



Estimating historical PM_{2.5} exposures for three decades (1987–2016) in Japan using measurements of associated air pollutants and land use regression[☆]

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

ARTICLE INFO

Article history:

Received 2 October 2019

Received in revised form

25 March 2020

Accepted 25 March 2020

Available online 5 April 2020

Keywords:

Air pollution

Machine learning

Temporal trend

Spatial distribution

ABSTRACT

Accurate estimation of historical PM_{2.5} exposures for epidemiological studies is challenging when extensive monitoring data are limited in duration. Here, we develop a national-scale PM_{2.5} exposure model for Japan using measurements recorded between 2014 and 2016 to estimate monthly means for 1987 through 2016. Our objective is to obtain accurate PM_{2.5} estimates for years prior to implementation of extensive PM_{2.5} monitoring, using observations from a limited period. We utilize a neural network to convey the non-linear relationship between the target pollutant and predictors, while incorporating the associated air pollutants. We obtain high R^2 values of 0.76 and 0.73 through spatial and temporal cross validation, respectively. We evaluate estimation accuracy using an independent data set and achieve an R^2 of 0.75. Moreover, monthly variations for 2000–2013 are well reproduced with correlation coefficients of greater than 0.78, obtained through a comparison with observations. We estimate monthly means at 1×1 km resolution from 1987 through 2016. The estimates show decreases in the area and population weighted means beginning in the 1990s. We successfully estimate monthly mean PM_{2.5} concentrations over three decades with outstanding predictive accuracy. Our findings illustrate that the presented approach achieves accurate long-term historical estimations using observations limited in duration.

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Exposure to PM_{2.5} (particulate matter with aerodynamic diameters equal to or less than 2.5 μm) has been associated with adverse health outcomes in many epidemiological studies (Rich et al., 2009; Faiz et al., 2013; Puett et al., 2014; Fleischer et al., 2014; Cohen et al., 2017). Proper exposure assessment requires accurate estimates of the spatial and temporal variations in the study domain and period. Statistical models, referred to as land use regression (LUR) models and their derivatives, have typically been used for this purpose (Beelen et al., 2013; Sampson et al., 2013; Beckerman et al., 2013; Di et al., 2016a). Because these models are dependent on observations, many previous models estimated exposures in modeling periods for which measurements are available

(Knibbs et al., 2014; Bechle et al., 2015; Proietti et al., 2016; Huang et al., 2018).

In cohort studies, accurate estimation of historical exposures is important in assessing the longitudinal impact of PM_{2.5} on chronic health outcomes such as cancer and cardiovascular disease. However, for areas where monitoring networks have been installed only recently, studying PM_{2.5} exposures and health outcomes will be difficult until sufficient monitoring data are available to properly estimate exposures. Meanwhile, epidemiological knowledge specific to the area is required for proper implementation of environmental policies such as emission controls and air quality standards.

Previous studies have presented historical estimation of PM_{2.5} for time periods prior to the implementation of extensive monitoring networks. Lall et al. (2004) built a model based on the ratio of PM_{2.5} to PM₁₀ or TSP (total suspended particle) obtained from observations recorded between 1987 and 2000, and estimated quarterly PM_{2.5} in the U.S. from 1979 through 1986. Li et al. (2017b) trained a generalized additive model using measurements recorded between 1999 and 2013, and predicted daily PM_{2.5} in California,

[☆] This paper has been recommended for acceptance by Admir Créso Targino.

* Corresponding author.

E-mail address: araki@ea.see.eng.osaka-u.ac.jp (S. Araki).

U.S., from 1990 through 1998. Kim et al. (2017) constructed a model using the estimated long-term trend in observed $PM_{2.5}$ between 1999 and 2010 to predict annual $PM_{2.5}$ in the U.S. from 1990 through 1998. These attempts used relatively long-term data sets (i.e., spanning more than ten years), which is reasonable because historical trends can be better estimated using observations recorded over longer periods. This would be possible in a country or area, such as the U.S., where a comprehensive monitoring network was implemented early on. Conversely, back-extrapolation would be infeasible for countries or areas where extensive networks developed more recently, because available observations are limited in duration.

In the statistical model, predictors such as land use, traffic-related, and meteorological variables are generally used for model constructions and predictions. Although these predictors represent the spatial variations at least in a certain time window, they would not sufficiently represent temporal variability when pollution levels change over time. Thus, variables sufficiently reflecting temporal variation are essential for back-extrapolation. Satellite-retrieved data has been widely used as a predictor in statistical models for NO_2 (Hoek et al., 2015; Xu et al., 2019) and $PM_{2.5}$ (Mao et al., 2012; Xu et al., 2019). Because they can be a useful predictor for capturing temporal as well as spatial variations, they are utilized in models estimating long-term variations in NO_2 (Knibbs et al., 2014; Bechle et al., 2015) and $PM_{2.5}$ (Beckerman et al., 2013; Di et al., 2016a; Zhang et al., 2018; Xu et al., 2018). However, this approach cannot be applied to estimates for the time period before Terra was launched in 1999 (Li et al., 2017b).

Meanwhile, PM_{10} could be a predictor for modeling historical $PM_{2.5}$ concentrations in earlier years, because it was monitored extensively before the comprehensive $PM_{2.5}$ network was implemented. Li et al. (2017b) used PM_{10} measurements recorded between 1999 and 2013 as one of the predictors for a generalized additive model, and estimated daily $PM_{2.5}$ in California, U.S., from 1990 through 2013 with a cross-validated R^2 of 0.79. Sulfate and nitrate, which are secondary constituents of $PM_{2.5}$, are generated by the oxidation of SO_2 and NO_2 , respectively. Thus, in addition to coarse particulate matter, these air pollutants share emission sources with $PM_{2.5}$ and can be informative predictors in the $PM_{2.5}$ model. Moreover, similar to PM_{10} , SO_2 and NO_2 were monitored extensively before the launch of the comprehensive $PM_{2.5}$ network, which raises the possibility of pre-1999 estimations. However, to our best knowledge, applications of these pollutants to model predictors for historical exposure assessments have been limited thus far.

In Japan, the national air quality standard for $PM_{2.5}$ was implemented in 2009. National-scale $PM_{2.5}$ monitoring started in 2010 with a small number of stations, and was gradually extended until an extensive network was established after 2014 when the number of monitors generally reached more than 700. At the time of this study, the extensive data set only covers the years 2014–2016. Before the launch of the national monitoring network, a limited number of monitoring stations (approximately ten, depending on the year) performed governmental monitoring starting in 2000. Thus, detailed spatial variation is unclear before 2010, and no information is available prior to 2000.

In this study, we develop a national-scale monthly $PM_{2.5}$ exposure model for Japan using measurements recorded between 2014 and 2016, incorporating the associated air pollutants as predictors. The overall objective is to obtain accurate $PM_{2.5}$ estimations for the years before the launch of the extensive monitoring network, using monitoring data from a limited time period. We utilize a neural network to model the complex, non-linear relationships between the target pollutant and predictors. We validate our model through cross validation, as well as external

validation using an independent data set containing 14 years (2000–2013) of data, and estimate monthly mean $PM_{2.5}$ from 1987 through 2016. Further, we discuss the accuracy of our model and the estimated historical trend.

2. Methods

2.1. Study area

The study area includes the main islands of Japan (130°E – 146°E, 31°N – 46°N); remote or small islands are omitted (Fig. S1).

2.2. Air quality measurements

We obtained $PM_{2.5}$ observations from two databases: the national air quality monitoring network (AQMN) database and the governmental monitoring database. The AQMN was developed in the early 1980s and $PM_{2.5}$ monitoring was started in 2010. The AQMN stations are maintained by local governments. Detail of the AQMN database usage is presented in Supplementary Material.

Governmental $PM_{2.5}$ monitoring was started by the Japanese Ministry of the Environment with four stations in 2000, ten years before the implementation of the national-scale network. The number of governmental sites increased gradually to 18 in 2009, one year before the implementation of the national network. We obtained data from four stations that are continuously operated: Nonodake (NND), Kawasaki (KWS), Osaka (OSK), and Amagasaki (AMG). Station NND is located in the rural area of northern Japan. Station KWS is installed in the Tokyo metropolitan area. Stations OSK and AMG are both positioned in the Osaka metropolitan area. These monitoring sites have been continuously operated since 2000. The locations of the stations are presented in Fig. S1. The availability and usage of the $PM_{2.5}$ monitoring data are summarized in Table S2.

We used the monitoring data recorded between 2014 and 2016 for model building, the latest available three years, when the number of monitors are more than 700 and became generally constant (Fig. S2), in order to avoid heterogeneity of observations in space and time and to better represent the study area and period. The remaining data from 2010 to 2013 are used as an independent test dataset for model validation. The governmental monitoring data are used to evaluate the correlation and trend. The model validation is described in more detail below. Fig. S3 shows the time series of the monthly means we used for model construction. The temporal variation is not clearly recognized, although it is generally high during spring.

2.3. Data set

Associated air pollutants, including suspended particulate matter (SPM), SO_2 , and NO_2 , were used as predictors. SPM has been measured in Japan since early 1970s as airborne particulate matter and is defined as particulate matter with a diameter of less than 10 μm by the Air Pollution Control Act. Thus, SPM is different from PM_{10} and corresponds to approximately to PM_7 (Kasahara, 2002), that is, particulate matter with an aerodynamic diameter of less than 7 μm . We obtained the associated pollutant monitoring data from the AQMN database. We spatially interpolated the monthly mean values of the three pollutants by ordinary kriging to derive gridded data at 1 km resolution and extracted the kriged values at the location and time of the observed $PM_{2.5}$. We did not use observations from collocated monitors because $PM_{2.5}$ monitors are not always collocated with all three pollutant monitors.

We include the emission intensity of large point sources as a predictor. Emission intensity was derived from the Emissions

Database for Global Atmospheric Research (EDGAR v4.3.2) (Crippa et al., 2018), which is a global emission inventory at 0.1° resolution covering an extensive time series (1970–2012). We retrieved annual grid data from road transport and large point sources for the years corresponding to our study period. The sectors we used for the latter sources (Table S3) were selected on the basis of estimated emission amounts in Japan (Crippa et al., 2018). Because emission data are not available for years after 2012, we used 2012 emission data for the analyses between 2013 and 2016 without any correction. For large point sources, we consider the distance decay effect (Araki et al., 2018) using the moving window approach (Vienneau et al., 2009).

We prepared gridded monthly mean precipitation, temperature, relative humidity, and wind speed at 1 km resolution by interpolation of the observations via ordinary kriging (Araki et al., 2018). The build-up area and green area ratio were derived using land use data (Araki et al., 2018). The land use and population data we used are originally at 1 km resolution. These data are available only for specific years, as shown in Table S4. For the years for which no data are available, we used the data of the closest year. We used the nearest straight-line distance from the observation and prediction points to the coastline as a predictor (Araki et al., 2017). We utilized longitude as a predictor representing the effect of long-range transport from the Asian continent (Araki et al., 2017).

We summarized the predictor variables described above in Table 1. Details of the data sets and sources are given in Table S4.

2.4. Neural networks

Artificial neural networks are mathematical models inspired by brain activity, consisting of artificial neurons (nodes) in layers (Gardner and Dorling, 1998; Agirre-Basurko et al., 2006; Cobourn et al., 2000). The network is formed by three layers of nodes: an input layer, one or more hidden layers, and an output layer. Neural networks make no prior assumptions regarding data distribution, unlike other statistical techniques (Gardner and Dorling, 1998; Cobourn et al., 2000). Neural networks can handle complex non-linear relationships between input and output variables (Di et al., 2016a), and are used in many applications related to atmospheric problems (Agirre-Basurko et al., 2006; Cobourn et al., 2000; Russo et al., 2013; Biancofiore et al., 2017; Di et al., 2016a, b; Ganesh et al., 2018).

Using the above listed variables, we trained a neural network against the measured PM_{2.5} concentrations. We used one hidden layer with seven nodes, because additional layers could have made the model unnecessarily complex and difficult to train (Cobourn et al., 2000). To eliminate randomness in the initial weights, we

trained 100 models (unless otherwise noted) with different random number seeds, and averaged the output from each network to obtain predictions (Kuhn et al., 2019). The weight decay was set to 0.01. The input data were normalized to the range of [0,1] by $y = (x - x_{min}) / (x_{max} - x_{min})$, where y is the normalized vector of the input variable and x is the raw input variable.

2.5. Evaluation

2.5.1. Cross validation

We conducted five-fold cross validation to evaluate the accuracy of our model. The monitoring locations were evenly split into five groups at random. The observations from one group of locations were removed for the whole period, and the 10 networks were trained with different random number seeds using the observations from the remaining four splits. Predictions at the time and location of the removed observations were obtained by averaging the output from each network. This procedure was repeated for the four groups. We iterated the five-fold cross validation 50 times with different random number seeds to extinguish the randomness of the splits. Finally, the cross validation predictions were obtained by averaging the 50 outputs and compared to the observed values. Thus, 500 networks in total were trained. We calculated R^2 and root mean squared error (RMSE) between the predictions and observations as the prediction accuracy indexes. R^2 was obtained by $1 - MSE/var(Y)$, where Y is a vector of observations and MSE is the mean squared error. We refer to this procedure as spatial cross validation.

We also performed temporal cross validation. In this process, observations were removed for a particular year (12 months), and the model was built using the data for the remaining two years (24 months). Values at the monitoring locations in the selected year were estimated by the model. This procedure was repeated for the other two years, and R^2 and RMSE values were obtained. Because we intended to back-extrapolate for more than 25 years, we removed observations year by year, not month by month, in order to evaluate the temporal predictive power of our model in a manner better suited to our purpose.

2.5.2. External validation

In addition to cross validation, we evaluated our model using an independent data set that was not included in the model building process. We refer to this process as external validation. From the AQMN database, we obtained PM_{2.5} data from 2010 to 2013 and used them for external validation. We estimated monthly PM_{2.5} at the time and locations of the external validation data. We computed R^2 as well as the RMSE between the estimated and measured values.

2.5.3. Reproducibility in monthly variations

We compared our model estimates against the observed monthly PM_{2.5} at the four continuously operated stations to further validate our model performance in terms of the reproducibility of monthly variations and trends. The observations were obtained from governmental monitoring performed during 2000–2013. Because the monitors at the four sites do not incorporate the standard analytical method for the current national-scale network set by the Japanese Ministry of the Environment, we only evaluate the correlation and trend, rather than comparing the absolute concentrations.

2.6. Estimation

We estimated the monthly mean PM_{2.5} at 1 × 1 km resolution in the whole study area from 1987 through 2016. To examine the

Table 1
Predictor variables.

Variables	Unit
SPM	$\mu\text{g m}^{-3}$
SO ₂	ppb
NO ₂	ppb
Emission from transportation	$\text{kg m}^{-2} \text{s}^{-1}$
Emission from large point sources	$\text{kg m}^{-2} \text{s}^{-1}$
Precipitation	mm hr^{-1}
Wind speed	m sec^{-1}
Temperature	°C
Relative humidity	%
Built-up area ratio	unitless
Green area ratio	unitless
Population	number
Distance to coastline	km
Longitude	Degree

long-term trend of PM_{2.5}, we calculated the monthly arithmetic average of the estimations at all the prediction grids, as well as the population weighted means. The former can be seen as area weighted means. The latter is obtained by

$$pwm = \frac{\sum_{i=1}^n C_i \cdot p_i}{\sum_{i=1}^n p_i}, \tag{1}$$

where *pwm* is the population weighted mean, *C_i* and *p_i* is the estimation and population in the *i* th prediction grid, respectively, and *n* is the number of prediction grids. The population data from the closest year is used for the calculation, because the data are not available for every year, as mentioned earlier.

2.7. Computation

All the computations were done using R statistical software 3.5.3 (R Core Team, 2019); the **raster** (Hijmans, 2019) and **gstat** (Pebesma, 2004) packages were used for preparing the predictor variables, and the **nnet** (Venables and Ripley, 2002) and **caret** (Kuhn et al., 2019) packages were used for training the neural network.

3. Results

3.1. Cross validation

The scatter plot of the observations and the predictions obtained by the spatial cross validation is given in Fig. 1 a). The colors of the dots represent the density of the points in the plot; red represents higher density and blue represents lower density. *R*² and RMSE are shown in the panel. We achieved an *R*² value of 0.76, with an RMSE value of 1.9 μg m^{−3}. Fig. S4 presents the spatial cross validation results plotted year by year. *R*² and RMSE values are similar between years.

Fig. 1 b) presents a scatter plot of the observations and the predictions obtained by the temporal cross validation. We obtained an *R*² value of 0.73, with an RMSE value of 2.0 μg m^{−3}. Fig. S5

presents the annual plots of the temporal cross validation. The differences in *R*² and RMSE values for the three years are marginal.

3.2. External validation

Fig. 2 shows the external validation result obtained by using the independent data set for 2010–2013. We achieved an *R*² of 0.75 with an RMSE value of 2.2 μg m^{−3}. These statistical indicators are generally similar to those for the spatial and temporal cross validation, and are comparable between years (Fig. S6). The estimations are not significantly over- or underestimated, which is apparent from Fig. 2.

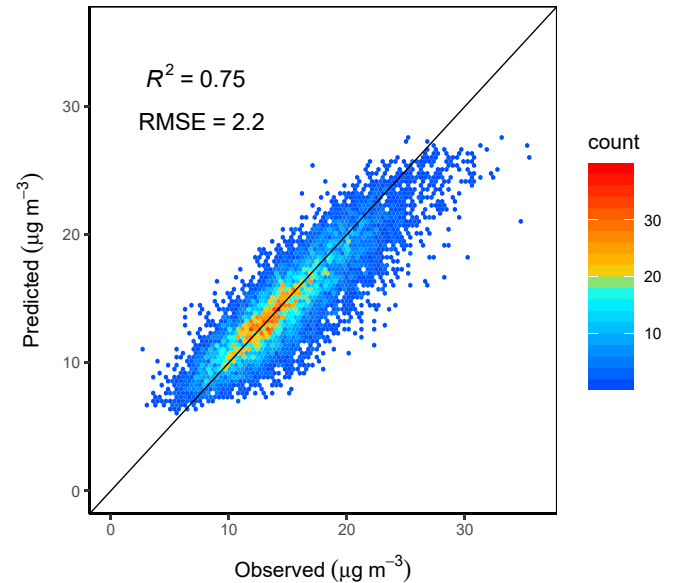


Fig. 2. Scatter plots of the observed and predicted concentrations obtained from external validation. Red represents higher point density and blue represents lower density. The line represents the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

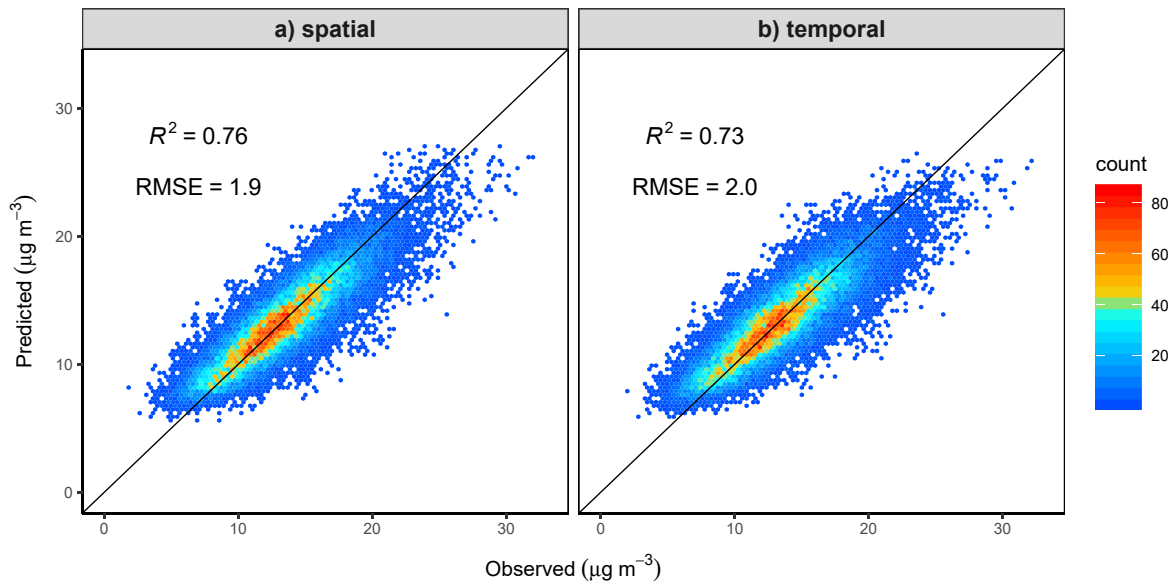


Fig. 1. Scatter plots of the observed and predicted concentrations obtained from a) spatial and b) temporal cross validation. Red represents higher point density and blue represents lower density. The lines in each panel represent the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.3. Reproducibility in monthly variations

Fig. 3 presents the temporal variation in the observed and estimated monthly means for the four continuously operated monitoring stations: NND, KWS, OSK, and AMG. The plotted lines for the observed means are disconnected where observations are missing. The temporal variations are generally well reproduced, with high Pearson's correlation coefficients (r) of greater than 0.78. We aggregated the observed and estimated monthly means in each calendar year into annual means. The estimations also conform to the long-term trends of the annual observations at the four locations, which is apparent from Fig. S7.

3.4. Estimations

Fig. 4 presents the annual prediction maps at 1×1 km resolution obtained by aggregating the estimated monthly means in each calendar year into annual means. While the concentrations in the southern part of Japan are consistently higher than those in the northern part throughout the study period, the estimations clearly decreased over the three decades in the whole study area. Fig. 5 shows the temporal variations in the area weighted and population weighted monthly mean estimations. The bold line represents the smoothed line obtained via locally estimated scatterplot smoothing (LOESS). The population weighted means are

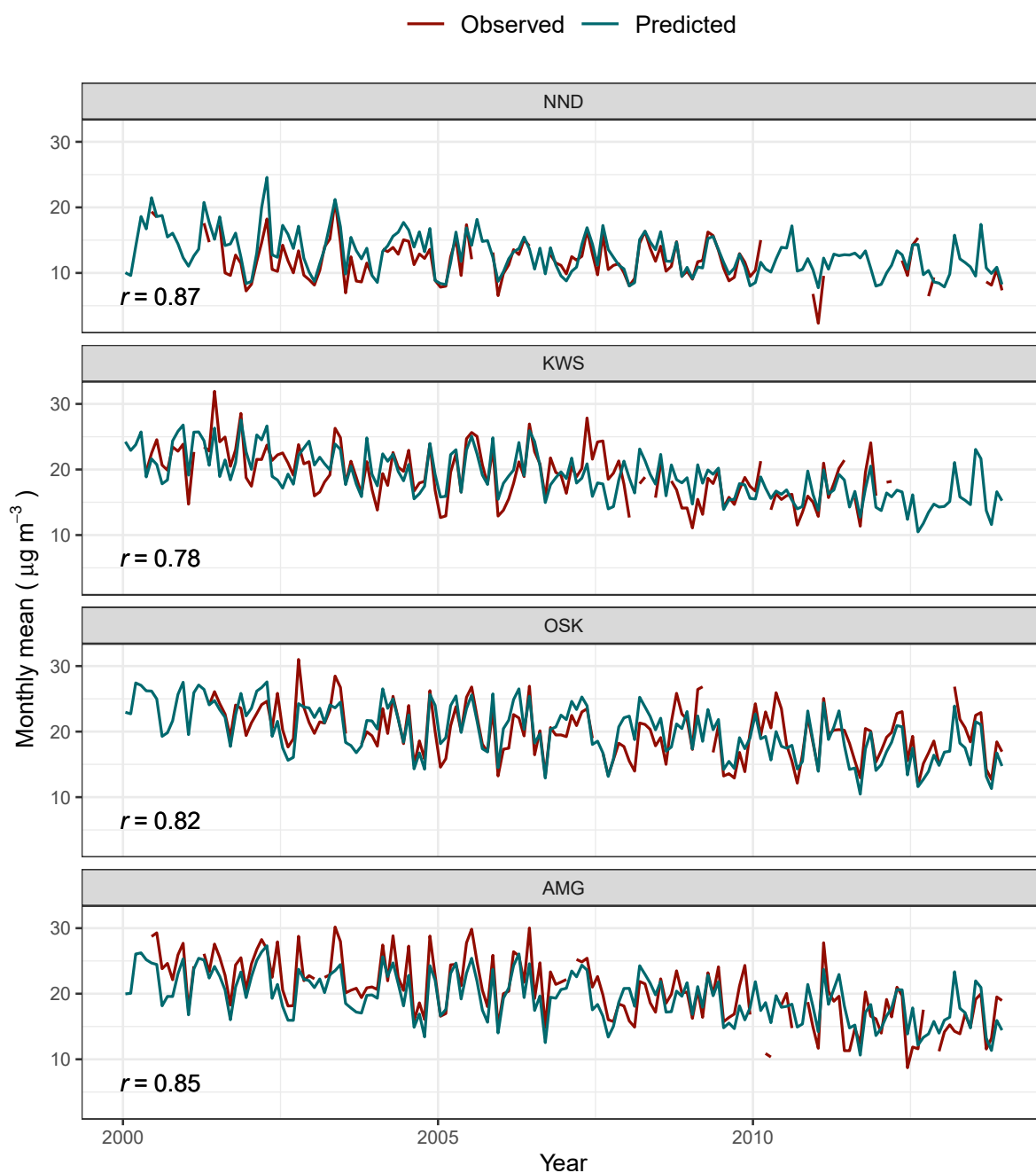


Fig. 3. Time series of the predicted and observed monthly $\text{PM}_{2.5}$ concentrations during 2000–2013 at stations NND, KWS, OSK, and AMG. The Pearson's correlation coefficients (r) between the estimated and measured values are presented in each panel.

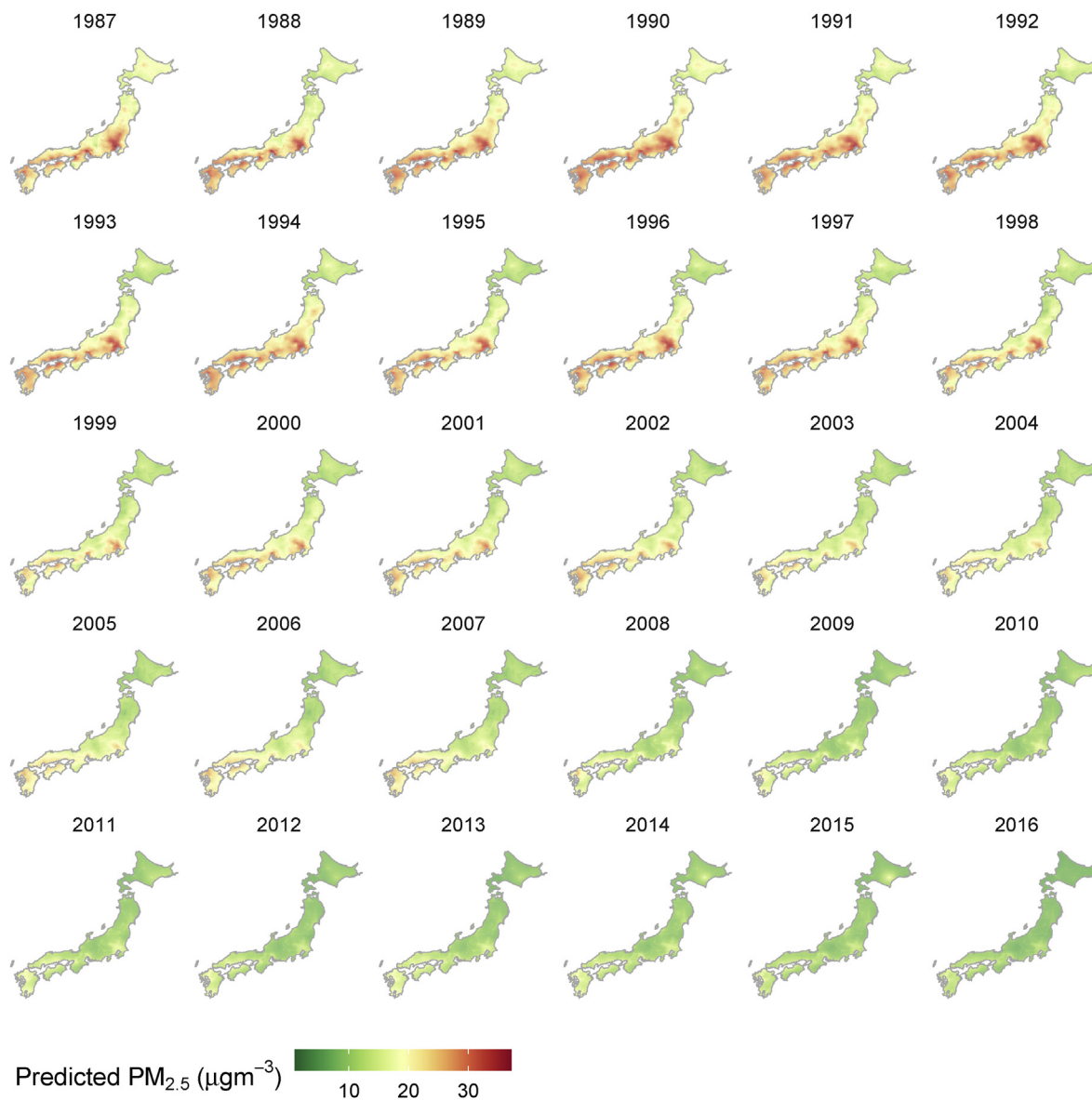


Fig. 4. Annual prediction maps obtained by aggregating the estimated monthly means from 1987 through 2016.

consistently higher than the area weighted means, although the difference decreases in later years. The decreasing trend is exhibited in both the area weighted and population weighted means. The decrement in concentrations for both of the means became remarkable in the 1990s.

4. Discussion

We developed the national-scale $PM_{2.5}$ exposure model to estimate the monthly means for three decades (1987–2016). Our model is accurate, with spatial cross validated R^2 and RMSE values of 0.76 and $1.9 (\mu g m^{-3})$, respectively, and temporal cross validated R^2 and RMSE values of 0.73 and $2.0 (\mu g m^{-3})$, respectively. Making predictions for the years beyond the modeling period might involve larger uncertainty due to the potential change in the relationship between $PM_{2.5}$ and predictors. Thus, in addition to the cross validation, we evaluated our model using two independent data sets recorded before the year 2014: the AQMN data and the governmental monitoring data. For external validation, we used AQMN

data recorded between 2010 and 2013 and achieved R^2 and RMSE values of 0.76 and $2.2 (\mu g m^{-3})$, respectively, which are nearly identical to the indicators obtained by the spatial and temporal cross validation. This indicates that our model is robust without overfitting. No significant over- or underestimation is recognized in the three validation results, as illustrated in Figs. 1 and 2. The spatial and temporal cross validations yield similar R^2 and RMSE values between years. This shows the homogeneity in the model accuracy between years. Meanwhile, we assessed the reproducibility of the long-term trend and the monthly variations between 2000 and 2013 using the independent data sets obtained from governmental monitoring. The monthly variations and the long-term trend are well reproduced with r values of greater than 0.78, as presented in Fig. 3. These results indicate that our model accurately reproduces the spatial and temporal variability of $PM_{2.5}$ concentrations for the times when observations are available.

Meanwhile, we were unable to validate the prediction accuracy for the years without recorded observations (1987–1999). Levy et al. (2015) estimated annual NO_x concentrations in Israel

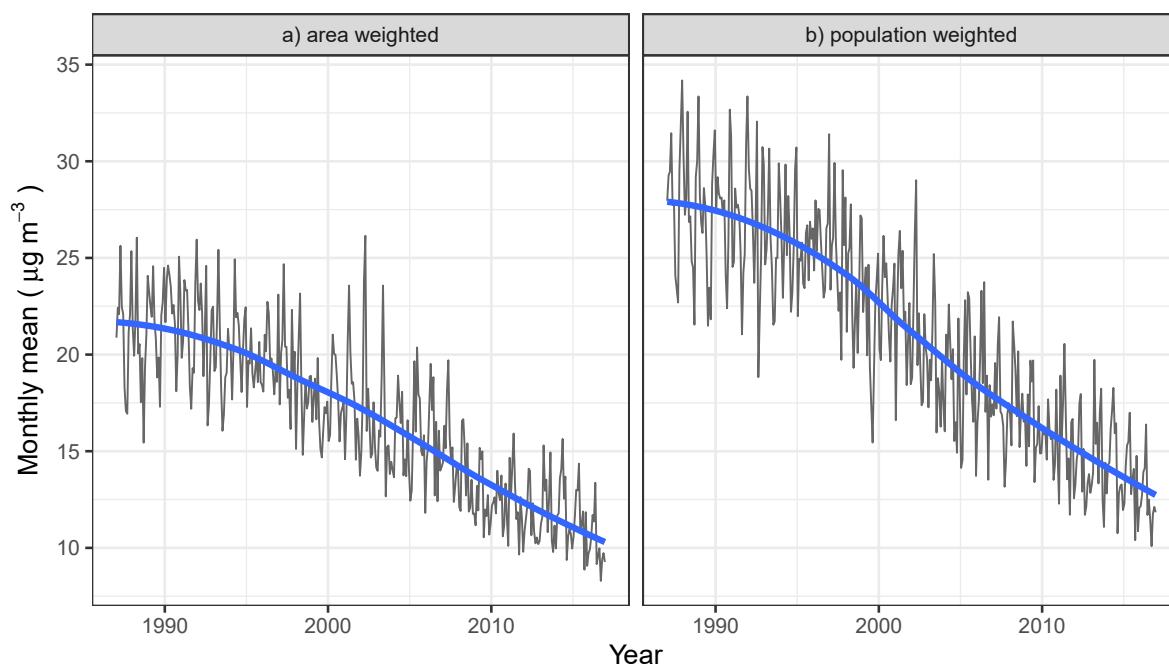


Fig. 5. Time series of the estimated monthly $\text{PM}_{2.5}$ concentrations of a) area weighted and b) population weighted means and their smoothed trends (bold lines) from 1987 through 2016.

between 1961 and 2011, employing a LUR model developed using observations recorded between 1991 and 2011. They validated the estimations against the independent annual NO_x emission estimates from road transport sources. Unlike NO_x , $\text{PM}_{2.5}$ consists of primary and secondary pollutants such as sulfate and nitrate, which are produced by oxidation of SO_2 and NO_2 , respectively. Long-range transport from the Asian continent has been reported to contribute to $\text{PM}_{2.5}$ concentrations in Japan (Aikawa et al., 2010; Shimadera et al., 2016). In addition, we used $\text{PM}_{2.5}$ emissions from large point sources as one of the predictors. Therefore, using emission data for evaluating our estimations would not be appropriate. Similarly, the associated air pollutants could not be used for validation of the model. Satellite derived AOD cannot be applied for evaluations for years prior to 1999 (Li et al., 2017b).

Alternatively, we compare our trend estimations with that of a previous study. Li et al. (2017a) estimated gridded global $\text{PM}_{2.5}$ trends between 1989 and 2013 using a chemical transport model. In their results, the grids over Japan exhibit a decreasing trend of approximately $0.3\text{--}0.4 \mu\text{g m}^{-3} \text{yr}^{-1}$. This is slightly lower than, but in a similar range as, our result of $0.4 \mu\text{g m}^{-3} \text{yr}^{-1}$, which is calculated from the area weighted means presented in Fig. 5. This comparison shows that our trend estimates are generally consistent with the previous study.

These evaluation results suggest that our model estimations can be expected to be accurate not only for the later period (2000–2016) but also for the years (1987–1999) for which monitoring data are not available, thus sufficiently capturing the spatial and temporal variation of $\text{PM}_{2.5}$.

The population weighted means are consistently higher than the area weighted means, as indicated in Fig. 5; this implies that areas with higher population densities have higher $\text{PM}_{2.5}$ concentrations due to local anthropogenic emissions. Similar results were reported in a previous study (Li et al., 2017a), confirming that population weighted $\text{PM}_{2.5}$ is higher than area weighted $\text{PM}_{2.5}$ in populated regions worldwide. The downward trend in $\text{PM}_{2.5}$ may owe to the decrease in local emissions. Itahashi and Hayami (2018)

conducted $\text{PM}_{2.5}$ source apportionment in Japan for the years 2000–2008 using a chemical transport model, and found that decreases in domestic anthropogenic emissions, particularly those from road transport, contributed to the downward trend. The decline in local anthropogenic emissions may reduce the difference between the population and area weighted means. This might be the effect of governmental regulations on automobile exhaust and volatile organic compound (VOC) emissions.

We utilized the associated air pollutant information as the model predictors. These predictors exhibit high correlation with the monitored $\text{PM}_{2.5}$ (Fig. S8) and could contribute to the performance of the model. These pollutants share some emission sources with $\text{PM}_{2.5}$, and in turn are influenced by meteorological conditions in a similar manner as with $\text{PM}_{2.5}$. Thus, they can be expected to reflect the spatial and temporal variation of $\text{PM}_{2.5}$ to some extent. Additionally, these variables have the same temporal resolution as the observed $\text{PM}_{2.5}$. The relationship between the associated pollutants and $\text{PM}_{2.5}$ is not causality, but correlation. Therefore, the associated pollutants are essentially different from commonly used predictors such as land use, meteorological, and traffic variables, which directly or indirectly affect $\text{PM}_{2.5}$ concentrations. Although satellite derived data are similar in this regard and have higher spatial coverage, the associated air pollutant information provides better temporal representativeness with significantly less missing data because it is recorded by onsite monitoring stations in continuous operation. Given these features, the information from the associated air pollutants can contribute to the predictive accuracy of statistical models and should be further examined for its applicability in a different context.

In this regard, building a daily model would be technically possible using the associated air pollutant information as well as meteorological parameters. More accurate monthly estimates could be derived by aggregating daily predictions obtained by a successfully trained daily model which capture spatial variations at a daily level. Meanwhile, the performance of a daily model would be crucial for improving the monthly estimation accuracy.

Preparing predictors which sufficiently represent PM_{2.5} daily variations is required to successfully build a daily model. However, it is particularly difficult for variables representing emission intensity. Therefore, constructing a daily model aiming at improving monthly estimation accuracy should be further examined.

There are some limitations in this study. First, we did not use some predictors that are unavailable in data from previous years. For instance, we used PM_{2.5} emissions from large point sources obtained from EDGAR v4.3.2 (Crippa et al., 2018), which is available from 1970 through 2012 at $0.1^\circ \times 0.1^\circ$ resolution. Although there is an emission inventory for Japan at 1×1 km resolution (Kannari et al., 2007; Fukui et al., 2014), this database does not contain data for years prior to 2000. Additionally, we did not use road network related variables as our model predictors, because historical road network data are not publicly available. Instead, we used road transport emission data obtained from EDGAR v4.3.2 (Crippa et al., 2018) at $0.1^\circ \times 0.1^\circ$ resolution, which may be too coarse for our model. Therefore, the accuracy of the model for recent years might be compromised to maintain the accuracy for the whole study period. We prioritized the temporal consistency of the model predictors over the estimation accuracy of a limited number of years, which may affect the prediction accuracy for the overall period.

Second, our estimations may not well represent the PM_{2.5} concentrations in non-residential areas. To generate predictors for our model, we spatially interpolated the monitored concentrations of the associated air pollutants by ordinary kriging. The monitoring sites are not homogeneously distributed; in fact, they are mostly located in residential areas where the concentrations of such air pollutants are generally higher than those in non-residential areas. The kriged concentrations, the predictors of our model, in the non-residential areas are likely to be over-influenced by the observed values of the associated pollutants in residential areas. Our model estimations in non-residential areas are in turn strongly influenced by the monitored values of the associated pollutants in residential areas. Thus, our estimations in non-residential areas may not well represent the actual pollution levels, rather reflecting the concentrations in residential areas. Therefore, our model may over-estimate PM_{2.5} concentrations in non-residential areas. Conversely, the potential uncertainty in our predictions is limited to non-residential areas. This is acceptable because we focus on PM_{2.5} exposures. To better estimate PM_{2.5} concentrations over an entire area, a more accurate spatial interpolation of the associated pollutants should be considered.

5. Conclusion

Despite these limitations, we successfully developed our model for estimating national-scale monthly mean PM_{2.5} concentrations for three decades with outstanding predictive accuracy. We demonstrated that an accurate long-term historical estimation can be achieved with observations limited in duration by our innovative approach, allowing to help fill the gap between limited data availability and prompt epidemiological knowledge accumulation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRedit authorship contribution statement

Shin Araki: Formal analysis, Visualization, Writing - original draft, Writing - review & editing. **Masayuki Shima:** Validation,

Writing - review & editing, Project administration. **Kouhei Yamamoto:** Conceptualization, Validation, Writing - review & editing, Funding acquisition.

Acknowledgement

This work was supported by a Grant for Environmental Research Projects from The Sumitomo Foundation, Japan, Grant Number 183141 and JSPS, Japan, KAKENHI Grant Number JP19K12370 and JP18H03060.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envpol.2020.114476>.

References

- Agirre-Basurko, E., Ibarra-Berastegi, G., Madariaga, I., 2006. Regression and multi-layer perceptron-based models to forecast hourly O₃ and NO₂ levels in the Bilbao area. *Environ. Model. Software* 21, 430–446. <https://doi.org/10.1016/j.envsoft.2004.07.008>.
- Aikawa, M., Ohara, T., Hiraki, T., Oishi, O., Tsuji, A., Yamagami, M., Murano, K., Mukai, H., 2010. Significant geographic gradients in particulate sulfate over Japan determined from multiple-site measurements and a chemical transport model: impacts of transboundary pollution from the Asian continent. *Atmos. Environ.* 44, 381–391. <https://doi.org/10.1016/j.atmosenv.2009.10.025>.
- Araki, S., Shima, M., Yamamoto, K., 2018. Spatiotemporal land use random forest model for estimating metropolitan NO₂ exposure in Japan. *Sci. Total Environ.* 634, 1269–1277. <https://doi.org/10.1016/j.scitotenv.2018.03.324>.
- Araki, S., Shimadera, H., Yamamoto, K., Kondo, A., 2017. Effect of spatial outliers on the regression modelling of air pollutant concentrations: a case study in Japan. *Atmos. Environ.* 153, 83–93. <https://doi.org/10.1016/j.atmosenv.2016.12.057>.
- Bechle, M.J., Millet, D.B., Marshall, J.D., 2015. National spatiotemporal exposure surface for NO₂: monthly scaling of a satellite-derived land-use regression, 2000–2010. *Environ. Sci. Technol.* 49, 12297–12305. <https://doi.org/10.1021/acs.est.5b02882>.
- Beckerman, B.S., Jerrett, M., Serre, M., Martin, R.V., Lee, S.J., van Donkelaar, A., Ross, 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. *Environ. Sci. Technol.* 47, 7233–7241. <https://doi.org/10.1021/es400039u>.
- 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., Gražulevičienė, 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., Strömberg, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., de Hoogh, K., 2013. Development of NO₂ and NO₂ land use regression models for estimating air pollution exposure in 36 study areas in Europe - the ESCAPE project. *Atmos. Environ.* 72, 10–23. <https://doi.org/10.1016/j.atmosenv.2013.02.037>.
- Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., Di Tommaso, S., Colangeli, C., Rosatelli, G., Di Carlo, P., 2017. Recursive neural network model for analysis and forecast of PM₁₀ and PM_{2.5}. *Atmos. Pollut. Res.* 8, 652–659. <https://doi.org/10.1016/j.apr.2016.12.014>.
- Coburn, W.G., Dolcine, L., French, M., Hubbard, M.C., 2000. A comparison of nonlinear regression and neural network models for ground-level ozone forecasting. *J. Air Waste Manage.* 50, 1999–2009. <https://doi.org/10.1080/10473289.2000.10464228>.
- Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C.A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C.J.L., Forouzanfar, M.H., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* 389, 1907. [https://doi.org/10.1016/S0140-6736\(17\)30505-6](https://doi.org/10.1016/S0140-6736(17)30505-6).
- Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Van Aardenne, J.A., Monni, S., Doering, U., Olivier, J.G., Pagliari, V., Janssens-Maenhout, G., 2018. Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2. *Earth Syst. Sci. Data* 10, 1987–2013. <https://doi.org/10.5194/essd-10-1987-2018>.
- 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. *Environ. Sci. Technol.* 50, 4712–4721. <https://doi.org/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. *Atmos. Environ.* 131, 390–399. <https://doi.org/10.1016/j.atmosenv.2016.02.002>.
- 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. <https://doi.org/10.1097/EDE.0b013e3182949ce5>.
- Fleischer, N.L., Merilä, M., van Donkelaar, A., Vadiello-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. *Environ. Health Perspect.* 122, 425–430. <https://doi.org/10.1289/ehp.1306837>.
- Fukui, T., Kokuryo, K., Baba, T., Kannari, A., 2014. Updating EAGrid2000-Japan emissions inventory based on the recent emission trends. *J. Jpn. Soc. Atmos. Environ.* 49, 117–125. <https://doi.org/10.11298/taiki.49.117>.
- Ganesh, S.S., Arulmozhiarman, P., Tatavarti, V.S., 2018. Prediction of PM_{2.5} using an ensemble of artificial neural networks and regression models. *J. Amb. Intel. Hum. Comp.* <https://doi.org/10.1007/s12652-018-0801-8>.
- Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmos. Environ.* 32, 2627–2636. [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0).
- Hijmans, R.J., 2019. Raster: geographic data analysis and modeling. <https://CRAN.R-project.org/package=raster>. r package version 2.9-5.
- Hoek, G., Eeftens, M., Beelen, R., Fischer, P., Brunekreef, B., Boersma, K.F., Veeckind, P., 2015. Satellite NO₂ data improve national land use regression models for ambient NO₂ in a small densely populated country. *Atmos. Environ.* 105, 173–180. <https://doi.org/10.1016/j.atmosenv.2015.01.053>.
- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D., Liu, Y., 2018. Predicting monthly high-resolution PM_{2.5} concentrations with random forest model in the North China Plain. *Environ. Pollut.* 242, 675–683. <https://doi.org/10.1016/j.envpol.2018.07.016>.
- Itahashi, S., Hayami, H., 2018. Insights into the gradual decreasing trend of PM_{2.5}: before the notification of environmental standard (2000–2008). *J. Jpn. Soc. Atmos. Environ.* 219, 219–236. <https://doi.org/10.11298/taiki.53.219>.
- Kannari, A., Tonooka, Y., Baba, T., Murano, K., 2007. Development of multiple-species 1 km × 1 km resolution hourly basis emissions inventory for Japan. *Atmos. Environ.* 41, 3428–3439. <https://doi.org/10.1016/j.atmosenv.2006.12.015>.
- Kasahara, M., 2002. The present state and future assignments of air pollution by particulate matter - the meaning of innovation of PM_{2.5}. *J. Jpn. Soc. Atmos. Environ.* 37, 96–107. <https://doi.org/10.11298/taiki.1995.37.96>.
- Kim, S.Y., Olives, C., Sheppard, L., Sampson, P.D., Larson, T.V., Keller, J.P., Kaufman, J.D., 2017. Historical prediction modeling approach for estimating long-term concentrations of PM_{2.5} in cohort studies before the 1999 implementation of widespread monitoring. *Environ. Health Perspect.* 125, 38–46. <https://doi.org/10.1289/EHP131>.
- 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. *Environ. Res.* 135, 204–211. <https://doi.org/10.1016/j.envres.2014.09.011>.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2019. Caret: classification and regression training. <https://CRAN.R-project.org/package=caret>. r package version 6.0-84.
- Lall, R., Kendall, M., Ito, K., Thurston, G.D., 2004. Estimation of historical annual PM_{2.5} exposures for health effects assessment. *Atmos. Environ.* 38, 5217–5226. <https://doi.org/10.1016/j.atmosenv.2004.01.053>.
- Levy, I., Levin, N., Yuval, Y., Schwartz, J.D., Kark, J.D., 2015. Back-extrapolating a land use regression model for estimating past exposures to traffic-related air pollution. *Environ. Sci. Technol.* 49, 3603–3610. <https://doi.org/10.1021/es505707e>.
- Li, C., Martin, R.V., Van Donkelaar, A., Boys, B.L., Hammer, M.S., Xu, J.W., Marais, E.A., Reff, A., Strum, M., Ridley, D.A., Crippa, M., Brauer, M., Zhang, Q., 2017a. Trends in chemical composition of global and regional population-weighted fine particulate matter estimated for 25 years. *Environ. Sci. Technol.* 51, 11185–11195. <https://doi.org/10.1021/acs.est.7b02530>.
- Li, L., Wu, A.H., Cheng, I., Chen, J.C., Wu, J., 2017b. Spatiotemporal estimation of historical PM_{2.5} concentrations using PM₁₀ meteorological variables, and spatial effect. *Atmos. Environ.* 166, 182–191. <https://doi.org/10.1016/j.atmosenv.2017.07.023>.
- Mao, L., Qiu, Y., Kusano, C., Xu, X., 2012. Predicting regional space-time variation of PM_{2.5} with land-use regression model and MODIS data. *Environ. Sci. Pollut. Res.* 19, 128–138. <https://doi.org/10.1007/s11356-011-0546-9>.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30, 683–691.
- Proietti, E., Delgado-Eckert, E., Vienneau, D., Stern, G., Tsai, M.Y., Latzin, P., Frey, U., Röösl, M., 2016. Air pollution modelling for birth cohorts: a time-space regression model. *Environ. Health: Global Access Sci. Sour.* 15, 1–9. <https://doi.org/10.1186/s12940-016-0145-9>.
- Puett, R.C., Hart, J.E., Yanosky, J.D., Spiegelman, D., Wang, M., Fisher, J.A., Hong, B., Laden, F., 2014. Particulate matter air pollution exposure, distance to road, and incident lung cancer in the nurses' health study cohort. *Environ. Health Perspect.* 122, 926–932. <https://doi.org/10.1289/ehp.1307490>.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <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. *J. Epidemiol. Community Health* 63, 488–496. <https://doi.org/10.1136/jech.2008.082792>.
- Russo, A., Raischel, F., Lind, P.G., 2013. Air quality prediction using optimal neural networks with stochastic variables. *Atmos. Environ.* 79, 822–830. <https://doi.org/10.1016/j.atmosenv.2013.07.072>.
- 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. *Atmos. Environ.* 75, 383–392. <https://doi.org/10.1016/j.atmosenv.2013.04.015>.
- Shimadera, H., Kojima, T., Kondo, A., 2016. Evaluation of air quality model performance for simulating long-range transport and local pollution of PM_{2.5} in Japan. *Adv. Meteorol.* 2016. <https://doi.org/10.1155/2016/5694251>.
- Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth ed. Springer, New York, ISBN 0-387-95457-0 <http://www.stats.ox.ac.uk/pub/MASS4>.
- Vienneau, D., de Hoogh, K., Briggs, D., 2009. A GIS-based method for modelling air pollution exposures across Europe. *Sci. Total Environ.* 408, 255–266. <https://doi.org/10.1016/j.scitotenv.2009.09.048>.
- Xu, H., Bechle, M.J., Wang, M., Szpiro, A.A., Vedal, S., Bai, Y., Marshall, J.D., 2019. National PM_{2.5} and NO₂ exposure models for China based on land use regression, satellite measurements, and universal kriging. *Sci. Total Environ.* 655, 423–433. <https://doi.org/10.1016/j.scitotenv.2018.11.125>.
- Xu, Y., Ho, H.C., Wong, M.S., Deng, C., Shi, Y., Chan, T.C., Knudby, A., 2018. Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level PM_{2.5}. *Environ. Pollut.* 242, 1417–1426. <https://doi.org/10.1016/j.envpol.2018.08.029>.
- Zhang, Z., Wang, J., Hart, J.E., Laden, F., Zhao, C., Li, T., Zheng, P., Li, D., Ye, Z., Chen, K., 2018. National scale spatiotemporal land-use regression model for PM_{2.5}, PM₁₀ and NO₂ concentration in China. *Atmos. Environ.* 192, 48–54. <https://doi.org/10.1016/j.atmosenv.2018.08.046>.