

Effect of spatial outliers on the regression modelling of air pollutant concentrations: A case study in Japan

Shin Araki^{a,c,*}, Hikari Shimadera^a, Kouhei Yamamoto^b, Akira Kondo^a

^a*Graduate School of Engineering, Osaka University, Yamadaoka 2-1, Suita, Osaka 565-0871, Japan*

^b*Graduate School of Energy Science, Kyoto University, Yoshidahonmachi, Sakyo, Kyoto 606-8501*

^c*Otsu City Public Health Center, Goryochō-3-1, Otsu, Shiga 520-8575, Japan*

Abstract

Land use regression (LUR) or regression kriging have been widely used to estimate spatial distribution of air pollutants especially in health studies. The quality of observations is crucial to these methods because they are completely dependent on observations. When monitoring data contain biases or uncertainties, estimated map will not be reliable. In this study, we apply the spatial outlier detection method, which is widely used in soil science, to observations of PM_{2.5} and NO₂ obtained from the regulatory monitoring network in Japan. The spatial distributions of annual means are modelled both by LUR and regression kriging using the data sets with and without the detected outliers respectively and the obtained results are compared to examine the effect of spatial outliers. Spatial outliers remarkably deteriorate the prediction accuracy except for that of LUR model for NO₂. This discrepancy of the effect might be due to the difference in the characteristics of PM_{2.5} and NO₂. The difference in the number of observations makes a limited contribution to it. Although further investigation at different spatial scales is required, our study demonstrated that the spatial outlier detection method is an effective procedure for air pollutant data and should be applied to it when observation based prediction methods are used to

*Corresponding author

Email addresses: araki@ea.see.eng.osaka-u.ac.jp (Shin Araki), shimadera@see.eng.osaka-u.ac.jp (Hikari Shimadera), yamamoto@energy.kyoto-u.ac.jp (Kouhei Yamamoto), kondo@see.eng.osaka-u.ac.jp (Akira Kondo)

generate concentration maps.

Keywords: land use regression, variogram, kriging, PM_{2.5}, NO₂

1. Introduction

An accurate estimate of spatial distribution of air pollutants is the essential piece of information to evaluate the risks to human health and/or the air quality policy quantitatively. To obtain the distribution, the chemical transport model (CTM) has been extensively used in the field of air quality study (e.g., Emmons et al., 2010; Chatani et al., 2014; Shimadera et al., 2016). CTM simulates physical and chemical processes including emission, advection, transformation and depositions, and reproduces the temporal and spatial variation of air pollutant concentrations by complicated and demanding computation. On the other hand, empirical methods are widely used in health studies (e.g., Briggs et al., 2000; Ross et al., 2007; Wu et al., 2014). This approach is often called land use regression (LUR) and develops regression model for observed data and predictor variables that may influence the air pollutant concentrations such as land use, traffic related variables, and/or meteorological parameters. The concentrations at the locations with no observations are predicted by the obtained regression model. In some studies, residuals of a regression model are interpolated by the kriging method and summed up to the predictions by the regression model (e.g., Beelen et al., 2009; Pearce et al., 2009; Sampson et al., 2013; Araki et al., 2015). This method is called regression kriging or universal kriging. These approaches based on measurements are not computationally demanding compared to CTM especially for long-term statistics such as annual mean. On the contrary, the quality of observations is crucial to these methods because they are completely dependent on observations, which may contain biases and uncertainties.

Spatial outliers can be defined as an observation that is unusual compared to their neighbours (Lark et al., 2012). In soil science, spatial outliers have been widely discussed in previous studies (e.g., Lark, 2000; Zhao et al., 2007; Sun et al., 2012), because such observations could lead to exaggerated estimates of

mapping uncertainty (Sun et al., 2012). In the air quality data, measurements might be spatially outlying due to influences of nearby emission sources, specific terrain of the surrounding area and/or biased monitoring devices due to mechanical or electrical malfunction. These observations represent the concentrations in limited spatial extent, or almost no extent, compared to non-outliers. Although the quality of observations from monitoring network is usually controlled by its respective protocol and erroneous values are eliminated consequently, some spatial outliers might still remain in the data set because they are difficult to identify by such usual procedure. Regression model obtained with observations including spatial outliers may generate an air pollutant map significantly affected by outliers, which could result in biased health effect estimates.

One might argue that spatial outliers could be modelled properly by regression models with appropriate predictor variables. However, it is difficult to achieve because of the following reasons. Firstly, proper modelling of spatial variations of air pollutants at much finer spatial scale than the resolution of covariates could never be achieved. Secondly, observations in a data set should represent the concentrations in the similar spatial extent, or cannot be treated equivalently. Thirdly, biased observations can never be modelled using predictor variables. Therefore, spatial outliers should be properly treated before analyses. However, they have not been paid close attention to when observation-based method is applied to estimate spatial distribution of air pollutants.

In this study, we apply the spatial outlier detection method that is used in soil science to the regulatory monitoring network data of $\text{PM}_{2.5}$ and NO_2 in Japan. The spatial distributions of these pollutants are modelled by LUR and regression kriging respectively using the data sets inclusive and exclusive of the detected outliers respectively and the obtained results are compared. The aim of this study is to examine the effect of spatial outliers on the estimation of air pollutant concentrations using regression methods and gain some insight into how to deal with observations that may include spatial outliers.

57 **2. Methodology**

58 *2.1. Study area and air quality data*

59 The study area includes the main islands of Japan (129.1-145.8°E, 31.0-
60 45.5°N) but remote or small islands are excluded. Air quality observations are
61 obtained from the database of the regulatory monitoring network in Japan. The
62 monitoring stations are categorized into two types: road side stations and gen-
63 eral environment stations. The former are located at crossroads or road sides to
64 monitor air pollutants from automobile traffic, and the latter are located where
65 they are not directly affected by specific emission sources. Only the general envi-
66 ronment station data are utilized because of the difficulty in modelling the small
67 scale spatial variation near the road sides with our potential predictor variables
68 with spatial resolution of 500 m at the finest. The estimated maps with the data
69 exclusive of spatial outliers could thus be interpreted as background or baseline
70 concentration maps. The daily mean concentrations of PM_{2.5} and NO₂ for the
71 Japanese fiscal year 2013 (i.e., from April 2013 to March 2014) are used for the
72 analysis. The number of the general environment stations under operation for
73 PM_{2.5} and NO₂ are 649 and 1295 respectively in the year 2013. The remarkable
74 difference in number of stations is mainly due to the fact that the national air
75 quality standard for PM_{2.5} in Japan was set in the year 2009 and development
76 of the monitoring network started after that, which is more than 30 years after
77 the development of the NO₂ network. The difference in number of observations
78 is evaluated discussed in terms of the effect of spatial outliers.

79 The annual mean concentrations of PM_{2.5} remain approximately at the same
80 level and those of NO₂ marginally decrease in recent years in Japan. Therefore,
81 the annual means of PM_{2.5} and NO₂ are generally considered as stationary in
82 these few years, and the results obtained in this study are not specific to the
83 year to be studied.

84 *2.2. Data set*

85 The data sets used to construct grid data of predictor variables are pre-
86 sented in Table 1 and described in detail below. The selection of datasets is

87 made principally in consideration of the key factors in the spatial distribution
88 of air pollutants including emission, advection, transformation and deposition.
89 The accessibility and usability are also considered. If necessary, we spatially ag-
90 gregate or resample the original data to conform with a prediction grid and/or
91 calculate the annual means for the fiscal year 2013 from the data with finer
92 temporal resolution (e.g., monthly).

93 For the determination of the resolution of the prediction grid, we calculate
94 the distance to the nearest monitoring station for each station in the air quality
95 data because the prediction grid with much finer resolution than the distances
96 to the closest stations is not appropriate for a reliable estimation. The median
97 of the nearest distance for $\text{PM}_{2.5}$ and NO_2 are 7.2 and 4.1 km respectively. In
98 consideration of these distances, we construct a 4×4 km resolution prediction
99 grid on the land area in the study area. The predictor variables are also prepared
100 as a 4×4 km resolution grid data.

101 As for the emission sources, build-up and agricultural area ratio in a grid
102 cell are calculated from land use data obtained from Global Map Japan version
103 1.2.1 downloaded from Geospatial Information Authority of Japan (GSI). The
104 population data is obtained from the National Census of the year 2010 through
105 the Statistics Bureau of Japan.

106 Transport is one of the emission sources of NO_x ($\text{NO} + \text{NO}_2$) as well as
107 $\text{PM}_{2.5}$, and the distance to a road is provided as a predictor variable. The
108 road network data is obtained from Global Map Japan version 2 downloaded
109 from GSI. In this data, road types are classified into three categories: highway,
110 primary and secondary. The distance to a road is calculated for each grid cell
111 centroid for each of these three categories. Likewise, road length is obtained
112 from the National Land Numeric Information Data downloaded through the
113 Japanese Ministry of Land, Infrastructure, Transportation and Tourism. This
114 road length data is classified into 10 categories depending on the road width.
115 We reclassify them into three new categories: road A (road width ≥ 19.5 m),
116 road B ($13 \leq \text{road width} < 19.5$ m) and road C ($5.5 \text{ m} \leq \text{road width} < 13$ m).
117 Only road B and C are provided as predictor variables because most grid cells

118 in the study area have no value of road A.

119 When typical land and sea breezes dominated, polluted air parcels are trans-
120 ported from industrial or urban areas in coastal regions to inland areas and
121 O_3 concentrations increase via a photochemical reaction during transporta-
122 tion (Kannari and Ohara, 2010). A portion of $PM_{2.5}$ is also formed via a
123 photochemical reaction. Therefore, we use distance to coastline as a predictor
124 variable for $PM_{2.5}$. This distance is calculated for each grid cell centroid as the
125 nearest straight-line distance to coastline, which is obtained from Global Map
126 Japan version 2.

127 The relationship between the ground-level concentrations of $PM_{2.5}$ and satel-
128 lite based aerosol optical depth (AOD) has been widely investigated and used
129 to estimate the spatial distribution of $PM_{2.5}$ (e.g., Wang and Christopher, 2003;
130 van Donkelaar et al., 2010). AOD is also utilized as a predictor variable for LUR
131 models (e.g., Kloog et al., 2011; Mao et al., 2012; Xie et al., 2015). We obtain
132 daily AOD (500 nm) from Japan Aerospace Exploration Agency (JAXA) Satel-
133 lite Measurements for Environmental Studies (JASMES) products courtesy of
134 JAXA/Tokai University.

135 As for the meteorological parameters, we utilize daily mean observations
136 of precipitation, temperature and wind speed from Automated Meteorologi-
137 cal Data Acquisition System (AMeDAS) maintained by Japan Meteorological
138 Agency. The monitoring stations of AMeDAS are densely and homogeneously
139 distributed. The number of stations monitoring precipitation, temperature and
140 wind speed in the study area are 1235, 843 and 871 respectively. The mean dis-
141 tance to the nearest neighbouring station is approximately 16 km with the range
142 from 1 to 42 km for the three parameters. We interpolate the measurements
143 of each of the parameters by ordinary kriging to obtain 4×4 km resolution grid
144 data.

145 Aikawa et al. (2010) observed negative correlation between longitude and
146 particulate sulfate in Japan, which is one of the constituents of $PM_{2.5}$, and re-
147 produced this longitudinal gradient by chemical transport model. Shimadera
148 et al. (2016) also showed the longitudinal gradient both in the observed and

149 simulated concentrations of PM_{2.5}. In both studies, the influence of long range
 150 transport from the Asian continent was suggested. Therefore, longitude is pro-
 151 vided as a potential predictor variable for PM_{2.5}.

152 2.3. Spatial outlier detection

153 We use the spatial outlier detection method proposed by Lark (2000, 2002)
 154 to identify spatial outliers.

155 Firstly, the data are checked if transformation is necessary. We follow the
 156 method proposed by Rawlins et al. (2005); octile skewness (OC) (Brys et al.,
 157 2004) is calculated and if it is smaller than -0.2 or larger than 0.2, then natural
 158 logarithm transformation is applied. Octile skewness is a measure of asymmetry
 159 that is insensitive to outlying values (Rawlins et al., 2005), obtained by

$$OC = \frac{(Q_{0.875} - Q_{0.5}) - (Q_{0.5} - Q_{0.125})}{Q_{0.875} - Q_{0.125}}, \quad (1)$$

160 where Q_q is q-quantile of the data. Next, variogram is estimated using Math-
 161 eron's estimator (Matheron, 1962),

$$2\hat{\gamma}_M(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2, \quad (2)$$

162 where $z(\mathbf{x}_i)$ is an observed value at location $\mathbf{x}_i, i = 1, 2, \dots, N(\mathbf{h})$, $\mathbf{h}$ is a separa-
 163 tion vectors. We set the cut-off distance to 80 km consisting of 15 lags (meaning
 164 that each lag width is approximately 5 km) with the intention to detect spatial
 165 outliers at a similar spatial scale as our prediction grid size of 4 km. Spherical
 166 and exponential models are fitted to the estimated variogram by weighted least
 167 squares, and one model is selected based on the residual mean square from the
 168 fitting (Lark, 2000). Leave-one-out cross validation is then carried out with the
 169 selected model. In this method, one measurement point is removed and then
 170 the concentration at that point is predicted by using the rest of the points. This
 171 procedure is repeated for all measurement points. The statistic $\theta(\mathbf{x})$ is defined
 172 as

$$\theta(\mathbf{x}_i) = \frac{\{z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)\}^2}{\sigma^2(\mathbf{x}_i)}, \quad (3)$$

173 where $\hat{Z}(\mathbf{x}_i)$ is the kriged estimate and $\sigma^2(\mathbf{x}_i)$ is an associated kriging vari-
 174 ance (Lark, 2000). If the variogram is correct, $\theta(\mathbf{x})$ will be distributed as χ^2
 175 with one degree of freedom and the median of $\theta(\mathbf{x})$ is 0.455 (Lark, 2000). The
 176 upper and lower confidence limit for the median of $\theta(\mathbf{x})$ is calculated using
 177 variance,

$$\sigma_{\theta}^2 = \frac{1}{8nf(\tilde{x})^2}, \quad (4)$$

178 where $f(\tilde{x})$ is a probability function of $\theta(\mathbf{x})$ with a sample of $2n + 1$ data (Lark,
 179 2000). If the median of $\theta(\mathbf{x})$ is inside a 95% confidence interval, the Math-
 180 eron's estimator is used during the following steps. Otherwise, it is significantly
 181 influenced by spatial outliers and robust estimators are used instead.

182 We use three robust estimators (Lark, 2000, 2002; Rawlins et al., 2005): The
 183 first is Cressie and Hawkins' estimator (Cressie and Hawkins, 1980),

$$\hat{\gamma}_{CH}(\mathbf{h}) = \frac{\left\{ \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} |z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})|^{\frac{1}{2}} \right\}^4}{0.457 + \frac{0.494}{N(\mathbf{h})} + \frac{0.045}{N^2(\mathbf{h})}}. \quad (5)$$

184 The second is Dowd's estimator (Dowd, 1984),

$$2\hat{\gamma}_D(\mathbf{h}) = 2.198 \left\{ \text{median} \left(|z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})| \right) \right\}^2, \quad (6)$$

185 where 2.198 is a scale estimator, and the third is Genton's estimator (Genton,
 186 1998),

$$2\hat{\gamma}_G(\mathbf{h}) = \left(2.219 \left\{ |y_i(\mathbf{h}) - y_j(\mathbf{h})|; i < j \right\}_{\binom{H}{2}} \right)^2, \quad (7)$$

187 where 2.219 is a scale estimator, $y_i(\mathbf{h}) = z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})$, $i = 1, 2, \dots, N(\mathbf{h})$ and
 188 H is integer part $(n/2) + 1$.

189 Model fitting and selection is carried out for each estimator in the same
 190 way for the Matheron's described above. The median of $\theta(\mathbf{x})$ is obtained for
 191 each estimator by leave-one-out cross validation. The robust estimator with a
 192 median value of $\theta(\mathbf{x})$ closest to 0.455 is selected.

193 Rawlins et al. (2005) classified an observation as a spatial outlier (large) if

194 the standardized kriging error,

$$SKE = \frac{\hat{Z}(\mathbf{x}_i) - z(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}, \quad (8)$$

195 is less than -1.96 , that is, if it falls below the lower 95% confidence limit.
196 Because air quality data may contain both large and small outliers, we identify
197 an observation as a spatial outlier if $\theta(\mathbf{x}_i)$ i.e., squared SKE, is larger than 3.84.

198 *2.4. Application of spatial outlier detection method*

199 We apply the spatial outlier detection method to every daily mean value
200 throughout a year and exclude the identified spatially outlying daily means
201 from the data set. The annual means are calculated from these outlier removed
202 daily values for each of the monitoring stations and the number of effective
203 daily values for each station is counted as well. The annual means with the
204 data coverage of more than 250 days a year remain in the data set, but others
205 are discarded to ensure the temporal representativeness. The remaining annual
206 values are in turn processed by the spatial outlier detection method again and
207 the identified outliers are removed. This is required because these annual means
208 are not automatically assured to be exclusive of spatial outliers especially when
209 a certain number of daily values are removed. The procedure described thus far
210 has an advantage of correcting annual means in addition to removing outlying
211 values, which would not be possible when the spatial outlier detection method
212 is applied only to annual means. In addition, annual means are also calculated
213 from the daily means including the detected outliers. In this case, the threshold
214 value of the data coverage of more than 250 days a year is also applied. The
215 data excluding spatial outliers as well as the raw annual mean data, which
216 may include spatial outliers, are provided for the analyses to evaluate the effect
217 of spatial outliers. The two data sets, one including spatial outliers and the
218 other excluding them, are hereinafter referred to as the inclusive data and the
219 exclusive data respectively.

220 2.5. LUR modelling and regression kriging

221 We build LUR models in a similar way as Araki et al. (2015). Candidates for
222 predictor variables of linear regression models for each pollutant are presented
223 in Table 2 with the pre-specified direction of effect according to the physical or
224 chemical relationship between the pollutants and the predictor variables (Beelen
225 et al., 2009). A linear regression model is developed using backward stepwise
226 procedure to select the significant variables (Hengl, 2007). The selected vari-
227 ables that have coefficients that conformed to the pre-specified direction of effect
228 are retained in the final linear regression model, but others are discarded (Bee-
229 len et al., 2009). The residuals of the LUR model are interpolated by ordinary
230 kriging. Empirical variogram of the residuals is obtained by Matheron’s estima-
231 tor with a cut-off distance of 80 km consisting of 15 lags in consideration of the
232 resolution of our prediction grid size of 4 km. Spherical and exponential models
233 are fitted to the estimated variogram by weighted least squares, and one model
234 is selected based on the residual mean square from the fitting (Lark, 2000).
235 The concentrations of pollutants are transformed to a natural logarithmic scale
236 before analysis, and the predictions are back transformed after analysis. This
237 procedure has the advantage that predicted concentrations are positive, which
238 is found not to be the case when analyses are performed without transforma-
239 tion (Beelen et al., 2009).

240 2.6. Evaluation

241 For evaluating the effect of spatial outliers, we carry out leave-one-out cross
242 validation and compute root mean squared error (RMSE) and r^2 between the
243 predicted and measured values as indicators of the prediction accuracy. RMSE
244 should be as small as possible. In the case of the exclusive data, the results
245 at every point are used to calculate the indicators. In the case of the inclusive
246 data, on the other hand, only the results at non-outlying points are used to
247 compute the indicators. That is, the prediction accuracy at non-outlying points
248 is assessed using non-outliers as well as spatial outliers, but accuracy at spatially
249 outlying points are not considered. When the corresponding indicators differ

250 between the two cases, the difference can be interpreted as the effect of spatial
251 outliers on the quality of prediction.

252 The difference is statistically evaluated using standard F -test, that evaluates
253 whether the two cases have the same variance, i.e. RMSE, assuming that the
254 mean error (ME) are the same (Hengl et al., 2015). The ME of the two cases
255 are evaluated by standard t -test if they are the same (Hengl et al., 2015).

256 Data analysis is carried out using R statistical software 3.2.5 (R Core Team,
257 2016) with the raster package (Hijmans, 2015) for the integration and construc-
258 tion of the grid data of predictor variables and with the gstat package (Pebesma,
259 2004) for the performance of kriging.

260 3. Results

261 3.1. Spatial outlier detection

262 The results of the spatial outlier detection are presented in Table 3. The
263 number of valid observations in the inclusive and exclusive data is 500 and 457
264 respectively for $\text{PM}_{2.5}$, and 1278 and 1155 respectively for NO_2 . Thus, the num-
265 ber of spatial outliers in the inclusive data is 43 and 123 for $\text{PM}_{2.5}$ and NO_2
266 respectively. The number of monitoring locations where annual mean observa-
267 tions of $\text{PM}_{2.5}$ and NO_2 are simultaneously detected as spatial outlier is 5, and
268 no clear correlation in the locations of outliers between $\text{PM}_{2.5}$ and NO_2 is rec-
269 ognized. The ratio of spatial outliers are similar between the two pollutants: 8.6
270 and 9.6% for $\text{PM}_{2.5}$ and NO_2 respectively. The distributions of the spatial out-
271 liers and non-outliers for both pollutants are presented in Fig. 1. Although the
272 ratio of the detected spatial outliers is higher in the lower and higher concentra-
273 tions, they are generally distributed throughout the range of the concentrations
274 for both pollutants. That is, some observations in midrange in the data are de-
275 tected as spatial outliers. This can be realized because spatial relationship and
276 dissimilarity of observations in neighbourhood areas are considered: absolute
277 differences in concentrations between observations are evaluated based on their
278 relative distances in kriging framework. This result demonstrates the advantage

279 of the method applied here over a statistical method where spatial positions are
280 not considered.

281 The comparison of the annual means between the inclusive and exclusive
282 data are given in Fig. 2. RMSE denotes root squared mean error and MAE
283 denotes mean absolute error. The differences between the inclusive and exclu-
284 sive data are basically small for most of the values, but remarkable for some
285 observations.

286 3.2. $PM_{2.5}$

287 The retained predictor variables and their coefficients, and statistical indi-
288 cators for $PM_{2.5}$ for each of the two data sets are given in Table 4. Distance
289 to highway is retained in the final regression models, but other traffic related
290 variables such as distance to primary/secondary road and road length B/C are
291 discarded. On the other hand, the meteorological variables such as precipi-
292 tation, temperature and wind speed are all retained in the models. AOD is
293 discarded during the backward stepwise procedure in spite of some successful
294 applications in LUR modelling (e.g., Kloog et al., 2011; Mao et al., 2012; Xie
295 et al., 2015). We calculate annual mean AOD by simply averaging daily values
296 and missing values are omitted from the calculation. Consequently, an aver-
297 aged value at a pixel with a lot of missing daily values may not appropriately
298 represent the annual mean. Moreover, calibration might be necessary to better
299 correlate with $PM_{2.5}$ concentrations because the relationship between AOD and
300 $PM_{2.5}$ concentrations can vary over space and time (Kloog et al., 2012). The
301 retained variables are the same for the both data sets, although no restriction is
302 implemented to select the same variables. The coefficients of the variables are
303 generally similar to the corresponding ones in the other data set.

304 Empirical and fitted variograms of the residuals of LUR models for both data
305 are given in Fig.3. The clearer spatial correlation is identified for the exclusive
306 data set. The semivariance ($\hat{\gamma}(\mathbf{h})$) at the corresponding distances is larger for
307 the inclusive data than that for the exclusive data.

308 The scatter plots of the predicted and observed concentrations obtained by

cross validation are presented in Fig. 4. The left and right panels are obtained with the inclusive and exclusive data respectively. The upper and lower panels are the results by LUR model and regression kriging respectively. The light and dark dots represent non-spatial outliers and spatial outliers respectively. RMSE and r^2 between the predicted and observed values for non-outlying points are presented in each panel.

Spatial outliers increase RMSE by 17% and decrease r^2 by 0.07 for the predictions by LUR model, and increase RMSE by 40% and decrease r^2 by 0.15 for the predictions by regression kriging. The t -test results show that the differences in ME between the two cases are not statistically significant ($p > 0.05$) both for LUR model and regression kriging. The F -test results indicate that the differences in RMSE between the two cases are statistically significant at the 5% level both for LUR model and regression kriging. These results indicate that spatial outliers degrade the prediction quality of LUR as well as regression kriging. No remarkable over or under estimation is recognized for the results obtained with the exclusive data.

The spatial distribution of $PM_{2.5}$ is estimated by LUR and regression kriging respectively, for each of the data set. ME and absolute mean error (AME) between the estimation with inclusive and exclusive data are calculated for LUR and regression kriging respectively. ME is 0.3 and AME is $0.4 \mu g m^{-3}$ for LUR, and ME is 0.1 and AME is $1.1 \mu g m^{-3}$ for regression kriging. These values are biases in the estimations brought by spatial outliers. Fig. 5 illustrates the spatial distribution of $PM_{2.5}$ predicted by regression kriging with the inclusive and exclusive data respectively. The locations of the detected spatial outliers are given in these maps. These maps share features in common with those obtained by LUR (not shown here). The estimation map obtained using the exclusive data is more smoothed than that using the inclusive data due to the removal of spatial outliers.

3.3. NO_2

The retained predictor variables and their coefficients for NO_2 for each of the two data sets are given in Table 5. The retained variables in the final model are the same for both data sets, although no constraint is imposed to select the same variables; all the potential predictor variables are retained except for distance to highway and road length C. The coefficients of the predictor variables are similar to the corresponding ones in the other cases.

Empirical and fitted variograms of the residuals of LUR models for the two data sets are given in Fig. 6, where the spatial correlation is clearly identified. Semivariance at the corresponding distance is generally similar between the two data sets, but that for the exclusive data is smaller.

The scatter plots of the predicted and observed concentrations of NO_2 obtained by cross validation are given in Fig. 7. The left and right panels are obtained with the inclusive and exclusive data respectively. The upper and lower panels are the results using LUR model and regression kriging respectively. The light and dark dots represent non-spatial outliers and spatial outliers respectively. RMSE and r^2 between the predicted and observed values only for non-outlying points are presented in each panel.

Spatial outliers increase RMSE by 3% and decrease r^2 by 0.01 for the predictions using LUR model, and increase RMSE by 19% and decrease r^2 by 0.06 for the predictions using regression kriging. The t -test results show that the differences in ME between the two cases are not statistically significant ($p > 0.05$) both for LUR model and regression kriging. The F -test results indicate the difference in RMSE between the two cases are statistically significant at the 5% level for regression kriging, but not for LUR model. These results indicate that the spatial outliers provide limited influence on the estimation by LUR model but rather degrade the quality of prediction of regression kriging. From the result obtained by regression kriging with the exclusive data, no over or under estimation is recognized.

The spatial distribution of NO_2 is estimated by LUR and regression kriging respectively, for each of the data set. ME and AME between the estimation

with inclusive and exclusive data are calculated for LUR and regression kriging respectively. ME is 0.1 and AME is 0.1 ppb for LUR, and ME is 0.2 and AME is 0.6 ppb for regression kriging. The spatial outliers cause these biases in the estimations. Fig. 8 illustrates the spatial distribution of NO_2 predicted by regression kriging with the inclusive and exclusive data respectively. These maps also show the locations of the detected spatial outliers. These maps share features in common with those obtained by LUR (not shown here). There is little qualitative difference in the predicted maps.

4. Discussion

4.1. Difference between $\text{PM}_{2.5}$ and NO_2

Although the spatial outliers influence the prediction quality both of $\text{PM}_{2.5}$ and NO_2 , there are some differences in the effects. First, spatial outliers degrade the prediction accuracy of LUR model for $\text{PM}_{2.5}$, but not for NO_2 . Second, spatial outliers considerably increase semivariance at the corresponding distance for $\text{PM}_{2.5}$, but marginally for NO_2 . Third, spatial outliers deteriorate the prediction quality of regression kriging for $\text{PM}_{2.5}$ more than that for NO_2 .

Some of the spatially outlying observations of $\text{PM}_{2.5}$ are outlying in the regression model as well (upper right panel of Fig 4). These outlying values worsen the statistical indicators of the LUR model. On the contrary, the spatial outliers of NO_2 are not necessarily outliers in the regression model (upper right panel of Fig 7). Hence, spatial outliers do not affect the resulting LUR model and, consequently, the statistical indicators of LUR models are almost identical between the inclusive and exclusive data as shown in Fig 7. Also, the difference in the estimation maps is minor. Similar LUR models of NO_2 result in similar residuals, and the variograms of the residuals are generally alike. On the other hand, the better LUR model of $\text{PM}_{2.5}$ with the exclusive data result in the more distinct spatial dependency in the residuals of the regression model. This leads to larger difference in the quality of prediction of regression kriging for $\text{PM}_{2.5}$ than that for NO_2 .

397 There are differences in characteristics between $\text{PM}_{2.5}$ and NO_2 . NO_2 is a
398 single substance, while $\text{PM}_{2.5}$ consists of various substances such as elemental
399 carbon, organic carbon, sulfate, nitrate, and metal compounds. Because of this
400 feature, positive and negative artifacts have been reported (e.g., Chow et al.,
401 2010; Liu et al., 2014). Therefore, observations of $\text{PM}_{2.5}$ could be more biased
402 than those of NO_2 .

403 The feature of the spatial distribution of the two pollutants is somewhat
404 different because of their inherent characteristics. High concentration areas for
405 $\text{PM}_{2.5}$ are widely distributed (Fig. 5). On the other hand, those for NO_2 are
406 focused in urban areas such as metropolitan Tokyo and along major highways
407 (Fig. 8) generally reflecting the distribution of emission sources, and the spatial
408 variability at a local scale is larger than that of $\text{PM}_{2.5}$. Hence, the spatial
409 resolution of 4 km could be better suited for $\text{PM}_{2.5}$ than for NO_2 and the effect
410 of spatial outlier for NO_2 might be different with a finer spatial resolution. These
411 differences in characteristics between $\text{PM}_{2.5}$ and NO_2 might contribute to the
412 discrepancies in the effects of the spatial outliers on the prediction quality of
413 LUR model and regression kriging.

414 Regarding the temporal trend in a year, both $\text{PM}_{2.5}$ and NO_2 show gen-
415 eral tendency of higher concentrations in winter possibly due to frequent stable
416 conditions. The concentrations of $\text{PM}_{2.5}$ increase via a photochemical reaction
417 during summer, which is not the case for NO_2 . Also, the contribution of long
418 range transport from the Asian continent to $\text{PM}_{2.5}$ concentrations in Japan
419 is substantial particularly in winter and spring, which is attributed in part to
420 higher concentrations of $\text{PM}_{2.5}$ in these seasons (Shimadera et al., 2016). On the
421 other hand, the contribution to NO_2 is negligible throughout a year (Shimadera
422 et al., 2016). Thus, the temporal trend of $\text{PM}_{2.5}$ is not consistent with that
423 of NO_2 . However, we use annual means and the dissimilarity of the temporal
424 variability in a year between $\text{PM}_{2.5}$ and NO_2 might be averaged out and have
425 limited influence on the effect of outliers studied.

4.2. Number of observations

The other remarkable difference between $\text{PM}_{2.5}$ and NO_2 is the number of valid observations in the study area; 500 for $\text{PM}_{2.5}$, while 1278 for NO_2 . In order to examine whether the number of observations differentiate the effect of spatial outliers on the quality of prediction, we extract the NO_2 monitoring stations where $\text{PM}_{2.5}$ is monitored simultaneously from the inclusive and exclusive data, and obtain the statistical indicators by leave-one-out cross validation for each of the two data sets.

The number of NO_2 observations in the subset are 478 and 402 for the inclusive and exclusive data respectively. These numbers are smaller than the corresponding ones of $\text{PM}_{2.5}$. This is because some of the stations monitor only $\text{PM}_{2.5}$. The results are given in Table 6. The retained variables in the final models are slightly different from those obtained by each of the full NO_2 data sets. Spatial outliers increase RMSE by 7% and decrease r^2 by 0.02 for the predictions by LUR model, and increase RMSE by 32% and decrease r^2 by 0.08 for the predictions by regression kriging. The marginal influence of spatial outliers on the indicators of LUR model and moderate effect on those of regression kriging are also observed with the full data set as described in 4.1. Therefore, the number of observations has limited influence on the effect of spatial outliers and the discrepancies in the effects between $\text{PM}_{2.5}$ and NO_2 is not explained by the difference in the number of observations.

4.3. Further requirements

We applied the spatial outlier detection method to a large number of observations and successfully detected spatial outliers. A sufficient number of observations are necessary for the application of this method because it is based on variogram analysis. With insufficient number of observations, variogram would not appropriately capture the spatial dependency in the domain of interest, which could lead to a false detection of spatial outlier. There is no threshold or guideline for the necessary number of observations to estimate proper variogram; it generally depends on each specific case. Therefore, it should be applied

carefully to a smaller number of observations, which is often the case with epidemiological studies for evaluating the individual exposure level at an urban or intra-urban scale. Meanwhile, spatial outliers could be more influential for data with a smaller number of observations and they should be excluded to gain an overall mapping accuracy as long as appropriate detection is possible. Thus, further investigation and evaluation of the application to a smaller network at smaller spatial scale is required. Also, examination with a finer prediction grid might be required.

Spatial outliers have little influence on the quality of NO_2 prediction by LUR model. However, this does not necessarily suggest that removing spatial outliers is unneeded in this case. The LUR predictions of NO_2 correlate less with observations than those of $\text{PM}_{2.5}$ as given in Fig 4 and Fig 7. Therefore, the effect of spatial outliers needs to be further evaluated using better LUR model obtained with additional or alternative covariates.

As already noted, the estimated map using the data excluding spatial outliers can be interpreted as background or baseline concentration map. Observations at "hot spots" are probably excluded by the spatial outlier detection method. Observations might be spatially outlying due to influences of nearby emissions, local terrain, meteorology and/or biased monitors due to mechanical or electrical malfunction. When a monitor is biased, observations obtained by the monitor should be removed because it does not correctly measure concentrations. In the other cases mentioned above, concentrations are correctly measured but represent smaller spatial extent compared to non-outliers, thus cannot be treated equally as non-outliers. The estimation with the data including outliers could degrade the LUR model quality and, consequently, exaggerate the entire estimation uncertainty. Although removing such outliers could result in over/under estimation around the locations of the removed points, this procedure can reduce the overall mapping uncertainty and improve the total estimation accuracy. Therefore, excluding spatial outliers is a reasonable approach. This does not mean that those observations are unimportant, but they may contain important information and can be useful in a different context.

487 The locations of the detected spatial outliers are inspected, but a potential
488 reason such as a near-by emission source, local topology or meteorology is not
489 clear. The possible reasons should further be investigated, which could be of
490 benefit for a better design of a monitoring network.

491 5. Conclusion

492 We applied the spatial outlier detection method to the observations of $\text{PM}_{2.5}$
493 and NO_2 obtained from the regulatory monitoring network in Japan, and spatial
494 outliers were identified. Some observations in midrange are detected as outliers
495 because dissimilarity of observations in neighbourhood is evaluated in kriging
496 framework. The effect of spatial outliers was assessed by comparison of the
497 prediction performance of LUR and regression kriging on the data inclusive and
498 exclusive of spatial outliers respectively. Spatial outliers deteriorate the quality
499 of prediction except for LUR model of NO_2 . Although further investigation is
500 required, our study demonstrated that the spatial outlier detection method is an
501 effective procedure for air pollutant data when certain spatial representativeness
502 is required and that it should be applied when observation based prediction
503 methods are used to generate concentration maps. The observations exclusive
504 of spatial outliers are also of benefit for validation of CTMs, where simulated
505 concentrations are mean values in each grid cell and observations are required
506 for the equivalent spatial representativeness.

507 Appendix

Data sources.

Air quality data	http://www.nies.go.jp/igreen/
Global Map Japan	http://www.gsi.go.jp/kankyochiri/gm_japan_e.html
Population	http://e-stat.go.jp/SG2/eStatGIS/page/download.html
Road length	http://nlftp.mlit.go.jp/ksj-e/gml/datalist/KsjTmplt-N04.html
AOD	http://kuroshio.eorc.jaxa.jp/JASMES/index.html
Meteorological data	http://www.data.jma.go.jp/gmd/risk/obsdl/index.php

508

509 References

- 510 Aikawa, M., Ohara, T., Hiraki, T., Oishi, O., Tsuji, A., Yamagami, M., Murano,
511 K., Mukai, H., 2010. Significant geographic gradients in particulate sulfate
512 over Japan determined from multiple-site measurements and a chemical trans-
513 port model: Impacts of transboundary pollution from the Asian continent.
514 Atmos. Environ. 44, 381–391. doi:10.1016/j.atmosenv.2009.10.025.
- 515 Araki, S., Yamamoto, K., Kondo, A., 2015. Application of Regression Kriging to
516 Air Pollutant Concentrations in Japan with High Spatial Resolution. Aerosol
517 and Air Quality Research 15, 234–241. doi:10.4209/aaqr.2014.01.0011.
- 518 Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., de Hoogh, K., Briggs, D.J.,
519 2009. Mapping of background air pollution at a fine spatial scale across the
520 European Union. Science of the Total Environment 407, 1852–1867. doi:10.
521 1016/j.scitotenv.2008.11.048.
- 522 Briggs, D.J., De Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham,
523 S., Smallbone, K., 2000. A regression-based method for mapping traffic-
524 related air pollution: Application and testing in four contrasting urban en-
525 vironments. Science of the Total Environment 253, 151–167. doi:10.1016/
526 S0048-9697(00)00429-0.

- 527 Brys, G., Hubert, M., Struyf, A., 2004. A robust measure of skewness. Jour-
528 nal of Computational and Graphical Statistics 13, 996–1017. doi:10.1198/
529 106186004X12632.
- 530 Chatani, S., Morino, Y., Shimadera, H., Hayami, H., Mori, Y., Sasaki, K.,
531 Kajino, M., Yokoi, T., Morikawa, T., Ohara, T., 2014. Multi-model analyses
532 of dominant factors influencing elemental carbon in Tokyo Metropolitan Area
533 of Japan. Aerosol and Air Quality Research 14, 396–405. doi:10.4209/aaqr.
534 2013.02.0035.
- 535 Chow, J.C., Watson, J.G., Chen, L.W.A., Rice, J., Frank, N.H., 2010.
536 Quantification of PM_{2.5} organic carbon sampling artifacts in US net-
537 works. Atmospheric Chemistry and Physics 10, 5223–5239. doi:10.5194/
538 acp-10-5223-2010.
- 539 Cressie, N., Hawkins, D.M., 1980. Robust estimation of the variogram: I. Jour-
540 nal of the International Association for Mathematical Geology 12, 115–125.
541 doi:10.1007/BF01035243.
- 542 van Donkelaar, A., Martin, R.V., Brauer, M., Kahn, R., Levy, R., Verduzco, C.,
543 Villeneuve, P.J., 2010. Global estimates of ambient fine particulate matter
544 concentrations from satellite-based aerosol optical depth: Development and
545 application. Environmental Health Perspectives 118, 847–855. doi:10.1289/
546 ehp.0901623.
- 547 Dowd, P.A., 1984. Geostatistics for Natural Resources Characterization:
548 Part 1. Springer Netherlands, Dordrecht. chapter The Variogram and
549 Kriging: Robust and Resistant Estimators. pp. 91–106. doi:10.1007/
550 978-94-009-3699-7_6.
- 551 Emmons, L.K., Walters, S., Hess, P.G., Lamarque, J.F., Pfister, G.G., Fill-
552 more, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando,
553 J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S.L., Kloster, S.,

2010. Description and evaluation of the model for ozone and related chemical tracers, version 4 (mozart-4). *Geoscientific Model Development* 3, 43–67. doi:10.5194/gmd-3-43-2010.

Genton, M.G., 1998. Highly Robust Variogram Estimation. *Mathematical Geology* 30, 213–221.

Hengl, T., 2007. A Practical Guide to Geostatistical Mapping of Environmental Variables. Office for Official Publications of the European Communities, Luxembourg.

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of africa at 250 m resolution: Random forests significantly improve current predictions. *PLOS ONE* 10, 1–26. doi:10.1371/journal.pone.0125814.

Hijmans, R.J., 2015. raster: Geographic Data Analysis and Modeling. URL: <https://CRAN.R-project.org/package=raster>. r package version 2.5-2.

Kannari, A., Ohara, T., 2010. Theoretical implication of reversals of the ozone weekend effect systematically observed in Japan. *Atmospheric Chemistry and Physics* 10, 6765–6776. doi:10.5194/acp-10-6765-2010.

Kloog, I., Koutrakis, P., Coull, B.A., Lee, H.J., Schwartz, J., 2011. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment* 45, 6267–6275. doi:10.1016/j.atmosenv.2011.08.066.

Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2012. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the mid-atlantic states. *Environmental Science and Technology* 46, 11913–11921. doi:10.1021/es302673e.

580 Lark, R.M., 2000. A comparison of some robust estimators of the variogram
581 for use in soil survey. *European Journal of Soil Science* 51, 137–157. doi:10.
582 1046/j.1365-2389.2000.00280.x.

583 Lark, R.M., 2002. Modelling complex soil properties as contaminated region-
584 alized variables. *Geoderma* 106, 173–190. doi:10.1016/S0016-7061(01)
585 00123-9.

586 Lark, R.M., Dove, D., Green, S.L., Richardson, A.E., Stewart, H., Stevenson,
587 A., 2012. Spatial prediction of seabed sediment texture classes by cokriging
588 from a legacy database of point observations. *Sedimentary Geology* 281, 35–
589 49. doi:10.1016/j.sedgeo.2012.07.009.

590 Liu, C.N., Lin, S.F., Awasthi, A., Tsai, C.J., Wu, Y.C., Chen, C.F., 2014.
591 Sampling and conditioning artifacts of PM_{2.5} in filter-based samplers. *Atmo-*
592 *spheric Environment* 85, 48–53. doi:10.1016/j.atmosenv.2013.11.075.

593 Mao, L., Qiu, Y., Kusano, C., Xu, X., 2012. Predicting regional space-
594 time variation of PM_{2.5} with land-use regression model and MODIS data.
595 *Environmental Science and Pollution Research* 19, 128–138. doi:10.1007/
596 s11356-011-0546-9.

597 Matheron, G., 1962. *Traité de géostatistique appliquée*, Tome 1. Editions Tech-
598 nip. p. 333.

599 Pearce, J.L., Rathbun, S.L., Aguilar-Villalobos, M., Naeher, L.P., 2009. Charac-
600 terizing the spatiotemporal variability of PM_{2.5} in Cusco, Peru using kriging
601 with external drift. *Atmospheric Environment* 43, 2060–2069. doi:10.1016/
602 j.atmosenv.2008.10.060.

603 Pebesma, E.J., 2004. Multivariable geostatistics in s: the gstat package. *Com-*
604 *puters and Geosciences* 30, 683–691.

605 R Core Team, 2016. R: A Language and Environment for Statistical Computing.
606 R Foundation for Statistical Computing. Vienna, Austria. URL: [https://](https://www.R-project.org/)
607 www.R-project.org/.

608 Rawlins, B., Lark, R., O'Donnell, K., a.M. Tye, Lister, T., 2005. The as-
609 sessment of point and diffuse metal pollution of soils from an urban geo-
610 chemical survey of Sheffield, England. *Soil Use and Management* , 353–
611 362doi:10.1079/SUM2005335.

612 Ross, Z., Jerrett, M., Ito, K., Tempalski, B., Thurston, G.D., 2007. A land use
613 regression for predicting fine particulate matter concentrations in the New
614 York City region. *Atmospheric Environment* 41, 2255–2269. doi:10.1016/j.
615 *atmosenv*.2006.11.012.

616 Sampson, P.D., Richards, M., Szpiro, A.A., Bergen, S., Sheppard, L., Lar-
617 son, T.V., Kaufman, J.D., 2013. A regionalized national universal kriging
618 model using Partial Least Squares regression for estimating annual PM_{2.5}
619 concentrations in epidemiology. *Atmospheric Environment* 75, 383–392.
620 doi:10.1016/j.*atmosenv*.2013.04.015, *arXiv:NIHMS150003*.

621 Shimadera, H., Kojima, T., Kondo, A., 2016. Evaluation of Air Quality Model
622 Performance for Simulating Long-Range Transport and Local Pollution of
623 PM_{2.5} in Japan. *Advances in Meteorology* 2016. doi:10.1155/2016/5694251.

624 Sun, X.L., Zhao, Y.G., Wu, Y.J., Zhao, M.S., Wang, H.L., Zhang, G.L., 2012.
625 Spatio-temporal change of soil organic matter content of Jiangsu Province,
626 China, based on digital soil maps. *Soil Use and Management* 28, 318–328.
627 doi:10.1111/j.1475-2743.2012.00421.x.

628 Wang, J., Christopher, S., 2003. Intercomparison between satellite-derived
629 aerosol optical thickness and PM 2.5 mass: Implications for air quality stud-
630 ies. *Geophysical Research Letters* 30, 2095. doi:10.1029/2003GL018174.

631 Wu, J., Li, J., Peng, J., Li, W., Xu, G., Dong, C., 2014. Applying land use
632 regression model to estimate spatial variation of PM_{2.5} in Beijing, China.
633 *Environmental Science and Pollution Research* 3, 7045–7061. doi:10.1007/
634 *s11356-014-3893-5*.

- 635 Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B., Bai, Y., 2015. Daily Estima-
636 tion of Ground-Level PM_{2.5} Concentrations over Beijing Using 3 km Resolu-
637 tion MODIS AOD. *Environmental Science and Technology* 49, 12280–12288.
638 doi:10.1021/acs.est.5b01413.
- 639 Zhao, Y., Xu, X., Huang, B., Sun, W., Shao, X., Shi, X., Ruan, X.,
640 2007. Using robust kriging and sequential Gaussian simulation to delin-
641 eate the copper- and lead-contaminated areas of a rapidly industrialized
642 city in Yangtze River Delta, China. *Environmental Geology* 52, 1423–1433.
643 doi:10.1007/s00254-007-0667-0.

644 **Figure captions**

645 Fig. 1 The distributions of spatial outliers and non-outliers in the
646 annual means for 1) $\text{PM}_{2.5}$ and 2) NO_2 .

647 Fig. 2 The comparison of the annual means of the inclusive and
648 exclusive data for $\text{PM}_{2.5}$ and NO_2 . The concentrations, RMSE
649 and MAE are in unit of $\mu\text{g m}^{-3}$ for $\text{PM}_{2.5}$ and ppb for NO_2 .
650 RMSE donates root mean squared error. MAE donates mean
651 absolute error.

652 Fig. 3 Empirical (dot) and fitted (line) Variograms of the residuals
653 of LUR model of $\text{PM}_{2.5}$ estimated by Matheron's estimator for
654 the 1) inclusive and 2) exclusive data.

655 Fig. 4 Scatter plot of the observed and predicted concentrations of
656 $\text{PM}_{2.5}$ for each data set and for each estimation method obtained
657 by cross validation results. RMSE represents root mean squared
658 error in unit of $\mu\text{g m}^{-3}$. The light and dark dots represent non-
659 spatial outliers and spatial outliers respectively. r^2 and RMSE
660 are calculated by the results at non-outlying points.

661 Fig. 5 The prediction map of $\text{PM}_{2.5}$ obtained by regression kriging
662 with the inclusive and exclusive data. Unit is $\mu\text{g m}^{-3}$. The
663 symbols on the maps show the locations of the detected spatial
664 outliers.

665 Fig. 6 Empirical (dot) and fitted (line) Variograms of the residuals
666 of LUR model of NO_2 estimated by Matheron's estimator for the
667 1) inclusive and 2) exclusive data.

668 Fig. 7 Scatter plot of the observed and predicted concentrations of
 669 NO₂ for each data set and for each estimation method obtained
 670 by cross validation results. RMSE represents root mean squared
 671 error in unit of ppb. The light and dark dots represent non-
 672 spatial outliers and spatial outliers respectively. r^2 and RMSE
 673 are calculated by the results at non-outlying points.
 674 Fig. 8 The prediction map of NO₂ obtained by regression kriging
 675 with the inclusive and exclusive data. Unit is ppm. The symbols
 676 on the maps show the locations of the detected spatial outliers.

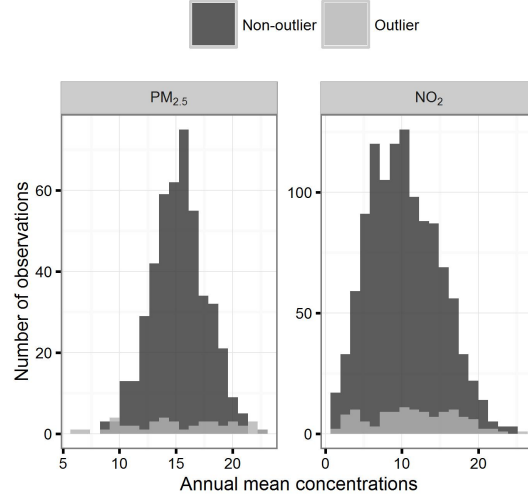


Figure 1: The distributions of spatial outliers and non-outliers in the annual means for 1) $PM_{2.5}$ and 2) NO_2 .

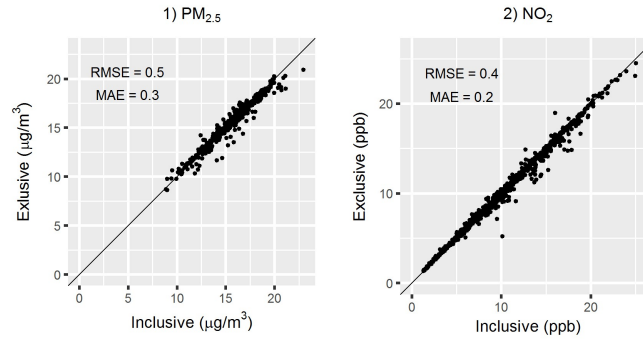


Figure 2: The comparison of the annual means of the inclusive and exclusive data for $PM_{2.5}$ and NO_2 . The concentrations, RMSE and MAE are in unit of $\mu g m^{-3}$ for $PM_{2.5}$ and ppb for NO_2 . RMSE donates root mean squared error. MAE donates mean absolute error.

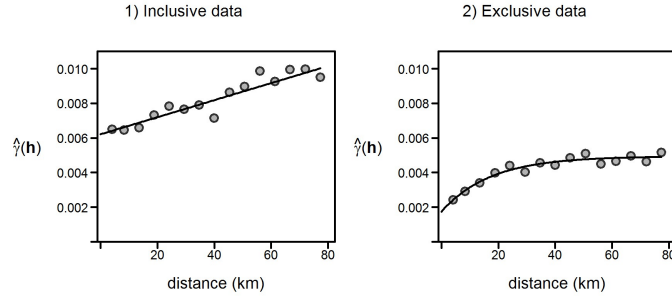


Figure 3: Empirical (dot) and fitted (line) Variograms of the residuals of LUR model of $\text{PM}_{2.5}$ estimated by Matheron's estimator for the 1) inclusive and 2) exclusive data.

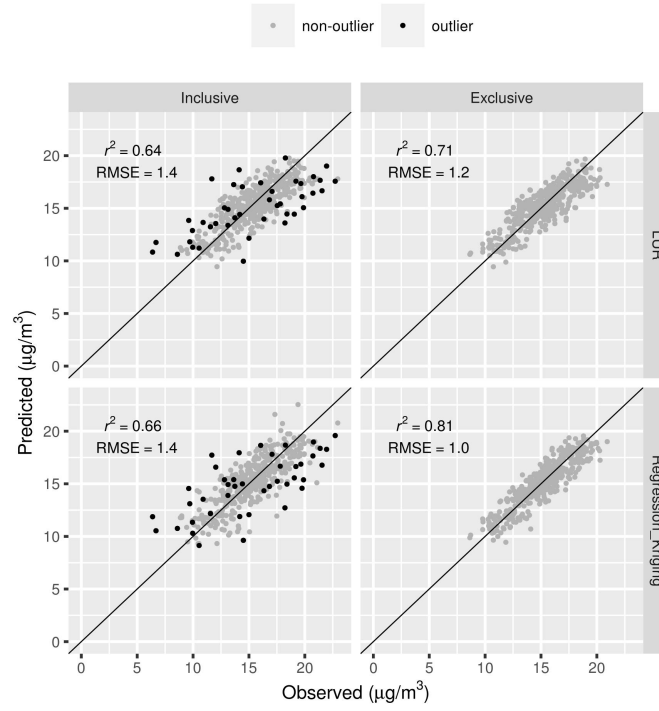


Figure 4: Scatter plot of the observed and predicted concentrations of $\text{PM}_{2.5}$ for each data set and for each estimation method obtained by cross validation results. RMSE represents root mean squared error in unit of $\mu\text{g m}^{-3}$. The light and dark dots represent non-spatial outliers and spatial outliers respectively. r^2 and RMSE are calculated by the results at non-outlying points.

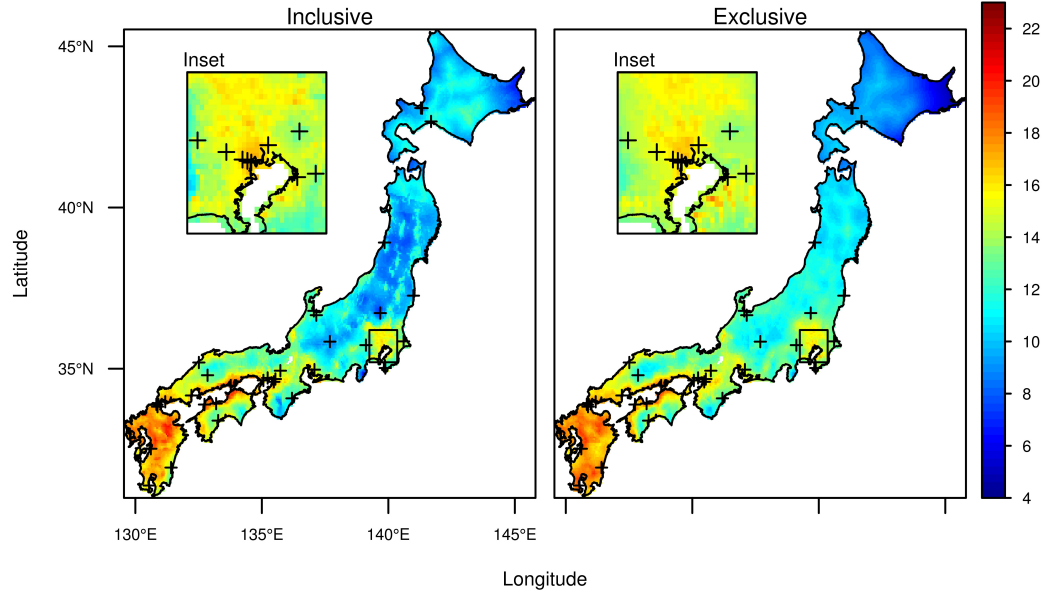


Figure 5: The prediction map of $PM_{2.5}$ obtained by regression kriging with the inclusive and exclusive data. Unit is $\mu g m^{-3}$. The symbols on the maps show the locations of the detected spatial outliers.

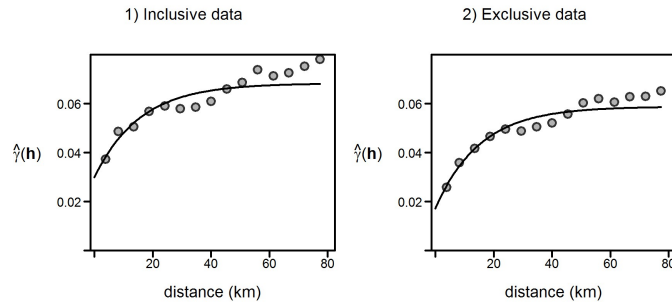


Figure 6: Empirical (dot) and fitted (line) Variograms of the residuals of LUR model of NO_2 estimated by Matheron's estimator for the 1) inclusive and 2) exclusive data.

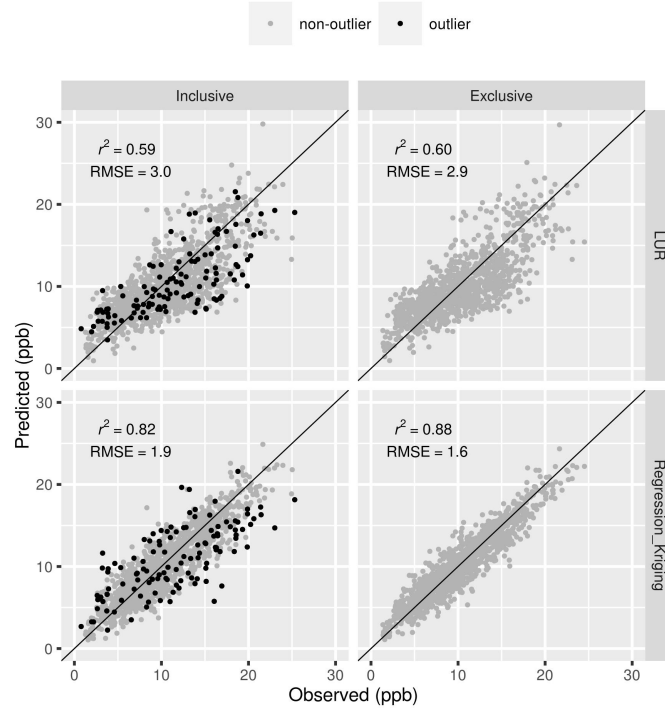


Figure 7: Scatter plot of the observed and predicted concentrations of NO₂ for each data set and for each estimation method obtained by cross validation results. RMSE represents root mean squared error in unit of ppb. The light and dark dots represent non-spatial outliers and spatial outliers respectively. r^2 and RMSE are calculated by the results at non-outlying points.

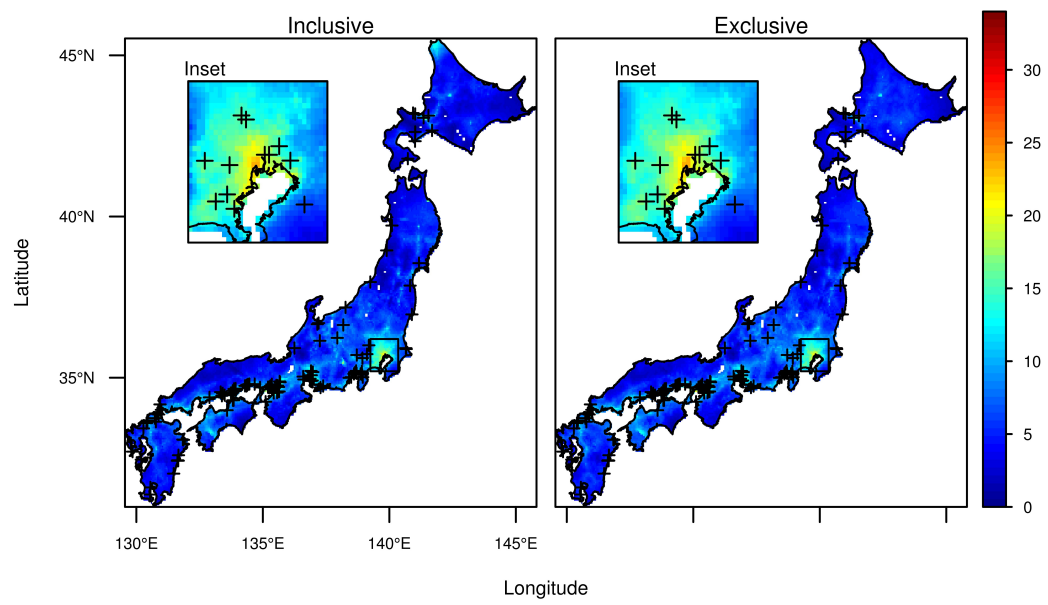


Figure 8: The prediction map of NO₂ obtained by regression kriging with the inclusive and exclusive data. Unit is ppm. The symbols on the maps show the locations of the detected spatial outliers.

Table 1: Summary of the data used in this study

Description	Source	Field	Spatial scale	Time periode
Monitored air quality data	Ministry of Environment	PM _{2.5} , NO ₂	point	2013
Global Map Japan	Geographical Information Authority of Japan	Land use	1 km	2006(ver.1.1)
		Road lines	Vector	2011(ver.2)
		Coast lines	Vector	2011(ver.2)
National Census	Statistics Bureau of Japan	Population	500 m	2010
National Land Numerical Information	Ministry of Land, Infrastructure, Transportation and Tourism	Road length	1 km	2010
JASMES Products	JAXA/Tokai University	AOD	1 km	2013
Amedas	Japan Meteorological Agency	Precipitation	point	2013
		Temperature		
		Wind speed		

Table 2: Predictor variables and predefined directions of effect.

Predictor variables	Unit	Air pollutants	
		PM _{2.5}	NO ₂
Built-up area ratio ²	unitless	+	+
Agriculture area ratio ²	unitless	+	
Population	person	+	+
Distance to highway	km	−	−
Distance to primary road	km	−	−
Distance to secondary road	km	−	−
Road length B	m/km ²	+	+
Road length C	m/km ²	+	+
Distance to coastline	km	+ / −	
AOD	unitless	+	
Precipitation	mm/hr	−	−
Temperature	°C	+	
Wind speed	m/sec	−	−
Longitude	degree	+	

¹ +:positive direction, -:negative direction² ratio of land use type

Table 3: The number of observations in the inclusive and exclusive data set, and the spatial outliers for PM_{2.5} and NO₂.

Pollutant	Inclusive	Exclusive	Spatial outliers	Outlier ratio (%)
PM _{2.5}	500	457	43	8.6
NO ₂	1278	1155	123	9.6

Table 4: Obtained LUR models for PM_{2.5}.

Variabes	Data set	
	Inclusive data	Exclusive data
Intercept	5.6	5.6
Bulid-up area ratio	1.0×10^{-1}	5.6×10^{-2}
Agriculture area ratio	1.2×10^{-1}	7.6×10^{-2}
Population	3.3×10^{-6}	6.0×10^{-6}
Distance to highway	-3.3×10^{-3}	-2.7×10^{-3}
Distance to coastline	-1.6×10^{-3}	-7.5×10^{-4}
Precipitation	-7.6×10^{-5}	-5.6×10^{-5}
Temperature	3.6×10^{-2}	3.8×10^{-2}
Wind speed	-6.0×10^{-2}	-5.4×10^{-2}
Longitude	-2.4×10^{-2}	-2.4×10^{-2}

Table 5: Obtained LUR models for NO₂.

Variabes	Data set	
	Inclusive data	Exclusive data
Intercept	2.7	2.7
Bulid-up area ratio	4.3×10^{-1}	3.5×10^{-1}
Population	3.8×10^{-5}	4.5×10^{-5}
Distance to highway	-2.4×10^{-2}	-2.3×10^{-2}
Distance to secondary road	-2.2×10^{-2}	-2.5×10^{-2}
Road Length B	7.1×10^{-5}	6.5×10^{-5}
Precipitation	-3.0×10^{-4}	-2.9×10^{-4}
Wind speed	-7.6×10^{-2}	-5.6×10^{-2}

Table 6: The LUR model and validation results using NO₂ observations which are collocated with PM_{2.5} monitors. RMSE represents root mean squared error. RMSE and r^2 are obtained by leave-one-out cross validation.

variables	Data set	
	Inclusive data	Exclusive data
Intercept	2.7	2.7
Bulid-up area ratio	3.1×10^{-1}	2.5×10^{-1}
Population	4.2×10^{-5}	4.6×10^{-5}
Distance to highway	-2.8×10^{-2}	-2.9×10^{-2}
Road Length B	6.9×10^{-5}	6.7×10^{-5}
Precipitation	-3.3×10^{-4}	-3.6×10^{-4}
RMSE of LUR model	2.9	2.7
r^2 of LUR model	0.65	0.67
RMSE of regression kriring	2.5	1.9
r^2 of regression kriging	0.75	0.83
n	478	402