Environmental Pollution 242 (2018) 605–613

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Spatiotemporal patterns of PM_{10} concentrations over China during 2005–2016: A satellite-based estimation using the random forests approach☆

Gongbo Chen [a, 1], Yichao Wang [b, 1], Shanshan Li [a], Wei Cao [c], Hongyan Ren [c], Luke D. Knibbs [d], Michael J. Abramson [a], Yuming Guo [a, *]

[a] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[b] Murdoch Children's Research Institute, Parkville, Victoria, Australia, Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia
[c] Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[d] Department of Epidemiology and Biostatistics, School of Public Health, The University of Queensland, Brisbane, Australia

ARTICLE INFO

Article history:
Received 8 February 2018
Received in revised form
4 July 2018
Accepted 4 July 2018
Available online 11 July 2018

Keywords:
PM_{10}
AOD
Random forests
China

ABSTRACT

Background: Few studies have estimated historical exposures to PM_{10} at a national scale in China using satellite-based aerosol optical depth (AOD). Also, long-term trends have not been investigated.
Objectives: In this study, daily concentrations of PM_{10} over China during the past 12 years were estimated with the most recent ground monitoring data, AOD, land use information, weather data and a machine learning approach.
Methods: Daily measurements of PM_{10} during 2014–2016 were collected from 1479 sites in China. Two types of Moderate Resolution Imaging Spectroradiometer (MODIS) AOD data, land use information, and weather data were downloaded and merged. A random forests model (non-parametric machine learning algorithms) and two traditional regression models were developed and their predictive abilities were compared. The best model was applied to estimate daily concentrations of PM_{10} across China during 2005–2016 at 0.1° (≈10 km).
Results: Cross-validation showed our random forests model explained 78% of daily variability of PM_{10} [root mean squared prediction error (RMSE) = 31.5 μg/m^3]. When aggregated into monthly and annual averages, the models captured 82% (RMSE = 19.3 μg/m^3) and 81% (RMSE = 14.4 μg/m^3) of the variability. The random forests model showed much higher predictive ability and lower bias than the other two regression models. Based on the predictions of random forests model, around one-third of China experienced with PM_{10} pollution exceeding Grade II National Ambient Air Quality Standard (>70 μg/m^3) in China during the past 12 years. The highest levels of estimated PM_{10} were present in the Taklamakan Desert of Xinjiang and Beijing-Tianjin metropolitan region, while the lowest were observed in Tibet, Yunnan and Hainan. Overall, the PM_{10} level in China peaked in 2006 and 2007, and declined since 2008.
Conclusions: This is the first study to estimate historical PM_{10} pollution using satellite-based AOD data in China with random forests model. The results can be applied to investigate the long-term health effects of PM_{10} in China.

1. Introduction

Particles with a aerodynamic diameter <10 μm (PM_{10}), are associated with short-term and long-term health effects (Brunekreef and Holgate, 2002). Increasing epidemiological evidence has demonstrated associations between PM_{10} air pollution and a wide range of health outcomes including respiratory and cardiovascular diseases, lung cancer, birth outcomes and mental health (Dockery and Stone, 2007; Lim et al., 2012; Raaschou-

☆ This paper has been recommended for acceptance by Haidong Kan.

* Corresponding author. Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria 3004, Australia.
E-mail address: yuming.guo@monash.edu (Y. Guo).

[1] These two authors contributed equally to the paper.

https://doi.org/10.1016/j.envpol.2018.07.012
0269-7491/

Nielsen et al., 2013; Stieb et al., 2012; Stockfelt et al., 2017; Yitshak-Sade et al., 2017). Exposure to ambient fine particulate matter ($PM_{2.5}$, a component of PM_{10}) accounted for 7.6% of total global deaths and 4.2% of global disability-adjusted life years (DALYs) in 2015 and more than half that burden was in east and south Asia (Cohen et al., 2017).

China faces severe problems with air pollution as a consequence of the rapid development of economy and urbanization (Kan et al., 2009). The relationships between PM_{10} and health outcomes have been widely investigated in China. However, current studies exploring these relationships are restricted to either localized areas or short time periods (Lai et al., 2013; Shang et al., 2013), with potentially less representative and biased results. Estimating exposure to PM_{10} in areas and time periods with sparse ground level monitoring data could benefit the future studies of PM_{10} in terms of research scale and study period.

Aerosol optical depth (AOD), is a measure of light attenuation by aerosols in the atmosphere (Lee et al., 2011). Previous studies have demonstrated that AOD could be well correlated with measured ground-level PM, with significant correlation coefficients around 0.70 (Nordio et al., 2013), indicating AOD is a good surrogate of particulate air pollution. Satellite-retrieved AOD has been used to predict concentrations of $PM_{2.5}$ in China (Fang et al., 2016; You et al., 2016). However, few studies have estimated the concentrations of PM_{10} over China employing PM-AOD association at the national scale and the long-term trend of PM_{10} has not been investigated. In addition, previous studies on prediction of PM in China relied on traditional parametric regression models [e.g., generalized additive models (GAM), geographically weighted regression models (GWR)] (Fang et al., 2016; Ma et al., 2016; Zhang et al., 2016b). Although these studies reported good performances, a recent study found the random forests approach, as a non-parametric regression based on machine learning algorithms, had advantages compared with traditional regression models and showed higher predictive ability (Hu et al., 2017). Moreover, the random forests approach involves fewer restrictive assumptions and is more flexible to use (Breiman, 2001).

In this study, we compared the predictive ability of the random forests model with two traditional regression models developed using satellite-retrieved AOD data, in-situ measurements of PM_{10}, land use information, and meteorological data in China. Then the best model of them was employed to estimate the spatial and temporal distribution of PM_{10} concentrations over the whole country during 2005–2016.

2. Material and methods

2.1. In-situ measurements of PM_{10}

In-situ measurements of PM_{10} from May 13 2014 to December 31 2016 were obtained from China National Environmental Monitoring Center (CNEMC) (http://www.cnemc.cn/). Daily average concentrations of PM_{10} were collected from 1479 ground monitoring stations located in 31 provinces and municipalities in China (Fig. 1). Real-time concentrations of PM_{10} were measured by Tapered Element Oscillating Microbalances (TEOM) and beta-attenuation methods. Details of the measurements have been reported elsewhere (Zhao et al., 2016). The calibration and quality control of PM_{10} data were performed by CNEMC (Jiang et al., 2015). Hourly concentrations of PM_{10} were converted into daily averages after data cleaning.

2.2. Satellite-retrieved AOD data

Daily data from the Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6 AOD during the study period were downloaded from Level 1 and Atmosphere Archive & Distribution System of NASA (https://ladsweb.modaps.eosdis.nasa.gov/). Two MODIS Aqua Level-2 AOD products, Dark Target (DT) and Deep Blue (DB), were downloaded at a spatial resolution of 10 km. DT and DB AOD have different aerosol retrieval algorithms and show different performances for dark and bright targets. As neither DT nor DB AOD consistently outperforms each other, a merged data product of them is recommended (Sayer et al., 2014). We did not use the existing merged MODIS AOD product (MODIS Aerosols Merged Dark Target Deep Blue Product), as this product discards some DB AOD values over dark vegetated surfaces and it has been challenged by several assumptions (Levy et al., 2013). To improve the spatial coverage of AOD data, we combined DT and DB AOD using the method previously reported (Ma et al., 2016).

First, the following linear regressions between these two products were developed and used to fill in their gaps;

$$AOD_{DT} = \alpha + \beta^{*}AOD_{DB} \quad (1)$$

$$AOD_{DB} = \alpha + \beta^{*}AOD_{DT} \quad (2)$$

where AOD_{DT} and AOD_{DB} refer to DT and DB AOD, respectively; α is the intercept and β is the coefficient. Missing DT with valid DB values were interpolated with formula (1), and similarly, missing DB with valid DT values were estimated with formula (2). Then, ground measures of AOD were obtained from Aerosol Robotic Network (AERONET) (https://aeronet.gsfc.nasa.gov/new_web/index.html). The downloading and processing of AERONET AOD data are shown in the "Interpolation of AOD at 550 nm" section of the Supplementary Material. DT and DB AOD data were combined by comparison with AERONET AOD using previously reported Inverse Variance Weighting (IVW) method (Chen et al., 2017). Compared to the merged dark target-deep blue product in Collection 6 MODIS AOD, the combined AOD data with our method remarkably increased the spatial coverage of satellite data in China and showed similar performance (Ma et al., 2016).

2.3. Meteorological data

Daily meteorological data in China during the study period were collected from the 824 weather stations operated by the China Meteorological Data Sharing Service System (CMDSSS) (http://data.cma.gov.cn). The locations of all weather stations are shown in Fig. S2 in the Supplementary Material. Daily mean temperature (°C), relative humidity (%), barometric pressure (kPa) and wind speed (km/h) were collected for model development and prediction. For ground monitoring stations or grid cells not covered by CMDSSS, daily values of four meteorological variables were interpolated with kriging (Furrer et al., 2009). The interpolation of meteorological variables is shown in details in the "Interpolation of meteorological variable" section of the Supplementary Material.

2.4. Land cover data

Monthly average data of MODIS Level 3 Normalized Difference Vegetation Index (NDVI) at 0.1-degree ($\approx$10 km) resolution were downloaded from the NASA Earth Observatory (http://neo.sci.gsfc.nasa.gov/). Annual land cover data (Collection 5.1) at 500-m resolution during 2005–2012 were obtained from Global Mosaics of the standard MODIS land cover type data of Global Land Cover Facility (http://glcf.umd.edu/) (Friedl et al., 2010). Annual percentage of urban cover within a buffer size was calculated and used for prediction. Particularly, the 2012 urban cover data (the most recent data) were used for the prediction during 2012–2016.

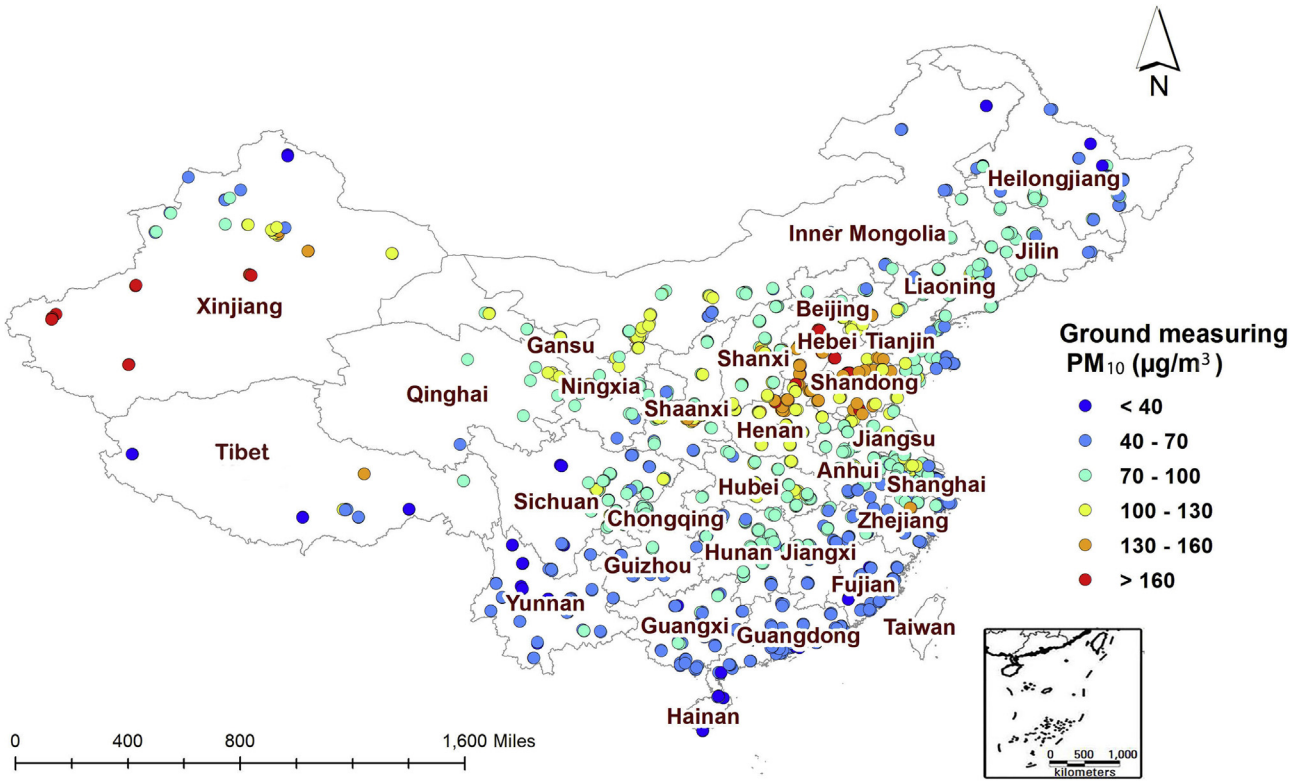

Fig. 1. Mean concentrations of PM_{10} (μg/m^3) in 1479 ground monitoring stations during 2014–2016 in China.

2.5. Other spatial predictors

MODIS active fire data (Collection 6) during the study period were downloaded from NASA Fire Information for Resource Management System (FIRMS) (https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data). Daily counts of fire spots were calculated within the buffer radius of 75 km (Hu et al., 2014). High-resolution elevation data covering China were downloaded from The CGIAR Consortium for Spatial Information at a resolution of 3 arc-seconds (≈90 m) (http://srtm.csi.cgiar.org/).

2.6. Model development

In contrast to traditional regression, the random forests algorithm for regression randomly draws bootstrap samples from the training data set and builds regression trees for each bootstrap sample. At each node of the regression tree, a random sample of all predictors was selected to choose the best split among them (Liaw and Wiener, 2002). Apart from AOD, a range of potential predictors relevant to PM_{10} were considered at the model development stage (Fang et al., 2016; Ma et al., 2014; Meng et al., 2016a). Firstly, all potential predictors were included in the preliminary random forests regression model and measures of variable importance [two parameters: increase in mean square errors (Inc.MSE) and increase in node purities (Inc.NodePurity)] were calculated for each of them. Then, predictors included in the final model were selected according to measures of variable importance (Hu et al., 2017). In this study, predictors with Inc.MSE >500 and Inc.NodePurity >2,000,000 were included in the final model and those below the thresholds were not included as they made no contribution to predictive ability. The final random forests model is shown as following:

$$PM_{10ij} = AOD_{ij} + TEMP_{ij} + RH_{ij} + BP_{ij} + WS_{ij} + NDVI + Urban\_cover + doy + log(elev) \quad (3)$$

where PM_{10ij} is the concentration of PM_{10} on day i at station j; AOD_{ij} is the combined AOD on day i at station j; *TEMP*, *RH*, *BP* and *WS* are daily mean temperature, relative humidity, barometric pressure and wind speed; *NDVI* is the monthly average NDVI value; *Urban_cover* is the percentage of urban cover with a buffer radius of 10 km; *doy* is of day of year; *log(elev)* is log transformed elevation.

Random forests are user-friendly non-parametric machine learning algorithms without the need to define the complex relationships between predictors and dependent variable (e.g., linear and non-linear relationships and degrees of freedom). Two parameters, the number of predictors in the random subset of each node (m_{try}) and the number of trees in the forest (n_{tree}), were set as the default value and 100, respectively, for model development. The optimal buffer radius of percentage of urban cover and NDVI values were selected according to median R^2 and mean square errors (mse). Details of buffer radius selections are shown in Table S2 in the Supplementary Material.

To compare the predictive ability of random forests model with traditional regression models, we also developed a widely used GAM (Liu et al., 2009), and a non-linear exposure-lag-response model which takes into account the potential lagged effects of meteorological variables on PM_{10} (Chen et al., 2018b). These two models are:

$$PM_{10ij} = AOD_{ij} + ns(TEMP_{ij}, 3) + ns(RH_{ij}, 3) + ns(BP_{ij}, 3) + ns(WS_{ij}, 3) + NDVI + ns(Urban\_cover, 3) + ns(doy,8) + log(elev) \quad (4)$$

$$PM_{10ij} = AOD_{ij} + cb\_TEMP_{ij} + cb\_RH_{ij} + cb\_BP_{ij} + cb\_WS_{ij} + NDVI + ns(Urban\_cover, 3) + ns(doy,8) + log(elev) \quad (5)$$

Model 4 and Model 5 are GAM and the non-linear exposure-lag-response model, respectively. In model 4, we fitted meteorological variables, percentage of urban cover and day of the year using natural cubic splines (*ns*) with 3, 3 and 8 degrees of freedom, respectively, considering their potential non-linear effects (Chen et al., 2018a). Model 5 was developed by incorporating distributed lag non-liner model (DLNM) into GAM (Gasparrini et al., 2010), in which, *cb_TEMP*, *cb_RH*, *cb_BP* and *cb_WS* are mean temperature, relative humidity, barometric pressure and wind speed on the current day and previous two days (lag 0–2 days) fitted using *crossbasis* function (3 degrees of freedom) of DLNM R package (Gasparrini, 2011). The selections of degrees of freedom for non-linear variables and lag days of meteorological variables depended on adjusted R^2 and Generalized Cross Validation (GCV) values. Details of the selections are shown in Table S3 in the Supplementary Material.

2.7. Model validation and prediction

The predictive abilities of these three models were evaluated with a 10-fold cross-validation (CV) method. The validation dataset was built by randomly selecting 148 (10%) out of 1479 ground monitoring stations and the rest of the stations (90%) were used as the training dataset. This process was repeated 200 times. The linear regression was built between predicted and ground measured PM_{10} to check their difference. The adjusted R^2, Root Mean Square Error (RMSE), regression intercept and coefficients were calculated.

The prediction of daily PM_{10} concentrations began by creating a 0.1-degree grid, including 96,103 grid cells, covering the entire China. Data on all spatial and temporal predictors included in the final model were linked to each grid cell by longitude, latitude and calendar date. Grid cells were assigned mean values of spatial predictors (e.g., AOD, elevation) where multiple values of them fell in one grid cell. For predictors with temporal resolution other than daily, e.g., NDVI and land cover, they were linked with other predictors by calendar month or year, as those values are unlikely to change within a month or a year.

The final best model (random forests model), developed using ground measurements during 2014–2016, was applied to predict the daily concentrations of PM_{10} over China during 2005–2016. Daily results for prediction were aggregated into monthly, seasonal and annual averages. To take into account regional variations of PM-AOD relationship (Zhang et al., 2009), model development and prediction were performed by each province separately. In addition, we plotted the trend lines of predicted PM_{10} in the entire China and several heavily-polluted regions during the study period.

3. Results

3.1. Ground measured concentrations of PM_{10}

The distribution of the 1479 ground monitoring stations included in the present study was asymmetrical, with several clusters appeared in Beijing-Tianjin metropolitan region (including Beijing, Tianjin and Hebei), the Yangtze River and Pearl River Deltas, while decentralized stations were located in eastern Inner Mongolia and western China (Fig. 1). From these monitors, the overall mean concentration of PM_{10} in China during the period from May 2014 to December 2016 was 85.9 $\mu g/m^3$. In total, 68% (1005/1479) of the monitoring stations reported levels of PM_{10} exceeding 70 $\mu g/m^3$, the Grade II concentration limit of National Ambient Air Quality Standard in China (CNAAQS) (Li et al., 2017). The highest levels of ground monitored PM_{10} (>160 $\mu g/m^3$) were present in the Taklamakan Desert of Xinjiang and Beijing-Tianjin metropolitan region, while the lowest (<40 $\mu g/m^3$) were observed in Tibet, Yunnan and Hainan. A summary of provincial level measured PM_{10} is shown in Table S4 in the Supplementary Material. Patterns of ground measured PM_{10} across China in our study were consistent with those previously reported. For example, previous studies indicated the high levels of PM_{10} in local cities of Beijing-Tianjin metropolitan region and Xinjiang (e.g, Shijiazhuang and Wulumuqi), while the low levels of PM_{10} in remote areas of western China (e.g., Lasa and Kunming) (He et al., 2017; Zhao et al., 2016).

3.2. Results of model fitting and 10-fold cross-validation

At the model development stage, 12 predictors were considered and assessed for their importance (Table S5 in the Supplementary Material) and 9 of them were included in our final model. Among those predictors, day of year was the most important predictor, followed by AOD and temperature. In our 10-fold cross-validation results (Fig. 2), the final daily random forests model captured most of the variability in PM_{10} (CV $R^2 = 78\%$, RMSE = 31.5 $\mu g/m^3$). When aggregated into monthly, seasonal and annual averages, the models captured 82% (RMSE = 19.3 $\mu g/m^3$), 81% (RMSE = 18.2 $\mu g/m^3$) and 81% (RMSE = 14.4 $\mu g/m^3$) of the variability, respectively. Daily GAM and the non-linear exposure-lag-response model showed less predictive ability, respectively explaining only 50% (RMSE = 47.4 $\mu g/m^3$) and 47% (RMSE = 29.1 $\mu g/m^3$) of the variability. Table 1 lists the results of 10-fold cross-validation in each province of China using 3 different models. Compared to GAM and non-linear exposure-lag-response model, random forests showed much higher CV R^2 in each province. Across China, random forests had high CV R^2 in northern China (Hebei, Tianjin, Henan and Beijing), but low CV R^2 in western China (Qinghai and Tibet). The strong relationship between MODIS AOD and PM in Beijing-Tianjin-Hebei Region were also shown in previous study (Zheng et al., 2016). The poor accuracy of AOD over Tibet may due to the clear air, low value of AOD and plateau climate (Wang et al., 2007). Overall, the daily random forests model showed much higher predictive ability than the two traditional regression models. Aggregated into monthly, seasonal and annual averages, higher CV R^2s of the random forests model were also shown. As the random forests model had much higher CV R^2 and lower RMSE than the traditional regression models, daily concentrations of PM_{10} across China during 2005–2016 were estimated from the final random forests model.

3.3. Predicted spatial patterns of PM_{10} concentrations

The results of prediction revealed the severe PM_{10} pollution in China during the study period. The annual mean concentration of predicted PM_{10} during the past 12 years was 63.2 $\mu g/m^3$ (Fig. 3). By prediction, around one third of territory of China (33% of grid cells) experienced PM_{10} pollution exceeding the CNAAQS Grade II standard (>70 $\mu g/m^3$). The most polluted areas (PM_{10} > 135$\mu g/m^3$) were observed in the Taklimakan Desert, Gobi Desert and Beijing-Tianjin metropolitan region, followed by Shandong, Henan and Sichuan Basin. Less polluted areas (PM_{10} < 45$\mu g/m^3$) were present in Yunnan, Hainan and north-eastern Inner Mongolia. The severe PM_{10} pollution in north-western China and the Beijing-Tianjin metropolitan region and the clean air in southern China were also previously reported (Zhang et al., 2016b).

3.4. Predicted temporal patterns of PM_{10} concentrations

Seasonal variations of the PM_{10} concentration during the 12-year prediction period are illustrated in Fig. 4. The most polluted

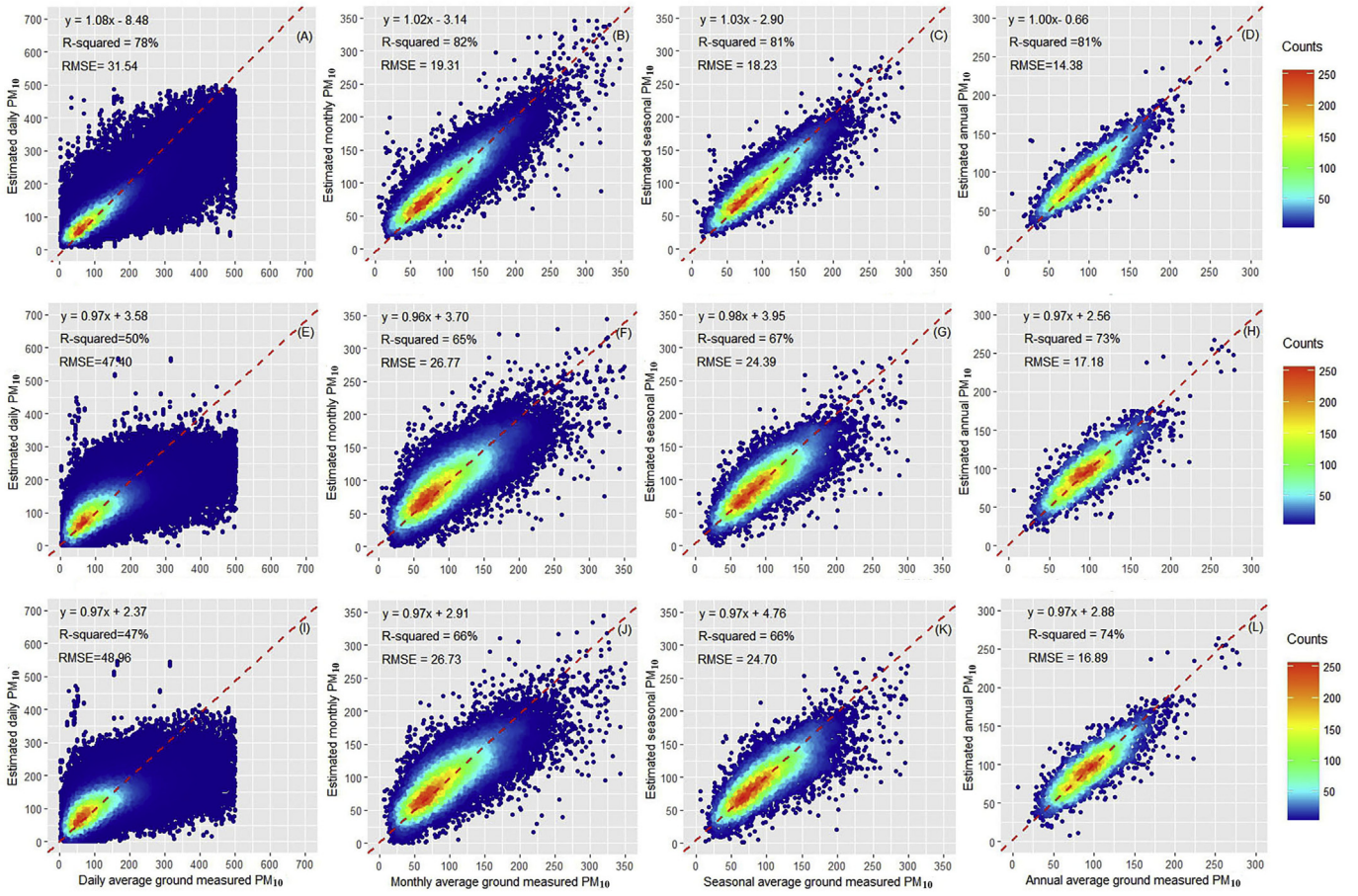

Fig. 2. Density scatterplots of 10-fold cross-validation. (A), (B), (C) and (D) are daily, monthly, seasonal and annual results for the random forests model, respectively; (E), (F), (G) and (H) are daily, monthly, seasonal and annual results the generalized additive model, respectively. (I), (J), (K) and (L) are daily, monthly, seasonal and annual results for the non-linear exposure-lag-response model, respectively.

season was winter (mean $PM_{10} = 74.1\ \mu g/m^3$), while the least polluted season was summer (mean $PM_{10} = 47.4\ \mu g/m^3$). For heavily-polluted areas (e.g., Beijing-Tianjin metropolitan region, North China Plain and Sichuan Basin), PM_{10} rose to the highest levels in winter from autumn and decreased in spring and summer. Unlike other areas of China where the severe PM_{10} occurred in winter, the Taklimakan Desert in Xinjiang province showed an uninterrupted increasing trend of PM_{10} from winter to spring. The annual changes of predicted PM_{10} in China are illustrated in Fig. 5. During the predictive period from 2005 to 2016, most areas of China showed modest changes in PM_{10}.

For the entire China and some heavily-polluted regions (e.g., Beijing-Tianjin-Hebei region and Taklamakan Desert of Xinjiang), levels of PM_{10} peaked in 2006 and 2007, and kept decreasing since 2008. (Fig. 5). Similar to our results, the decreasing trends of AOD and predicted $PM_{2.5}$ after 2008 were previously revealed (He et al., 2016; Ma et al., 2016).

4. Discussion

In this study, we first developed a daily random forests model of ground-level PM_{10} based on satellite-retrieved AOD and other spatial and temporal predictors for 2014–2016. The validated model was then applied to estimate historical PM_{10} concentrations in China over a 12-year period (2005–2016). The final model indicated our predicted PM_{10} could capture the ground measurements with relatively low prediction error. During the study period, heavily-polluted regions were mainly located in the deserts and North China Plain. We reported spatial patterns of PM_{10} in four seasons in China. The annual changes showed most of the territory of China experienced a reduction of PM_{10} pollution since 2008.

Most of the published studies on satellite-based prediction of PM have focused on $PM_{2.5}$, and few studies have reported the spatiotemporal trends of predicted PM_{10} concentrations in China. Meng et al. (2016b) estimated the ground level PM_{10} with satellite data, but results were limited to one city, Shanghai. However, the seasonal patterns of PM_{10} they reported in Shanghai were in line with our results in the Yangze River Delta Region that winter and summer were the most-polluted and least-polluted season, respectively. Zhang et al. (2016b) reported the national-scale spatial distribution of PM_{10} using a satellite-based geographically weighted regression model. However, they only developed a seasonal model, and did not have daily or monthly prediction results. Our study has built on the results of these previous studies, by predominantly providing a long-term prediction of daily PM_{10} concentrations over China.

MODIS AOD have been widely used to predict ground-level $PM_{2.5}$ and PM_{10} concentrations in recent years. We compared the accuracy of MODIS AOD data in our study with previous studies. Kong et al. (2016) reported that the linear AOD-PM_{10} relationship could explain 54% of the variability of PM_{10} ($R^2 = 54\%$) over Beijing during 2009–2010 and it was 57% in our study. We also found the regional difference of AOD-PM_{10} relationship that higher accuracy of AOD was observed in Central China (e.g., Henan and Hubei) and

Table 1
The results of 10-fold cross-validation in each province of China.

Province	Random forests		GAM		Non-linear exposure-lag-response model	
	CV R^2	RMSE	CV R^2	RMSE	CV R^2	RMSE
Hebei	86%	34.2	55%	44.5	50%	48.9
Tianjin	83%	32.0	51%	37.8	40%	42.1
Henan	82%	31.3	45%	33.9	41%	35.6
Beijing	82%	31.4	51%	37.9	47%	40.1
Hubei	81%	22.9	55%	19.4	51%	20.3
Sichuan	80%	22.3	51%	16.4	50%	16.1
Jiangsu	79%	24.0	47%	21.7	40%	22.6
Shaanxi	79%	32.3	45%	34.2	42%	34.6
Chongqing	77%	19.5	45%	12.7	50%	12.2
Liaoning	77%	27.8	37%	31.6	31%	33.5
Jilin	77%	28.1	43%	25.4	41%	27.3
Xinjiang	77%	51.8	52%	58.5	51%	52.1
Shanghai	77%	21.5	40%	19.0	34%	19.1
Heilongjiang	77%	27.4	43%	23.6	40%	23.6
Shanxi	76%	33.5	38%	36.2	31%	38.3
Zhejiang	75%	21.7	42%	16.6	44%	16.1
Shandong	74%	38.0	46%	33.6	42%	35.6
Hunan	72%	24.8	42%	16.9	42%	17.1
Guangdong	71%	16.3	42%	10.5	44%	10.1
Inner Mongolia	71%	33.9	37%	33.9	27%	36.8
Gansu	69%	40.4	31%	41.6	27%	40.4
Anhui	69%	27.4	37%	22.4	33%	22.9
Guangxi	67%	26.1	43%	14.6	47%	14.0
Jiangxi	66%	24.9	30%	18.7	30%	18.2
Guizhou	65%	22.7	25%	11.9	32%	12.0
Ningxia	62%	40.9	22%	42.1	15%	44.4
Fujian	56%	16.9	20%	10.3	22%	11.3
Yunnan	51%	19.8	24%	13.8	31%	12.2
Qinghai	40%	47.4	20%	33.0	18%	33.0
Tibet	35%	37.2	22%	16.5	19%	18.2

Abbreviations: Generalized additive models (GAM), Cross Validation (CV), Root Mean Square Error (RMSE).

lower accuracy in remote areas of western China (e.g., Qinghai and Tibet) (Wang et al., 2007). The mean values of AOD over some major cities in China (e.g., Beijing, Tianjin, Hebei and Nanjing) during the study period were similar to those previously reported (Shao et al., 2017; Zhang et al., 2017). In addition, a slight decreasing trend of AOD value over urban areas of Beijing reported by Zhang et al. (2016a) was also observed in our study.

The monitoring network of ground PM_{10} concentrations is expanding in China as consequence of better management and control of air pollution (Gao et al., 2017). However, the spatial coverage of current network is asymmetrical, with more stations located in eastern coastal cities or urban areas and fewer stations in middle and western regions of China. Estimating PM from satellite-retrieved AODdata is a promising and effective approach to fill in the gaps of ground monitoring. Compared with previous studies, the present study included the most in-situ ground monitoring stations (Ma et al., 2016; Zhang et al., 2016b). Also, the new stations included in this study were located in central and western China where little ground monitoring data were available in the past. The locations of these new stations are shown in Fig. S3 in the Supplementary Material. With better representation of the ground measurements, the results of our prediction may be more capable than previous studies to capture the spatial variations of PM_{10} across all of China. In addition, ground monitoring data included in this study covered more than two consecutive years, rather than the one year of data in most published studies, which could improve the representativeness of model predictions.

The spatial distribution of PM_{10} of our study is similar to the $PM_{2.5}$ estimations in roughly the same time period reported by previous studies (Ma et al., 2016), indicating $PM_{2.5}$ is an important component of PM_{10}. Zhao et al. (2016) also stated that $PM_{2.5}$ accounted for around 50% of PM_{10} concentrations in China. Seasonal patterns of PM_{10} shown in our study are consistent with ground monitoring data reported by previous studies (Wang et al., 2015; Zhao et al., 2016). Summer is the cleanest season throughout a year benefiting from the meteorological conditions, less anthropogenic emissions and summer monsoon (Wang et al., 2015). Spring and winter are the two worst seasons for PM_{10} pollution, but might be dominated by different sources.

The severe PM_{10} air pollution in the Taklimakan desert, one of the largest deserts in the world and located in north-western China, mainly originates from wind-blown dust (Wu et al., 2012). The level of PM_{10} peaks in spring when resuspended dust episodes occur in the absence of snow and ice cover (Vallius et al., 2000). By contrast, high levels of PM_{10} in North China Plain, characterized with dense steel and power industries, are dominated by fuel combustion and traffic emission sources (Hu et al., 2010; Wang et al., 2014). In addition, domestic heating is provided on the North China Plain

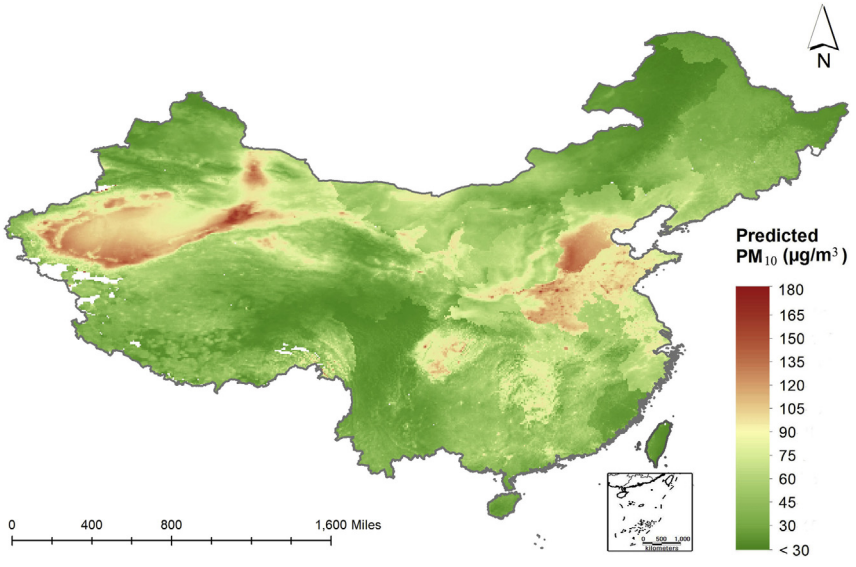

Fig. 3. Mean concentrations of predicted PM_{10} ($\mu g/m^3$) over China during 2005–2016.

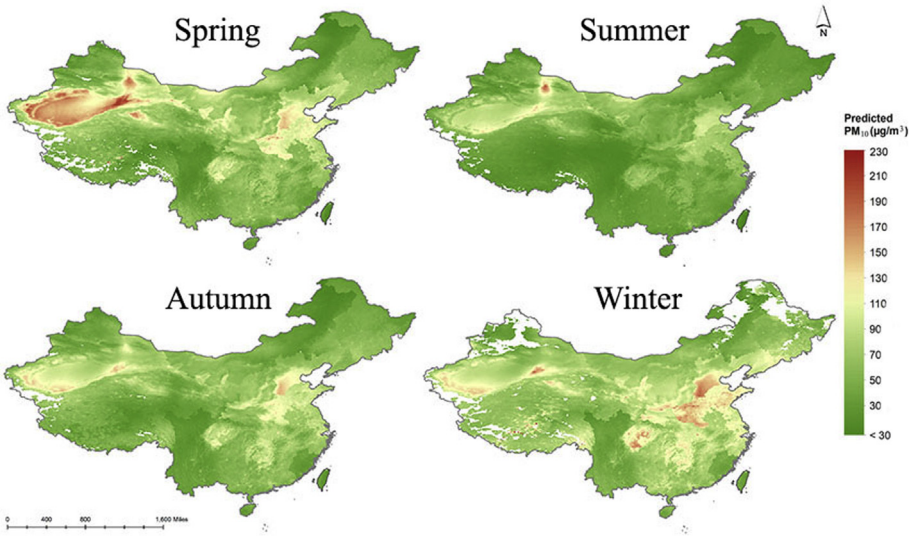

Fig. 4. Mean concentrations of predicted PM_{10} (μg/m^3) in four seasons of China during 2005–2016.

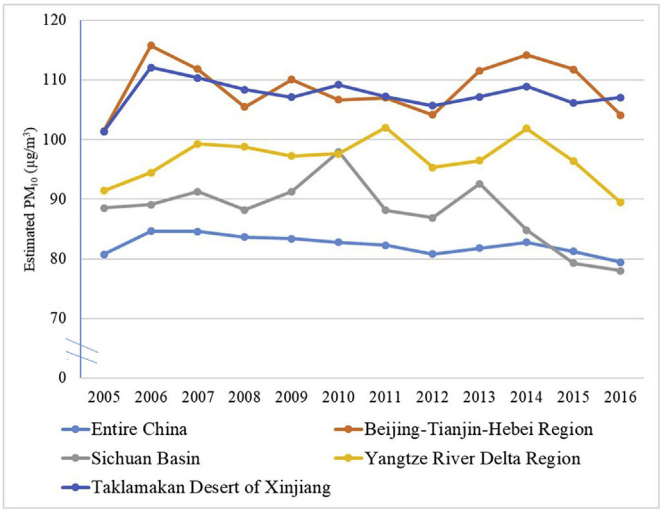

Fig. 5. Annual changes of estimated PM_{10} (μg/m^3) in the entire China and heavily-polluted regions during the study period.

throughout the whole winter season, which contributes to more coal combustion and generates more pollutant emissions (Sun et al., 2013).

The annual changes of PM_{10} in our study were characterized with the peak of PM_{10} pollution in 2006–2007 and continued decreases since 2008. This time trend of PM_{10} is consistent with previous studies (Ma et al., 2016; Yao and Lu, 2014). The annual changes during the study period reflected the preliminary achievement of air pollution control in China over recent years. Since 2009, China's government has taken a series of strict actions to control air pollution, including national plans and policies to tackle air pollution, updating the national ambient air quality standard and extending national air quality monitoring network (Gao et al., 2017). Although challenges of air pollution reduction in China remain, the benefits of these actions have been recognized (Li et al., 2016).

In addition to including the most recent ground monitoring data for PM_{10}, we employed non-parametric machine learning algorithms in this study. The high predictive ability we observed in our study may due to several strengths of the random forests approach. For example, the use of randomness contributes to the reduced bias and increased predictive ability (Breiman, 2001). The best split at each node of the random forest is performed using the best subset of predictors (Liaw and Wiener, 2002). Thus, this approach is capable of dealing with the complex relationships between predictors and may be more robust to noise and overfitting. The internal estimates of random forests are used to evaluate the importance of each predictor included in the model and facilitate the process selecting the best predictors (Breiman, 2001).

However, there were still some limitations in our study. Most notably, we did not have ground monitoring data to validate the PM_{10}-AOD before 2014 and assumed that the relationship we observed could be applied back to 2005. This may affect the long-term trends of predicted PM_{10}. Traffic is often the primary source of pollutant emissions and traffic data can provide information for the distribution of air pollutants (Hamra et al., 2015; Künzli et al., 2000). Unfortunately, we had no access to the traffic data at the national scale of China. However, we considered the variable of distance to the nearest highway at the model development stage. This variable was not included in the final model, as it did not contribute to model performance. Also, missing AOD values are a challenge for satellite-based prediction of PM concentrations (Xiao et al., 2017). Advanced methods to offset missing values need more exploration.

5. Conclusions

Using satellite-based AOD, land use inforamtion, and meteorological data to estimate the PM_{10} level in China can compensate for the insufficient coverage of ground monitoring stations and lack of historical exposure data. This is the first study to estimate the historical PM_{10} concentrations at 0.1-degree ($\approx$10 km) resolution over a 12-year period (2005–2016) in China using a daily prediction model and satellite-based AOD data. By linking with health data, the predicted concentraionscan be applied to investigate the long-term health effects of PM_{10} in China.

Conflicts of interest

The authors have declared that no competing interests exist.

Acknowledgements

Y.G. was supported by the Career Development Fellowship of Australian National Health and Medical Research Council (APP1107107). S.L. was supported by the Early Career Fellowship of Australian National Health and Medical Research Council (APP1109193) and Seed Funding from the National Health and Medical Research Council (NHMRC) Centre of Research Excellence (CRE) – Centre for Air quality and health Research and evaluation (CAR) (APP1030259). G.C. was supported by China Scholarship Council (CSC).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envpol.2018.07.012.

References

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Brunekreef, B., Holgate, S.T., 2002. Air pollution and health. Lancet 360, 1233–1242.
Chen, G., Knibbs, L.D., Zhang, W., Li, S., Cao, W., Guo, J., Ren, H., Wang, B., Wang, H., Williams, G., 2017. Estimating spatiotemporal distribution of PM 1 concentrations in China with satellite remote sensing, meteorology, and land use information. Environ. Pollut.
Chen, G., Li, S., Knibbs, L.D., Hamm, N.A.S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M.J., Guo, Y., 2018a Sep 15. A machine learning method to estimate $PM_{2.5}$ concentrations across China with remote sensing, meteorological and land use information. Sci. Total Environ. 636, 52–60.
Chen, Z., Zhang, T., Zhang, R., Zhu, Z., Ou, C., Guo, Y., 2018b. Estimating PM2. 5 concentrations based on non-linear exposure-lag-response associations with aerosol optical depth and meteorological measures. Atmos. Environ. 173, 30–37.
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 3rd, 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, 1907–1918.
Dockery, D.W., Stone, P.H., 2007. Cardiovascular risks from fine particulate air pollution. Mass Medical Soc 356 (5), 511–513.
Fang, X., Zou, B., Liu, X., Sternberg, T., Zhai, L., 2016. Satellite-based ground PM 2.5 estimation using timely structure adaptive modeling. Rem. Sens. Environ. 186, 152–163.
Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. Rem. Sens. Environ. 114, 168–182.
Furrer, R., Nychka, D., Sain, S., Nychka, M.D., 2009. Package 'fields'. R Foundation for Statistical Computing, Vienna, Austria. http://www.idg.pl/mirrors/CRAN/web/packages/fields/fields.pdf. (Accessed 22 December 2012).
Gao, J., Woodward, A., Vardoulakis, S., Kovats, S., Wilkinson, P., Li, L., Xu, L., Li, J., Yang, J., Cao, L., 2017. Haze, public health and mitigation measures in China: a review of the current evidence for further policy response. Sci. Total Environ. 578, 148–157.
Gasparrini, A., 2011. Distributed lag linear and non-linear models in R: the package dlnm. J. Stat. Software 43, 1.
Gasparrini, A., Armstrong, B., Kenward, M.G., 2010. Distributed lag non-linear models. Stat. Med. 29, 2224–2234.
Hamra, G.B., Laden, F., Cohen, A.J., Raaschou-Nielsen, O., Brauer, M., Loomis, D., 2015. Lung cancer and exposure to nitrogen dioxide and traffic: a systematic review and meta-analysis. Environ. Health Perspect. 123, 1107.
He, J., Gong, S., Yu, Y., Yu, L., Wu, L., Mao, H., Song, C., Zhao, S., Liu, H., Li, X., 2017. Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major Chinese cities. Environ. Pollut. 223, 484–496.
He, Q., Zhang, M., Huang, B., 2016. Spatio-temporal variation and impact factors analysis of satellite-based aerosol optical depth over China from 2002 to 2015. Atmos. Environ. 129, 79–90.
Hu, Waller, L.A., Lyapustin, A., Wang, Y., Liu, Y., 2014. Improving Satellite-driven PM Less than 2.5 Micrometer Models with Moderate Resolution Imaging Spectroradiometer Fire Counts in the Southeastern US.
Hu, H., Yang, Q., Lu, X., Wang, W., Wang, S., Fan, M., 2010. Air pollution and control in different areas of China. Crit. Rev. Environ. Sci. Technol. 40, 452–518.
Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L., Strickland, M., Liu, Y., 2017. Estimating PM2. 5 concentrations in the conterminous United States using the random forest approach. Environ. Sci. Technol. 51 (12), 6936–6944.
Jiang, J., Zhou, W., Cheng, Z., Wang, S., He, K., Hao, J., 2015. Particulate matter distributions in China during a winter period with frequent pollution episodes (January 2013). Aerosol Air Qual. Res 15, 494–503.
Kan, H., Chen, B., Hong, C., 2009. Health impact of outdoor air pollution in China: current knowledge and future research needs. Environ. Health Perspect. 117, A187.
Kong, L., Xin, J., Zhang, W., Wang, Y., 2016. The empirical correlations between PM2. 5, PM10 and AOD in the Beijing metropolitan region and the PM2. 5, PM10 distributions retrieved by MODIS. Environ. Pollut. 216, 350–360.
Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., Herry, M., Horak, F., Puybonnieux-Texier, V., Quénel, P., 2000. Public-health impact of outdoor and traffic-related air pollution: a European assessment. Lancet 356, 795–801.
Lai, H.-K., Tsang, H., Wong, C.-M., 2013. Meta-analysis of adverse health effects due to air pollution in Chinese populations. BMC Publ. Health 13, 360.
Lee, H., Liu, Y., Coull, B., Schwartz, J., Koutrakis, P., 2011. A novel calibration approach of MODIS AOD data to predict PM2. 5 concentrations. Atmos. Chem. Phys. 11, 7991.
Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., Hsu, N., 2013. The Collection 6 MODIS aerosol products over land and ocean. Atmos. Meas. Tech 6, 2989–3034.
Li, S., Williams, G., Guo, Y., 2016. Health benefits from improved outdoor air quality and intervention in China. Environ. Pollut. 214, 17–25.
Li, Y., Tang, Y., Fan, Z., Zhou, H., Yang, Z., 2017. Assessment and comparison of three different air quality indices in China. Environ. Eng. Res. 23 (1), 21–27.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R. News 2, 18–22.
Lim, Y.-H., Kim, H., Kim, J.H., Bae, S., Park, H.Y., Hong, Y.-C., 2012. Air pollution and symptoms of depression in elderly adults. Environ. Health Perspect. 120, 1023.
Liu, Y., Paciorek, C.J., Koutrakis, P., 2009. Estimating regional spatial and temporal variability of PM2. 5 concentrations using satellite data, meteorology, and land use information. Environ. Health Perspect. 117, 886.
Ma, Z., Hu, X., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L., Liu, Y., 2016. Satellite-based spatiotemporal trends in PM2. 5 concentrations: China, 2004–2013. Environ. Health Perspect. 124 (2), 184.
Ma, Z.W., Hu, X.F., Huang, L., Bi, J., Liu, Y., 2014. Estimating ground-level PM2.5 in China using satellite remote sensing. Environ. Sci. Technol. 48, 7436–7444.
Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., Xue, W., Wang, J., Wang, D., Kan, H., 2016a. Estimating ground-level PM 10 in a Chinese city by combining satellite data, meteorological information and a land use regression model. Environ. Pollut. 208, 177–184.
Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., Xue, W., Wang, J., Wang, D., Kan, H., Liu, Y., 2016b. Estimating ground-level PM(10) in a Chinese city by combining satellite data, meteorological information and a land use regression model. Environ. Pollut. 208, 177–184.
Nordio, F., Kloog, I., Coull, B.A., Chudnovsky, A., Grillo, P., Bertazzi, P.A., Baccarelli, A.A., Schwartz, J., 2013. Estimating spatio-temporal resolved PM10 aerosol mass concentrations using MODIS satellite data and land use regression over Lombardy, Italy. Atmos. Environ. 74, 227–236.
Raaschou-Nielsen, O., Andersen, Z.J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M.J., Brunekreef, B., Xun, W.W., Katsouyanni, K., Dimakopoulou, K., Sommar, J., Forsberg, B., Modig, L., Oudin, A., Oftedal, B., Schwarze, P.E., Nafstad, P., De Faire, U., Pedersen, N.L., Ostenson, C.G., Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K.T., Sorensen, M., Tjonneland, A., Ellermann, T., Eeftens, M., Peeters, P.H., Meliefste, K., Wang, M., Bueno-de-Mesquita, B., Key, T.J., de Hoogh, K., Concin, H., Nagel, G., Vilier, A., Grioni, S., Krogh, V., Tsai, M.Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Trichopoulou, A., Bamia, C., Vineis, P., Hoek, G., 2013. Air pollution and lung cancer incidence in 17 european cohorts: prospective analyses from the european study of cohorts for air pollution effects (ESCAPE). Lancet Oncol. 14, 813–822.
Sayer, A., Munchak, L., Hsu, N., Levy, R., Bettenhausen, C., Jeong, M.J., 2014. MODIS Collection 6 aerosol products: comparison between Aqua's e-Deep Blue, Dark Target, and "merged" data sets, and usage recommendations. J. Geophys. Res.: Atmosphere 119.
Shang, Y., Sun, Z., Cao, J., Wang, X., Zhong, L., Bi, X., Li, H., Liu, W., Zhu, T., Huang, W., 2013. Systematic review of Chinese studies of short-term exposure to air pollution and daily mortality. Environ. Int. 54, 100–111.
Shao, P., Xin, J., An, J., Kong, L., Wang, B., Wang, J., Wang, Y., Wu, D., 2017. The empirical relationship between PM2. 5 and AOD in nanjing of the Yangtze River Delta. Atmos. Pollu. Res. 8, 233–243.
Stieb, D.M., Chen, L., Eshoul, M., Judek, S., 2012. Ambient air pollution, birth weight and preterm birth: a systematic review and meta-analysis. Environ. Res. 117, 100–111.
Stockfelt, L., Andersson, E.M., Molnár, P., Gidhagen, L., Segersson, D., Rosengren, A., Barregard, L., Sallsten, G., 2017. Long-term effects of total and source-specific particulate air pollution on incident cardiovascular disease in Gothenburg, Sweden. Environ. Res. 158, 61–71.
Sun, Y., Zhou, X., Wai, K., Yuan, Q., Xu, Z., Zhou, S., Qi, Q., Wang, W., 2013. Simultaneous measurement of particulate and gaseous pollutants in an urban city in North China Plain during the heating period: implication of source contribution. Atmos. Res. 134, 24–34.
Vallius, M.J., Ruuskanen, J., Mirme, A., Pekkanen, J., 2000. Concentrations and estimated soot content of PM1, PM2. 5, and PM10 in a subarctic urban atmosphere. Environ. Sci. Technol. 34, 1919–1925.
Wang, L., Wei, Z., Yang, J., Zhang, Y., Zhang, F., Su, J., Meng, C., Zhang, Q., 2014. The

2013 severe haze over southern Hebei, China: model evaluation, source apportionment, and policy implications. Atmos. Chem. Phys. 14, 3151–3173.
Wang, L., Xin, J., Wang, Y., Li, Z., Liu, G., Li, J., 2007. Evaluation of the MODIS aerosol optical depth retrieval over different ecosystems in China during EAST-AIRE. Atmos. Environ. 41, 7138–7149.
Wang, Y.Q., Zhang, X.Y., Sun, J.Y., Zhang, X.C., Che, H.Z., Li, Y., 2015. Spatial and temporal variations of the concentrations of PM10, PM2.5 and PM1 in China. Atmos. Chem. Phys. 15, 13585–13598.
Wu, F., Zhang, D., Cao, J., Xu, H., An, Z., 2012. Soil-derived sulfate in atmospheric dust particles at Taklimakan desert. Geophys. Res. Lett. 39.
Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., Liu, Y., 2017. Full-coverage high-resolution daily PM2. 5 estimation using MAIAC AOD in the Yangtze River Delta of China. Rem. Sens. Environ. 199, 437–446.
Yao, L., Lu, N., 2014. Spatiotemporal distribution and short-term trends of particulate matter concentration over China, 2006–2010. Environ. Sci. Pollut. Control Ser. 21, 9665–9675.
Yitshak-Sade, M., Yudovitch, D., Novack, V., Tal, A., Kloog, I., Goldbart, A., 2017. Air pollution and hospitalization forbronchiolitis among young children. Ann Am Thorac Soc 14 (12), 1796–1802.
You, W., Zang, Z., Zhang, L., Li, Y., Wang, W., 2016. Estimating national-scale ground-level PM25 concentration in China using geographically weighted regression based on MODIS and MISR AOD. Environ. Sci. Pollut. Control Ser. 23, 8327–8338.
Zhang, H., Hoff, R.M., Engel-Cox, J.A., 2009. The relation between Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth and PM2. 5 over the United States: a geographical comparison by US Environmental Protection Agency regions. J. Air Waste Manag. Assoc. 59, 1358–1369.
Zhang, J., Xin, J., Zhang, W., Wang, S., Wang, L., Xie, W., Xiao, G., Pan, H., Kong, L., 2017. Validation of MODIS C6 AOD products retrieved by the dark target method in the Beijing–Tianjin–Hebei urban agglomeration, China. Adv. Atmos. Sci. 34, 993–1002.
Zhang, Q., Xin, J., Yin, Y., Wang, L., Wang, Y., 2016a. The variations and trends of modis c5 & c6 products' errors in the recent decade over the background and urban areas of north China. Rem. Sens. 8, 754.
Zhang, T., Gong, W., Zhu, Z., Sun, K., Huang, Y., Ji, Y., 2016b. Semi-physical estimates of national-scale PM10 concentrations in China using a satellite-based geographically weighted regression model. Atmosphere 7, 88.
Zhao, S., Yu, Y., Yin, D., He, J., Liu, N., Qu, J., Xiao, J., 2016. Annual and diurnal variations of gaseous and particulate pollutants in 31 provincial capital cities based on in situ air quality monitoring data from China National Environmental Monitoring Center. Environ. Int. 86, 92–106.
Zheng, Y., Zhang, Q., Liu, Y., Geng, G., He, K., 2016. Estimating ground-level PM2. 5 concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. Atmos. Environ. 124, 232–242.