Spatiotemporal land use random forest model for estimating metropolitan NO_2 exposure in Japan

Shin Araki[a,*], Masayuki Shima[b], Kouhei Yamamoto[c]

[a] *Graduate School of Engineering, Osaka University, Yamadaoka 2-1, Suita, Osaka 565-0871, Japan*
[b] *Department of Public Health, Hyogo College of Medicine, Mukogawa-cho 1-1, Nishinomiya, Hyogo 663-8501, Japan*
[c] *Graduate School of Energy Science, Kyoto University, Yoshidahonmachi, Sakyo, Kyoto 606-8501, Japan*

Abstract

Adequate spatial and temporal estimates of NO_2 concentrations are essential for proper prenatal exposure assessment. Here, we develop a spatiotemporal land use random forest (LURF) model of the monthly mean NO_2 over four years in a metropolitan area of Japan. The overall objective is to obtain accurate NO_2 estimates for use in prenatal exposure assessments. We use random forests to convey the non-linear relationship between NO_2 concentrations and predictor variables, and compare the prediction accuracy with that of a linear regression. In addition, we include the distance decay effect of emission sources on NO_2 concentrations for more efficient model construction. The prediction accuracy of the LURF model is evaluated through a leave-one-monitor-out cross validation. We obtain a high R^2 value of 0.79, which is better than that of the conventional land use regression model using linear regression (R^2 of 0.73). We also evaluate the LURF model via a temporal and overall cross validation and obtain R^2 values of 0.84 and 0.92, respectively. We successfully integrate temporal and spatial components into our model, which exhibits higher accuracy than spatial models constructed individually for each month. Our findings illustrate the advantage of using a LURF to model the spatiotemporal variability of NO_2 concentrations.

*Corresponding author
Email addresses: araki@ea.see.eng.osaka-u.ac.jp (Shin Araki), shima-m@hyo-med.ac.jp (Masayuki Shima), yamamoto@energy.kyoto-u.ac.jp (Kouhei Yamamoto)

Keywords: air pollution, machine learning, distance decay effect, prenatal exposure, land use regression

1. Introduction

Exposure to air pollutants has been associated with adverse pregnancy outcomes in many epidemiological studies (Maroziene and Grazuleviciene, 2002; Rich et al., 2009; Faiz et al., 2012, 2013; Malmqvist et al., 2013; Fleischer et al., 2014; Stieb et al., 2016). Spatially and temporally adequate estimates of air pollutant concentrations are essential for proper exposure assessments in order to avoid potential misclassification or biased risk estimates. Land use regression (LUR) models have typically been used to satisfy this demand. In that approach, a linear regression model is developed incorporating both monitored concentrations, as the objective variable, and predictor variables that may affect the concentrations. The obtained regression model is then applied to unmonitored locations to estimate target air pollutant concentrations. LUR models are often applied for estimation of long-term averages, such as annual means of NO_2 (Beelen et al., 2013; Vienneau et al., 2013), NO_x (Beelen et al., 2013), $PM_{2.5}$(Sampson et al., 2013), and PM_{10}(Vienneau et al., 2013). Further, monthly averages of NO_2 (Knibbs et al., 2014; Bechle et al., 2015; Proietti et al., 2016) and $PM_{2.5}$ (Beckerman et al., 2013), biweekly means of NO_2 (Ross et al., 2013; Proietti et al., 2016) and $PM_{2.5}$ (Ross et al., 2013), and daily NO_2 (Lee and Koutrakis, 2014; Cordioli et al., 2017), $PM_{2.5}$ (Di et al., 2016a), and PM_{10} (Alam and McNabola, 2015) concentrations have been estimated in some studies.

In many LUR studies, multiple linear regression has been applied to model pollutant concentrations(e.g., Beelen et al., 2013; Vienneau et al., 2013; Knibbs et al., 2014; Proietti et al., 2016). However, the relationship between the concentrations and potential predictor variables is often complicated and not necessarily linear. Another problem with linear regression is the difficulty in capturing the complex interactions between predictors. To handle these disadvantages, machine learning has been successfully applied in some recent studies. For example, Di et al. (2016a) used a neural network to model daily $PM_{2.5}$ concentrations across the continental United States with a cross validated R^2 of more than 0.8. Further, Di et al. (2016b) estimated the $PM_{2.5}$ constituents in the northern United States and obtained a cross validated R^2 of 0.6–0.8 for the major components. Brokamp et al. (2017) compared the performance of random forest and multiple linear regression techniques by applying them to the prediction of $PM_{2.5}$ elemental components, reporting that the random forest method was more accurate and precise.

Random forests, proposed by Breiman (2001), are a non-parametric statistical method that can handle non-linear relationships. The method is based on decision trees; it constructs each tree using a bootstrap sample of the data and splits each point in the tree according to the best of a subset of randomly chosen predictors at each point (Liaw and Wiener, 2002). This method can be applied to both regression and classification problems. The advantage of random forests is better performance compared to other machine learning methods such as support vector machines and neural networks (Liaw and Wiener, 2002). Moreover, random forests are robust against overfitting (Breiman, 2001). Another advantage is that random forests have only two user-defined parameters: the number of variables in the subset at each node and the number of trees in the forest (Liaw and Wiener, 2002). Furthermore, the random forest cross validated accuracy is typically very insensitive to the values of these parameters (Liaw and Wiener, 2002).

Variable selection is an important step in LUR model construction that excludes irrelevant or colinear predictors, which would otherwise generate unstable estimates (Brokamp et al., 2017). Several buffer sizes are usually defined to represent the range of influence of the predictors. The concentration at the center of a buffer is regressed on the summed values in the buffer. This approach increases the number of potential variables to be considered by multiplying the number of variables by the number of buffer sizes. Given that some predictors represent emission intensity, the buffer approach assumes that emission sources of the same intensity in a buffer equally contribute to the concentration at the buffer centroid, regardless of the distance to the center. This assumption seems to be contradictory to the air pollutant behavior; air pollutant concentrations decrease with distance from its source due to diffusion. Vienneau et al. (2009) introduced the distance decay effect to the LUR framework. They applied the focal-sum approach and successfully modeled monitored NO_2 concentrations using the inverse distance-weighted sum of the emissions in the surrounding area. The clear advantage of this approach is that a large number of potential buffer sizes are not required. Furthermore, this approach is consistent with air pollutant behavior. Note that some studies have already included inverse distance-weighted variables, but several buffer sizes are simultaneously defined (Li et al., 2012, 2013; Eeftens et al., 2016). Su et al. (2009) proposed a variable selection method based on the distance decay effect, but did not use distance-weighted predictors for LUR model construction. Extending the focal-sum with the distance decay effect to all potential predictors representing the emission intensity constitutes a reasonable attempt at higher-efficiency model construction in the LUR framework.

In this study, we develop a spatiotemporal land use random forest (LURF) model of monthly mean NO_2 in a metropolitan area of Japan, where a birth cohort study has been conducted. The overall objective is to obtain accurate NO_2 estimates for use in prenatal exposure assessments. We use random forests to capture the non-linear relationship between the NO_2 concentrations and predictors. We consider the distance decay effect and apply a focal-sum approach to the preparation of potential predictors with the aim of constructing the model in the most efficient manner. We then evaluate the developed model using cross validation and compare the performance of our model to that of the LUR model using multiple linear regression and the same potential variables. Furthermore, we discuss the advantages of a spatiotemporal model using random forests.

2. Methods

2.1. Study area

The Japan Environment and Children's Study (JECS) is an ongoing nationwide birth cohort study implemented in January 2011 to evaluate the effects of various environmental factors on child health and development (Kawamoto et al., 2014). The JECS incorporates 15 regions across Japan in which pregnant women were recruited as study participants from 2011 to 2014 (Kawamoto et al., 2014). Our study area included one of the JECS regions, Amagasaki City (135.4°E, 34.7°N), and its surrounding area (Fig. S1). Amagasaki City has a population of 430,000 and an area of 50 km^2. We extended the study area outside Amagasaki City by approximately 20 km, because only three observations were available in the city. The study area covered approximately 46 km from east to west and 55 km from north to south. This area, containing more than 10 million inhabitants, includes mega cities such as Osaka and Kobe. In the future, we intend to conduct an exposure assessment in Amagasaki City.

2.2. Air quality measurements

We obtained air quality observations from 2011 to 2014 from the database of the regulatory monitoring network in Japan. Network data are collected and stored in this database by the Japanese Ministry of Environment. Data quality is controlled according to the uniform national standard. The monitoring stations are categorized into two types, automobile exhaust stations and general environment stations, and are located according to their specific purpose. That is, the former are located at intersections or roads with heavy traffic to monitor severe air pollution, i.e., at hot spots. The latter are located such that they are not directly affected by specific emission sources, in order to measure the representative concentrations over a certain spatial extent. Accordingly, we utilized observations from general environment stations only in this work. In our study area, stations of this type are located at various distances from major traffic roads (highway, primary, and secondary roads defined in the road network data used in this study, as described below). Specifically, some monitoring sites are located close to major roads (the shortest distance is less than 20 m), whereas others are positioned very far from major roads (the shortest distance is more than 2 km). Further, note that the shortest distances from the monitoring sites to the major roads are distributed relatively homogeneously, as apparent from Table S1. We believe the observations at the latter stations well represent the concentrations in our study area, covering not only the urban background, but also the areas where the concentrations are influenced by traffic.

We used hourly mean concentrations to calculate the monthly mean values over a four-year period. Data with a temporal coverage of more than 80% on both daily and monthly bases were used for the analysis to ensure that it was temporally representative. The number of general environment stations under operation for NO_2 monitoring was 81 in 2014, but only three monitors were located in Amagasaki City.

Fig. 1 presents a plot of the monthly mean concentrations used in this study. The seasonal variation in concentration is clear: high in winter and low in summer.

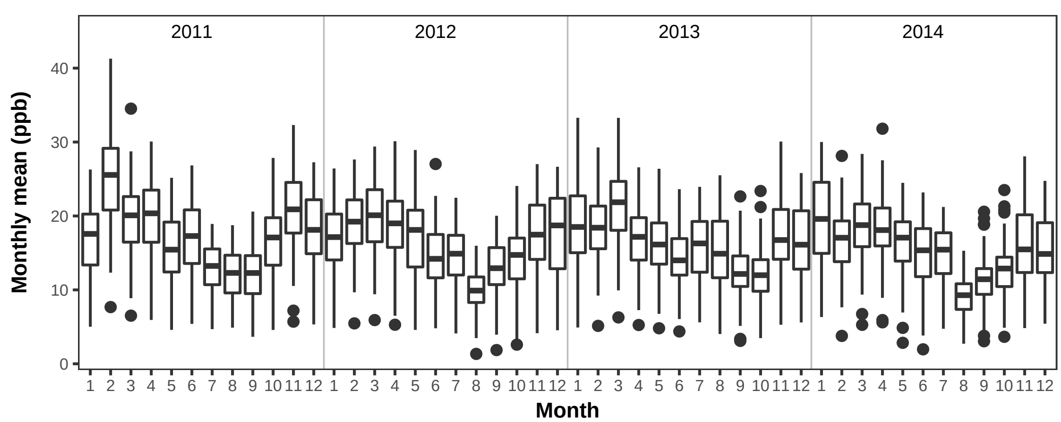

Figure 1: Box plot of monthly NO_2 concentrations used in this study.

2.3. Data set

We selected data sets considering key factors affecting the spatial distribution of air pollutants, including emission, advection, and deposition. Some of the gridded data were resampled to conform to an origin and resolution of 500 m. The built-up area ratio in a grid cell was calculated from land use data. The green area ratio was obtained by summing the ratio of rice fields, agricultural fields, and forest from the land use data.

We calculated the road length in a grid cell using road network data instead of readily available road length data. This is because the spatial resolution of the publicly available road length data is, to the best of our knowledge, coarser than that of our 500-m resolution grid. In the road network data set, road types are classified into three categories: highway, primary, and secondary; in this study, the road length in a grid cell was calculated for each of these categories. We also calculated the shortest distances from a grid cell centroid to each road type and employed these values as predictors.

We included the emission intensities of large point sources as a predictor. The emission intensity was obtained from EAGrid2010 (Fukui et al., 2014), which is a widely used emission inventory in Japan, being specially compiled for air quality models. This inventory has a spatial resolution of 1 km and a temporal resolution of 1 month. We excluded the emission intensity of transportation in the EAGrid2010 database, because the road length used as a transportation proxy had a finer spatial resolution of 500 m.

As for meteorological parameters, we utilized daily mean observations of precipitation and wind speed from the Automated Meteorological Data Acquisition System (AMeDAS), the monitoring stations of which are densely and homogeneously distributed throughout the country. We calculated the monthly means and interpolated them using ordinary kriging to obtain gridded data of monthly means with a 500-m resolution.

Satellite-derived NO_2 data have a wide temporal and spatial coverage. This feature is useful for constructing a spatiotemporal LUR model. In recent studies, the NO_2 tropospheric column abundance has been introduced as a predictor variable for LUR, and good prediction performance has been reported (Knibbs et al., 2014; Bechle et al., 2015). The Ozone Monitoring Instrument (OMI) flown on the Aura satellite measures the spectrum in the ultraviolet/visible wavelength range with a very high spatial resolution and daily global coverage (Levelt et al., 2006). We obtained the daily NO_2 tropospheric column abundance from the version 3.0 release of the gridded OMINO2d product and calculated the monthly means. Because the spatial resolution of OMI NO_2 data is 0.25° (approximately 25 km) and coarse compared to our prediction grid size of 500 m, we simply disaggregated these data into a 5-km-resolution data set through bilinear interpolation. Details of the data sets are presented in Table S2.

2.4. Implementation of distance decay effect

To consider the distance decay effect, we applied the focal-sum approach (Vienneau et al., 2013) to the potential predictor variables that indicate emission intensity: land use, road length, population, and large point sources. In this approach, a moving window passes over all grid cells. The values inside the window are multiplied by the corresponding factors defined by the inverse distance to the central cell. The sum of the products is assigned to the central cell (Vienneau et al., 2013). This new value is the distance-weighted measure for the central cell. Previously, Vienneau et al. (2013) examined various window shapes and weighting factors, and reported similar accuracies for NO_2 concentration estimates. Here, we used a simple circular window and the squared inverse distance as a weighting factor, which was obtained by:

$$w = \frac{1}{(d+1)^2}, \tag{1}$$

where w is the weighting factor and d is the distance (km) from the central cell. We used *d+1* rather than d in the denominator to avoid division by zero. The central cell has a value of 1. Note that the radius of the moving window can be set to infinity to include all emission sources; however, for practicality, we set the window radius to 15 km so that the minimum weighting factor was approximately 1 % of the largest factor at the central cell. Other variables such as OMI NO_2 and meteorological parameters were supplied without implementation of the focal-sum process, because these variables do not represent emission intensity.

We also included the month and year as predictor variables in order to capture the temporal variations. These variables were treated as categorical variables. The monthly or annual trends were not considered. Table 1 presents the potential predictor variables.

Table 1: Potential predictor variables.

Predictor variables	Unit	Direction of effect
Built-up area ratio	unitless	+
Green area ratio	unitless	−
Population	number	+
Road length, highway	km/km^2	+
Road length, primary road	km/km^2	+
Road length, secondary road	km/km^2	+
Distance to highway	m	-
Distance to primary road	m	-
Distance to secondary road	m	-
Point source	Tg/year	+
OMI NO_2	10^{-15} molecules/cm^2	+
Precipitation	mm/h	−
Wind speed	m/s	−
Month	none	not specified
Year	none	not specified

2.5. Land use random forest model

We constructed a spatiotemporal LURF model using the variable selection method proposed by Genuer et al. (2015). First, we ran an initial random forest with all potential variables, repeating it 50 times. The potential predictors were then ranked by sorting the variable importance measure, averaged over the repetitions, in descending order. Random forest models were constructed with *k* first predictors for *k=1,2,..., m*, where *m* is the number of potential predictors, with each being repeated 25 times. We selected the model with the smallest out-of-bag error averaged over the repetitions for each predictor combination. Next, the variables in the selected model were sequentially introduced to the random forest model in order of variable importance, as determined in the first step. A variable was retained in the model only if the out-of-bag error decreased by a greater degree than the averaged variations of the noisy variables removed in the second step. The variables in the last model were selected. During this process, the number of variables in the subset at each node (m_{try}) was set to *2p/3*, where *p* is the number of predictor variables in the entire data set. Finally, the random forest model with the selected predictors was optimized for m_{try}. The number of trees (*ntree*) was consistently set to 500. The other parameters were set to the default values of the ranger package (Wright and Ziegler, 2017) used in this study, including a minimum node size of 5. The model R^2 was calculated as 1-*MSE/var(Y)* where Y is the observed values and *MSE* is the mean of the out-of-bag errors for all the prediction points (Brokamp et al., 2017).

2.6. Land use regression model

We constructed a spatiotemporal LUR model based on a supervised stepwise selection procedure used to develop LUR models for NO_2 in Europe (Beelen et al., 2013). The potential predictor variables of the LUR models were identical to those of the LURF model presented in Table 1. We specified the direction of effect according to the relationship between the pollutants and predictor variables (Beelen et al., 2013). First, univariate regression analyses were conducted for all potential predictors. The initial regression model was constructed using the predictor giving the highest adjusted R^2 with the defined direction of effect. Second, the remaining variables were consecutively tested through addition to the model. The predictor with the highest additional increase in adjusted R^2 was retained, if the following conditions were fulfilled: 1) the predictor increased the adjusted R^2 by more than 0.01; 2) its coefficient conformed to the specified direction of effect for the variable; 3) it did not change the direction of effect for the predictors already in the model. This variable test was repeated until there were no more variables that increased the adjusted R^2 by more than 0.01. Third, variables with a p-value greater than 0.1 were removed and the regression model was reconstructed using the retained variables. For categorical variables, a likelihood ratio test was conducted between models with and without the variable; hence, a p-value was obtained. Finally, the variance inflation factors (VIF) were checked to determine whether they were less than or equal to 3. In addition, the Cook's D statistics for all the observations were assessed to determine whether they were less than or equal to 1.

The pollutant concentrations were transformed to a natural logarithmic scale before analysis and the predictions were back-transformed after analysis. This procedure has the advantage that the predicted concentrations are positive, which is not the case when analyses are performed without transformation (Beelen et al., 2009).

2.7. Evaluation

We performed leave-one-monitor-out cross validation to assess the accuracy of the obtained models. The observed data were removed from one location for the entire period and the model was constructed using the remaining location data. This process was repeated for the remaining locations. The R^2 and root mean squared error (RMSE) between the predicted and measured values were computed as indicators of the prediction accuracy. Note that the RMSE values are desired to be as small as possible. We refer to this validation process as spatial cross validation. We also conducted temporal and overall cross validations. For the temporal cross validation, monitoring data were omitted for a particular month and the model was constructed using the remaining 47 months of data. Concentrations at the monitored locations in the selected month were then predicted using the model. This process was repeated for the remaining 47 months and R^2 and RMSE values were computed. For the overall cross validation, we performed 5-fold cross validation. The observations were evenly divided into five splits at random. One observation split was omitted and the model was constructed using the remaining four observation splits. The concentrations at the times and locations of the selected observations were predicted by the model. This process was repeated for the remaining four splits. The R^2 and RMSE values were computed. For a fair comparison, the splits were identical for the LURF and LUR model evaluations. These statistical indicators were also calculated separately for monitoring locations in Amagasaki City. We refer to these stations and all stations in the study area as "inside stations" and "all stations", respectively. We also construct LURF and LUR models using all the potential predictors and conducted spatial cross validation in order to compare R^2 values with the corresponding ones obtained for the LURF and LUR model constructed using the selected variables (i.e., the final models) for a sensitivity analysis.

To assess the advantages of a spatiotemporal LURF model over a spatial LURF model, we constructed spatial models for each month; thus, 48 monthly models were obtained. We then individually evaluated the spatial models through leave-one-monitor-out cross validation and calculated R^2 and RMSE values for each model. We constructed the spatial models with the same variables as those for the spatiotemporal model, except for month and year. We did not apply the variable selection process to the monthly models because of the computation costs . For the temporally variable predictors, we extracted the data from the corresponding year and month. We also constructed and evaluated spatial LUR models in the same manner for comparison purposes.

We statistically evaluated the differences between the spatiotemporal LURF and LUR models using a paired t- test and F-tests (Hengl et al., 2015). The paired t-test evaluates whether two models have the same mean errors (ME). The F-test evaluates whether two models have the same variance, i.e. RMSE, assuming that the MEs are the same.

2.8. Computation

All spatial and statistical calculations were performed using R statistical software (3.4.3) (R Core Team, 2017), with the raster package (Hijmans, 2016) for integration and construction of the potential predictor variables, and the ranger package (Wright and Ziegler, 2017) for implementation of the random forests.

3. Results

3.1. Spatiotemporal LURF model

Fig. 2 shows the variable importance plot of the final LURF model. In this plot, the selected variables are listed in order of importance from top to bottom. The horizontal axis represents the measure of importance. The green area ratio is the best predictor. Satellite-based NO_2 is the second most influential variable, followed by point emission sources and month, reflecting the clear seasonality in the concentrations, as shown in Fig. 1. Highway road length and distance to highway are also important covariates. The remaining variables, including the meteorological parameters, built-up area ratio, and year, are ranked as less important. The variables removed from the model by the variable selection process are primary road length, secondary road length, distance to primary road, distance to secondary road, and population. The model R^2 value is 0.92.

Scatter plots of the predicted and observed concentrations obtained through cross validation are presented in panels (a)–(c) of Fig. 3. Panels (a) and (d), (b) and (e), and (c) and (f) show the results of the spatial, temporal, and overall cross validation, respectively. The dot color indicates the point density in the plot: red and green indicate higher and lower density, respectively. The triangles indicate the results for inside stations. R^2 and RMSE values are given in each panel for all stations and inside stations.

The R^2 values for the spatial and temporal validation are 0.79 and 0.84, respectively. The RMSE values are 2.6 and 2.2 (ppb), respectively. A high R^2 value of 0.92 is obtained for the overall cross validation, with an RMSE value

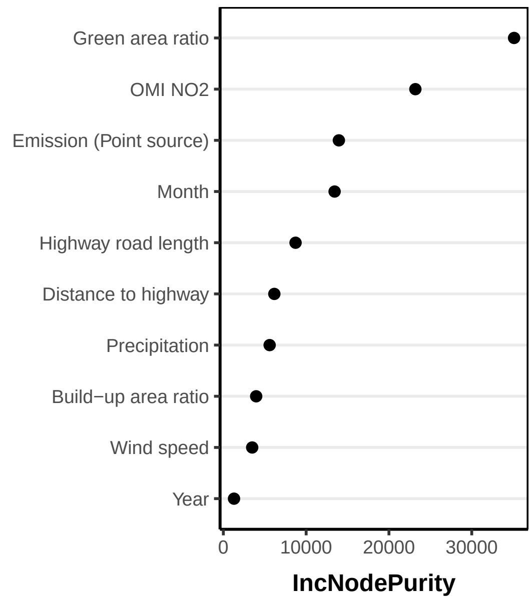

Figure 2: Variable importance plot for the LURF model. The variables are listed in order of importance from top to bottom. The horizontal axis represents the measure of importance.

of 1.6 (ppb). Compared to the corresponding values for all stations, the R^2 values for the inside stations are lower for the overall and temporal cross validations, and higher for spatial cross validation. The RMSE values are similar for all stations and inside stations for the three types of cross validation. The LURF model constructed using all the potential predictors gives a cross validated R^2 value of 0.79 and RMSE of 2.6 (ppb), which are almost identical to those obtained for the final LURF model using the selected variables.

The statistical indicators of the spatial LURF models for 48 months are presented as box plots in panels (a) and (c) of Fig. 4, showing the R^2 and RMSE values, respectively. The indicators of the spatiotemporal LURF model are also presented for comparison, as a horizontal line on the left side of each panel. The median R^2 values for the spatial models are 0.73 and 2.4 (ppb), respectively, indicating that the spatiotemporal model outperforms the spatial models in terms of R^2.

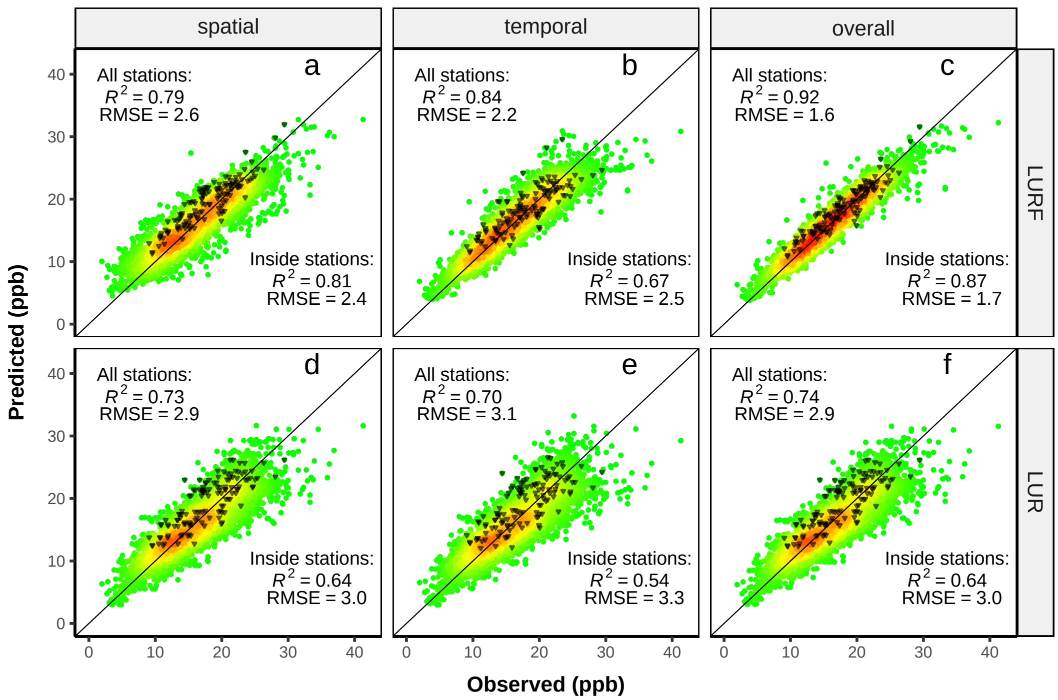

Figure 3: Scatter plots of predicted and observed concentrations obtained from cross validation. (a)–(c) and (d)–(f) show LURF and LUR results, respectively. (a) and (d), (b) and (e), and (c) and (f) show the spatial, temporal, and overall cross validation results, respectively. The red and green colors indicate higher and lower point density, respectively. The triangles indicate the results for the inside stations.

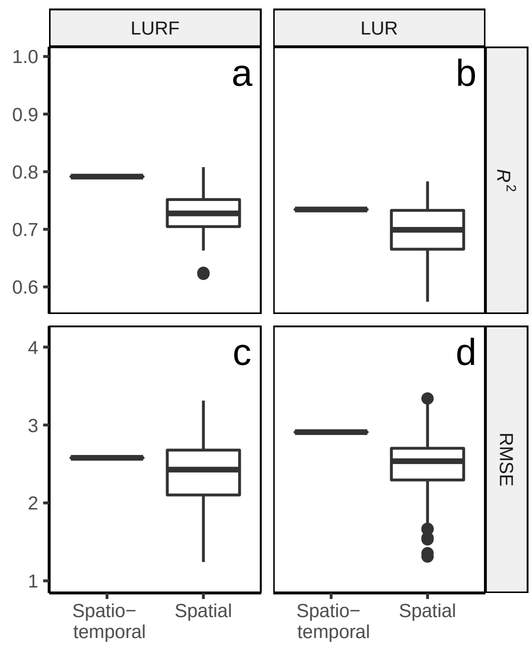

Figure 4: Box plots of statistical indicators for spatiotemporal and spatial models. (a) and (b) show R^2 values and (c) and (d) show RMSE values. (a) and (c), and (b) and (d), present LURF and LUR results, respectively.

3.2. Spatiotemporal LUR model

The selected variables in the final model are the green area ratio, month, and highway road length. These predictors are ranked as important in the LURF model, but other important predictors such as OMI NO_2 and point emission sources are discarded. The model adjusted R^2 value is 0.77. Table S3 presents the details of the final spatiotemporal LUR model.

Scatter plots of the predicted and observed values obtained via cross validation are presented in panels (d)–(f) of Fig. 3. A R^2 value of 0.73 is obtained for the spatial cross validation. The R^2 and RMSE values are similar between the three cross validation results. A comparison of the results from the inside stations against all stations shows that the R^2 values for the former are smaller than those for the latter, while the RMSE values are similar. The LUR model constructed using all the potential variables gives a cross validated R^2 value of 0.76 and RMSE of 2.8 (ppb), which are similar to those obtained for the final LUR model using the selected variables.

The R^2 and RMSE values obtained for the 48 spatial LUR models are presented as box plots in panels (b) and (d) of Fig. 4. The median R^2 is 0.70, which is slightly lower than that of the spatiotemporal LUR model. The median RMSE is smaller.

3.3. Comparison

The spatial cross validated R^2 value of 0.79 for the spatiotemporal LURF model is higher than that of the spatiotemporal LUR model. The paired t-test results show that the differences in ME between LURF and LUR are not statistically significant (p >0.01). The F-test result indicates that the differences in RMSE between the two models are statistically significant at the 1% level. In both the temporal and overall cross validation results, the differences in RMSE between the two models are significant at the 1% level, while the differences in ME are not significant (p >0.01). The temporal and overall cross validated R^2 values for the LURF model are 0.84 and 0.92, respectively, which are higher than those for the LUR model, at 0.70 and 0.74, respectively. These results show that the LURF model outperforms the LUR model.

We report higher cross validated R^2 values and similar RMSE values for the spatiotemporal LURF model than for the spatial LURF models (Figs. 4(a) and (c)). Meanwhile, the R^2 and RMSE values are marginally higher and larger for the spatiotemporal LUR model than for the spatial LUR models, respectively, as shown in Figs. 4(b) and (d). A comparison of the spatial LURF and LUR models shows that the median R^2 of the LURF models is slightly higher and the median RMSE is slightly smaller than those of the LUR models, although the F-test result indicates that the differences in RMSE are not statistically significant at the 1% level (p=0.02).

3.4. Mapping

Fig. 5 is a prediction map of the NO_2 concentrations averaged over the study period. This map was produced by averaging the monthly estimations over the four-year study period, and disaggregated to 100-m resolution via bilinear interpolation for presentation purposes.

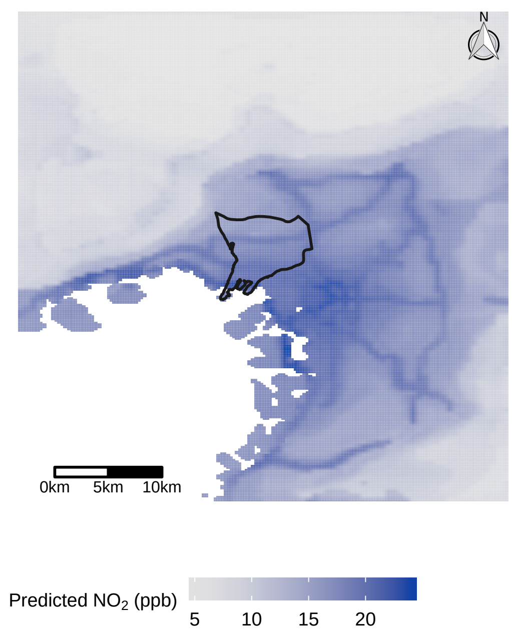

Figure 5: Prediction map of four-year mean concentrations of NO_2, disaggregated to 100-m resolution by bilinear interpolation for presentation purposes.

4. Discussion

We developed the spatiotemporal LURF model of NO_2 reported in this study to predict the monthly mean NO_2 concentrations for the consecutive four-year study period. Our spatiotemporal LURF model is accurate, with a spatial cross validated R^2 and RMSE value of 0.79 and 2.6 (ppb), respectively. No significant over or under estimation is apparent in the cross validation results, as shown in Fig. 3. Thus, when applying of our LURF model to exposure assessments, the estimations at participant addresses can be expected to be accurate. The overall cross validation provides better R^2 and smaller RMSE values than the temporal and spatial cross validation. This suggests that we have successfully combined the temporal and spatial components in our spatiotemporal LURF model. The overall cross validated R^2 is almost identical to the model R^2, while the spatial and temporal cross validated R^2 are smaller. This indicates that our LURF model is not over-fitted overall, but is over-fitted especially in the spatial aspect. The LURF model constructed with all the potential variables shows almost identical cross validated R^2 and RMSE values to those for the final LURF model, which indicates that the variable selection process worked properly and successfully removed irrelevant variables. This result also demonstrates that the random forests are robust to noise variables (Breiman, 2001).

Prenatal exposure assessment requires several NO_2 estimates at a fine temporal scale over a certain time period. Estimation models developed for this purpose should, therefore, be extended from two-dimensional space to three dimensions by adding a temporal axis. This can readily be achieved by constructing individual two-dimensional (i.e., spatial) models for each time step, with no interaction between models. However, this involves cumbersome repetition of the model construction process, including variable selection. A probably more popular solution is the temporal scaling approach, where spatial estimates for a particular time step are temporally scaled according to the measurements obtained from fixed continuous monitors (e.g., Slama et al., 2007; Ghosh et al., 2012). This approach assumes the spatial distribution pattern of air pollutants is constant over a certain period. Air pollutant concentrations are affected by meteorological parameters and/or emissions. Consequently, their spatial distribution pattern changes over time according to the temporal changes in the spatial pattern of the influential factors. For instance, wind direction and wind speed change in time and space, as do emissions from different types of sources such as automobile and power plants, resulting in variation in the spatial pattern of emissions over time. These spatial variations in the influential factors may be averaged out when mean concentrations over relatively longer timescales are considered. Thus, the scaling approach may be applicable to estimation of annual means, where the spatial pattern of the factors, and consequently the concentrations, are constant between years. However, it would be difficult to apply it to finer temporal scales when the spatial variation in the influential factors may not be averaged out and, accordingly, the spatial pattern of pollutant concentrations may temporally change. Our spatiotemporal model, on the other hand, is a three-dimensional model that implements a temporal component and integrates individual two-dimensional models into a three-dimensional model. This enables model construction and estimation without iteration for each time step. Further, this allows for temporal variation in the spatial distribution pattern. Hence, our spatiotemporal modeling is advantageous in its simplicity and flexibility. Clearly, estimation accuracy is of principal importance, and our spatiotemporal LURF model gives accurate predictions, which are better than those of spatial models. Therefore, our spatiotemporal LURF model has advantages over spatial models for estimating monthly mean NO_2 concentrations.

The estimation accuracy for inside stations is satisfactory, and the statistical indicators of the LURF model are similar to those for all stations, as shown in Fig. 3. This result indicates that our spatiotemporal LURF model has sufficient predictive power for future exposure assessment in smaller areas, despite having been developed based on larger areas.

With respect to the spatiotemporal LUR model, the statistical indicators obtained for the overall cross validation are comparable with those for the temporal and spatial validation, in contrast to the LURF model. This may be because random forests are powerful classifiers and can handle the temporal component, implemented as categorical variables in this study, more effectively than a linear regression. Hence, accurate predictions are provided by the spatiotemporal LURF model, which outperforms the spatiotemporal LUR model considered in this study. We note, however, that the implementation of the temporal component as categorical variables may not be optimal for the LUR model and that better modeling of the temporal component could improve the performance of the spatiotemporal LUR. Our spatiotemporal LURF model consistently gives higher R^2 and significantly smaller RMSE values than the LUR model for spatial, temporal, and overall cross validation. This may be due to the ability of random forests to handle non-linear relationships between the predictors and outcome. On the other hand, the advantage of random forests is not as clearly demonstrated for the spatial models compared to the spatiotemporal models. One possible explanation is that we did not conduct a variable selection process for each monthly model, which would be required for a fair comparison of spatial models, for both LURF and LUR.

Although the prediction accuracy of the spatiotemporal LUR model is inferior to the spatiotemporal LURF models, the performance of the LUR model is still satisfactory, with an R^2 value of 0.73. The difference in the model and cross validated R^2 values is small, meaning that the spatiotemporal LUR model is not significantly over-fitted. The marginal difference in the cross validated R^2 and RMSE values between the final LUR model and the LUR model constructed using all the potential variables indicates that the variable selection process worked properly to discard unimportant predictors. The predictors prepared using the focal-sum with distance decay effect may contribute to the performance, although evaluation of the focal-sum approach is outside the scope of this study (a simple comparison of LURF and LUR models with and without the distance decay effect shows improvement in the model performance especially for LUR, as given in Table S4 and Fig.S3). We note, however, that the optimal weighting factor may be specific to each predictor depending on emission source characteristics, because the pollutants emitted from a high stack diffuse differently from those emitted from ground level sources like traffic. We require further investigation of the optimal selection of the weighting factor, other than the inverse distance squared approach, to improve the estimation accuracy, as well as a detailed evaluation of the approach. Nonetheless, in this study, we efficiently constructed land use models, reducing the effort required for the variable selection process through this method.

Some predictors such as OMI NO_2 and point emission sources, which are ranked as important for the LURF model, are not retained in the final LUR model. While the LURF model selects variables based on the prediction error, the LUR model chooses predictors based on R^2. In addition, random forests and linear regression are inherently different procedures. These differences may explain the different predictors of the LURF and LUR models. Although satellite NO_2 has been the focus of many LUR studies (e.g., Knibbs et al., 2014; Bechle et al., 2015), OMI NO_2 was discarded in our LUR model. This may be due to the coarse spatial resolution of the original data and/or our simple bilinear interpolation approach for downscaling. In addition, we calculated monthly means by simply averaging daily values and missing values are omitted from the calculation. Consequently, an averaged value at a pixel with many missing daily values may not be an appropriate representation of a monthly value. Kim et al. (2016) noted that the spatial resolution of OMI NO_2 is too coarse to capture the spatial distribution in urban areas, with possible underestimation at urban centers and overestimation outside. Satellite data at a finer resolution could provide improved estimation accuracy for both LURF and LUR. In addition, Kuhlmann et al. (2014) developed a new gridding algorithm for OMI NO_2, demonstrating that this method improves the accuracy of the obtained spatial distribution of regional NO_2. Thus, a more accurate downscaling method is required to improve the accuracies of LURF and LUR.

Brokamp et al. (2017) noted the difficulty in interpreting the results of random forests. Unlike the LUR model, the LURF model lacks coefficients representing the directions and magnitudes of the effects of predictor variables on air pollutant concentrations (Brokamp et al., 2017). This may be a trade-off for the improved performance of random forests (Brokamp et al., 2017). However, LUR models are not constructed based on a cause-consequence relationship, but on correlation. When a variable equally contributes to concentrations in the area of interest, the variable is most likely to be excluded in the resulting LUR model. This is because it contributes to the concentrations, but not to the spatial difference in concentrations. Precipitation, for instance, is generally an influential parameter for NO_2 concentrations, but is not retained in our final LUR model. Therefore, the LUR model is unfit for elucidation of the physical or chemical processes of air pollutants. LUR model results may be useful for obtaining a basic understanding of the factors influencing the spatial distribution of air pollutants, but this model is not suitable for achieving detailed comprehension or performing quantitative analysis. Therefore, the difficulty in interpreting random forests can be more than compensated for by their prediction ability.

Although our spatiotemporal LURF model exhibits remarkable prediction accuracy, there are some limitations. Firstly, the high prediction accuracy may be specific to the monthly spatiotemporal LURF model. The high R^2 value of the overall cross validation may arise because the spatial variation pattern is relatively similar between months, with only the concentration level changing. This may also explain the finding that the month serves as a key predictor in our spatiotemporal LURF model. The spatial variation pattern may have higher variance on a finer temporal scale, e.g., weekly or daily, for which the temporal indicator variable is less important. Further investigation of the application of the LURF model to a finer temporal scale, which is preferable for prenatal exposure assessments, is required because we hope to extend our LURF model to a finer temporal scale as well as to a larger area and to other pollutants based on the results of this study. In addition, higher-spatial-resolution satellite data could play a more important role in improving the prediction accuracy of the LURF model on such a temporal scale. Secondly, concentration estimates at intersections or busy roads and their adjacent areas are likely to be underestimated. We constructed our spatiotemporal model without observations from automobile exhaust stations. These stations monitor potentially severe air pollution in limited areas (hot spots) at intersections or busy roads. Actually, the estimations at automobile exhaust stations via the spatiotemporal LURF model exhibit underestimations of 7.1 (ppb) on average (Supplementary material, pp-S7). The road structure in a metropolitan area is complicated, and primary or secondary roads are often located beneath elevated highways. The vertical and horizontal positions of the monitors of the automobile exhaust stations at such locations may influence the observed pollution level. Monitors are sometimes installed in a building, and the measurements differ depending on the side of the building at which the monitor inlets are placed. This information is not available in the database used in this study. Moreover, it is difficult to model a three-dimensional structure using LURF or LUR. Although exclusion of automobile exhaust stations is a reasonable decision, use of our LURF model to predict concentrations in such potential hot spots would require caution.

Despite these limitations, in this study, we successfully developed a spatiotemporal LURF model for estimating accurate monthly mean NO_2 concentrations. We demonstrated the important advantages of using random forests to handle non-linearity and to capture temporal variation for the three-dimensional model. Our study also illustrates the potential for random forests to be incorporated into the LUR framework for epidemiological studies.

Acknowledgement

This work was supported in part by JSPS KAKENHI Grant Number 15H04790.

References

Alam, M.S., McNabola, A., 2015. Exploring the modeling of spatiotemporal variations in ambient air pollution within the land use regression framework: Estimation of PM_{10} concentrations on a daily basis. Journal of the Air & Waste Management Association 65, 628–640. doi:10.1080/10962247.2015.1006377.

Bechle, M.J., Millet, D.B., Marshall, J.D., 2015. National Spatiotemporal Exposure Surface for NO_2: Monthly Scaling of a Satellite-Derived Land-Use Regression, 2000-2010. Environmental Science and Technology 49, 12297–12305. doi:10.1021/acs.est.5b02882.

Beckerman, B.S., Jerrett, M., Serre, M., Martin, R.V., Lee, S.J., van Donkelaar, A., Ross ev, Z., Su, J., Burnett, R.T., 2013. A Hybrid Approach to Estimating National Scale Spatiotemporal Variability of $PM_{2.5}$ in the Contiguous United States. Environmental Science and Technology 47, 7233–7241. doi:10.1021/es400039u.

Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., de Hoogh, K., Briggs, D.J., 2009. Mapping of background air pollution at a fine spatial scale across the European Union. Science of the Total Environment 407, 1852–1867. doi:10.1016/j.scitotenv.2008.11.048.

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K.T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nádor, G., Varró, M.J., Dedele, A., Gražulevičiene, R., Mölter, A., Lindley, S., Madsen, C., Cesaroni, G., Ranzi, A., Badaloni, C., Hoffmann, B., Nonnemacher, M., Krämer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Strömgren, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., de Hoogh, K., 2013. Development of NO_2 and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project. Atmospheric Environment 72, 10–23. doi:10.1016/j.atmosenv.2013.02.037.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32. doi:10.1023/A:1010933404324.

Brokamp, C., Jandarov, R., Rao, M.B., LeMasters, G., Ryan, P., 2017. Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. Atmospheric Environment 151, 1–11. doi:10.1016/j.atmosenv.2016.11.066.

Cordioli, M., Pironi, C., De Munari, E., Marmiroli, N., Lauriola, P., Ranzi, A., 2017. Combining land use regression models and fixed site monitoring to reconstruct spatiotemporal variability of NO_2 concentrations over a wide geographical area. Science of the Total Environment 574, 1075–1084. doi:10.1016/j.scitotenv.2016.09.089.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016a. Assessing $PM_{2.5}$ Exposures with High Spatiotemporal Resolution across the Continental United States. Environmental Science and Technology 50, 4712–4721. doi:10.1021/acs.est.5b06121.

Di, Q., Koutrakis, P., Schwartz, J., 2016b. A hybrid prediction model for $PM_{2.5}$ mass and components using a chemical transport model and land use regression. Atmospheric Environment 131, 390–399. doi:10.1016/j.atmosenv.2016.02.002.

Eeftens, M., Meier, R., Schindler, C., Aguilera, I., Phuleria, H., Ineichen, A., Davey, M., Ducret-Stich, R., Keidel, D., Probst-Hensch, N., Künzli, N., Tsai, M.Y., 2016. Development of land use regression models for nitrogen dioxide, ultrafine particles, lung deposited surface area, and four other markers of particulate matter pollution in the Swiss SAPALDIA regions. Environmental Health 15:53. URL: http://ehjournal.biomedcentral.com/articles/10.1186/s12940-016-0137-9, doi:10.1186/s12940-016-0137-9.

Faiz, A.S., Rhoads, G.G., Demissie, K., Kruse, L., Lin, Y., Rich, D.Q., 2012. Ambient air pollution and the risk of stillbirth. American Journal of Epidemiology 176, 308–316. doi:10.1093/aje/kws029.

Faiz, A.S., Rhoads, G.G., Demissie, K., Lin, Y., Kruse, L., Rich, D.Q., 2013. Does ambient air pollution trigger stillbirth? Epidemiology 24, 538–544. doi:10.1097/EDE.0b013e3182949ce5.

Fleischer, N.L., Merialdi, M., van Donkelaar, A., Vadillo-Ortega, F., Martin, R.V., Betran, A.P., Souza, J.P., O'Neill, M.S., 2014. Outdoor air pollution, preterm birth, and low birth weight: Analysis of the world health organization global survey on maternal and perinatal health. Environmental Health Perspectives 122, 425–430. doi:10.1289/ehp.1306837.

Fukui, T., Kokuryo, K., Baba, T., Kannari, A., 2014. Updating EAGrid2000-Japan emissions inventory based on the recent emission trends. Journal of Japan Society for Atmospheric Environment 49, 117–125. doi:10.11298/taiki.49.117.

Genuer, R., Poggi, J.M., Tuleau-Malot, C., 2015. Vsurf: Variable selection using random forests. The R Journal 7, 19–33. URL: https://journal.r-project.org/archive/2015/RJ-2015-018/index.html.

Ghosh, J.K.C., Wilhelm, M., Su, J., Goldberg, D., Cockburn, M., Jerrett, M., Ritz, B., 2012. Assessing the influence of traffic-related air pollution on risk of term low birth weight on the basis of land-use-based regression models and measures of air toxics. American Journal of Epidemiology 175, 1262–1274. doi:10.1093/aje/kwr469.

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(6). doi:10.1371/journal.pone.0125814.

Hijmans, R.J., 2016. raster: Geographic Data Analysis and Modeling. URL: https://CRAN.R-project.org/package=raster. r package version 2.5-8.

Kawamoto, T., Nitta, H., Murata, K., Toda, E., Tsukamoto, N., Hasegawa, M., Yamagata, Z., Kayama, F., Kishi, R., Ohya, Y., Saito, H., Sago, H., Okuyama, M., Ogata, T., Yokoya, S., Koresawa, Y., Shibata, Y., Nakayama, S., Michikawa, T., Takeuchi, A., Satoh, H., 2014. Rationale and study design of the Japan environment and children's study (JECS). BMC Public Health 14:25. doi:10.1186/1471-2458-14-25.

Kim, H.C., Lee, P., Judd, L., Pan, L., Lefer, B., 2016. OMI NO2 column densities over North American urban cities: The effect of satellite footprint resolution. Geoscientific Model Development 9, 1111–1123. doi:10.5194/gmd-9-1111-2016.

Knibbs, L.D., Hewson, M.G., Bechle, M.J., Marshall, J.D., Barnett, A.G., 2014. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. Environmental Research 135, 204–211. doi:10.1016/j.envres.2014.09.011.

Kuhlmann, G., Hartl, A., Cheung, H.M., Lam, Y.F., Wenig, M.O., 2014. A novel gridding algorithm to create regional trace gas maps from satellite observations. Atmospheric Measurement Techniques 7, 451–467. doi:10.5194/amt-7-451-2014.

Lee, H.J., Koutrakis, P., 2014. Daily ambient NO_2 concentration predictions using satellite OMI NO_2 data and land use regression. Environmental science & technology 48, 2305–2311. doi:10.1021/es404845f.

Levelt, P.F., van den Oord, G.H.J., Dobber, M.R., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J.O.V., Saari, H., 2006. The ozone monitoring instrument. Ieee Transactions on Geoscience and Remote Sensing 44, 1093–1101. doi:Urn:nbn:nl:ui:25-648485.

Li, L., Wu, J., Kay, J., Ritz, B., 2013. Estimating spatiotemporal variability of ambient air pollutant concentrations with a hierarchical model. Atmospheric Environment 71, 54–63. doi:10.1016/j.atmosenv.2013.01.038.

Li, L., Wu, J., Wilhelm, M., Ritz, B., 2012. Use of generalized additive models and cokriging of spatial residuals to improve land-use regression estimates of nitrogen oxides in Southern California. Atmospheric Environment 55, 220–228. doi:10.1016/j.atmosenv.2012.03.035.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R News 2, 18–22. URL: http://CRAN.R-project.org/doc/Rnews/.

Malmqvist, E., Jakobsson, K., Tinnerberg, H., Rignell-Hydbom, A., Rylander, L., 2013. Gestational diabetes and preeclampsia in association with air pollution at levels below current air quality guidelines. Environmental Health Perspectives 121, 488–493. doi:10.1289/ehp.1205736.

Maroziene, L., Grazuleviciene, R., 2002. Maternal exposure to low-level air pollution and pregnancy outcomes: a population-based study. Environmental Health 1:6. doi:10.1186/1476-069X-1-6.

Proietti, E., Delgado-Eckert, E., Vienneau, D., Stern, G., Tsai, M.Y., Latzin, P., Frey, U., Röösli, M., 2016. Air pollution modelling for birth cohorts: a time-space regression model. Environmental Health 15:51. doi:10.1186/s12940-016-0145-9.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rich, D.Q., Demissie, K., Lu, S.E., Kamat, L., Wartenberg, D., Rhoads, G.G., 2009. Ambient air pollutant concentrations during pregnancy and the risk of fetal growth restriction. Journal of Epidemiology and Community Health 63, 488–496. doi:10.1136/jech.2008.082792.

Ross, Z., Ito, K., Johnson, S., Yee, M., Pezeshki, G., Clougherty, J.E., Savitz, D., Matte, T., 2013. Spatial and temporal estimation of air pollutants in New York City: exposure assignment for use in a birth outcomes study. Environmental Health 12:51. doi:10.1186/1476-069X-12-51.

Sampson, P.D., Richards, M., Szpiro, A.A., Bergen, S., Sheppard, L., Larson, T.V., Kaufman, J.D., 2013. A regionalized national universal kriging model using Partial Least Squares regression for estimating annual $PM_{2.5}$ concentrations in epidemiology. Atmospheric Environment 75, 383–392. doi:10.1016/j.atmosenv.2013.04.015.

Slama, R., Morgestern, V., Cyrys, J., Zutavern, A., Herbarth, O., Wichmann, H.E., Heinrich, J., 2007. Traffic-related atmospheric pollutants levels during pregnancy and offspring's term birth weight: A study relying on a land-use regression exposure model. Environmental Health Perspectives 115, 1283–1292. doi:10.1289/ehp.10047.

Stieb, D.M., Chen, L., Hystad, P., Beckerman, B.S., Jerrett, M., Tjepkema, M., Crouse, D.L., Omariba, D.W., Peters, P.A., van Donkelaar, A., Martin, R.V., Burnett, R.T., Liu, S., Smith-Doiron, M., Dugandzic, R.M., 2016. A national study of the association between traffic-related air pollution and adverse pregnancy outcomes in Canada, 1999-2008. Environmental Research 148, 513–526. doi:10.1016/j.envres.2016.04.025.

Su, J.G., Jerrett, M., Beckerman, B., 2009. A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. Science of the Total Environment 407, 3890–3898. doi:10.1016/j.scitotenv.2009.01.061.

Vienneau, D., De Hoogh, K., Bechle, M.J., Beelen, R., Van Donkelaar, A., Martin, R.V., Millet, D.B., Hoek, G., Marshall, J.D., 2013. Western European land use regression incorporating satellite- and ground-based measurements of NO_2 and PM_{10}. Environmental Science and Technology 47, 13555–13564. doi:10.1021/es403089q.

Vienneau, D., de Hoogh, K., Briggs, D., 2009. A GIS-based method for modelling air pollution exposures across Europe. Science of the Total Environment 408, 255–266. doi:10.1016/j.scitotenv.2009.09.048.

Wright, M.N., Ziegler, A., 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software 77, 1–17. doi:10.18637/jss.v077.i01.