temporal scale. However, RMSE of all the pollutants for all network sizes are relatively small compared to the ranges of the simulated daily or hourly means of the corresponding pollutants. Although not every daily or hourly result is checked manually, these results demonstrates that the

optimized network obtained using the annual mean is also optimal, at least to some extent, to produce maps of daily or hourly mean concentrations.

4. Conclusions

In this paper, the performance and the applicability of HGS that is a hybrid of GA and SA is evaluated and compared to the single GA and SA using CTM simulated values to derive variograms for the computation of the mean kriging variance as a design criterion for $PM_{2.5}$, NO_2 and O_3 in the Kinki region of Japan for three network sizes. HGS consistently outperforms the rest of the two algorithms for all the network sizes and pollutants. This demonstrates that HGS successfully combines the advantages of GA and SA. The optimized network for one pollutant is optimal for other pollutants and the relative goodness of a network compared to others does not change between the pollutants. This is because the variograms of the pollutants are similar. This implies that only one pollutant might be necessary to be considered when a network with multiple pollutants of similar variograms is optimized simultaneously. The concentration fields are generated by ordinary kriging using the simulated values at the selected sites in the optimized networks obtained by GA, SA and HGS respectively, and compared to the simulated fields to obtain interpolation errors. The difference between the algorithms is minute and the optimized networks obtained by the three algorithms represent the spatial distribution of the concentrations

in the region of interest sufficiently. The optimized network to well map the annual mean is optimal, to some extent, to produce maps of daily or hourly mean concentrations.

It is demonstrated that HGS is a valid and feasible method for the optimization of an air monitoring network. The combination of HGS and CTM simulation to derive a variogram for a criterion calculation is of benefit especially when a network with a limited number of stations or network under development is considered, and provide information of vital use on the air monitoring network optimization.

Acknowledgements

This research was partly supported by JSPS KAKENHI Grant Number 26740038.

References

Araki, S., Iwahashi, K., Shimadera, H., Yamamoto, K., Kondo, A., 2015. Optimization of the $PM_{2.5}$ monitoring network using hybrid genetic algorithm. Journal of Japan Society for Atmospheric Environment 50, 35-43 [in Japanese].

Baume, O.P., Gebhardt, A., Gebhardt, C., Heuvelink, G.B.M., Pilz, J., 2011. Network optimization algorithms and scenarios in the context of automatic mapping. Computers & Geosciences, 37, 289-294.

Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., de Hoogh, K., Briggs, D. J., 2009. Mapping of

358 background air pollution at a fine spatial scale across the Europe Union. Science of the Total
359 Environment 407, 1852-1867.

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

362 Brus, D.J., Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging for
363 environmental variables. Geoderma 138, 86-95.

364 Chatani, S., Morino, Y., Shimadera, H., Hayami, H., Mori, Y., Sasaki, K., Kajino, M., Yokoi, T.,
365 Morikawa, T., Ohara, T., 2014. Multi-model analyses of dominant factors influencing
366 elemental carbon in Tokyo metropolitan area of Japan. Aerosol and Air Quality Research 14,
367 396-405.

368 Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C.,
369 Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C.,
370 Baughcum, S. L., Kloster, S., 2010. Description and evaluation of the Model for Ozone and
371 Related chemical Tracers, version 4 (MOZART-4), Geoscientific Model Development 3, 43-
372 67.

373 Fuentes, M., Chaudhuri, A., Holland, D.M., 2007. Bayesian entropy for spatial sampling design
374 of environmental data. Environmental and Ecological Statistics 14, 323-340.

375 Hiemstra, P.H., Pebesma, E.J., Twenhofel, C.J.W., Heuvelink, B.M., 2009. Real-time automatic
376 interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network.
377 Computers & Geosciences 35, 1711-1721.

378 Kumral, M., Ozer, U., 2013. Planning additional drilling campaign using two-space genetic
379 algorithm: A game theoretical approach. Computers & Geosciences 52, 117-125.

380 Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers &
381 Geosciences 30, 683-691.

382 R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for
383 Statistical Computing, Vienna, Austria, URL http://www.R-project.org/.

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

386 Scrucca, L., 2013. GA: A package for genetic algorithms in R, Journal of Statistical Software,
387 53(4), 1-37.

388 Shimadera, H., Hayami, H., Ohara, T., Morino, Y., Takami, A., Irei, S., 2014. Numerical
389 simulation of extreme air pollution by fine particulate matter in China in winter 2013. Asian
390 Journal of Atmospheric Environment 8, 25-34.

391 Shimadera, H., Kojima, T., Kondo, A., Inoue, Y., 2015. Performance comparison of CMAQ and

392 CAMx for one-year $PM_{2.5}$ simulation in Japan. International Journal of Environment and
393 Pollution, in press.

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

397 Van Groenigen, J.W., 2000. The influence of variogram parameters on optimal sampling schemes
398 for mapping by kriging. Geoderma 97, 223-236

399 Wu, L., Bacquet, M., Chevallier, M., 2010. Optimal reduction of the ozone monitoring network
400 over France. Atmospheric Environment 44, 3071-3083.

401 Wu, L., Bocquet, M., 2011. Optimal redistribution of the background ozone monitoring stations
402 over France. Atmospheric Environment 45, 772-783.

403 Zidek, J.V., Sun, W., Le, N.D., 2000. Designing and integrating composite networks for
404 monitoring multivariate Gaussian pollution fields. Applied Statistics, 49, 63-79.

405

406

Figure Captions

Figure 1. The area of interest and the spatial distribution of $PM_{2.5}$, NO_2 and O_3 concentrations obtained by CTM for the Japanese fiscal year 2010. The values for $PM_{2.5}$ and NO_2 are annual means and that for O_3 is annual mean of daily maximum 8-hr means. The circles represent the 15 extra points used to derive variograms. Units are $\mu g\ m^{-3}$ for $PM_{2.5}$ and ppb for NO_2 and O_3.

Figure 2. The scatter plots of the simulated and observed values for annual means of $PM_{2.5}$ and NO_2, and annual means of daily maximum 8-hr mean concentrations for O_3. Units are $\mu g\ m^{-3}$ for $PM_{2.5}$ and ppb for NO_2 and O_3.

Figure 3. Histograms of the simulated values for $PM_{2.5}$, NO_2 and O_3. Units are $\mu g\ m^{-3}$ for $PM_{2.5}$ and ppb for NO_2 and O_3.

Figure 4. The variograms derived for $PM_{2.5}$, NO_2 and O_3.

Figure 5. The flowchart of the HGS algorithm.

Figure 6. The mean fitness values obtained from 30 random trials of each algorithm for $PM_{2.5}$ (in the left row panels), NO_2 (in the middle row panels) and O_3 (in the right row panels) for the network size of n=57 (in the upper column panels), n=117 (in the middle column panels) and n=150 (in the lower column panel). The fitness values are normalized by the mean values of GA of the corresponding network size and pollutant. The error bars represent SD.

Figure 7. The optimized network by HGS for $PM_{2.5}$ (in the left row panels), NO_2 (in the middle row panels) and O_3 (in the right row panels) for the network size of n=57 (in the upper column panels), n=117 (in the middle column panels) and n=150 (in the lower column panel). The crosses represent the selected sites. The circles represent the 15 extra points used to derive variograms.

Figure 8. The mean fitness values of a set of 30 trials for $PM_{2.5}$ (in the left row panels), NO_2 (in the middle row panels) and O_3 (in the right row panels) obtained using the network optimized for each pollutant for the network size of n=57 (in the upper column panels), n=117 (in the middle column panels) and n=150 (in the lower column panel). The top legend refers to the pollutant for which fitness value is computed using the optimized network of the three pollutants and the bottom legend refers to the pollutant for which the network is optimized for. The shaded areas

indicate that the fitness values are calculated for a pollutant using the optimized network of the same pollutant. The fitness values in each panel are normalized by mean values of $PM_{2.5}$-optimized network in the same panel. The error bars represent SD.

Figure 9. RMSE between CTM simulated field and kriged field with the simulated values at the selected sites of the optimized network of each set of 30 trials by GA, SA and HGS for the network size of n=57, 117 and 150, and for $PM_{2.5}$, NO_2 and O_3. The error bars represent SD. Units of RMSE for $PM_{2.5}$ are $\mu g\ m^{-3}$ and ppb for NO_2 and O_3.

Figure 10. The error map between the CTM simulated fields and the kriged fields with the simulated values at the selected sites of the optimized network of a) $PM_{2.5}$, b) NO_2 and c) O_3 by HGS for the network size of n=57. The positive values represent overestimates and the negative values represent underestimates. The crosses represent the selected sites. The circles represent the 15 extra points used to derive variograms. Units of concentrations are $\mu g\ m^{-3}$ for $PM_{2.5}$ and ppb for NO_2 and O_3.