



Article

An Ensemble Spatiotemporal Model for Predicting PM_{2.5} Concentrations

Lianfa Li ^{1,2,*}, Jiehao Zhang ^{1,2,†}, Wenyang Qiu ^{1,2}, Jinfeng Wang ^{1,2} and Ying Fang ^{1,2}

¹ State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, No A11, Datun Road, Beijing 100101, China; zhangjh@reis.ac.cn (J.Z.); qiuwy@reis.ac.cn (W.Q.); wangjf@reis.ac.cn (J.W.); fangy.16s@igsrr.ac.cn (Y.F.)

² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: lilf@reis.ac.cn; Tel.: +86-10-6488-8362

† The authors contributed equally to this work.

Academic Editors: Louise Ryan and Craig Anderson

Received: 11 March 2017; Accepted: 9 May 2017; Published: 22 May 2017

Abstract: Although fine particulate matter with a diameter of $<2.5 \mu\text{m}$ (PM_{2.5}) has a greater negative impact on human health than particulate matter with a diameter of $<10 \mu\text{m}$ (PM₁₀), measurements of PM_{2.5} have only recently been performed, and the spatial coverage of these measurements is limited. Comprehensively assessing PM_{2.5} pollution levels and the cumulative health effects is difficult because PM_{2.5} monitoring data for prior time periods and certain regions are not available. In this paper, we propose a promising approach for robustly predicting PM_{2.5} concentrations. In our approach, a generalized additive model is first used to quantify the non-linear associations between predictors and PM_{2.5}, the bagging method is used to sample the dataset and train different models to reduce the bias in prediction, and the variogram for the daily residuals of the ensemble predictions is then simulated to improve our predictions. Shandong Province, China, is the study region, and data from 96 monitoring stations were included. To train and validate the models, we used PM_{2.5} measurement data from 2014 with other predictors, including PM₁₀ data, meteorological parameters, remote sensing data, and land-use data. The validation results revealed that the R² value was improved and reached 0.89 when PM₁₀ was used as a predictor and a kriging interpolation was performed for the residuals. However, when PM₁₀ was not used as a predictor, our method still achieved a CV R² value of up to 0.86. The ensemble of spatial characteristics of relevant factors explained approximately 32% of the variance and improved the PM_{2.5} predictions. The spatiotemporal modeling approach to estimating PM_{2.5} concentrations presented in this paper has important implications for assessing PM_{2.5} exposure and its cumulative health effects.

Keywords: PM_{2.5}; PM₁₀ predictor; exposure estimation; kriging; ensemble model

1. Introduction

Studies show that air pollution has negative health impacts on humans [1], and this issue has received considerable attention in recent years [2,3]. The air pollutants that are most dangerous to humans include sulfur dioxide, nitrogen dioxide, ozone, and particulate matter with an aerodynamic diameter of less than $2.5 \mu\text{m}$ (PM_{2.5}) [4]. Because of its small size, particulate matter can adhere to the deep respiratory tract and affect blood circulation by penetrating lung cells [5,6]. Several studies have shown that particulate matter increases the risk of developing airway obstructive disease, chronic bronchitis [7], asthma (in children) [8–10], lung cancer [11], and various other cardiovascular diseases [12–14]. Thus, when assessing PM_{2.5} pollution and the cumulative health effects, an accurate method of estimating PM_{2.5} exposure at fine spatial and temporal resolutions is essential, even when

actual measurements are unavailable. However, the global monitoring of PM_{2.5} remains in a nascent phase and is characterized by limited spatial coverage, as observed in China [15,16].

Due to missing early PM_{2.5} monitoring data, and the insufficient spatiotemporal coverage of existing PM_{2.5} monitoring data, the approaches to estimating PM_{2.5} exposure have not been developed fully until recently. Early practitioners [17,18] used simple methods to predict PM_{2.5} at time slices when measurements of co-located pollutants (PM₁₀ or total suspended particulates (TSPs)) were available. These approaches involved the use of long-term ratios of PM_{2.5} to the co-located pollutants, which introduced considerable limitations because of the lack of available data for these pollutants.

Advanced approaches have recently been applied for PM_{2.5} exposure estimations, including the land-use regression (LUR) [19] and kriging methods [20]. Further, the increased availability of satellite data has led to strategies that employ normalized difference vegetation index (NDVI), surface temperature, and aerosol optical thickness (AOT) data in combination with other non-satellite variables, such as meteorological parameters, traffic indices and elevation levels [21]. Of these recent approaches, linear regressions have been applied extensively, although non-linear approaches have also been used [22]. Kriging is often used separately from LUR because of the model's complexities in combining other covariates. In addition, several spatiotemporal models have been developed for the estimation of PM_{2.5} concentrations at high spatiotemporal resolutions. Kloog et al. [23] proposed a number of spatiotemporal models using satellite-derived AOT and they obtained an out-of-sample average R² of 0.81. Xie et al. [24] and Zheng et al. [25] employed satellite-derived AOT data for China to predict the daily PM_{2.5} levels and achieved a R² of roughly 0.80. However, the two AOT-based approaches measure pollution at a coarse spatial resolution (three kilometers) and cannot reliably estimate the within-community variability of PM_{2.5} at higher spatial resolutions. For the kriging methods, measurement data are usually used to train variogram models to perform PM_{2.5} concentration predictions without considering other covariates [20] that might limit the model's predictive power.

Due to the limited temporal and spatial coverage of PM_{2.5} monitoring data, as well as the limitation of the previous estimation methods, we propose an ensemble spatiotemporal modeling approach to be an improvement in estimating PM_{2.5} concentrations. Compared with the previous methods, our approach integrates non-linear associations, ensemble learning, and residual kriging methods to predict the within-community variability of PM_{2.5} with an improved accuracy. Our approach employed generalized additive models (GAM) to consider the variability of spatial and spatiotemporal predictors with non-linear effects to capture associations between predictors and PM_{2.5} [26]. Ensemble learning can be used to generate stable predictions with less extreme values based on multiple GAM models while also outputting an uncertainty indicator (standard deviation). A kriging interpolation of the daily residuals derived from ensemble learning predictions can be used to capture residual spatial patterns and considerably improve estimations even without use of PM₁₀ or other co-pollutant predictor. We also examined the proposed approach for several scenarios of different combinations of predictors, and we demonstrated how our approach could achieve optimal accuracy for these different scenarios. Supplemental Materials Table S1 presents an annex table for technical terms used in this paper.

2. Materials and Methods

2.1. Study Domain

Shandong Province, China, was the study region, and it is located between latitude 34°25' and 38°23' N and longitude 114°35' and 112°43' E. In 2014, the province covered an area of 157,900 km² and had a population of roughly 97.89 million. The study domain is a typical northern province of China that is heavily polluted by PM_{2.5}. We have a complete monitoring data of PM_{2.5} and PM₁₀ with a lot of covariates available. Thus, we chose this region as the study domain for our approach.

2.2. Monitoring Data

PM_{2.5} and PM₁₀ concentration data (unit: $\mu\text{g}/\text{m}^3$) were obtained from the Ministry of Environmental Protection of the People's Republic of China (<http://datacenter.mep.gov.cn/>). The data were gathered at 96 national air quality monitoring stations located in 17 cities (Figure 1) from 1 January to 31 December 2014 (365 days). Because 15 days of data were missing, we imputed the omitted data using the singular value decomposition-based missing-value method [27] using the `pcaMethods` package (Max-Planck Institute for Molecular Plant Physiology, Golm, Germany), which is available through the R statistical software program (version 3.2.2). PM₁₀ and PM_{2.5} concentrations were monitored hourly, and we calculated the daily average for each site using the 75% measurement standard (measurements covering 75% of a 24-h period were used to calculate the daily means). The daily mean concentration for 2014 is shown in Figure 2. In 2014, monitoring stations in the northwest region of Shandong Province presented lower average PM_{2.5} concentrations than those of the central and southern regions (Figure 1).

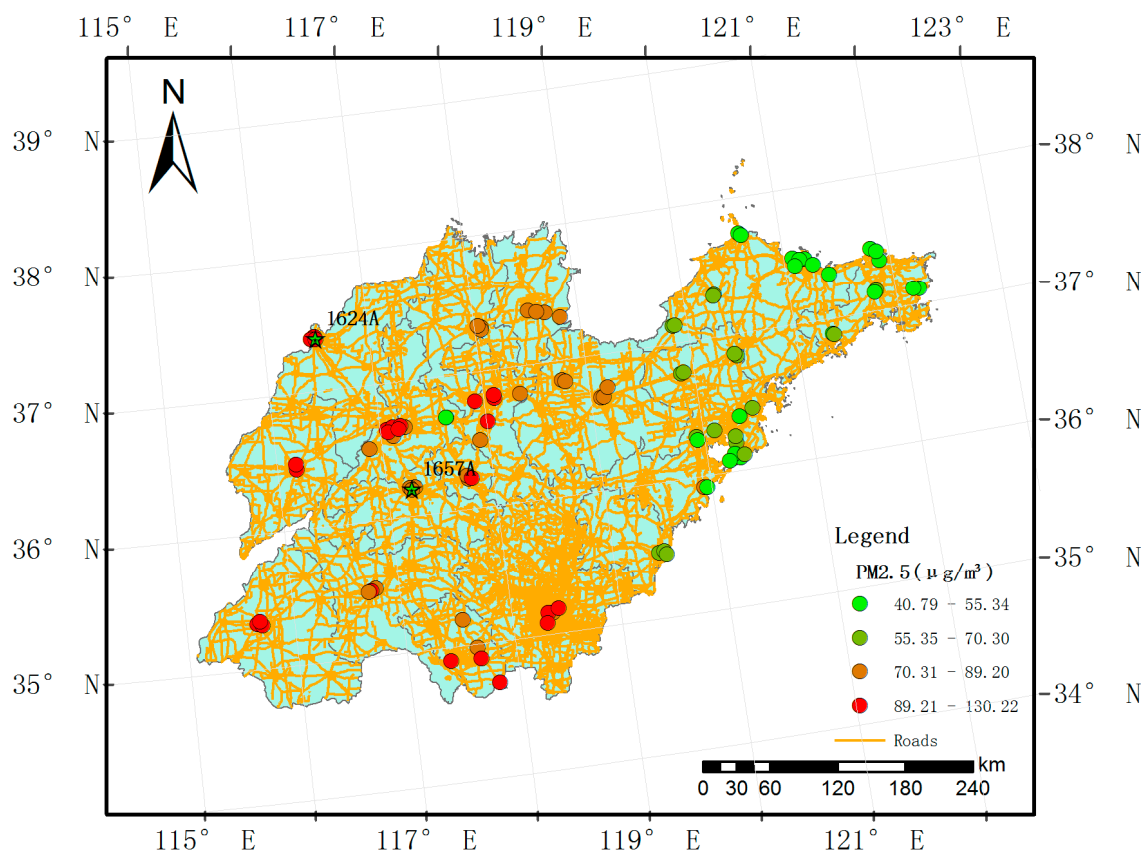


Figure 1. Study region (Shandong Province in China) with monitoring stations.

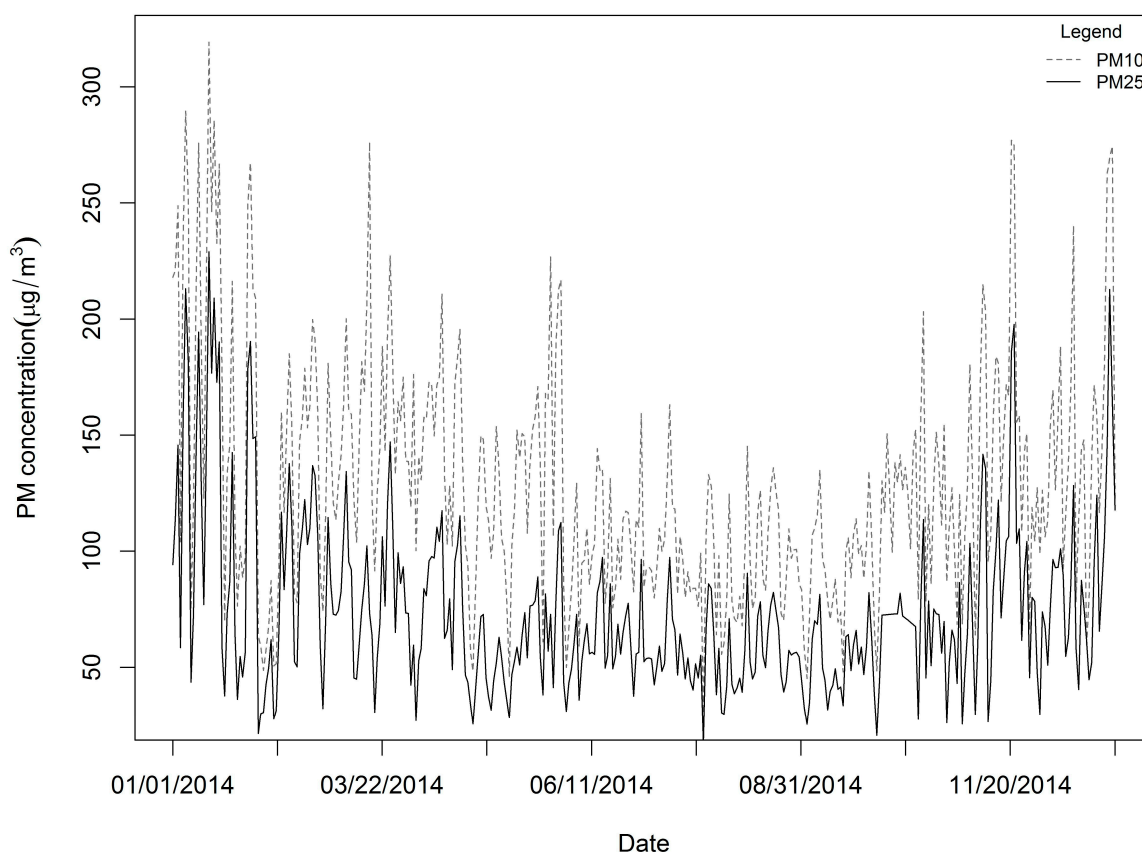


Figure 2. Daily averages of PM_{2.5} and PM₁₀ over all the monitoring stations in 2014.

2.3. Predictors

We collected predictor data from multiple sources, including meteorological parameters, remote sensing derived variables, traffic indices, point emissions, land uses, and elevation levels. Meteorological parameters such as air temperature, precipitation, humidity, and wind speed are closely associated with concentrations of particulate matter, particularly PM_{2.5} [28–32]. Data from the Modern-Era Retrospective Analysis for Research and Applications were used to extract the meteorological parameters, including the surface air temperature (°C), effective surface specific humidity (kg/kg), bias-corrected total precipitation (kg/m²s), surface eastward wind speed (m/s), and surface northward wind speed levels (m/s). The Modern-Era Retrospective Analysis dataset is a NASA reanalysis dataset for the satellite era that integrates a variety of observational systems to produce a temporally and spatially consistent synthesis [33,34]. The spatial resolution of the meteorological parameters was 0.5° latitude by 0.6° longitude, and an hourly temporal resolution was used [34]. For the meteorological predictors, we calculated the average of the 24-h measurements for each day.

Transportation emissions constitute one source of air pollution emissions [26,35–37]. We extracted two index variables as predictors of transportation: the road lengths of the 10 km buffer zones around each monitoring station and the shortest distance from each station to the nearest road. We used the Open Street Map (<http://www.openstreetmap.org/>) as the road network for deriving traffic predictors.

In addition to traffic predictors, industrial plants, such as coal, oil-burning, chemical, smelting, power, paper, and mining plants, may constitute major sources of PM_{2.5} emissions [38]. To reflect the emission sources for PM_{2.5}, we obtained location information on 326 plants in the study domain via Baidu map points of interest technologies (a major map provider in China). We measured the number of plants potentially releasing PM_{2.5} pollution within the optimal 10 km buffer of each monitoring station as well as the shortest distance between each station and its closest plant.

Land-use data for 2010 were obtained from the Chinese Academy of Sciences Data Center for Resources and Environmental Sciences [39]. The dataset consists of six first-class types (plow areas, forest land, grassland, water bodies, residential areas, and unutilized areas) and 25 second-class types with a 1 km resolution. We extracted the proportion of the forest land-use areas as an environmental factor and the proportion of construction land-use areas (referring to land used for factories and mines, oil fields, and stone pits) as an indirect emission factor within the optimal 10 km buffer (with the highest correlation to $PM_{2.5}$) of the monitoring stations.

Aerosol optical thickness (AOT) data were extracted from satellite images [40]. More specifically, AOT data (spatial resolution: 1 km; temporal resolution: daily) derived from the Moderate Resolution Imaging Spectroradiometer MODIS Terra/Aqua satellites were used, and we observed a strong correlation between the satellite AOT data and $PM_{2.5}$ levels in certain regions of China [41,42]. In addition, we extracted the monthly NDVI (normalized difference vegetation index) from the MODIS MODND1D dataset (spatial resolution: 500 m; temporal resolution: monthly) for each sampling location. The dataset was provided by the International Scientific and Technical Data Mirror Site Computer Network Information Center of the Chinese Academy of Sciences [43].

In addition to the predictors described above, we considered GDP and digital elevation model (DEM) values. GDP denotes the economic level of the study domain, and DEM values reflect the terrain characteristics that may influence pollution diffusion [44]. However, the data were not significantly correlated with $PM_{2.5}$; therefore, the two variables were not considered in the models.

Variables were selected based on the following two steps. First, we calculated the variance inflation factor (VIF) to avoid multicollinearity [45]. Second, we used the backward-stepwise method to select variables for each combination by removing a variable once until the Akaike's information criterion (AIC) [46] remained the same. Finally, we selected variable combinations with the lowest AIC values as optimal inputs.

2.4. Modeling Approach

This section describes our modeling process from the generation of the predictors of temporal basis trends (Section 2.4.1), non-linear additive modeling (Section 2.4.2), ensembling learning (Section 2.4.3), kriging of the residuals (Section 2.4.4), and validation and independent test (Section 2.4.5).

2.4.1. Generation of the Predictors of Temporal Trends

Besides the meteorological, remote sensing, traffic, point emission, land-use and elevation predictors listed in Section 2.3, we also extracted the predictors of temporal trends from the temporally continuous monitoring $PM_{2.5}$ data. Air pollutant levels, including $PM_{2.5}$, follow a regular temporal trend of high values in winter and low values in summer [47]. Thus, we can use the temporal variables extracted to model seasonal trends for the study region. Of the two methods, empirical mode decomposition [48] and singular value decomposition (SVD) [49,50], we used the SVD method to obtain the temporal basis functions and capture dominant seasonal trends because of its better generalizability and missing data processing capabilities. Cubic smoothing splines were used to fit the temporal basis function, and the first and second temporal basis functions for the Shandong province of China are shown in Figure 3. The first and second temporal basis functions were used as predictors as they account for a majority of the seasonal variability of $PM_{2.5}$ for the study region [51].

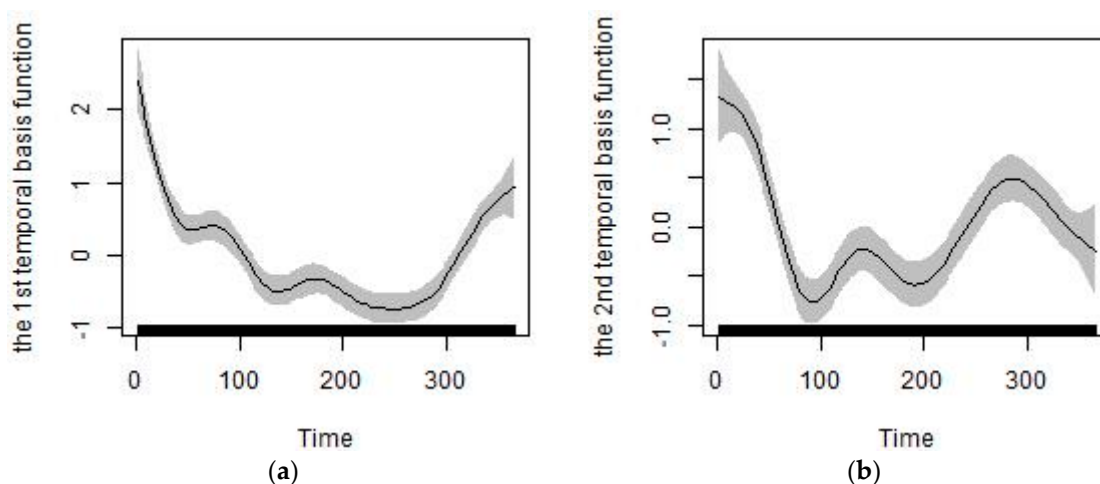


Figure 3. Temporal basis functions for Shandong Province of China: (a) The first temporal basis function; (b) The second temporal basis function.

2.4.2. Non-Linear Additive Modeling

The following equation illustrates our non-linear modeling framework:

$$y(s, t) = \mu(s, t) + \varepsilon(s, t), \quad y(s, t) \sim N(\mu, \sigma) \quad (1)$$

$$\mu(s, t) = f_1(t) + f_2(t) + \sum_{i=1}^n s(x_i(s, t)) + \sum_{k=1}^m s(p_k(s)) \quad (2)$$

In Equation (1), s and t represent the location (latitude and longitude) and time (day index), respectively, of the monitoring sample; $y(s, t)$ represents the log-transformed concentration of measured or estimated $\text{PM}_{2.5}$ ($\mu\text{g}/\text{m}^3$), because the histograms of the $\text{PM}_{2.5}$ and PM_{10} levels (Supplemental Materials Figure S1) showed a right-skewed distribution; thus, log transformation was necessary for normalization; and $\mu(s, t)$ represents the mean for $y(s, t)$ with spatiotemporal residual $\varepsilon(s, t) \sim N(0, \sigma^2)$. In Equation (2), the mean of $\mu(s, t)$ are modeled based on temporal variables (the first and second temporal basis functions $f_1(t)$ and $f_2(t)$) and spatiotemporal variables ($x_i(s, t)$ including co-located PM_{10} , meteorological parameters, AOT, and NDVI) and spatial variables ($p_k(s)$, including traffic indices and land-use predictors). We used generalized additive models (GAMs) to perform the non-parametric non-linear associations ($s(\dots)$ refers to smooth functions), and we set the maximum degree of freedom to 10 to prevent overfitting in the non-linear model [49].

The “mgcv” package available through the R statistical software program (version 3.2.2) was used to model the non-linear additive modeling in Equation (2).

2.4.3. Ensemble Learning

Section 2.4.1 presents the individual GAM models, which may be sensitive to new datasets when performing practical predictions. The ensemble method of machine learning [52] was used to enhance the predictions of the multiple GAM models included in Equations (1) and (2) and to avoid the potential over-fitting problem; meanwhile, standard deviation can be obtained from multiple predictions as an uncertainty indicator. Because of the strong performance of bootstrap aggregating (bagging) methods [53–57], we used bagging to conduct integrated learning for the individual models. In this method, input data are first sampled with replacements to generate multiple sets of input data of the same size as the original input data. Using different sample subsets, various regression models were trained. Cross-validations were used to determine the optimal number of models, and we ultimately obtained 1000 models to capture the dataset’s different variability levels based on the varied sample subsets. Each trained model was then used to make predictions for the remaining sample to train the model [57]. Because the original dataset included 35,040 data points, we obtained a re-sampled training

dataset (63.2% of the original dataset) and remaining test data (36.8%) for independent validation [58]. The final integrated prediction was the mean weighted by the model’s performance (measured by R²):

$$\mu_f(s, t) = \sum_i \mu_i(s, t) w_i \tag{3}$$

where $\mu_f(s, t)$ is determined by individual predictions, $\mu_i(s, t)$, from the GAMs generated by bootstrap sampling and their weights, w_i that is proportion to the square of R²:

$$w_i = \frac{R_i^2}{\sum_i R_i^2} \tag{4}$$

Further, the predictions outputted the uncertainty of predictions by calculating the standard deviation of all the predictions from all the models:

$$\sigma_f(s, t) = \sqrt{\frac{\sum_i w_i (\mu_i(s, t) - \mu_f(s, t))^2}{\frac{M-1}{M} \sum_i w_i}} \tag{5}$$

where $\sigma_f(s, t)$ is the standard deviation from the output of multiple models, and M is the number of nonzero weights.

2.4.4. Kriging of Spatiotemporal Residuals

The daily residuals were modeled for their spatiotemporal patterns. A reasonable assumption is the stable spatial domain after removal of local means [59]. Based on this assumption, the residuals after removal of the means present spatial pattern (autocorrelation) and the variogram could be used to capture such autocorrelation. Consequently, based on the variogram fit, we can infer the residuals for any location under the study domain. Kriging is a classical approach and the readers can refer to [60] for details. Here, based on the stable domain theory after removal of local means [49], kriging was used to estimate the residuals:

$$\varepsilon(s_0, t) = \sum_{i=1}^N \lambda_i \varepsilon(s_i, t) \tag{6}$$

where $\varepsilon(s_0, t)$ is the residual for a target location to be estimated, s_0 with time point, t , $\varepsilon(s_i, t)$ is the residual from the estimated mean of the training sample at the location s_i and the same time point t , from the ensemble models. λ_i is the weight for $\varepsilon(s_i, t)$ determined by the fitted variogram.

Due to the complexity and instability of directly fitting the spatiotemporal variogram, we simplified variogram fitting by independently fitting daily residuals. Thereby, we can avoid the complex assumption of temporal stability besides spatial stability for directly fitting and improved the effectiveness of parameter estimates for the variogram.

By leave-one-out-cross-validation, we selected the exponential model from multiple models (spherical, circular, exponential, Gaussian, linear models) [61] to simulate the variogram with their optimal performance.

$$\gamma(d, t) = \begin{cases} c_0(t) + c(t) \left(1 - \exp\left(-\frac{d}{a_0(t)}\right) \right) & , h > 0 \\ 0, & h = 0 \end{cases} \tag{7}$$

where $\gamma(d, t)$ is the variogram function for the residuals, $\varepsilon(s, t)$ to be fit, d is spatial distance, t is the temporal index (day index), $c_0(t)$ is nugget, $c(t)$ is the sill and a_0 is the parameter to be estimated.

For variogram fitting, three important parameters—range, nugget and sill—were estimated. Nugget reflects the variance at the discontinuity at the origin, sill is the limit of the variogram tending to infinity lag distances, and range is the distance in which the difference of the variogram from the

sill becomes negligible [58]. In our model, the variogram was used to derive the covariance between different locations at the same time that was used in the weighted regression in Equation (6).

Using the daily residual variogram models, we simulated temporal changes in the parameters (range, partial sill, and nugget) for the residual variogram models for 2014. For the predictions, we created grid interpolations for daily residuals for a year across the study region and then added the corresponding estimate of the residual to the bagging prediction estimate to arrive at our final predictions. Therefore, we improved the final predictions by incorporating non-linear associations and residual spatial autocorrelations. The variogram and prediction residuals were completed in ArcGIS (version 10.3, ESRI, Sacramento, CA, USA) using the geostatistical analyst module [62], and the nugget and sill variogram parameters were optimized through cross validations with a focus on the estimation of range parameters.

2.4.5. Validation and Independent Test

To examine the predictive performance of our models under different scenarios, we designed six models with different predictor combinations: Model 1, GAM with no use of the PM₁₀ predictor and residual kriging; Model 2, GAM with the PM₁₀ predictor but without residual kriging; Model 3, bagging with no use of the PM₁₀ predictor and residual kriging; Model 4, bagging with no use of the PM₁₀ predictor but with residual kriging; Model 5, bagging with the PM₁₀ predictor but without residual kriging; and Model 6, bagging with the PM₁₀ predictor and residual kriging. In the six models, the PM₁₀ and PM_{2.5} variables were log transformed and transferred back by the exponential function. Bagging predictions for the training dataset were similar to the 63.2–36.8% cross validation. We also conducted a 10 × 10-fold cross-validation for Models 1 and 2. The R² value, root-mean-square error (RMSE), and residual plots between the predicted and observed values were used to compare the performances of the different models.

Independent test was also conducted using Model 4 trained by the 2014 data to predict the 2016 monthly mean PM_{2.5} levels of the 30 sub-regions (Figure S2 of Supplemental Materials) of Shandong Province. Due to the missing PM_{2.5} monitoring station based data for 2016, we just compared the predicted monthly sub-regional mean level of PM_{2.5} with the measured PM_{2.5} level derived from the Ministry of Environmental Protection of People's Republic of China to illustrate our method's applicability. Monthly averages of PM_{2.5} were summarized over daily predicted or measured values. Block kriging was used to derive the regional means based on the point estimated or measured values.

3. Results

3.1. Data Summary

We obtained 35,040 measurements for all 96 sample locations, with a rate of missing daily values of 4%, and the SVD method was used to complete the training sample. The daily mean PM_{2.5} and PM₁₀ values in the study domain are shown in Figure 2. For all measurements, the mean PM_{2.5} concentration was 73.86 µg/m³, while the mean PM₁₀ was 128.80 µg/m³. The summer and autumn concentrations were obviously lower than the spring and winter concentrations (Figure 2). The PM_{2.5} across the 96 stations varied considerably from approximately 1 µg/m³ to nearly 600 µg/m³ (standard deviation: 50.79 µg/m³), whereas the PM₁₀ levels across the 96 stations also varied greatly from roughly 3 µg/m³ to nearly 690 µg/m³ (standard deviation: 77.37 µg/m³). The highest PM_{2.5} mean concentration measured for the period was 130.22 µg/m³, and a relatively high PM₁₀ value (158.55 µg/m³) was measured at the 1624A station located in the northwest part of the province (Figure 1). This figure is much higher than the other observed levels. The highest daily PM_{2.5} concentration for the period was 592.33 µg/m³ (PM₁₀: 658.92 µg/m³), and it was recorded at the 1657A station located in the central part of the province on January 19, 2014. We found 50 stations in the study domain with PM_{2.5} annual mean concentrations exceeding 75 µg/m³. We also examined the means of the PM_{2.5} to PM₁₀ ratios across daily timelines for 2014 (Figure 4), and the ratios appear to follow a dominant trend with minor

variations (mean: $0.57 \mu\text{g}/\text{m}^3$ with a standard deviation of $0.10 \mu\text{g}/\text{m}^3$, with slightly lower levels in summer than winter).

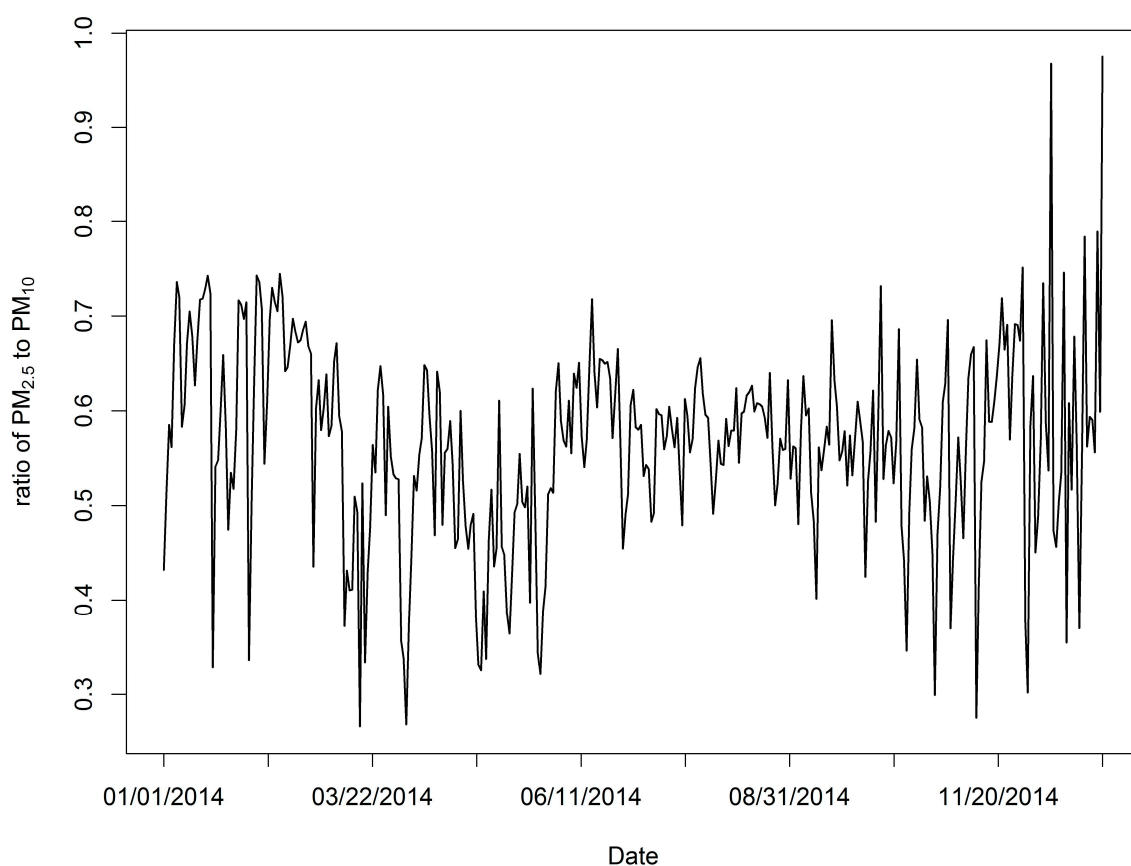


Figure 4. Daily averages of the ratio of $\text{PM}_{2.5}$ to PM_{10} over all the monitoring stations across 2014.

3.2. Predictors

Using univariate and multivariate GAMs (generalized additive models), we explored the non-linear associations of each predictor with $\text{PM}_{2.5}$ as well as their respective contributions to the variance. Table 1 presents the variance explained by each predictor of the univariate and multivariate GAM models with and without the inclusion of PM_{10} . The results showed that the highest contributions were observed with the co-located PM_{10} (accounting for 67.97–73% of the variance) and the second highest contributions were observed with the first temporal basis function (accounting for 4.84–37% of the variance) in the univariate and multivariate models. For the univariate model, the day of the year and humidity levels accounted for the third and fourth highest contributions to $\text{PM}_{2.5}$ variability, respectively. In addition to the co-located PM_{10} and the first temporal basis function, the other predictive variables varied in their contributions in terms of the variance between the univariate model and the multivariate model. For example, although the day of the year explained 14.7% of the variance in the univariate model, it only accounted for a small portion (0.73–2.38%) of the variance in the multivariate model. Similarly, the contributions of AOT (aerosol optical thickness), mean specific humidity and the number of emission plants were also lower for the multivariate models than the univariate models (AOT: 7.38% vs. 0.48–4.77%; humidity: 9.08% vs. 0.48%; number of emission plants: 4.48% vs. 0.97–1.67%). Finally, the traffic index (including roadway lengths within a 10 km buffer of a monitoring location and distances to a primary roadway) made various contributions to explaining the variance (2.62% vs. 1.45–2.86%; 1.73% vs. 1.21–2.15%, respectively).

Table 1. Variance explained for each predictive variable.

Predictive Variable (Unit)	Variance Explained in the Univariate Model	Variance Explained in the Multivariate Model (without PM ₁₀)	Variance Explained in the Multivariate Model (Including PM ₁₀)
PM ₁₀ (µg/m ³)	73.00%	-	67.97%
Aerosol optical thickness (AOT)	7.38%	4.77%	0.48%
Normalized difference vegetation index (NDVI)	3.14%	0.24%	0.24%
Precipitation (kg/m ² s)	1.75%	0.02%	0.02%
Temperature (°C)	1.08%	2.62%	0.48%
Mean specific humidity (kg/kg)	9.08%	0.48%	0.48%
Roadway length within the 10 km buffer of a monitoring station (m)	2.62%	2.86%	1.45%
Shortest distance of roadway to a monitoring station (m)	1.73%	2.15%	1.21%
Wind speed vector	3.76%	1.43%	0.48%
Area proportion of the factories and mines, oil fields and stone-pit land-use	2.29%	2.15%	0.73%
Area proportion of the forest land-use	2.06%	2.62%	0.48%
Number of the emission plants	4.48%	1.67%	0.97%
Shortest distance to the emission plants	1.70%	1.67%	0.24%
The first temporal basis function	37.00%	26.71%	4.84%
The second temporal basis function	5.79%	1.43%	0.48%
Time (day of year)	14.70%	2.38%	0.73%
Total		53.20%	81.30%

Figure 5 and Figure S3 of the Supplemental Materials section present the non-linear associations between each predictor and PM_{2.5}. The variance explained by the non-linear models was up to 10% than that of the linear models. As shown in Figure 5, the GAMs showed a good ability to capture the non-linear associations between the predictive variables and PM_{2.5}. The varying associations (e.g., PM₁₀ non-linear associations, including critical varying points, such as 148 µg/m³ with PM_{2.5} (also known as 5 log PM₁₀) in Figure 5a) captured by the GAMs were critical because they better captured the varying trends between influential factors and concentrations than the linear model. For example, the higher concentrations measured in winter compared with those in summer for the day of the year parameter (Figure 5g) could not be quantified in the linear models. Similarly, varying non-linear associations were also observed for the AOT, the number of emission plants, precipitation, and air temperature.

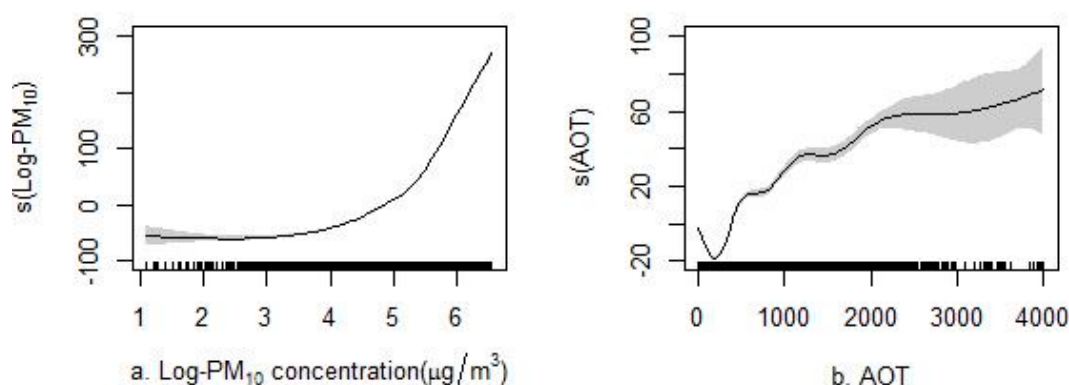


Figure 5. Cont.

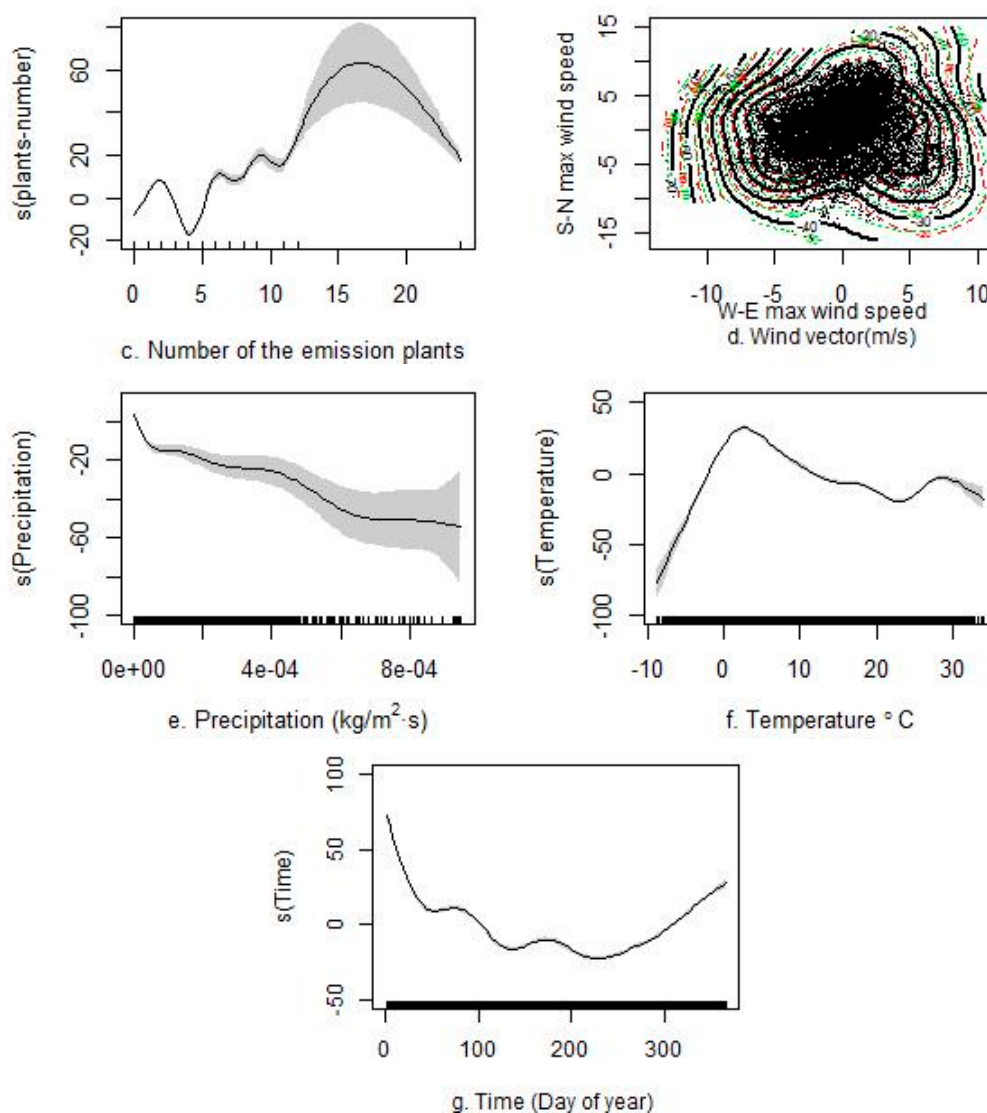


Figure 5. The non-linear associations between each predictor and PM_{2.5}. The non-linear association between (a) Log-PM₁₀ and PM_{2.5}; (b) aerosol optical thickness (AOT) and PM_{2.5}; (c) number of the emission plants and PM_{2.5}; (d) wind and PM_{2.5}; (e) precipitation and PM_{2.5}; (f) temperature and PM_{2.5}; (g) day of year and PM_{2.5}.

3.3. Comparison of the Different Models and Independent Test

Table 2 shows the results of the six models. Model 1 did not use PM₁₀ as a predictor or perform a kriging interpolation of the residuals, and it achieved the worst performance, generating a CV R² of 0.53 and the highest RMSE of 34.69 µg/m³. The bagging method (Model 3) results were also similar to that of the ten-fold cross-validation and achieved a similar predictive performance as Model 1; in addition, this method was able to output the standard deviation as a measure of uncertainty. Although the co-located PM₁₀ significantly contributed to explaining the variance (28–36%), the kriging interpolation of residuals was able to explain a large proportion (33%) of the variance (Model 4 vs. Model 3) when PM₁₀ was not used as a predictor. When PM₁₀ was used as a predictor, the kriging interpolation of the daily residuals improved the model's ability to explain the variance by 7% (Model 6 vs. Model 5). When the PM₁₀ predictor was not used, the final model (Model 4) achieved a CV R² of 0.86, and when the PM₁₀ predictor was used, the final model (Model 6) achieved a CV R² of 0.89. The plots of predicted vs. measured values and their residual (Models 4 and 6) are shown in Figures S4 and S5 of Supplemental Materials.

Table 2. Comparison of multiple models.

Model	Use of Predictive Variables and Residual Kriging	R ²	CV ^a R ²	CV RMSE ^b (µg/m ³)
Model 1	GAM with no use of PM ₁₀ data and residual kriging	0.53	0.53	34.69
Model 2	GAM with PM ₁₀ data but without residual kriging	0.81	0.81	21.87
Model 3	Bagging without PM ₁₀ data and residual kriging		0.53	34.79
Model 4	Bagging without PM ₁₀ data but with residual kriging		0.86	18.85
Model 5	Bagging with PM ₁₀ data but without residual kriging		0.82	21.82
Model 6	Bagging with PM ₁₀ data and residual kriging		0.89	17.06

^a CV: Cross Validation; ^b RMSE: Root Mean Squared Error.

Our independent test shows the R² of 0.73 for sub-regional monthly PM_{2.5} averages. Figures S6 of Supplemental Materials shows the scatter plot of predicted vs. measured values and their residual plot.

3.4. Variogram Modeling of the Residuals

As shown in Table 2, the daily residual kriging interpolation considerably improved the predictive performance and accounted for 33% of the explained variance in the model that did not use the PM₁₀ predictor. An exponential model was used to fit the variogram of the residuals. Three parameters for the variogram model (range, partial sill, and nugget) were measured over all 365 days of 2014 to explore the temporal patterns of daily residuals. Table 3 shows the variogram parameters for Models 4 and 6, and Figure 6 presents the temporal trends with the fitted curves of the three parameters for 2014. The results show lower range, partial sill, and nugget values in summer than in winter, and this trend may have been caused by the higher concentrations of PM_{2.5}, larger influential range, and stronger spatial auto-correlations in winter than summer.

Table 3. Summary of variogram parameters for Models 4 and 6.

Model	Parameter	Minimum	1st Qu. ^a	Median	Mean	3rd Qu. ^a	Maximum
Model 4	Range	5551	63,780	93,050	107,100	137,000	712,900
	Partial sill	1.65	96.18	208	448.7	507	6560
	Nugget	20.74	97.12	159.6	260.4	284.3	3096
Model 6	Range	4250	60,350	95,660	103,100	144,500	475,200
	Partial sill	0.0839	29.7	71.27	144.4	162.3	2866
	Nugget	0.0912	85.33	146.6	228.1	264.5	2650

^a Qu.: Quarter.

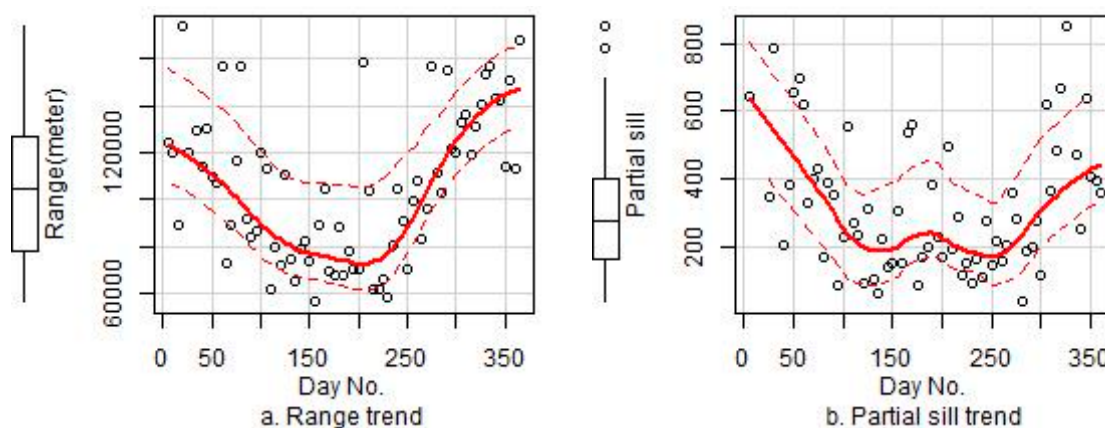


Figure 6. Cont.

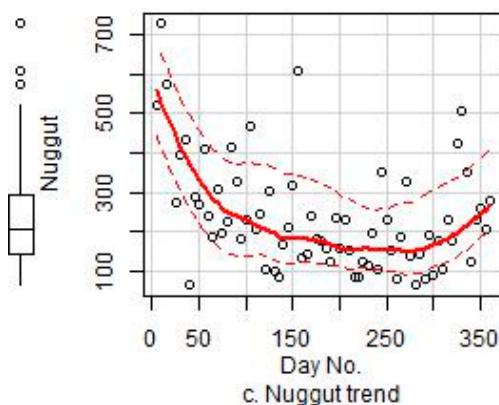


Figure 6. The temporal trends of the three variogram parameters for the residuals of Model 4: (a) the temporal trend of range; (b) the temporal trend of partial sill; (c) the temporal trend of nugget.

After quantifying the variogram from the exponential model, ordinary kriging was used to interpolate the residuals. The final PM_{2.5} prediction included the output of the bagging approach and the estimate of the specific-time residual. To extend our models to other years, we assumed limited changes in the variograms for PM_{2.5} across different years and overlaid the day-of-year interpolated surfaces of the residual (obtained by the kriging method in ArcGIS) with the target location within the study region to extract the residual estimate for a certain day of the target year.

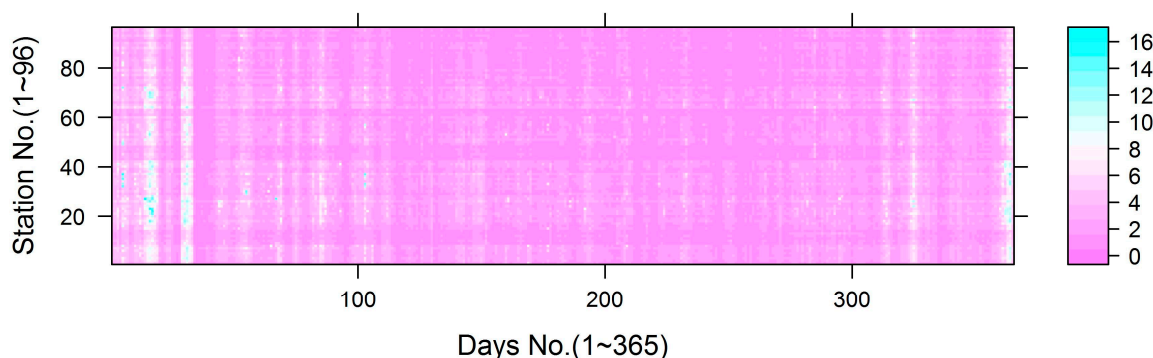
3.5. Uncertainty

As illustrated, the bagging approach integrated the predictions from 1000 models to generate a stable estimate with standard deviations as a measure of uncertainty. Using the estimated standard deviation, we could also generate 95% confident intervals for the estimate [63,64]. The standard deviation is summarized in Table 4, and the standard deviation distribution is shown in Figure 7. The results show small uncertainty (only 5.2% of the standard deviations exceeded 5 µg/m³ for Models 3 and 4, and only 1.9% of the standard deviation exceeded 5 µg/m³ for Models 5 and 6).

Table 4. Summary of standard deviation.

Model	Minimum	1st Qu. ^a	Median	Mean	3rd Qu. ^a	Max.
Models 3 and 4	0.43	1.33	1.80	2.18	2.54	21.43
Models 5 and 6	0.26	0.79	1.19	1.47	1.76	26.81

^a Qu.: Quarter.



(a)

Figure 7. Cont.

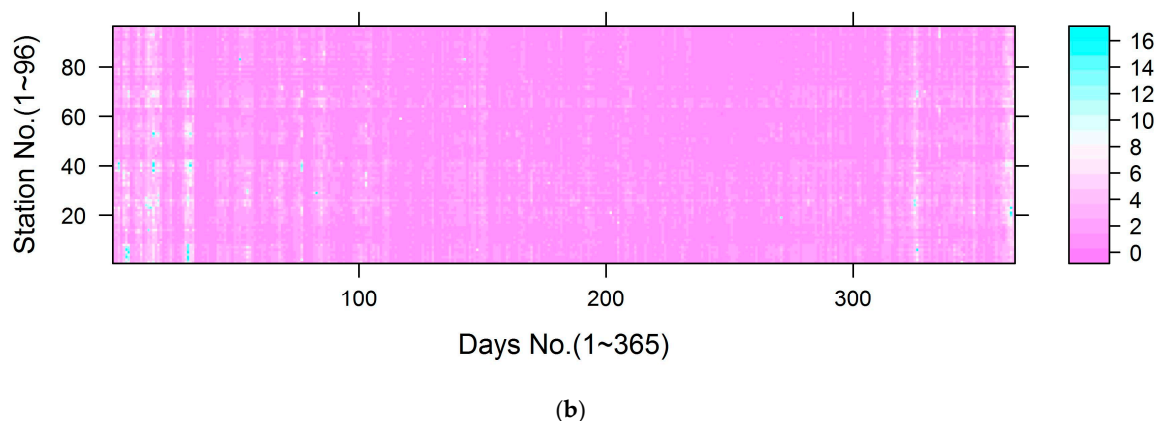


Figure 7. The distribution of standard deviation across monitoring stations: (a) the distribution of standard deviation for Models 3 and 4; (b) the distribution of standard deviation for Models 5 and 6.

4. Discussion

In this paper, we proposed an ensemble spatiotemporal modeling approach for robustly predicting $PM_{2.5}$ concentrations. For the individual model, we used a GAM (generalized additive models) to determine the non-linear association between $PM_{2.5}$ and multiple predictors (meteorological patterns, traffic indices, season variations, the number of emission sources, and land-use patterns). In particular, we extracted temporal basis functions from the monitoring stations to represent the seasonal $PM_{2.5}$ variability for the study region, which accounted for a large portion of the explained variance. We then used the bagging method to sample the dataset to train 1000 individual models and derived stable ensemble predictions with standard deviations as an uncertainty measure. Then, the kriging method was used to model the residuals from the ensemble predictions to estimate the daily $PM_{2.5}$ residuals throughout a year (365 days) for the study region. For the model that did not include PM_{10} as a predictor, the daily residual kriging achieved a similar predictive performance (R^2 : 0.86 vs. 0.89) as the model using PM_{10} as a predictor. Strong spatial autocorrelations of the residuals accounted for a considerable portion (33%) of the explained variance when co-located PM_{10} values were not included. These results denote the usefulness of residual spatial autocorrelations for predicting $PM_{2.5}$ when PM_{10} measurements are missing as observed for our study location of Shandong Province, China. To our knowledge, previous studies that have performed the same predictions [22,23,65] have only reported R^2 values of 0.54–0.81, and few studies have achieved a similar estimation accuracy for $PM_{2.5}$ without using PM_{10} as a predictor.

This study also illustrates the important contributions of non-linear associations in the models [19,20,22,23]. The final results show a considerable predictive improvement through the use of non-linear additive models. The results reveal a notable positive non-linear association between $PM_{2.5}$ and PM_{10} , AOT (aerosol optical thickness), the number of emission plants, and air temperature, with varying fluctuations found for different predictor intervals (Figure 5a–c,f), whereas a negative association was observed between $PM_{2.5}$ and precipitation (Figure 5e). For the wind vector terms, the $PM_{2.5}$ concentrations tended to decline as the wind speed increased in the south-north and east-west directions because of complex interactions (Figure 5d). Such associations are generally consistent with previous conclusions [29,49].

As a regional pollutant, $PM_{2.5}$ is affected by various factors, including meteorological parameters, emissions sources, and traffic indices [18,19,22,66]. PM_{10} consists of $PM_{2.5}$ and other components, and it is strongly correlated with $PM_{2.5}$; therefore, co-located PM_{10} is a primary predictor of $PM_{2.5}$ levels and explained most of the variance in the multivariate models. Following PM_{10} , the first temporal basis function was the second most important predictor. The temporal basis functions captured the seasonal variability of $PM_{2.5}$ levels for the study region. For the model with PM_{10} used as a predictor, the first temporal basis function accounted for roughly 5% of the total variance. However, in the

multivariate model (Model 3) without PM₁₀, meteorological parameters (including precipitation, wind speed, temperature, and humidity) accounted for only 4.55% of the variance. In addition, traffic indices accounted for 5.01% of the variance; emission plants accounted for roughly 3.34% of the variance; and AOT and NDVI (normalized difference vegetation index) accounted for 4.77% and 0.24% of the variance, respectively. Previous studies of certain regions show weak correlations between AOT and surface PM_{2.5} because the surface reflectance ratio between visible and shortwave infrared channels of the AOT product was underestimated [67]. Our study also illustrates the limited contributions of AOT and NDVI as predictors.

For the models using PM₁₀ as a predictor (Models 5 and 6), co-located PM₁₀ accounted for most (67.97%) of the variance, whereas the other predictors together accounted for only 13.33%. This finding illustrates that PM₁₀ captured a major part of the spatiotemporal variability in PM_{2.5}. Without the PM₁₀ predictor, the other variables accounted for roughly 53.2% of the variance. In addition to the first temporal basis function (accounting for 26.71%), the other predictors together accounted for only 26.49% of the variance.

Unfortunately, historical PM_{2.5} and PM₁₀ measurements are not available for China and many other countries across the globe [16,68]. Even today, the spatial coverage of the PM_{2.5} and PM₁₀ monitoring network are limited in China. To comprehensively monitor PM_{2.5} pollution levels and assess their cumulative health effects on humans, reliable estimates of PM_{2.5} concentrations at fine spatiotemporal resolutions should be performed using the limited available PM_{2.5} monitoring data for time periods without available data over extensive spatial areas. Unfortunately, accurately predicting PM_{2.5} levels at high spatiotemporal resolutions is difficult without relevant data on co-located pollutants. In this paper, we explored the use of kriging interpolations of daily residuals, and the results show considerable improvements in predictive performance. Whereas the GAM already took both spatial and temporal information by the covariates, such spatial and spatiotemporal covariates couldn't fully capture the spatiotemporal variability of PM_{2.5} and so the performance is not so good without use of spatiotemporal residuals or PM₁₀. The daily residual's kriging captured an important portion of PM_{2.5} spatiotemporal variability not captured by GAMs. Although a previous study also employed residual kriging interpolations [69], the variogram modeling of daily residuals remains poorly researched. Therefore, we analyzed the variogram patterns of residuals for a one-year period and performed a cross validation to illustrate the generalizability of the proposed method. As a regional pollutant, PM_{2.5} exhibited stronger effects and spatial autocorrelations than nitrogen oxide. Spatial autocorrelations of the residuals were better able to explain the spatial variability of PM_{2.5} than nitrogen oxide pollutants [49]. Because of the considerable effects of PM_{2.5}, particularly for winter in Shandong Province, the residual kriging method was applicable despite the limited number of PM_{2.5} monitoring stations examined and large spatial distances between the monitored samples. As shown in the results (Figure 6), PM_{2.5} in winter showed effects over a longer time period and a greater spatial area than those observed in summer. We expected PM_{2.5} to have an effect over a longer time period (a longer range) and to present higher concentrations (also resulting in higher partial sill and nugget values) in winter than summer. Our variogram modeling of the daily residuals captured these spatial autocorrelation trends to compensate for the gap in the variance explained due to the missing PM₁₀ predictor, and thus improved the accuracy of our final predictions. In practice, co-located PM₁₀ and other pollutant measurements are not usually available. Thus, Model 4 (using residual kriging interpolations but not PM₁₀ as a predictor) has the potential for use in a greater number of applications than the other models that used PM₁₀ as a predictor. Our cross-validation results show that Model 4 achieved high levels of accuracy and the results were slightly better than those of Model 5 (using PM₁₀ as a predictor but not the residual kriging interpolation) (R^2 : 0.86 vs. 0.82), although the results were slightly less accurate than those of Model 6 (using PM₁₀ and residual kriging) (R^2 : 0.86 vs. 0.89). The R^2 value of Model 6 was only 7% higher than that of Model 5, which may have been related to the spatiotemporal PM₁₀ values, which accounted for a major portion of the variance explained in

Model 6; whereas the remaining 7% that was not captured by the predictors was explained by the spatial autocorrelations of the residuals.

Based on individual GAMs, our approach introduced the bagging technique to obtain the stable prediction with standard deviation as an uncertainty indicator. Thereby, our prediction could be robust and less over-fitting although our performance is pretty similar to the individual model's output from the other approaches [13,24,25,65,70]. On the other hand, due to characteristics of strong spatial autocorrelation of PM_{2.5} concentration, our approach could considerably improve performance even without use of the PM₁₀ predictor. Thus, for practical prediction, our approach does not need the extraction of complicated covariates such as output of the chemical transport model (GEOS-CHEM) and is less costly with a similar performance at high spatiotemporal resolution. Production of PM_{2.5} is complicated and varies with different regions. While some approaches may have good performances in other regions [13,65,70], they may be not applied to China as our approach, due to unavailability of some predictive covariates or differences in geographical and meteorological factors and emission sources.

This study presents several limitations. First, we extracted the first and second temporal basis functions from regular monitoring data to represent the study region's seasonal variability in PM_{2.5}. This procedure could have resulted in overfitting. To determine the influence on the final results, we conducted a 10 × 10 cross-validation and independent test, and the results (CV R²: 0.86–0.89; 0.73 for independent test) show high levels of predictive accuracy, thereby illustrating the limited influence of the temporal basis functions extracted from the predictions. Second, our variogram modeling of the daily residuals was not systematic, which may have affected the applicability of the proposed spatiotemporal approach. Our validation and independent results show that such effects are limited for extensive applications. Third, our spatiotemporal modeling approach was based on monitoring data for Shandong Province, China; thus, its applicability may be limited. PM_{2.5} pollution has a larger sphere of influence in northern regions of China, including Shandong Province, and the spatial autocorrelation of residuals of PM_{2.5} concentrations was markedly strong. Thus, the use of the residual kriging method for less dense monitoring networks could still capture the spatiotemporal variability of PM_{2.5} and account for a large portion of the variance. For regions with lower PM_{2.5} pollution and sparser monitoring networks, the residual kriging method may not be applicable, and spatial effects modeling may represent a more preferable approach. We have explored and reported on such a method in another study [71].

5. Conclusions

This paper has proposed an ensemble spatiotemporal modeling approach for improved prediction of PM_{2.5} even with missing co-located pollutants such as PM₁₀. Our approach uses a generalized additive model to quantify the non-linear associations between the predictors and PM_{2.5}, and ensemble learning was used to obtain stable predictions, with standard deviations used as an uncertainty measure. In addition, the variogram of the daily residuals was modeled to capture day-of-year patterns for the residuals. The results showed the better CV R² of 0.89 (with co-located PM₁₀ used as a predictor) and 0.86 (without PM₁₀ used as a predictor) than the previous methods. The results demonstrate that our approach can be used to estimate PM_{2.5} exposure at a good accuracy and indicate its potential applicability for estimating PM_{2.5} exposure in Shandong Province or other similar regions of China.

Supplementary Materials: The following are available online at www.mdpi.com/1660-4601/14/5/549/s1, Figure S1: Histogram with normal density plot of PM_{2.5} and PM₁₀, Figure S2: Thirty sub-regions in Shandong province, Figure S3: Associations of predictive covariates and PM_{2.5}, Figure S4: Residual plots for the measured and fitted PM_{2.5} in Model 4 (no PM₁₀ used but including estimates of the residuals), Figure S5: Residual plots for the measured and predicted PM_{2.5} in Model 6 (PM₁₀ used and incorporation of estimates of the residuals), Figure S6: Residual plots for the measured and predicted monthly-mean PM_{2.5} in regions, Table S1: An annex table for technical terms.

Acknowledgments: This study was supported through the Natural Science Foundation of China (41471376 and 41171344) and National S&T Planed Projects of the Ministry of Science and Technology of China (2014FY121100).

Author Contributions: Lianfa Li conceived and designed the experiments, provided advice relating to analysis of the data and revised the manuscript; Jiehao Zhang collected the dataset, performed the experiments, analyzed the results and drafted the manuscript. The authors declare no competing interests; Wenyang Qiu provided the relate data; Jinfeng Wang provided advice relating to analysis of the data and methods; Ying Fang provided some useful suggestions about the manuscript structure.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bell, M.L. Assessment of the health impacts of particulate matter characteristics. *Res. Rep.* **2012**, *161*, 5–38.
2. World Health Organization. Review of Evidence on Health Aspects of Air Pollution—REVIHAAP Project: Final Technical Report. Available online: <http://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/2013/review-of-evidence-on-health-aspects-of-air-pollution-revihaap-project-final-technical-report> (accessed on 1 December 2016).
3. World Health Organization. Health Effects of Particulate Matter: Policy Implications for Countries in Eastern Europe, Caucasus and Central Asia; 2013. Available online: <http://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/2013/health-effects-of-particulate-matter-policy-implications-for-countries-in-eastern-europe,-caucasus-and-central-asia-2013> (accessed on 1 December 2016).
4. EPA. Criteria Air Pollutants; 2015. Available online: <https://www.epa.gov/criteria-air-pollutants> (accessed on 1 December 2016).
5. Perrone, M.; Gualtieri, M.; Consonni, V.; Ferrero, L.; Sangiorgi, G.; Longhin, E.; Ballabio, D.; Bolzacchini, E.; Camatini, M. Particle size, chemical composition, seasons of the year and urban, rural or remote site origins as determinants of biological effects of particulate matter on pulmonary cells. *Environ. Pollut.* **2013**, *176*, 215–227. [[CrossRef](#)] [[PubMed](#)]
6. Bono, R.; Tassinari, R.; Bellisario, V.; Gilli, G.; Pazzi, M.; Pirro, V.; Mengozzi, G.; Bugiani, M.; Piccioni, P. Urban air and tobacco smoke as conditions that increase the risk of oxidative stress and respiratory response in youth. *Environ. Res.* **2015**, *137*, 141–146. [[CrossRef](#)] [[PubMed](#)]
7. Abbey, D.E.; Ostro, B.E.; Petersen, F.; Burchette, R.J. Chronic respiratory symptoms associated with estimated long-term ambient concentrations of fine particulates less than 2.5 microns in aerodynamic diameter (PM_{2.5}) and other air pollutants. *J. Expo. Anal. Environ. Epidemiol.* **1994**, *5*, 137–159.
8. Mar, T.F.; Jansen, K.; Shepherd, K.; Lumley, T.; Larson, T.V.; Koenig, J.Q. Exhaled nitric oxide in children with asthma and short-term PM_{2.5} exposure in Seattle. *Environ. Health Perspect.* **2005**, *113*, 1791–1794. [[CrossRef](#)] [[PubMed](#)]
9. Nikasinovic, L.; Just, J.; Sahraoui, F.; Seta, N.; Grimfeld, A.; Momas, I. Nasal inflammation and personal exposure to fine particles PM_{2.5} in asthmatic children. *J. Allergy Clin. Immunol.* **2006**, *117*, 1382–1388. [[CrossRef](#)] [[PubMed](#)]
10. Tecer, L.H.; Alagha, O.; Karaca, F.; Tuncel, G.; Eldes, N. Particulate matter (PM_{2.5}, PM_{10-2.5}, and PM₁₀) and children’s hospital admissions for asthma and respiratory diseases: A bidirectional case-crossover study. *J. Toxicol. Environ. Health* **2008**, *71*, 512–520. [[CrossRef](#)] [[PubMed](#)]
11. Loomis, D.; Grosse, Y.; Lauby-Secretan, B.; El Ghissassi, F.; Bouvard, V.; Benbrahim-Tallaa, L.; Guha, N.; Baan, R.; Mattock, H.; Straif, K. The carcinogenicity of outdoor air pollution. *Lancet Oncol.* **2013**, *14*, 1262. [[CrossRef](#)]
12. Brunekreef, B.; Janssen, N.; de Hartog, J.J.; Oldenwening, M.; Meliefste, K.; Hoek, G.; Lanki, T.; Timonen, K.L.; Vallius, M.; Pekkanen, J. Personal, indoor, and outdoor exposures to PM_{2.5} and its components for groups of cardiovascular patients in Amsterdam and Helsinki. *Res. Rep.* **2005**, *127*, 1–79.
13. Dockery, D.W. Epidemiologic evidence of cardiovascular effects of particulate air pollution. *Environ. Health Perspect.* **2001**, *109* (Suppl. 4), 483–486. [[CrossRef](#)] [[PubMed](#)]
14. Liao, D.; Creason, J.; Shy, C.; Williams, R.; Watts, R.; Zweidinger, R. Daily variation of particulate air pollution and poor cardiac autonomic control in the elderly. *Environ. Health Perspect.* **1999**, *107*, 521–525. [[CrossRef](#)] [[PubMed](#)]

15. Yuan, Y.; Liu, S.; Castro, R.; Xubin, P. PM_{2.5} monitoring and mitigation in the cities of China. *Environ. Sci. Technol.* **2012**, *46*, 3627–3628. [[CrossRef](#)] [[PubMed](#)]
16. Statistical Center of Beijing University. *Assessment Report of Air Quality (In Chinese)*; Statistics Science Center of Beijing University: Beijing, China, 2015.
17. Blanchard, C.L.; Tanenbaum, S.; Motallebi, N. Spatial and temporal characterization of PM_{2.5} mass concentrations in California, 1980–2007. *J. Air Waste Manag. Assoc.* **2011**, *61*, 339–351. [[CrossRef](#)] [[PubMed](#)]
18. Lall, R.; Kendall, M.; Ito, K.; Thurston, D.G. Estimation of historical annual PM_{2.5} exposures for health effects assessment. *Atmos. Environ.* **2004**, *38*, 5217–5226. [[CrossRef](#)]
19. Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561–7578. [[CrossRef](#)]
20. Kim, S.Y.; Sheppard, L.; Kim, H. Health effects of long-term air pollution: Influence of exposure prediction methods. *Epidemiology* **2009**, *20*, 442–450. [[CrossRef](#)] [[PubMed](#)]
21. Di, Q.; Kloog, I.; Koutrakis, P.; Lyapustin, A.; Wang, Y.J.; Schwartz, J. Assessing PM_{2.5} exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol.* **2016**, *50*, 4712–4721. [[CrossRef](#)] [[PubMed](#)]
22. Yuval; Broday, D.M. Enhancement of PM_{2.5} exposure estimation using PM₁₀ observations. *Environ. Sci. Proc. Imp.* **2014**, *16*, 1094–1102. [[CrossRef](#)] [[PubMed](#)]
23. Kloog, I.; Coull, B.A.; Zanobetti, A.; Koutrakis, P.; Schwartz, J.D. Acute and chronic effects of particles on hospital admissions in New England. *PLoS ONE* **2012**, *7*, e34664. [[CrossRef](#)] [[PubMed](#)]
24. Xie, Y.; Wang, Y.; Zhang, K.; Dong, W.; Lv, B.; Bai, Y. Daily estimation of ground-level PM_{2.5} concentrations over Beijing using 3 km resolution MODIS AOD. *Environ. Sci. Technol.* **2015**, *49*, 12280–12288. [[CrossRef](#)] [[PubMed](#)]
25. Zheng, Y.X.; Zhang, Q.; Liu, Y.; Geng, G.N.; He, K.B. Estimating ground-level PM_{2.5} concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. *Atmos. Environ.* **2016**, *124*, 232–242. [[CrossRef](#)]
26. Li, L.; Wu, J.; Hudda, N.; Sioutas, C.; Fruin, S.A.; Delfino, R.J. Modeling the concentrations of on-road air pollutants in Southern California. *Environ. Sci. Technol.* **2013**, *47*, 9291–9299. [[CrossRef](#)] [[PubMed](#)]
27. Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R.B. Missing value estimation methods for DNA microarrays. *Bioinformatics* **2001**, *17*, 520–525. [[CrossRef](#)] [[PubMed](#)]
28. Bayraktar, H.; Turalioğlu, F.S.; Tuncel, G. Average mass concentrations of TSP, PM₁₀ and PM_{2.5} in Erzurum urban atmosphere, Turkey. *Stoch. Environ. Res. Risk Assess.* **2008**, *24*, 57–65. [[CrossRef](#)]
29. Tai, A.P.K.; Mickley, L.J.; Jacob, D.J. Correlations between fine particulate matter (PM_{2.5}) and meteorological variables in the United States: Implications for the sensitivity of PM_{2.5} to climate change. *Atmos. Environ.* **2010**, *44*, 3976–3984. [[CrossRef](#)]
30. Tai, A.P.K.; Mickley, L.J.; Jacob, D.J.; Leibensperger, E.M.; Zhang, L.; Fisher, J.A.; Pye, H.O.T. Meteorological modes of variability for fine particulate matter (PM_{2.5}) air quality in the United States: Implications for PM_{2.5} sensitivity to climate change. *Atmos. Chem. Phys.* **2011**, *13*, 3131–3145. [[CrossRef](#)]
31. Liu, Y.; Paciorek, C.J.; Koutrakis, P. Estimating regional spatial and temporal variability of PM_{2.5} concentrations using satellite data, meteorology, and land use information. *Environ. Health Perspect.* **2009**, *117*, 886. [[CrossRef](#)] [[PubMed](#)]
32. Dawson, J.P.; Adams, P.J.; Pandis, S.N. Sensitivity of PM_{2.5} to climate in the eastern us: A modeling case study. *Atmos. Chem. Phys.* **2007**, *7*, 4295–4309. [[CrossRef](#)]
33. Bosilovich, M.G.; Lucchesi, R.; Suarez, M. Merra-2: File Specification. 2015. Available online: <https://ntrs.nasa.gov/search.jsp?R=20150019760> (accessed on 1 December 2016).
34. Global Modeling and Assimilation Office (GMAO). M2T1NXFLX. Available online: https://disc.sci.gsfc.nasa.gov/uuu/datasets?keywords=M2T1NXFLX_V5.12.4 (accessed on 1 July 2016).
35. Wang, X.; Bi, X.; Sheng, G.; Fu, J. Chemical composition and sources of PM₁₀ and PM_{2.5} aerosols in Guangzhou, China. *Environ. Monit. Assess.* **2006**, *119*, 425–439. [[CrossRef](#)] [[PubMed](#)]
36. Giugliano, M.; Lonati, G.; Butelli, P.; Romele, L.; Tardivo, R.; Grosso, M. Fine particulate (PM_{2.5}–PM₁) at urban sites with different traffic exposure. *Atmos. Environ.* **2005**, *39*, 2421–2431. [[CrossRef](#)]

37. Querol, X.; Alastuey, A.; Ruiz, C.R.; Artiñano, B.; Hansson, H.C.; Harrison, R.M.; Buringh, E.; Brink, H.M.T.; Lutz, M.; Bruckmann, P. Speciation and origin of PM₁₀ and PM_{2.5} in selected European cities. *J. Aerosol Sci.* **2004**, *38*, 6547–6555. [CrossRef]
38. Fusheng, W.; Enjiang, T.; Guoping, W.; Wei, H.; Wilson, W.E.; Chapman, R.S.; Pau, J.C.; Zhang, J. Concentrations and elemental components of PM_{2.5} and PM₁₀ in ambient air in four large Chinese cities. *Environ. Monit. China* **2001**, *17*, 1–6. (In Chinese).
39. Chinese Academy of Sciences (RESDC). Available online: <http://www.resdc.cn> (accessed on 1 July 2016).
40. NASA Center for Climate Simulation. Available online: <ftp://dataportal.nccs.nasa.gov/DataRelease/> (accessed on 1 July 2016).
41. Ma, Z.; Hu, X.; Sayer, A.M.; Levy, R.; Zhang, Q.; Xue, Y.; Tong, S.; Bi, J.; Huang, L.; Liu, Y. Satellite-based spatiotemporal trends in PM_{2.5} concentrations: China, 2004–2013. *Environ. Health Perspect.* **2016**, *124*, 184–193. [CrossRef] [PubMed]
42. Wang, J. Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies. *Geophys. Res. Lett.* **2003**, *30*, 267–283. [CrossRef]
43. Geospatial Data Cloud. Available online: <http://www.gscloud.cn> (accessed on 1 July 2016).
44. Jiang, M.; Weiwei, S. Investigating Metrological and Geographical Effect in Remote Sensing Retrieval of PM_{2.5} Concentration in Yangtze River Delta. In Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 4108–4111.
45. O'brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **2007**, *41*, 673–690. [CrossRef]
46. Bozdogan, H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **1987**, *52*, 345–370. [CrossRef]
47. Yang, Y.; Christakos, G. Spatiotemporal characterization of ambient PM_{2.5} concentrations in Shandong province (China). *Environ. Sci. Technol.* **2015**, *49*, 13431–13438. [CrossRef] [PubMed]
48. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* **1998**, 903–995. [CrossRef]
49. Li, L.; Wu, J.; Ghosh, J.K.; Ritz, B. Estimating Spatiotemporal Variability of Ambient Air Pollutant Concentrations with a Hierarchical Model. *Atmos. Environ.* **2013**, *71*, 54–63. [CrossRef] [PubMed]
50. Onorati, R.; Sampson, P.; Guttorp, P. A spatio-temporal model based on the SVD to analyze daily average temperature across the Sicily region. *J. Environ. Stat.* **2013**, *5*.
51. Lindstrom, J.; Szpiro, A.A.; Sampson, P.D.; Sheppard, L.; Oron, A.; Richards, M.; Larson, T. A Flexible Spatio-Temporal Model for Air Pollution: Allowing for Spatio-Temporal Covariates. UW Biostatistics Working Paper Series. 2011. Available online: <http://lup.lub.lu.se/record/4730139> (accessed on 1 December 2016).
52. Dietterich, T.G. Ensemble methods in machine learning. *Mult. Classifier Syst.* **2000**, *1857*, 1–15. [CrossRef]
53. Bühlmann, P.; Hothorn, T. Boosting algorithms: Regularization, prediction and model fitting. *Stat. Sci.* **2007**, *22*, 477–505. [CrossRef]
54. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.
55. Peters, A.; Hothorn, T.; Lausen, B. Ipred: Improved predictors. *R News* **2002**, *2*, 33–36.
56. Ridgeway, G. The state of boosting. *Comput. Sci. Stat.* **1999**, 172–181.
57. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
58. Aslam, J.; Popa, R.; Rivest, R. On estimating the size and confidence of a statistical audit. *EVT* **2007**, *7*, 8.
59. Christakos, G. A Bayesian maximum entropy view to the spatial estimation problem. *Math. Geol.* **1990**, *22*, 763–777. [CrossRef]
60. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, NY, USA, 1997; ISBN 9780195115383.
61. Chiles, P.J.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*; John Wiley & Sons: New York, NY, USA, 1999; ISBN 0-471-08315-1.
62. Johnston, K.; Hoef, M.J.; Krivoruchko, K.; Lucas, N. *ArcGIS 9: Using ArcGIS Geostatistical Analyst*; ESRI Press: Sacramento, CA, USA, 2004; ISBN 9781589481060.
63. Bell, S. *A beginner's Guide to Uncertainty of Measurement*; National Physical Laboratory: Teddington, UK, 2001; ISBN 1368-6550.

64. Kalla, S. Measurement of Uncertainty: Standard Deviation. Available online: <https://explorable.com/measurement-of-uncertainty-standard-deviation> (accessed on 11 December 2016).
65. Kloog, I.; Koutrakis, P.; Coull, B.A.; Lee, H.J.; Schwartz, J. Assessing temporally and spatially resolved PM_{2.5} exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ.* **2011**, *45*, 6267–6275. [[CrossRef](#)]
66. Zang, Z.; Wang, W.; You, W.; Li, Y.; Ye, F.; Wang, C. Estimating ground-level PM_{2.5} concentrations in Beijing, China using aerosol optical depth and parameters of the temperature inversion layer. *Sci. Total Environ.* **2017**, *575*, 1219–1227. [[CrossRef](#)] [[PubMed](#)]
67. Raffuse, S.; McCarthy, M.; Craig, K.; DeWinter, J.; Jumbam, L.; Fruin, S.; Gauderman, J.; Lurmann, F. High-resolution MODIS aerosol retrieval during wildfire events in California for use in exposure assessment. *J. Geophys. Res. Atmos.* **2013**, *118*, 11242–11255. [[CrossRef](#)]
68. Watson, J.G.; Chow, J.C.; Shah, J.J. *Analysis of Inhalable and Fine Particulate Matter Measurements*; National Tech Information Service: Alexandria, VA, USA, 1981; ISBN 9781249430346.
69. Jerrett, M.; Burnett, R.; Pope, A.; Krewski, D.; Thurston, G.; Christakos, G.; Hughes, E.; Ross, Z.; Shi, Y.; Thun, M. *Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort: Final Report*; State of California Air Resources Board: Sacramento, CA, USA, 2012.
70. Robichaud, A.; Ménard, R. Multi-year objective analyses of warm season ground-level ozone and PM_{2.5} over North America using real-time observations and Canadian operational air quality models. *Atmos. Chem. Phys.* **2014**, *14*, 1769–1800. [[CrossRef](#)]
71. Li, L.; Wu, A.; Cheng, I.; Wu, J. Spatiotemporal Estimation of Historical 1 PM_{2.5} Concentrations Using PM₁₀, Meteorological Variables, and Spatial Effect. *Atmos. Environ.* **2017**. under review.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).