



Spatiotemporal continuous estimates of PM_{2.5} concentrations in China, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations[☆]



Tao Xue^{a,b}, Yixuan Zheng^b, Dan Tong^b, Bo Zheng^c, Xin Li^b, Tong Zhu^a, Qiang Zhang^{b,*}

^a BIC-ESAT and SKL-ESPC, College of Environmental Science and Engineering, Peking University, Beijing 100871, China

^b Department of Earth System Science, Tsinghua University, Beijing 100084, China

^c State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Handling Editor: Yong-Guan Zhu

Keywords:

Fine particulate matter
Satellite remote sensing
Aerosol optical depth
Machine learning

ABSTRACT

Ambient exposure to fine particulate matter (PM_{2.5}) is known to harm public health in China. Satellite remote sensing measurements of aerosol optical depth (AOD) were statistically associated with in-situ observations after 2013 to predict PM_{2.5} concentrations nationwide, while the lack of surface monitoring data before 2013 have created difficulties in historical PM_{2.5} exposure estimates. Hindcast approaches using statistical models or chemical transport models (CTMs) were developed to overcome this limitation, while those approaches still suffer from incomplete daily coverage due to missing AOD data or limited accuracy due to uncertainties of CTMs. Here we developed a new machine learning (ML) model with high-dimensional expansion (HD-expansion) of numerous predictors (including AOD and other satellite covariates, meteorological variables and CTM simulations). Through comprehensive characterization of the nonlinear effects of, and interactions among different predictors, the HD-expansion parameterized the association between PM_{2.5} and AOD as a nonlinear function of space and time covariates (e.g., planetary boundary layer height and relative humidity). In this way, the PM_{2.5}-AOD association can vary spatiotemporally. We trained the model with data from 2013 to 2016 and evaluated its performance using annually-iterated cross-validation, which iteratively held out the in-situ observations for a whole calendar year (as testing data) to examine the predictions from a model trained by the rest of the observations. Our estimates were found to be in good agreement with in-situ observations, with correlation coefficients (R^2) of 0.61, 0.68, and 0.75 for daily, monthly and annual averages, respectively. To interpolate the missing predictions due to incomplete AOD data, we incorporated a generalized additive model into the ML model. The two-stage estimates of PM_{2.5} sacrificed the prediction accuracy on a daily timescale ($R^2 = 0.55$), but achieved complete spatiotemporal coverage and improved the accuracy of monthly ($R^2 = 0.71$) and annual ($R^2 = 0.77$) averages. The model was then used to predict daily PM_{2.5} concentrations during 2000–2016 across China and estimate long-term trends in PM_{2.5} for the period. We found that population-weighted concentrations of PM_{2.5} significantly increased, by 2.10 (95% confidence interval (CI): 1.74, 2.46) $\mu\text{g}/\text{m}^3/\text{year}$ during 2000–2007, and rapidly decreased by 4.51 (3.12, 5.90) $\mu\text{g}/\text{m}^3/\text{year}$ during 2013–2016. In this study, we produced AOD-based estimates of historical PM_{2.5} with complete spatiotemporal coverage, which were evidenced as accurate, particularly in middle and long term. The products could support large-scale epidemiological studies and risk assessments of ambient PM_{2.5} in China and can be accessed via the website (<http://www.meicmodel.org/dataset-phd.html>).

1. Introduction

Globally, ambient exposure to fine particulate matter (PM_{2.5}, defined as the particles with a diameter < 2.5 μm) has been identified as one of the leading causes of harm to public health (Cohen et al., 2017).

Exposure-response functions that are derived from large-sample epidemiological studies (e.g., cohort studies) are essential to quantify the disease burden of PM_{2.5} (Burnett et al., 2014). Currently, in widely used exposure-response functions, the risks posed by high concentrations of PM_{2.5} remain insufficiently studied, or are approximated according to

[☆] The authors declare they have no actual or potential competing financial interests.

* Corresponding author at: Mengminwei Science and Technology Building, Tsinghua University, Haidian, Beijing 100084, China.

E-mail address: qiangzhang@tsinghua.edu.cn (Q. Zhang).

the effects of other particles (e.g., from tobacco smoking). During the first 10 years of the 21st century, China became one of the most polluted countries worldwide, in terms of PM_{2.5}, because of rapid urbanization (Cohen et al., 2017). To fight against the poor air quality, China initiated rapid reductions in anthropogenic emissions of PM_{2.5} from 2013 onwards (Zheng et al., 2017) via the China clean air action plan (China State Council, 2013). This policy-driven air quality change in China provides a quasi-experimental scenario to study the health effects of high doses of PM_{2.5}. However, the lack of historical PM_{2.5} data in China limits such studies, as few in-situ observations of PM_{2.5} were performed on a national scale before 2013 (Ma et al., 2016). Accurate estimate of historical air pollution is a key requirement for health-related studies of PM_{2.5} in China.

Satellite remote sensing of aerosol optical depth (AOD) has proven to be an effective measurement of particulate matter pollution at the surface, and has been recorded by multiple sensors, including the moderate-resolution imaging spectroradiometer (MODIS), which has been in operation since 1999, and started to officially release data since 2000 (Wang and Christopher, 2003; van Donkelaar et al., 2015). A variety of statistical approaches, including linear mixed effect (LME) regression (Ma et al., 2016), geographically weighted regression (Ma et al., 2014; van Donkelaar et al., 2016; You et al., 2016), generalized additive model (GAM) (Zou et al., 2017), statistical downscaling model (Chang et al., 2014; Lv et al., 2016), support vector machine (Hou et al., 2014), random forest (Hu et al., 2017), neural networks (Di et al., 2016; Gupta and Christopher, 2009; Zou et al., 2015), deep learning (Li et al., 2017), and other machine learning (ML) models (Jiang and Christakos, 2018; Reid et al., 2015; Zhan et al., 2017) have been utilized to associate AOD with in-situ observations of PM_{2.5}. Benefiting from the global coverage of earth observing satellites (e.g., Terra and Aqua), and the fine temporal resolution of geostationary satellites (e.g., Himawari-8), such methods can extend the spatiotemporal scope of ground monitoring networks, and have been used to evaluate ambient exposure to PM_{2.5} on national and global scales (Cohen et al., 2017; Ma et al., 2016; Zheng et al., 2017).

PM_{2.5} estimates derived from the statistical approaches usually depend on in-situ measurements or their spatiotemporal autocorrelations, such that their application is restricted to periods during which in-situ observations are available. For instance, in China, most of the AOD-based estimates were focused on PM_{2.5} variations occurring after 2013 (Ma et al., 2016). A few studies reconstructed historical concentrations of PM_{2.5} over China (Lin et al., 2018; Ma et al., 2016) using statistical approaches, which were more accurate at long timescale (e.g., annual averages) than a short period (e.g., daily or monthly averages), partly due to incomplete coverage of satellite-based AOD data. Continuous spatiotemporal coverage of satellite-based PM_{2.5} estimates could be derived with the input from global or regional chemical transport models (e.g., Boys et al., 2014; Geng et al., 2015; van Donkelaar et al., 2015), however, accuracy of those estimates are thought to be limited due to uncertainties in chemical transport models (Bravo et al., 2012).

Due to a lack of constraints from in-situ observations, statistical models that aimed to predict historical concentrations of PM_{2.5} based on other variables were less accurate than those that aimed to extend the spatiotemporal coverage of ground-surface networks through interpolation of monitoring values (Lin et al., 2018; Ma et al., 2016). Few existing statistical models were designed specifically to make historical predictions. Therefore, an advanced model for AOD with higher accuracy over short timescales is required to assess historical exposure (acute exposure in particular) to PM_{2.5} in China.

In this work, we developed a technique that we refer to as high-dimensional expansion (HD-expansion), which expands the linear terms of AOD and other covariates (e.g., meteorological variables) into a high-dimensional space to characterize complicated effects (e.g., nonlinear or interacting effects) of such predictors (Fig. 1). For example, the association between AOD and PM_{2.5} has been found to vary both spatially and temporally (Guo et al., 2017). Under HD-expansion, the coefficient

that links AOD to PM_{2.5} is parameterized as a function of multiple spatiotemporal covariates (e.g., meteorological variables and satellite nightlight) to mimic the space- and time-varying associations. Incorporating the HD-expansion terms into a ML model, elastic-net (Zou and Hastie, 2005), allowed regression of high-dimensional predictors with in-situ observations of daily PM_{2.5}. In accordance with previous studies, we evaluated the accuracy of historical PM_{2.5} estimates using annually iterated cross-validation (CV), which iteratively held out 1 calendar years' data as testing data, to examine the estimates produced from a model trained by the remaining data in other years. We trained and validated the model using daily PM_{2.5} observations from national networks across mainland China from 2013 to 2016, and reconstructed historical daily maps of PM_{2.5} at a spatial resolution of 0.1° × 0.1° from 2000 to 2016. After incorporating county-level census data from 2000 and 2010, we evaluated the spatiotemporal variations of both chronic and acute exposure to PM_{2.5} in China.

2. Materials & methods

2.1. Datasets

In this study, we used similar inputs (e.g., in-situ PM_{2.5}, satellite AOD, satellite covariates, Weather Research and Forecasting (WRF) and Community Multiscale Air Quality (CMAQ) simulations) during the studying period as employed in our previous work on PM_{2.5} estimation during 2014 (Xue et al., 2017). The sources and preparation of data are described briefly below. For details, please refer to Xue et al. (2017) and Zheng et al. (2017).

2.1.1. PM_{2.5} monitoring data

We obtained hourly observations of PM_{2.5} from multiple networks in China, i.e., the networks of the China Environmental Monitoring Center, Beijing Municipal Environmental Monitoring Center and Guangdong Environmental Monitoring Center. After excluding duplicate measurement sites, there were 1497 sites (as shown in Fig. S1) in total, distributed over the mainland of China (i.e., excluding Hong Kong, Macau, Taiwan and some islands in the South China Sea) during 2013–2016. The hourly PM_{2.5} data were converted into daily averages and then incorporated into the regression models. Due to the lack of nationwide monitoring networks of PM_{2.5} in China before 2013 (Ma et al., 2016), only the in-situ observations during the recent four years were involved into the model training procedures.

2.1.2. Satellite AOD and auxiliary variables

We obtained MODIS level 2 products of AOD at a spatial resolution of 3 km (MOD04_3K and MYD04_3K) from the earth observing satellites Terra (2000–2016) and Aqua (2002–2016), maintained by the National Aeronautics and Space Administration (NASA) (Levy et al., 2015). Satellite remote sensing technology can measure the total column concentration of aerosol from the earth surface to the top of atmosphere, and its product, AOD has been evidenced as a good predictor for ground surface PM_{2.5} (Geng et al., 2015). From the raw AOD products, this study extracted the “best estimate” retrievals, which passed quality assurance tests. Satellite auxiliary variables including NO₂ column density, normalized difference vegetation index (NDVI), and nightlight data, can act as surrogates for emissions, such as those from transportation (NO₂ and nightlight), power plant (NO₂) and residential usage (NDVI and nightlight) (Ma et al., 2014; Zhang and Hu, 2017; Zheng et al., 2016). Ozone Measurement Instrument level 2 products of NO₂ (Krotkov et al., 2018) were produced by satellite Aura (2004–2016); MODIS level 3 products of NDVI (Didan, 2015) were collected by Terra (2000–2016); and 2013 annual nightlight data (National Geophysical Data Center, 2018), at a spatial resolution of 1 km, were produced from the visible and infrared sensors of the Defense Meteorological Satellite Program, and were distributed by the National Centers for Environmental Information (NCEI). All satellite data, except nightlight data,

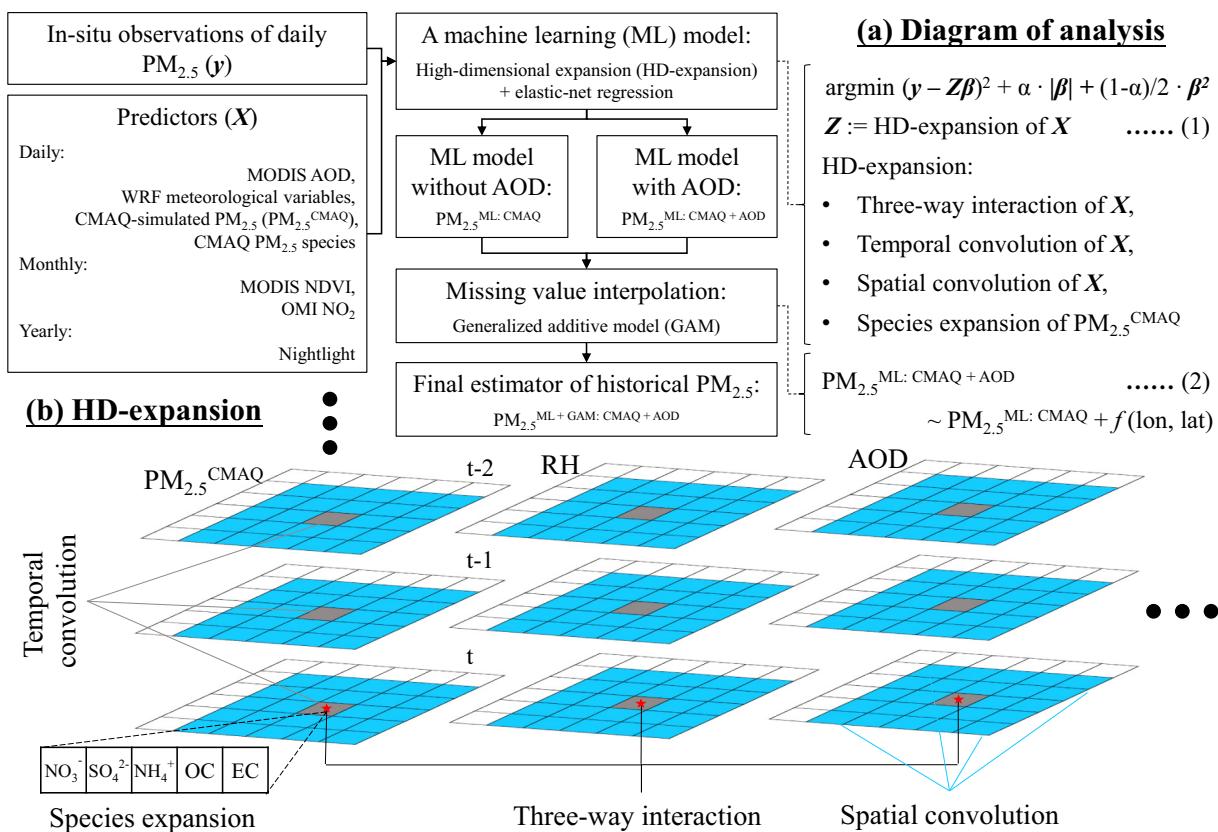


Fig. 1. Diagram of the statistical analysis (a) and illustration of the high-dimensional expansion (b).

were downloaded from <https://search.earthdata.nasa.gov>. Nightlight data were downloaded from the NCEI website (<https://ngdc.noaa.gov/eog/dmsp.html>).

2.1.3. WRF and CMAQ simulations

For the period 2000–2016, we simulated maps of meteorological variables including temperature, wind speed and direction, relative humidity, pressure, and planet boundary layer height using the WRF model (ver. 3.5.1) (Skamarock et al., 2008), which utilized the National Centers for Environmental Prediction Final Analysis (NCEP-FNL) reanalysis data as initial and boundary conditions. Driven by the outputs of the WRF model, we also simulated concentrations of PM_{2.5} and their five major components, i.e., NO₃⁻, SO₄²⁻, elemental carbon (EC), organic carbon (OC) and NH₄⁺, based on the 2000–2016 inventories from the Multi-resolution Emission Inventory of China model (<http://meicmodel.org/>) and using the CMAQ model (ver. 5.1) (US EPA, 2015). The WRF-CMAQ model system mimics the physical and chemical processes of multiple air pollutants, including their emissions, dispersions, transports, atmospheric chemical reactions, and depositions, using the technology of computer simulation. Therefore, the WRF simulates acted as the predictors for the climate field that affected PM_{2.5} concentrations, and the CMAQ simulates provided prior knowledge for the spatiotemporal variations of PM_{2.5}. The original horizontal resolution of both the WRF and CMAQ simulations was 36 km × 36 km. The hourly numerical outputs were first averaged into daily means by pixel. For further details of the WRF and CMAQ model settings, please refer to our previous publications (Xue et al., 2017; Zheng et al., 2017), which applied the same model settings.

2.2. Data preparation

To match all the inputs spatially, we first averaged (for AOD, NDVI and nightlight), resampled (for NO_2) or downscaled (for WRF and

CMAQ outputs) the raw data into a gridded map of $0.1^\circ \times 0.1^\circ$ resolution over China. The downscaling method was based on an inverse distance weighted average. In the previous study (Xue et al., 2017), we have evidenced that the downscaled CMAQ-simulated PM_{2.5} ($0.1^\circ \times 0.1^\circ$) is slightly better correlated with the in-situ observations, compared to the raw CMAQ output (36 km \times 36 km). The in-situ observations of PM_{2.5} were also assigned to each pixel of the grid for matching with the inputs. For the monitors that located in the same pixel, we further calculated average of their in-situ observations in each day. In terms of the temporal resolution, nightlight data were annual values; satellite NO₂ and NDVI data were in monthly means to increase the spatial coverage; all other input values were daily averages. After spatial resampling and temporal aggregation, there were still a few missing monthly NDVI and NO₂ values, which were interpolated using a Kriging method. After data preparation, all inputs had complete spatiotemporal coverage, except AOD. Aqua AOD data were first calibrated using Terra AOD with a linear regression model (i.e. Terra AOD $\sim \beta \times$ Aqua AOD + α , where β and α denote the regression coefficients derived from the co-located AOD data, and then were utilized to calibrate all Aqua AOD), and then a combined AOD product with improved spatial coverage (as shown in Fig. S2), was derived for regression analysis. NO₂ before 2004, and a proportion of the NDVI data for 2000, were not available and thus were replaced by other values in the same month (of 2004–2005 and 2000–2001 for NO₂ and NDVI, respectively). Monthly or yearly variable that was utilized as a temporally-constant predictor within the corresponding month or year, can capture the spatial patterns and long-term trends of PM_{2.5}.

2.3. Statistical analysis

The entire data analysis procedure is shown in Fig. 1. Briefly, to derive the final estimates of historical PM_{2.5}, we designed a two-stage method, and applied it to the whole study domain. In stage I, we

developed two separate ML models: with satellite AOD (ML: CMAQ + AOD) and without satellite AOD (ML: CMAQ). Estimates from the former model ($\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$) were more accurate, but the latter ($\text{PM}_{2.5}^{\text{ML: CMAQ}}$) had complete spatiotemporal coverage. In stage II, to interpolate the missing values of $\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$, we developed a generalized additive model (GAM) with an offset of $\text{PM}_{2.5}^{\text{ML: CMAQ}}$. The details of the statistical modeling for each stage are presented below.

2.3.1. Stage 1: ML model

In general, the associations between $\text{PM}_{2.5}$ and its predictors are known to be more complex than simple linear functions. Previous studies have utilized the spline expansion (also known as a GAM) to model the nonlinear association between $\text{PM}_{2.5}$ and AOD or other covariates (Zou et al., 2017). However, such methods may still remain too simplistic to properly characterize the associations. For instance, interactions between different predictors (Guo et al., 2017) are not covered by the technique, because a GAM usually assumes that the nonlinear effects of any two predictors are additive. To fully characterize the complexity involved in modeling $\text{PM}_{2.5}$, we developed an approach, referred to as HD-expansion (Eq. 1 in Fig. 1), which expanded the predictors in an analogous manner to a GAM. In this study, we used four types of expansion. First, we considered three-way interaction terms, using a predictor set of three variables ($x_i, i = 1, 2, 3$) as an example. Under the three-way interaction, the predictors are expanded as a set of $[x_1^3, x_1^2x_2, x_1^2x_3, x_1x_2^2, x_1x_3^2, x_1x_2x_3, x_2^3, x_2^2x_3, x_2x_3^2, x^3, x_1^2, x_1x_2, x_1x_3, x_2^2, x_2x_3, x_3^2, x_1, x_2, x_3]$. In this way, we captured the variance in, and nonlinearity of, the association between $\text{PM}_{2.5}$ and AOD. Putting all AOD-based predictors together, the regression equation is given by $y \sim \beta_1 \text{AOD} + (\beta_2 \text{RH}^2 + \beta_3 \text{PBL}^2 + \beta_4 \text{RH} \cdot \text{PBL} + \dots + \beta_{k+1} \text{RH} + \beta_{k+2} \text{PBL} + \dots) \text{AOD} + (\beta_{q+1} \text{RH} + \beta_{q+2} \text{PBL} + \dots) \text{AOD}^2 + \beta_{l+1} \text{AOD}^3 + \dots$, where k, q , or l denotes the number of coefficients located in the equation before the term β_{k+1}, β_{q+1} or β_{l+1} , respectively. The above equation illustrates the three-way interactions of AOD, RH and PBL, and the omitted part include interactions between AOD and the other pairs of predictors (i.e., CMAQ-WRF simulates, NO_2 , NDVI and night-light), which are parameterized analogously as a large number of linear terms. In this way, we parameterized the correlation between $\text{PM}_{2.5}$ and AOD as a nonlinear function of other spatiotemporal covariates, when linking all predictors (including AOD) to the in-situ observations. Second, we considered spatial convolution for all of the predictors. Convolution is equivalent to the weighted average of the nearby values of a predictor, and has been widely used to improve accuracy in neural networks or other advanced algorithms used to develop AOD-based estimators of $\text{PM}_{2.5}$ (Di et al., 2016). Instead of using kernel-based averages, in this study we used direct values of the site-collocated pixel and its nearest 24 neighbors, and let the regression model determine the weights of different pixels for all the predictors except for AOD, using the elastic-net regularization (Zou and Hastie, 2005) as described below. Considering the missing values of AOD, we separately derived the averages for the nearest 8 or 24 pixels as the spatial convolution predictors of AOD. Third, analogous to spatial convolution, we also applied temporal convolution by incorporating the measurements obtained 1 or 2 days previously, for all the predictors on a daily timescale. Finally, as $\text{PM}_{2.5}$ is a mixture of different species, its chemical composition may also influence the statistical association between in-situ observations and AOD. To incorporate such variability, we replaced CMAQ-simulated $\text{PM}_{2.5}$ by its five major components (i.e., NO_3^- , SO_4^{2-} , EC, OC and NH_4^+); we refer to this approach as species expansion. Since the other $\text{PM}_{2.5}$ components (e.g., dust) than the five species can be derived by a linear combination of the six variables ($\text{PM}_{2.5}$, NO_3^- , SO_4^{2-} , EC, OC and NH_4^+), and thus should not further involved into the model. After HD-expansion, > 1000 linear terms were generated. These linear terms may be highly correlated with each other, and some of them can be less predictive for $\text{PM}_{2.5}$. To deal with collinearity and redundancy simultaneously, we used the elastic-net

regression model (Zou and Hastie, 2005), which has been widely applied to the analysis of high-dimensional data. Through combining L-1 and L-2 regularization, the elastic-net model automatically selected the most effective predictors among many highly correlated variables. Therefore, the model is capable to regularize the HD-expansion terms, which contained information overlapped with each other.

In stage I, we developed two ML models, which were identical to each other in all aspects of model settings, except for involving AOD or not. The ML model with AOD produced more accurate estimates of $\text{PM}_{2.5}$ ($\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}, N = 407,049$). The model involved 1075 HD-expansion terms. According to the results of elastic-net regularization (i.e., absolute value of regression coefficients), the most predictive variable in the ML was the linear term of AOD, followed by the linear term of CMAQ-simulated $\text{PM}_{2.5}$, WRF-simulated variables (temperature, relative humidity, pressure, and wind), satellite NO_2 , nightlight, and the quadratic term of NDVI. However, the model's predictions depend on the existence of AOD measurements and are thus incomplete. To estimate $\text{PM}_{2.5}$ at the spatiotemporal coordinates for which AOD was missing, we developed another ML model without AOD ($\text{PM}_{2.5}^{\text{ML: CMAQ}}, N = 1,347,457$). In this way, we sacrificed model accuracy but achieved complete spatiotemporal coverage. The ML models were trained by in-situ observations of $\text{PM}_{2.5}$ and its predictors during 2013–2016, and were used to generate historical estimates (i.e., $\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$ and $\text{PM}_{2.5}^{\text{ML: CMAQ}}$) based on the predictors during 2000–2016.

2.3.2. Stage 2: GAM

To further decrease modeling errors in the complete estimator ($\text{PM}_{2.5}^{\text{ML: CMAQ}}$), we merged the estimates from the two ML models using a GAM. For each day, the differences between $\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$ and $\text{PM}_{2.5}^{\text{ML: CMAQ}}$ were modeled as a function of spatial coordinates (Eq. 2 in Fig. 1), such that differences at the coordinates without AOD could be interpolated and then added back to $\text{PM}_{2.5}^{\text{ML: CMAQ}}$. Therefore, the final estimator of the two-stage approach is denoted by $\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$.

2.4. Exposure and trend analysis

To calculate the long- and short-term exposure to $\text{PM}_{2.5}$ in China, we first derived county-level means of daily $\text{PM}_{2.5}$ using the area-weighted average approach. Combining the county-level $\text{PM}_{2.5}$ with detailed population data (i.e., number of residents) from the censuses of 2000 and 2010 (National Bureau of Statistics of China, 2003; National Bureau of Statistics of China, 2012), we calculated the population-weighted statistics on a national scale and for specific regions. To incorporate demographic changes, we linearly interpolated the population data for each county during 2000–2010. For 2011 and the years thereafter, we assumed that the populations were stable and did not account for any demographic changes. Considering the large uncertainty in the daily estimates, exposure statistics were only presented in monthly or yearly scale.

We also analyzed the trends in $\text{PM}_{2.5}$ during 2000–2016 based on the final estimator of the two-stage approach ($\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$). Because of the relative large errors in daily estimates, the trend analysis was based on monthly $\text{PM}_{2.5}$. To remove seasonality, we first derived the $\text{PM}_{2.5}$ anomalies by subtracting the long-term averages in the same month of different years from the monthly values, and then calculated the linear trend using the least-squares approach, as in the previous study (Ma et al., 2016).

To explore whether $\text{PM}_{2.5}$ exposures and their trends are sensitive to the demographic dynamics, we utilized alternative population estimations, and re-calculated the related results. According to the exploratory analysis (Table S1 and Fig. S3), the exposure estimates were not sensitive to the demographic settings, which is consistent to the previous worldwide study (Apte et al., 2015).

2.5. Model validation

The state-of-the-art approach to evaluate model performance for the historical prediction of $\text{PM}_{2.5}$, based on satellite AOD, is annually iterated CV (Liang et al., 2018; Ma et al., 2016). This approach examines how the model, trained by the data in one period, predicts the $\text{PM}_{2.5}$ in another period. In our annually iterated CV, we first non-randomly divided the in-situ $\text{PM}_{2.5}$ observations during 2013–2016 into four folds by the calendar year, and then used one fold to test the estimates of a model trained by the other three folds. To further explore the model performance over the historical period, we developed an offset validation using data from the $\text{PM}_{2.5}$ monitoring sites maintained by the US embassies in Beijing (2008–2016), Chengdu (2012–2016), Guangzhou (2011–2016), Shanghai (2011–2016) and Shenyang (2013–2016). Although the offset validation was less representative due to its limited spatial coverage, the US sites held the best available data (in terms of temporal coverage, between-site comparability, and routine maintenance) for the historical time series in our study domain.

Long-term trends derived from historical estimates of $\text{PM}_{2.5}$ may have large uncertainties and should be further validated. Considering that the historical $\text{PM}_{2.5}$ trends (2000–2016) were estimated from a model trained by a small set of recent observations (2013–2016), we reproduced the two-stage estimates for the period 2013–2015 using observed $\text{PM}_{2.5}$ concentrations and other inputs from 2016 alone, and compared the trends (η_0) (2013–2015) derived from in-situ observations to those derived from the collocated estimates (η_1). To statistically quantify the consistency between the two trends, we first calculated the difference between them ($D = \eta_0 - \eta_1$) and its variance [$V = \text{var.}(\eta_0 - \eta_1)$], based on the least-squares trend estimators (and their standard errors) for each pair of time series data located at each monitoring site, and calculated the root mean square difference (RMSD) as follows: $\text{RMSD} = [\text{mean}(D^2 + V)]^{1/2}$. This equation combines the biasness (D^2) and variance (V) of the estimate-based trends (η_1), when using the observation-based trends as reference standards. We also examined the hypothesis that there was no difference between the two trends ($D = 0$) using Wald tests, and calculated the rate of rejection (RoR) of the null hypothesis at different significance levels. RoR values are given by the fraction of pairs of trends detected as statistically different. Smaller RoR values reflect better agreement between two trends.

All data analysis was performed in R software (R Core Team, 2017), and statistical inference of the ML models was achieved using the R package *glmnet* (Friedman et al., 2010). The final estimates are accessed via the website (<http://www.meicmodel.org/dataset-phd.html>).

3. Results

3.1. Model validation results

Fig. 2 shows the results of the annually iterated CV for the historical estimates of daily $\text{PM}_{2.5}$, as produced by the original CMAQ simulation ($\text{PM}_{2.5}^{\text{CMAQ}}$) and the ML approaches without AOD ($\text{PM}_{2.5}^{\text{ML: CMAQ}}$) and with AOD ($\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$), at the spatiotemporal coordinates for which AOD data were available. The $\text{PM}_{2.5}$ estimates from the full ML model ($\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$) were in good agreement with the in-situ observations (CV correlation coefficient, $R^2 = 0.61$), with a root mean square error (RMSE) of $27.8 \mu\text{g/m}^3$ that accounted for 47% of the mean $\text{PM}_{2.5}$ observations (defined as the relative prediction error, RPE) and 63% of the standard deviation of the $\text{PM}_{2.5}$ observations (defined as the normalized RMSE, NRMSE). The estimator was also shown to be unbiased (mean bias = $2.02 \mu\text{g/m}^3$), but slightly over-smoothed (the slope of the regression of predictions against the observations was 0.61). The ML-based estimates (i.e., $\text{PM}_{2.5}^{\text{ML: CMAQ}}$ and $\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$) were shown to be more accurate than the CMAQ simulations ($R^2 = 0.38$ for $\text{PM}_{2.5}^{\text{CMAQ}}$), but removing AOD from the predictors decreased the model performance considerably ($R^2 = 0.53$ for $\text{PM}_{2.5}^{\text{ML: CMAQ}}$).

Detailed CV results for the monthly and yearly averages are given in the supplemental materials (Fig. S4). Averaging over time decreased the modeling errors. For $\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$, the CV R^2 was 0.68 and 0.75 for monthly and yearly averages, respectively.

Fig. 3 presents the results of annually iterated CV for the historical estimates of daily $\text{PM}_{2.5}$, produced by $\text{PM}_{2.5}^{\text{CMAQ}}$, $\text{PM}_{2.5}^{\text{ML: CMAQ}}$, and the two-stage-estimates of $\text{PM}_{2.5}$ ($\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$), at all spatiotemporal coordinates (i.e., for those having or missing AOD values). For $\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$, the daily CV had an R^2 of 0.55 and a RMSE of $30.2 \mu\text{g/m}^3$, demonstrating moderate accuracy that was lower than $\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$ ($R^2 = 0.61$) but slightly higher than $\text{PM}_{2.5}^{\text{ML: CMAQ}}$ ($R^2 = 0.53$). Although the two-stage estimator had the same inputs as the one-stage estimator, its large error might be caused by the uncertainty introduced by the GAM interpolation and more accumulation of the uncertainties of the inputs (e.g., the WRF-CMAQ simulations). The corresponding CV results for the monthly and yearly averages are shown in the supplemental materials (Fig. S5). Because of the complete spatiotemporal coverage, the monthly and yearly averages of $\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$ included more samples than those of $\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$, which partially explains why the average $\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$ (monthly $R^2 = 0.71$, yearly $R^2 = 0.77$) was more accurate than that of $\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$ (monthly $R^2 = 0.68$, yearly $R^2 = 0.75$).

Fig. 4 presents the evaluation results of the final estimates, $\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$, using the offset from the US embassy monitors. The results of the offset validation were comparable to the results of the CV. The R^2 varied from 0.51 (Guangzhou) to 0.67 (Beijing) for different cities, with an overall R^2 of 0.66 for the combined data of the five sites. In addition, the combined RPE was 48%, which ranged from 40% (Shanghai) to 49% (Beijing). Between cities, the variation in offset validation values suggests a geographic heterogeneity in the modeling error of $\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$, which depends on the spatially varying features of the predictors, such as the emission inventory accuracy or the fraction of missing AOD data. The offset validation results are also presented as time series data in the supplemental materials (Fig. S6). The time series indicates that $\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$ accurately predicted the temporal variation in $\text{PM}_{2.5}$ over long timescales, but only partially captured extreme $\text{PM}_{2.5}$ pollution episodes over short timescales.

Additionally, we contrasted the fitted results from the one-stage ($\text{PM}_{2.5}^{\text{ML: CMAQ+AOD}}$) or two-stage estimator ($\text{PM}_{2.5}^{\text{ML+GAM: CMAQ+AOD}}$) with the in-situ observations (Fig. S7), and visualized the comparisons of their long-term averages (Fig. S1). Roughly speaking, the performance of the fitted values were very similar to that of the cross-validated values (Figs. 2–3), which suggests no over-fitness in the ML models.

3.2. Estimated spatiotemporal patterns

Fig. 5 presents the spatiotemporal patterns of annual $\text{PM}_{2.5}$, estimated by the two-stage ML approach during 2000–2016. In 2000, the North China Plain was the most $\text{PM}_{2.5}$ -polluted region. Other $\text{PM}_{2.5}$ hotspots included the Gobi and Taklamakan Deserts, Sichuan Basin, Northeast China Plain and Pearl River Delta (PRD). The historical estimates exhibited a similar pattern of long-term exposure to $\text{PM}_{2.5}$ during 2000–2016. In most of the hotspots, except the PRD, residents were consistently exposed to higher $\text{PM}_{2.5}$ loadings ($> 60 \mu\text{g/m}^3$) than in other regions. For the most populous regions, particularly the three metropolitan areas of the Beijing-Tianjin-Hebei (BTH) region, Yangtze River Delta (YRD), and PRD, we also explored the spatiotemporal variations in $\text{PM}_{2.5}$ during 2000–2016, as shown in Fig. 5. Among these three key areas, the BTH region was the most polluted, followed by the YRD and PRD, where the population-weighted average $\text{PM}_{2.5}$ concentrations during 2000–2016 were $89.0 \mu\text{g/m}^3$, $60.7 \mu\text{g/m}^3$ and $49.5 \mu\text{g/m}^3$, respectively. The long-term level of exposure to $\text{PM}_{2.5}$ in the BTH region was considerably above the national average of

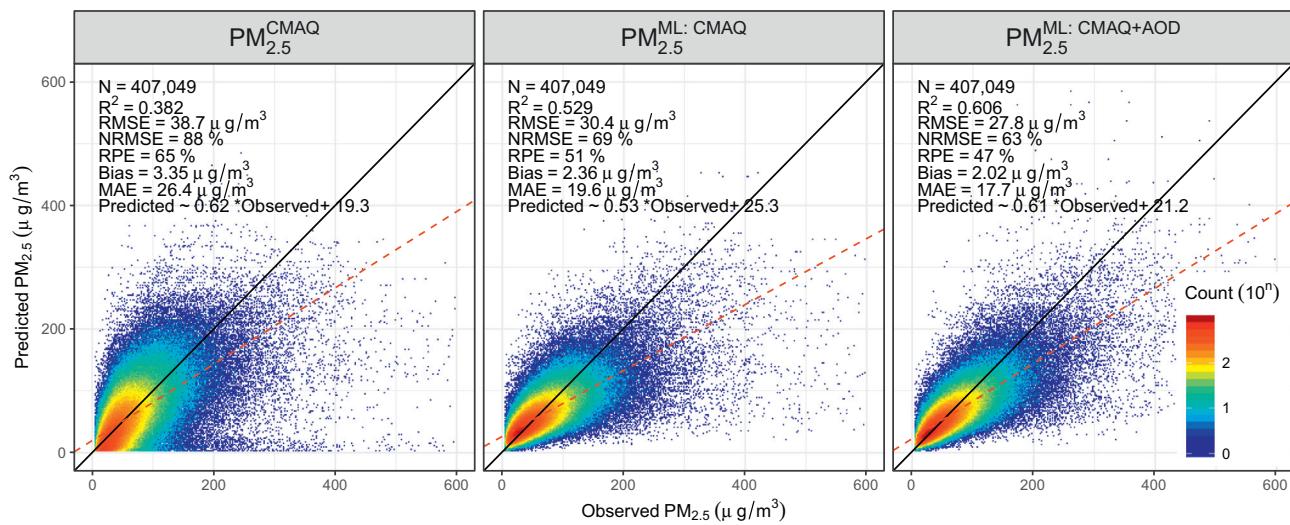


Fig. 2. Cross-validation results for different estimators on daily scale at the spatiotemporal coordinates for which aerosol optical depth (AOD) data were available.

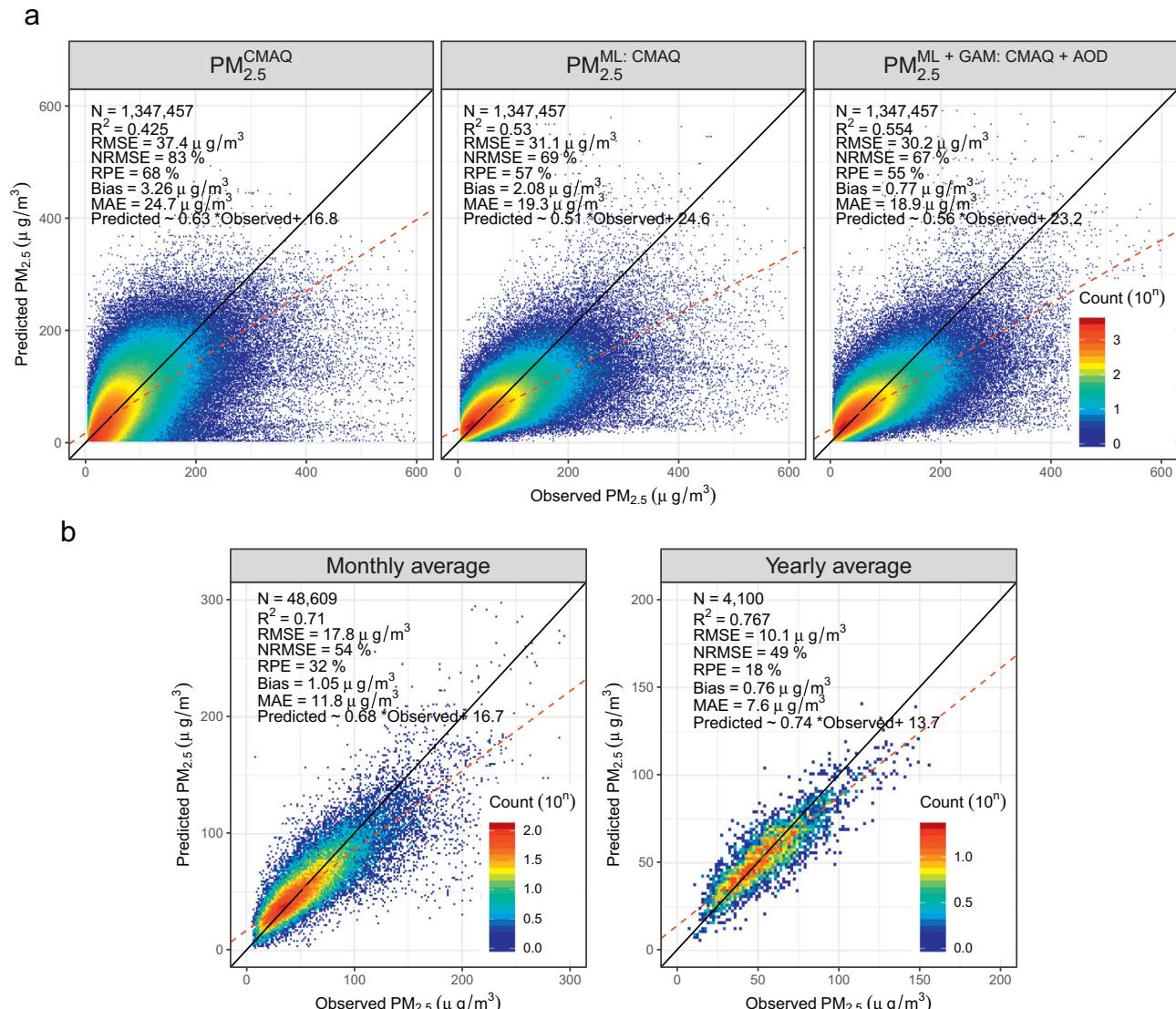


Fig. 3. Cross-validation results for different estimators at all spatiotemporal coordinates, including data for which AOD values were missing: (a) Performance of different estimators on a daily timescale; (b) Performance of the two-stage estimator on monthly and yearly timescales.

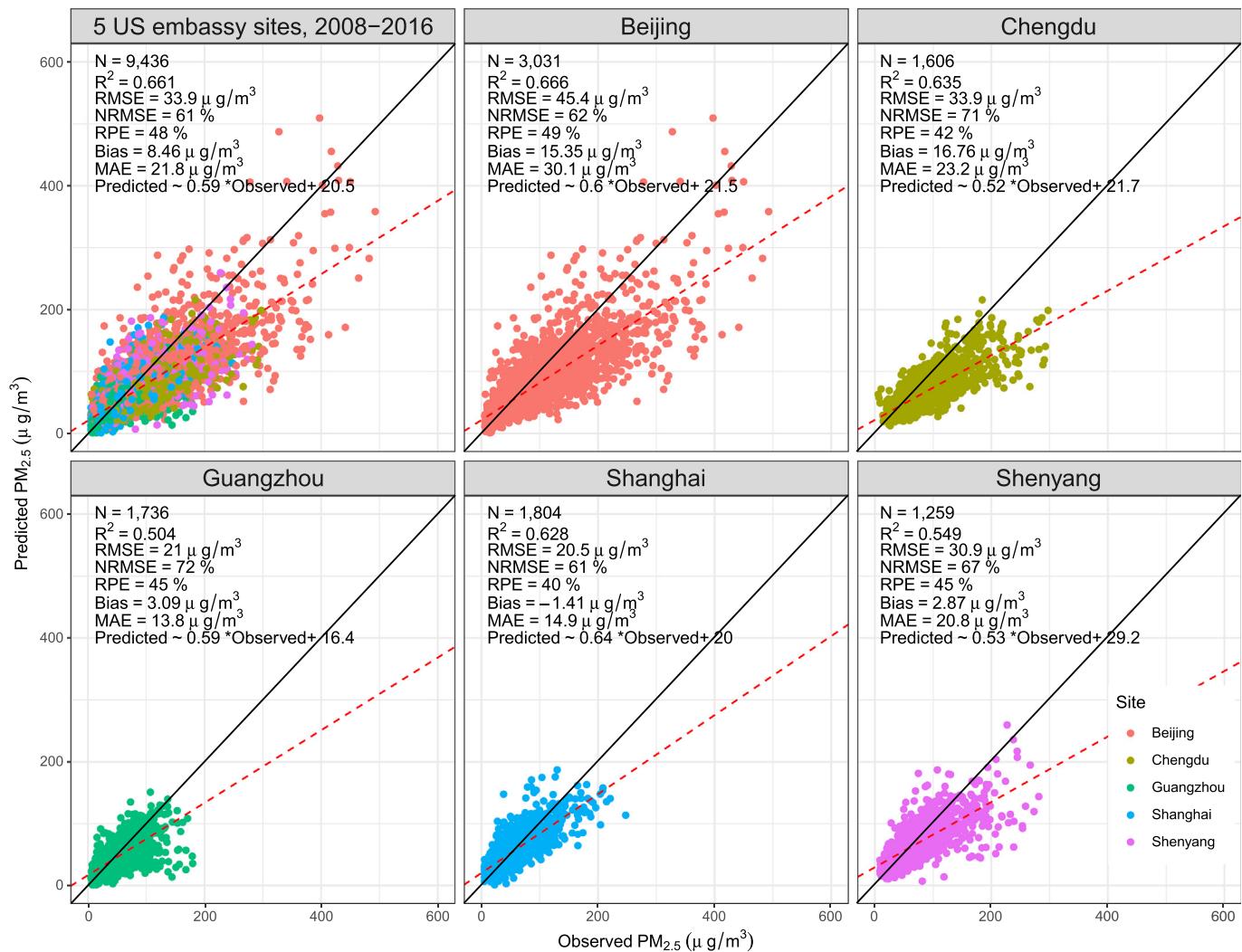


Fig. 4. Offset validation for the historical estimates of fine particulate matter (PM_{2.5}) (PM_{2.5}^{ML + GAM: CMAQ + AOD}).

59.6 µg/m³. In terms of annual average PM_{2.5}, the BTH region was most polluted in 2007 (105.3 µg/m³), followed by 2006 (103.2 µg/m³) and 2013 (100.9 µg/m³); YRD was most polluted in 2011 (67.1 µg/m³), followed by 2007 (67.0 µg/m³) and 2008 (65.6 µg/m³); PRD was most polluted in 2007 (62.9 µg/m³), followed by 2006 (59.9 µg/m³) and 2005 (59.1 µg/m³).

3.3. Long-term trends of PM_{2.5} in China

To further explore the temporal variations in PM_{2.5} concentrations, we generated time series of the population-weighted averages of monthly PM_{2.5} for the three key areas of BTH, YRD and PRD, as presented in Fig. 6. Clear seasonal patterns were found in monthly PM_{2.5}, in that levels were higher in cooler seasons and lower in warmer ones. The temporal trends in PM_{2.5} were found to be non-monotonous, and distributed heterogeneously across locations. Fig. 6 also shows the probability distribution of exposure to different categories of air quality (i.e., excellent to severely polluted) for each month, which exhibited a similar temporal profile to PM_{2.5} concentrations. For instance, during 2000–2016, there were 1495 (24.1% of the 17 years' period) polluted days and 179 (2.9%) heavily-or-above polluted days nationwide, in terms of population-weighted averages.

To statistically characterize the long-term trends in population-weighted average monthly PM_{2.5} concentrations, nationwide and for the three key areas, we calculated PM_{2.5} anomalies and derived linear

slopes for the periods 2000–2016, 2000–2007, 2008–2012 and 2013–2016, as presented in Fig. 7. According to these results, due to the rapid economic growth and urbanization that occurred during 2000–2007, the population-weighted exposure to PM_{2.5} increased at a rate of 3.78 (95% confidence interval (CI): 2.66, 4.91) µg/m³/year in the BTH region, 2.27 (1.57, 2.96) µg/m³/year in the PRD, 1.76 (1.30, 2.22) µg/m³/year in the YRD, and 2.10 (1.74, 2.46) µg/m³/year nationwide. Driven by emissions-reduction policies, such as the China Clean Air Act, ambient PM_{2.5} levels decreased from 2013 by 8.68 (4.17, 13.20) µg/m³/year, 4.91 (3.08, 6.74) µg/m³/year, 4.51 (2.78, 6.25) µg/m³/year, and 4.51 (3.12, 5.90) µg/m³/year in the BTH region, PRD, YRD and nationwide, respectively. During 2008–2012, no significant trends of PM_{2.5} were observed nationwide, and in the BTH region and YRD. In the PRD, PM_{2.5} began to decrease after 2007, by 2.30 (0.77, 3.83) µg/m³/year for the period 2007–2012. The plateau in PM_{2.5} concentrations during 2008–2012 may have been caused by the counteracting effects of the growth in industrial emissions (Klimont et al., 2013; Lu et al., 2011), and to the decreased power-plant emissions caused by installation of flue gas desulfurization devices (Xu, 2011). A gridded map of the estimated linear trends is presented in Fig. S8.

Based on in-situ observations of PM_{2.5} in 2016, the validation results of different methods for estimated PM_{2.5} concentration trends during 2013–2015 were documented in supplemental Table S2. The final estimator (PM_{2.5}^{ML + GAM: CMAQ + AOD}) was correlated with the in-situ observations, with an R² of 0.53 and 0.70 for daily and monthly averages,

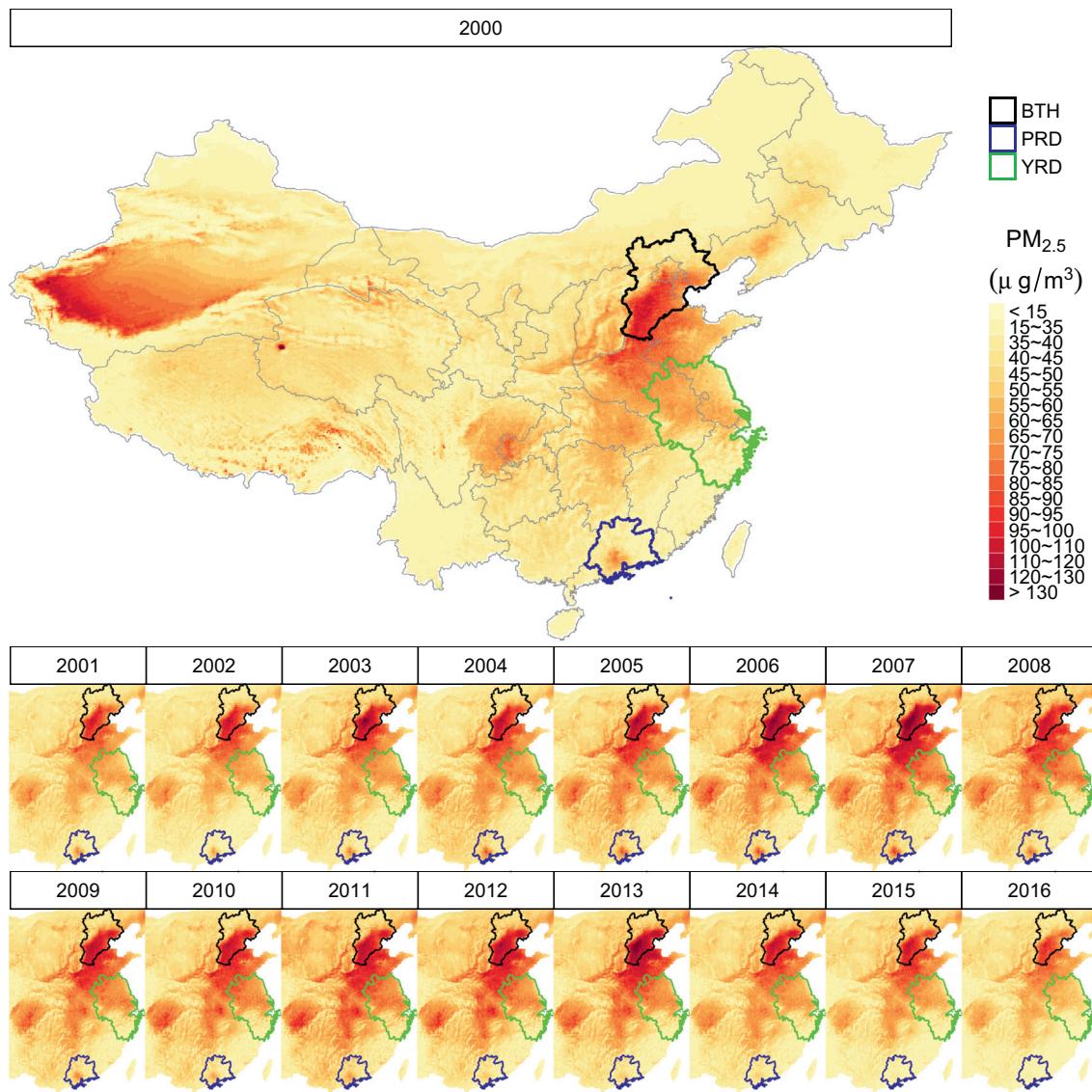


Fig. 5. Spatiotemporal patterns of PM_{2.5} estimated over Eastern China during 2000–2016.

respectively. The mean difference (D) between the least-square trend of in-situ observations and that of PM_{2.5} estimates was 0.16 $\mu\text{g}/\text{m}^3/\text{year}$, with a mean standard deviation ($V^{1/2}$) of 1.42 $\mu\text{g}/\text{m}^3/\text{year}$ and a RMSD of 2.00 $\mu\text{g}/\text{m}^3/\text{year}$ for PM_{2.5}^{ML + GAM: CMAQ + AOD}. The RoR displayed good agreement between PM_{2.5}^{ML + GAM: CMAQ + AOD} and observations of PM_{2.5}. At a significance level of 0.05, only 1% of all pairs of least-square trends were found to be statistically different. A comparison of the different estimators suggests that they performed equally well in terms of evaluating trends in historical PM_{2.5}. We further examined the consistency between the estimated trends of PM_{2.5} during 2000–2016 for the three megacity regions and nationwide, using different methods. Details of these results are presented in Fig. S9, and the national trends are briefly summarized in Fig. S10. To explore how CMAQ simulations affected the estimated trends, we also included another estimator (PM_{2.5}^{ML: AOD}), which was generated by the ML model with AOD data, other satellite covariates and meteorological variables only. We found no significant differences in the trends estimated by the different models.

4. Discussion

Reconstructing the historical time series of PM_{2.5} from satellite AOD

is difficult, because in the existing models, the PM_{2.5}-AOD association cannot be estimated accurately at different spatiotemporal coordinates without monitoring data. The PM_{2.5}-AOD association has been found to vary both temporally and spatially, and is too complex to characterize using a linear model with a restricted number of covariates. Therefore, in previous statistical models that estimated ground surface PM_{2.5} from AOD measurements (e.g., LME), the association was fitted locally at specific coordinates in the space and time dimensions, and then applied to PM_{2.5} predictions at neighboring coordinates. The local fit approach is not appropriate for estimating historical PM_{2.5} from AOD. To overcome this difficulty, Ma et al. (2014) assumed that the PM_{2.5}-AOD relationship was identical for the same days between different years, and estimated historical PM_{2.5} in China during 2004–2014 for the first time. The model performed well over long timescales (monthly and yearly), but not for daily averages; this limited the usage of AOD for the assessment of short-term exposure to PM_{2.5}. The logic underlying the assumption of Ma et al. (2014) is that hidden factors (e.g., meteorological fields) that can physically change the PM_{2.5}-AOD association tend to be similar on the same calendar day of different years, because of the general periodic patterns in atmospheric systems. Instead of assuming between-year similarity, we characterized the hidden factors that determined the PM_{2.5}-AOD relationship using HD-expansion of a

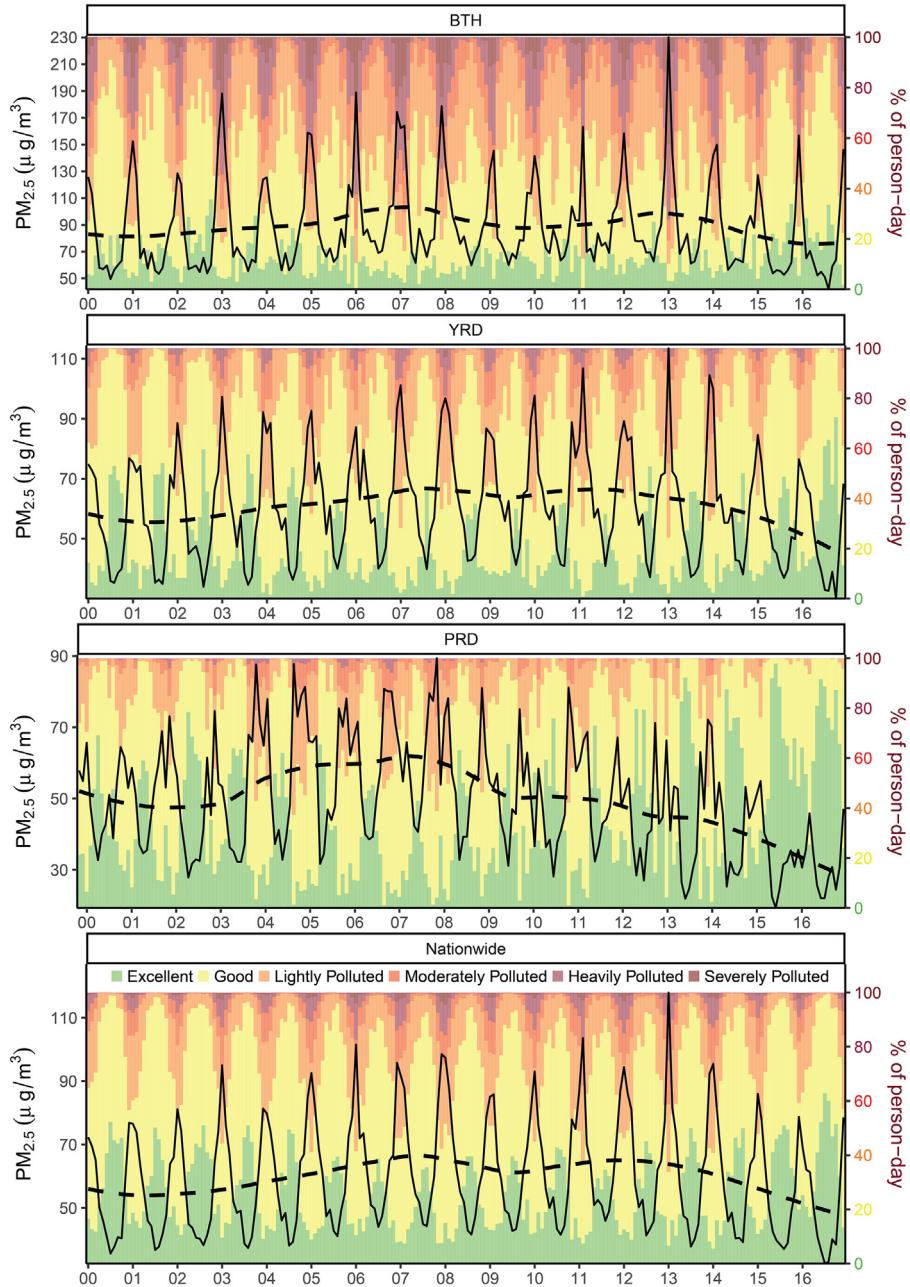


Fig. 6. Time series of population-weighted average monthly $\text{PM}_{2.5}$ concentrations and their smoothed trends (dashed lines) for the period 2000–2016. The colored bars (right axis) represent the probability distribution of exposure to different categories of air quality in each month.

few spatiotemporal and spatial covariates. Through modeling the higher order interactions between AOD and other covariates in a regression model, the $\text{PM}_{2.5}$ -AOD association was parameterized as a nonlinear function of such covariates. Driven by the spatiotemporal variations in the covariates, the $\text{PM}_{2.5}$ -AOD association in our ML model was changed for different coordinates and different periods. To illustrate how the HD-expansion improved the model performance, we developed a batch of alternative models and evaluated them by annually iterated CV (Fig. S11). The results showed that the set of three-way interaction terms, i.e., the technique used to model the varying association between AOD and $\text{PM}_{2.5}$, played a key role in reducing the modeling error.

Both satellite AOD measurements and CMAQ simulations have been applied independently to extend the spatiotemporal coverage of in-situ observations of $\text{PM}_{2.5}$. Historical estimates combined both of these approaches, similar to our previous study, which indicated that CMAQ-

simulated $\text{PM}_{2.5}$ was less predictive than AOD. In the current study, we compared models without AOD or CMAQ simulations to the full model, which consistently showed that AOD played a more significant role in achieving low modeling errors, especially with respect to the variance of errors (Fig. S11). This also explains why the historical estimates of daily $\text{PM}_{2.5}$ at the spatiotemporal coordinates where AOD data were available were more accurate than those at the coordinates where AOD values were missing. However, the advantages of the CMAQ simulations should not be overlooked: incorporating CMAQ outputs into a regression with AOD can further reduce the modeling error (Fig. S11), and CMAQ simulations provide prior knowledge on chemical species and the distribution of $\text{PM}_{2.5}$ with complete spatiotemporal coverage. Although the direct outputs of CMAQ ($\text{PM}_{2.5}^{\text{CMAQ}}$) had low accuracy, calibrating them by the ML model without AOD ($\text{PM}_{2.5}^{\text{ML-CMAQ}}$) didn't change the complete spatiotemporal coverage, and could moderately improve the model performances (daily $R^2 = 0.43$ vs. 0.53, Fig. 3a).

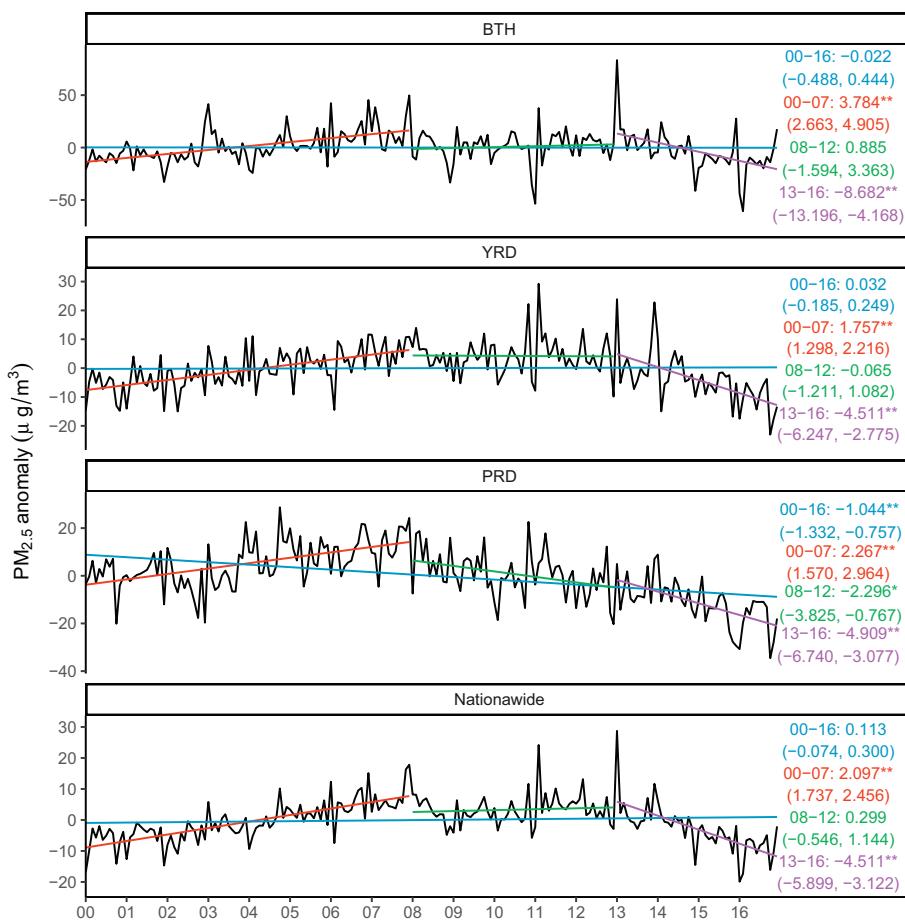


Fig. 7. Linear trends (with 95% confidence intervals) in estimated PM_{2.5} anomalies during 2000–2016 in three megacity regions of China, and nationwide. The *P*-value for the null hypothesis that there is no trend in PM_{2.5} anomalies is presented as * (0.001 < *P* < 0.05) or ** (*P* < 0.001).

particularly in medium term (monthly $R^2 = 0.54$ vs. 0.69, Fig. S5a) or long term (yearly $R^2 = 0.53$ vs. 0.75, Fig. S5b). In addition, due to the (non-random) incomplete satellite measurements, AOD-based estimates may ignore some PM_{2.5} hotspots in China. Therefore, the long-term average PM_{2.5} concentrations of AOD-based estimates could be slightly biased. This may partially explain why the complete estimator (PM_{2.5} ML + GAM: CMAQ + AOD) was less accurate than the incomplete estimator (PM_{2.5} ML: CMAQ + AOD) on a daily timescale ($R^2 = 0.55$ vs. 0.61), but more accurate on monthly ($R^2 = 0.71$ vs. 0.68) and yearly timescales ($R^2 = 0.77$ vs. 0.75).

We also compared our estimates to the results of previous studies (Table 1). To enable a valid comparison, we extracted a subset of our annually iterated CV according to the periods and study domains of

previous studies. The results indicate that our estimates were as good as, or slightly better, than previous approaches in terms of model accuracy. For instance, the CV R^2 was reported as 0.41 for the state-of-the-art model developed by Ma et al. (2016) for daily average PM_{2.5} concentrations, but we improved this to 0.60 in our ML-based estimates. However, we cannot conclude that our model over-performs versus previous estimates based on this comparison alone: the CV R^2 only reflects the linear correlation between the predictions and observations, and other studies had advantages in different aspects. For example, using advanced products of AOD, Liang et al. (2018) generated an estimator with a finer spatial resolution (1 km × 1 km), which was able to better characterize the intra-city variability in PM_{2.5} than our estimates. This comparison is only intended to validate our results and should not be over-interpreted. Different from the extant AOD-based PM_{2.5} estimates, which mainly included meteorological variables as auxiliary predictors, our approach further incorporated CMAQ-simulated PM_{2.5}. According to our sensitivity analyses (Fig. S11), the ML model without CMAQ simulates ($R^2 = 0.57$), which has similar model inputs with the previous ones, performed not as well as the full model ($R^2 = 0.61$). The result may partially explain different model performances between studies (Table 1).

Based on the estimates with complete spatiotemporal coverage, we evaluated the historical trends in PM_{2.5} in China, both nationwide and in three megacity areas, during 2000–2016. Because of data availability constraints, these trends have rarely been studied. Cohen et al. (2017) estimated annual and global maps of PM_{2.5} during 1990–2015, and reported that PM_{2.5} in China increased linearly from 2000 (population-weighted PM_{2.5} of 50 μg/m³) to 2010 (58 μg/m³), and then plateaued during 2011–2015. Ma et al. (2016) estimated historical PM_{2.5} in China

Table 1

Comparison of the accuracy of our models (#PM_{2.5}^{ML}: CMAQ + AOD and \$PM_{2.5}^{ML+GAM}: CMAQ + AOD) with that of previous estimates, based on annually iterated CV. *The study domain was defined as a rectangle around Beijing (longitude 114°–118.5°, latitude 38°–42°).

Validation of historical PM _{2.5} estimates in previous studies			CV R ² reported in this study			
Study	Domain	Period	Temporal resolution	CV R ²	ML [#]	ML + GAM ^{\$}
Ma et al., 2016	China	Jan–Jun, 2014	Daily	0.41	0.60	0.57
Liang et al., 2018	Beijing*	2013–2014	Monthly	0.42	0.85	0.78
				0.55	0.78	0.77

PM_{2.5}, particulate matter; CV, cross-validation; M, machine learning; GAM, generalized additive model.

during 2004–2013 based on an LME of AOD, and reported an increasing trend of 1.97 (95% CI: 1.22, 2.72) $\mu\text{g}/\text{m}^3/\text{year}$ during 2004–2007, followed by a decreasing trend of 0.46 (0.14, 0.78) $\mu\text{g}/\text{m}^3/\text{year}$. Liu et al. (2017) estimated the historical time series of PM_{2.5} during 1957–1964 and 1973–2014 using visibility data from 674 meteorological sites across China, and reported an increasing trend of 0.24 (0.14, 0.34) $\mu\text{g}/\text{m}^3/\text{year}$ during 1996–2005 and a decreasing trend of 0.34 (0.16, 0.52) $\mu\text{g}/\text{m}^3/\text{year}$ during 2006–2014. Lin et al. (2018) derived high-resolution PM_{2.5} in annual scale across China during 2001–2015 based on an empirical algorithm of MODIS AOD, and reported a nation-scale trend of 0.04 $\mu\text{g}/\text{m}^3/\text{year}$, $-0.65 \mu\text{g}/\text{m}^3/\text{year}$, $-2.33 \mu\text{g}/\text{m}^3/\text{year}$ in 2001–2005, 2005–2010 or 2011–2015, respectively. Boys et al. (2014) estimated global PM_{2.5} during 1998–2012 by scaling satellite AOD using a chemical transport model, and found that, in East Asia, PM_{2.5} increased until 2007 at a rate of 0.79 (0.26, 1.32) $\mu\text{g}/\text{m}^3/\text{year}$. Considering the differences in temporal periods, study domains, data availability and modeling accuracy, the long-term trends in PM_{2.5} of previous studies are comparable to those of the present study. Previous studies were in good agreement with respect to the increasing PM_{2.5} trends before 2005 or 2007. Recently, satellite-based estimates and in-situ measurements showed a rapidly decreasing trend in PM_{2.5} (by 4.7 $\mu\text{g}/\text{m}^3/\text{year}$ in 2013–2015 or 3.4 $\mu\text{g}/\text{m}^3/\text{year}$ in 2015–2017) after 2013 (Ben et al., 2018; Zheng et al., 2017), which is consistent with our model results.

In addition, the estimated historical trends in PM_{2.5} were consistent with the commencement of emissions reduction policies nationwide, and in certain individual regions. According to our results, the increasing trend in PM_{2.5} ended in 2007 and then PM_{2.5} plateaued during 2008–2012. During 2000–2010, energy consumption driven by economic growth, for example in capital formation and exports, was found to have significantly contributed to the increasing emissions of primary PM_{2.5} (Guan et al., 2014). For example, from 2005 to 2010, it was reported that thermal power generation increased by 63%, and vehicle production by 220% (Zhang et al., 2012). Meanwhile, China began to strengthen its emissions reduction policies to mitigate decreasing air quality. For instance, driven by the widespread installation of flue gas desulfurization systems, reductions in SO₂ emissions of 1.5 and 17.5 million tons were achieved in 2005 and 2010, respectively (Zhang et al., 2012). However, the benefits of such emissions reduction schemes may have been offset by increased energy usage across most of China, except for the PRD. PM_{2.5} in the PRD exhibited a decreasing trend after 2008, which was dominated by reductions in organic compounds and sulfate (Fu et al., 2014). The long-lasting episodes of PM_{2.5} in central and eastern China (including the BTH region and YRD) in winter 2012 and January 2013 (Fig. 6) attracted widespread attention in Chinese society (Wang et al., 2014). Triggered by this air pollution crisis, China implemented the most stringent policies to control air pollution to date, by introducing the China Clean Air Act. Under this act, multiple approaches, including production structure changes, clean energy adoption and the designation of new air quality criteria were applied to reduce pollutant emissions, and provincial governments achieved air quality targets before 2017. For example, the BTH region planned to reduce annual concentrations of PM_{2.5} by 25%. Driven by the Clean Air Act, PM_{2.5} levels began to decline. In 2016, the population-weighted PM_{2.5} was reduced to 49.9 $\mu\text{g}/\text{m}^3$, slightly lower than that in 2000 (52.9 $\mu\text{g}/\text{m}^3$). However, for approximately 79% of Chinese citizens, the annual exposure to PM_{2.5} in 2016 was still above the air quality limit of 35 $\mu\text{g}/\text{m}^3$, which indicates that emissions reductions are still required in China. Additionally, although our previous study has evidenced that the recent decreasing trend of PM_{2.5} was mainly attributable to the emission control policies rather than the meteorological variations, the air quality influences of the climate changes should not be ignored. However, most of extant evidences in China (e.g., Liang et al., 2017) were from local areas or focused on short/medium-term periods. The nexus of climate and air quality can be complexly depended on physical processes (e.g., diffusions and transfers) driven by

the meteorological field, optical effects of aerosols (e.g., radiative forcing), atmospheric chemical reactions, and etc. (Hong et al., 2017). Therefore, quantifying the PM_{2.5} variations attributable to the climate changes is beyond the capability and scope of this study. Future nationwide studies on this issue are warranted.

Due to issues with data availability (e.g., lack of routine monitoring data) or quality (e.g., incomplete coverage of AOD-based estimates), most previous health-related studies on a national scale in China focused on the chronic effects of PM_{2.5}. For instance, the data products of Ma et al. (2016) have been utilized for risk assessment and cohort studies of the mortality risks of PM_{2.5}. However, recent epidemiological findings have suggested that exposure to PM_{2.5} could cause both acute and chronic adverse health effects (Shi et al., 2015). Although the exposure-response functions between long-term exposure to PM_{2.5} and disease have been well studied, both globally and in China, ignoring the acute health effects of PM_{2.5} may lead to under-estimation of the health impacts of air pollution. Assessment of acute exposure to PM_{2.5} requires accurate historical estimates on daily timescales that, before the current study, were rarely available for China on a national scale. Our data products will support epidemiological studies on both the acute and chronic effects of heavily PM_{2.5}- polluted air, as well as other health-related studies on how public health changes in developing countries during rapid transitions in ambient air quality.

The major limitation of this study was the uncertainty in our estimates. Although the annually iterated CV was the best-available method to evaluate the historical predictions of PM_{2.5}, it could not directly quantify the model errors or uncertainty in the estimates. The offset validation using US embassy PM_{2.5} observations helped to quantify a proportion of the historical estimates after 2008, but this might be not particularly representative due to the restricted spatial coverage of the monitoring sites. Furthermore, due to massive computing burden of the complex ML model and knowledge gaps in uncertainties of the multiple inputs (e.g., CMAQ-WRF simulations), quantifying the pointwise standard errors or CIs for our historical PM_{2.5} estimates is beyond our capability or study scope. When applying our products to health-related studies of PM_{2.5}, including epidemiological studies and risk assessments, ignoring the uncertainty may bias results. In terms of epidemiology, large uncertainties in PM_{2.5} concentrations may result in exposure misclassification and can reduce statistical power. Our data products are more appropriate for the studies with samples sufficiently large to guarantee adequate statistical power to detect associations between PM_{2.5} and adverse health outcomes. Although as far as we know, our products may be the best-available daily PM_{2.5} estimates with complete spatiotemporal coverage during 2000–2016 in China, extra analysis on the impacts of uncertainty on the robustness of the associations is required, as done in our previous study (Xue and Zhang, 2018). For risk assessments, utilizing our products may lead to underestimation of the health burdens attributable to PM_{2.5} due to the slight bias in our estimates (e.g., a bias of 1.0 $\mu\text{g}/\text{m}^3$ in monthly averages). In addition, the AOD predictors were derived from MODIS level 2 products at an original resolution of 3 km, to allow estimates of PM_{2.5} at a finer resolution in future studies. However, only Dark Target AOD was incorporated into the 3 km products, which had lower spatial coverage than the combined AOD from both the Dark Target and Deep Blue products (10 km \times 10 km MODIS AOD). Considering the computational complexity, in this study we did not further examine the performance of the ML model by using the combined AOD data as alternative inputs.

5. Conclusions

This study describes an alternative method to model the deterministic association between PM_{2.5} and AOD with adjustments of other inputs. In comparison to models using only linear terms of the predictors, the HD-expansions herein were better able to characterize the complexities in PM_{2.5}-AOD associations and produced more accurate

AOD-based estimates of PM_{2.5} where monitoring networks did not exist. The two-stage estimator has been shown to be comparable to other AOD-based estimators for monthly average PM_{2.5} concentrations, and to perform slightly better over daily timescales (Table 1). Using the estimator, we quantified the population exposures to PM_{2.5} and derived their long-term trends in China, from 2000 to 2016. During the period, we reported a population-weighted average of 59.6 µg/m³ and found an inverse U-shape trend with the plateau between 2008 and 2012, in nationwide exposure to ambient PM_{2.5}. Our products will support studies on the health effects of both acute and chronic exposure to ambient PM_{2.5} in China, which are critical to fill knowledge gaps with respect to exposure-response functions for heavily-polluted air.

Acknowledgement

This work was supported by National Natural Science Foundation of China (41625020, 41571130032, 41701591, and 91744310).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envint.2018.11.075>.

References

- Apte, J.S., Marshall, J.D., Cohen, A., Brauer, M., 2015. Addressing global mortality from ambient PM_{2.5}. *Environ. Sci. Technol.* 49, 8057–8066.
- Ben, S., Carly, R., Stephen, A., Dominick, V.S., 2018. Substantial changes in air pollution across China during 2015 to 2017. *Environ. Res. Lett.* <https://doi.org/10.1088/1748-9326/aae718>.
- Boys, B.L., Martin, R.V., van Donkelaar, A., MacDonell, R., Hsu, N.C., Cooper, M., et al., 2014. Fifteen-year global time series of satellite-derived fine particulate matter. *Environ. Sci. Technol.* 48, 11109–11118.
- Bravo, M.A., Fuentes, M., Zhang, Y., Burr, M.J., Bell, M.L., 2012. Comparison of exposure estimation methods for air pollutants: ambient monitoring data and regional air quality simulation. *Environ. Res.* 116, 1–10.
- Burnett, R.T., Pope 3rd, C.A., Ezzati, M., Olives, C., Lim, S.S., Mehta, S., et al., 2014. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ. Health Perspect.* 122, 397–403.
- Chang, H.H., Hu, X.F., Liu, Y., 2014. Calibrating modis aerosol optical depth for predicting daily PM_{2.5} concentrations via statistical downscaling. *J. Expo. Sci. Environ. Epidemiol.* 24, 398–404.
- China State Council, 2013. Air Pollution Prevention and Control Action Plan (Available): http://www.gov.cn/zwgk/2013-09/12/content_2486773.htm 2018.
- Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., et al., 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.
- Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A.I., Wang, Y., Schwartz, J., 2016. Assessing PM_{2.5} exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol.* 50, 4712–4721.
- Didan, K., 2015. MOD13A3 MODIS/terra vegetation indices monthly L3 global 1 km SIN grid V006 [data set]. In: NASA EOSDIS LP DAAC, <https://doi.org/10.5067/MODIS/MOD13A3.006>.
- Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22.
- Fu, X., Wang, X., Guo, H., Cheung, K.L., Ding, X., Zhao, X., et al., 2014. Trends of ambient fine particles and major chemical components in the pearl river delta region: observation at a regional background site in fall and winter. *Sci. Total Environ.* 274–281.
- Geng, G.N., Zhang, Q., Martin, R.V., van Donkelaar, A., Huo, H., Che, H.Z., et al., 2015. Estimating long-term PM_{2.5} concentrations in China using satellite-based aerosol optical depth and a chemical transport model. *Remote Sens. Environ.* 166, 262–270.
- Guo, D., Su, X., Zhang, Q., Peters, G.P., Liu, Z., Lei, Y., et al., 2014. The socioeconomic drivers of China's primary PM_{2.5} emissions. *Environ. Res. Lett.* 9, 024010.
- Guo, J., Xia, F., Zhang, Y., Liu, H., Li, J., Lou, M., et al., 2017. Impact of diurnal variability and meteorological factors on the PM_{2.5}-AOD relationship: implications for PM_{2.5} remote sensing. *Environ. Pollut.* 221, 94–104.
- Gupta, P., Christopher, S.A., 2009. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *J. Geophys. Res.-Atmos.* 114 (D20).
- Hong, C., Zhang, Q., Zhang, Y., Tang, Y., Tong, D., He, K., 2017. Multi-year downscaling application of two-way coupled WRF v3.4 and CMAQ v5.0.2 over east Asia for regional climate and air quality modeling: model evaluation and aerosol direct effects. *Geosci. Model Dev.* 10, 2447–2470.
- Hou, W.Z., Li, Z.Q., Zhang, Y.H., Xu, H., Zhang, Y., Li, K.T., et al., 2014. Using support vector regression to predict PM₁₀ and PM_{2.5}. *IOP Conf. Ser. Earth Environ.* 17, 1.
- Hu, X.F., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., et al., 2017. Estimating PM_{2.5} concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol.* 51, 6936–6944.
- Jiang, Q.T., Christakos, G., 2018. Space-time mapping of ground-level PM_{2.5} and NO₂ concentrations in heavily polluted northern China during winter using the bayesian maximum entropy technique with satellite data. *Air Qual. Atmos. Health* 11, 23–33.
- Klimont, Z., Smith, S.J., Cofala, J., 2013. The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions. *Environ. Res. Lett.* 8, 014003.
- Krotkov, N.A., Lamsal, L.N., Marchenko, S.V., Celarier, E.A., Bucsela, E., Swartz, W.H., Veefkind, P., 2018. OMII/Aura Nitrogen Dioxide (NO₂) Total and Tropospheric Column 1-orbit L2 Swath 13 × 24 km V003. Goddard Earth Sciences Data and Information Services Center, Greenbelt, MD, USA.
- Levy, R., Hsu, C., et al., 2015. MODIS atmosphere L2 aerosol product. In: NASA MODIS Adaptive Processing System. Goddard Space Flight Center, USA. https://doi.org/10.5067/MODIS/MOD04_L2.006.
- Li, T., Shen, H., Yuan, Q., Zhang, X., Zhang, L., 2017. Estimating ground-level PM_{2.5} by fusing satellite and station observations: a geo-intelligent deep learning approach. *Geophys. Res. Lett.* 44, 11,985–911,993.
- Liang, P., Zhu, T., Fang, Y., Li, Y., Han, Y., Wu, Y., et al., 2017. The role of meteorological conditions and pollution control strategies in reducing air pollution in Beijing during APEC 2014 and parade 2015. *Atmos. Chem. Phys.* 17 (22), 13921–13940.
- Liang, F., Xiao, Q., Wang, Y., Lyapustin, A., Li, G., Gu, D., et al., 2018. Maiaac-based long-term spatiotemporal trends of PM_{2.5} in Beijing, China. *Sci. Total Environ.* 616–617, 1589–1598.
- Lin, C.Q., Liu, G., Lau, A.K.H., Li, Y., Li, C.C., Fung, J.C.H., et al., 2018. High-resolution satellite remote sensing of provincial PM_{2.5} trends in China from 2001 to 2015. *Atmos. Environ.* 180, 110–116.
- Liu, M.M., Bi, J., Ma, Z.W., 2017. Visibility-based PM_{2.5} concentrations in China: 1957–1964 and 1973–2014. *Environ. Sci. Technol.* 51, 13161–13169.
- Lu, Z., Zhang, Q., Streets, D.G., 2011. Sulfur dioxide and primary carbonaceous aerosol emissions in China and India, 1996–2010. *Atmos. Chem. Phys.* 11, 9839–9846.
- Lv, B.L., Hu, Y.T., Chang, H.H., Russell, A.G., Bai, Y.Q., 2016. Improving the accuracy of daily PM_{2.5} distributions derived from the fusion of ground-level measurements with aerosol optical depth observations, a case study in north China. *Environ. Sci. Technol.* 50, 4752–4759.
- Ma, Z.W., Hu, X.F., Huang, L., Bi, J., Liu, Y., 2014. Estimating ground-level PM_{2.5} in China using satellite remote sensing. *Environ. Sci. Technol.* 48, 7436–7444.
- Ma, Z.W., Hu, X.F., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y.G., et al., 2016. Satellite-based spatiotemporal trends in PM_{2.5} concentrations: China, 2004–2013. *Environ. Health Perspect.* 124, 184–192.
- National Bureau of Statistics of China, Department of Population and Employment Statistics, 2003. Tabulation on the 2000 Population Census of the people's Republic of China by County. China Statistics Press.
- National Bureau of Statistics of China, Department of Population and Employment Statistics, 2012. Tabulation on the 2010 Population Census of the People's Republic of China by County. China Statistics Press.
- National Geophysical Data Center, 2018. Version 4 DMSP-OLS Nighttime Lights Time Series. Boulder, CO, USA. <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html> (Accessed: Mar. 2018).
- R Core Team, 2017. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
- Reid, C.E., Jerrett, M., Petersen, M.L., Pfister, G.G., Morefield, P.E., Tager, I.B., et al., 2015. Spatiotemporal prediction of fine particulate matter during the 2008 northern California wildfires using machine learning. *Environ. Sci. Technol.* 49, 3887–3896.
- Shi, L., Zanobetti, A., Kloog, I., Coull, B.A., Koutrakis, P., Melly, S.J., et al., 2015. Low-concentration PM_{2.5} and mortality: estimating acute and chronic effects in a population-based study. *Environ. Health Perspect.* 124, 46.
- Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., et al., 2008. A Description of the Advanced Research WRF Version 3. National Center for Atmospheric Research, Boulder, CO, USA.
- US EPA Office of Research and Development, 2015. CMAQv5.1 (Version 5.1). Zenodoh:<https://doi.org/10.5281/zenodo.1079909>.
- van Donkelaar, A., Martin, R.V., Brauer, M., Boys, B.L., 2015. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ. Health Perspect.* 123, 135–143.
- van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., et al., 2016. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environ. Sci. Technol.* 50, 3762–3772.
- Wang, J., Christopher, S.A., 2003. Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: implications for air quality studies. *Geophys. Res. Lett.* 30 (21), 2095–2099.
- Wang, Y.S., Yao, L., Wang, L.L., Liu, Z.R., Ji, D.S., Tang, G.Q., et al., 2014. Mechanism for the formation of the January 2013 heavy haze pollution episode over central and eastern China. *Sci. China Earth Sci.* 57, 14–25.
- Xu, Y., 2011. Improvements in the operation of SO₂ scrubbers in China's coal power plants. *Environ. Sci. Technol.* 45, 380–385.
- Xue, T., Zhang, Q., 2018. Associating ambient exposure to fine particles and human fertility rates in China. *Environ. Pollut.* 235, 497–504.
- Xue, T., Zheng, Y., Geng, G., Zheng, B., Jiang, X., Zhang, Q., et al., 2017. Fusing observational, satellite remote sensing and air quality model simulated data to estimate spatiotemporal variations of PM_{2.5} exposure in China. *Remote Sens.* 9, 221.
- You, W., Zang, Z.L., Zhang, L.F., Li, Y., Pan, X.B., Wang, W.Q., 2016. National-scale estimates of ground-level PM_{2.5} concentration in China using geographically weighted regression based on 3 km resolution modis AOD. *Remote Sens.* 8 (3).
- Zhan, Y., Luo, Y.Z., Deng, X.F., Chen, H.J., Grieneisen, M.L., Shen, X.Y., et al., 2017. Spatiotemporal prediction of continuous daily PM_{2.5} concentrations across China

- using a spatially explicit machine learning algorithm. *Atmos. Environ.* 155, 129–139.
- Zhang, X., Hu, H., 2017. Improving satellite-driven PM_{2.5} models with VIIRS nighttime light data in the Beijing–Tianjin–Hebei region, China. *Remote Sens.* 9, 908.
- Zhang, Q., He, K., Huo, H., 2012. Policy: cleaning China's air. *Nature* 484, 161.
- Zheng, Y., Zhang, Q., Liu, Y., Geng, G., He, K., 2016. Estimating ground-level PM_{2.5} concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. *Atmos. Environ.* 232–242.
- Zheng, Y., Xue, T., Zhang, Q., Geng, G.N., Tong, D., Li, X., et al., 2017. Air quality improvements and health benefits from China's clean air action since 2013. *Environ. Res. Lett.* 12 (11).
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat Methodol.* 67, 301–320.
- Zou, B., Wang, M., Wan, N., Wilson, J.G., Fang, X., Tang, Y.Q., 2015. Spatial modeling of PM_{2.5} concentrations with a multifactorial radial basis function neural network. *Environ. Sci. Pollut. Res.* 22, 10395–10404.
- Zou, B., Chen, J.W., Zhai, L., Fang, X., Zheng, Z., 2017. Satellite based mapping of ground PM_{2.5} concentration using generalized additive modeling. *Remote Sens.* 9, 1.