Remote Sensing of Environment 140 (2014) 220–232

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Estimating ground-level $PM_{2.5}$ concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model

Xuefei Hu [a], Lance A. Waller [b], Alexei Lyapustin [c], Yujie Wang [c,d], Mohammad Z. Al-Hamdan [e], William L. Crosson [e], Maurice G. Estes Jr. [e], Sue M. Estes [e], Dale A. Quattrochi [f], Sweta Jinnagara Puttaswamy [a], Yang Liu [a,*]

[a] Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA
[b] Department of Biostatistics & Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA
[c] NASA Goddard Space Flight Center, Greenbelt, MD, USA
[d] University of Maryland Baltimore County, Baltimore, MD, USA
[e] Universities Space Research Association at the National Space Science and Technology Center, NASA Marshall Space Flight Center, Huntsville, AL 35805, USA
[f] Earth Science Office at NASA Marshall Space Flight Center, National Space Science and Technology Center, Huntsville, AL 35805, USA

ARTICLE INFO

Article history:
Received 26 February 2013
Received in revised form 31 July 2013
Accepted 20 August 2013
Available online 26 September 2013

Keywords:
Aerosol optical depth
MAIAC
MODIS
$PM_{2.5}$
Two-stage model

ABSTRACT

Previous studies showed that fine particulate matter ($PM_{2.5}$, particles smaller than 2.5 μm in aerodynamic diameter) is associated with various health outcomes. Ground in situ measurements of $PM_{2.5}$ concentrations are considered to be the gold standard, but are time-consuming and costly. Satellite-retrieved aerosol optical depth (AOD) products have the potential to supplement the ground monitoring networks to provide spatiotemporally-resolved $PM_{2.5}$ exposure estimates. However, the coarse resolutions (e.g., 10 km) of the satellite AOD products used in previous studies make it very difficult to estimate urban-scale $PM_{2.5}$ characteristics that are crucial to population-based $PM_{2.5}$ health effects research. In this paper, a new aerosol product with 1 km spatial resolution derived by the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm was examined using a two-stage spatial statistical model with meteorological fields (e.g., wind speed) and land use parameters (e.g., forest cover, road length, elevation, and point emissions) as ancillary variables to estimate daily mean $PM_{2.5}$ concentrations. The study area is the southeastern U.S., and data for 2003 were collected from various sources. A cross validation approach was implemented for model validation. We obtained R^2 of 0.83, mean prediction error (MPE) of 1.89 μg/m^3, and square root of the mean squared prediction errors (RMSPE) of 2.73 μg/m^3 in model fitting, and R^2 of 0.67, MPE of 2.54 μg/m^3, and RMSPE of 3.88 μg/m^3 in cross validation. Both model fitting and cross validation indicate a good fit between the dependent variable and predictor variables. The results showed that 1 km spatial resolution MAIAC AOD can be used to estimate $PM_{2.5}$ concentrations.

1. Introduction

Numerous epidemiological studies have shown that $PM_{2.5}$ (particle size less than 2.5 μm in the aerodynamic diameter) is associated with various adverse health outcomes including cardiovascular and respiratory diseases (Dominici et al., 2006; Gauderman et al., 2004; Gold et al., 2000; Peters, Dockery, Muller, & Mittleman, 2001; Schwartz & Neas, 2000). The estimation of population exposures to $PM_{2.5}$ has traditionally been done by assigning measurements of a central ground monitor to people living within a certain distance of it (e.g., a few kilometers (Laden, Schwartz, Speizer, & Dockery, 2006) to a few tens of kilometers (Samet, Dominici, Curriero, Coursac, & Zeger, 2000)). Exposure misclassification due to spatial misalignment causes biased and often reduced estimates of health risks. Thus, accurate, spatially resolved $PM_{2.5}$ exposure characterization is very important in effectively conducting air quality assessment and environmental epidemiologic studies.

Because ground monitoring networks are costly to maintain, even the United States, which has the most extensive regulatory monitoring programs, only has its most populated counties (less than 30% of over 3000 in total) covered with one or more monitors. Satellite remote sensing provides a potentially cost effective way to predict $PM_{2.5}$ concentrations by using aerosol optical depth (AOD) in areas where monitors are not available or too sparse (Hoff & Christopher, 2009). AOD measures light extinction by aerosol scattering and absorption in an atmospheric column and is related to the loadings of fine particles in the column. AOD products from several satellite sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Hu et al., 2013; Liu, Franklin, et al., 2007; Zhang, Hoff, & Engel-Cox, 2009), the

* Corresponding author. Tel.: +1 404 727 2131; fax: +1 404 727 8744.
E-mail address: yang.liu@emory.edu (Y. Liu).

0034-4257/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.rse.2013.08.032

Multiangle Imaging SpectroRadiometer (MISR) (Liu, Franklin, et al., 2007; Liu, Koutrakis, Kahn, Turquety, et al., 2007; Liu, Koutrakis, et al., 2007), and the Geostationary Operational Environmental Satellite Aerosol/Smoke Product (GASP) (Liu, Paciorek, & Koutrakis, 2009; Paciorek, Liu, Moreno-Macias, & Kondragunta, 2008) have been used in previous studies for estimating $PM_{2.5}$ concentrations. In addition, many previous studies have improved satellite AOD retrievals by using top-of-atmosphere reflectance and the GEOS-Chem chemical transport model to increase the accuracy of particle concentration estimation (Drury et al., 2010; van Donkelaar et al., 2013; Wang, Xu, Spurr, Wang, & Drury, 2010). However, one of the limitations of the current AOD products is the coarse spatial resolution. For example, the nominal spatial resolutions for AOD retrieved by MODIS, MISR, and GASP operational algorithms are 10 km, 17.6 km, and 4 km, respectively. Recently, a new Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm was developed. MAIAC uses time-series analysis and image-based processing techniques to make aerosol retrievals and atmospheric corrections over both dark vegetated land and brighter range of surfaces (Lyapustin, Wang, et al., 2011). Derived from MODIS radiances, the MAIAC AOD product has 1 km spatial resolution, and has been demonstrated to have strong correlations with $PM_{2.5}$ levels in New England region (Chudnovsky, Kostinski, Lyapustin, & Koutrakis, 2012).

Many previous studies established quantitative relationships between ground-level $PM_{2.5}$ concentrations and satellite-derived AOD using methods such as linear regression (Gupta & Christopher, 2009; Liu, Sarnat, Kilaru, Jacob, & Koutrakis, 2005; Wallace, Kanaroglou, & Ieee, 2007) without considering the day-to-day variations in the $PM_{2.5}$–AOD relationship. Lee, Liu, Coull, Schwartz, and Koutrakis (2011) developed a linear mixed effects model to consider the temporal variations in the $PM_{2.5}$–AOD relationship with AOD used as the only predictor. Kloog, Koutrakis, Coull, Lee, and Schwartz (2011) expanded Lee's method by incorporating other predictors and random-effects variables in the model. However, both models assume that there is little spatial variability in the relationship, which is not necessarily true, especially when the modeling domain gets larger. Previous studies showed that the correlation between $PM_{2.5}$ and AOD varies spatially (Engel-Cox, Holloman, Coutant, & Hoff, 2004; Hu, 2009). Hu et al. (2013) found that the $PM_{2.5}$–AOD relationship varies spatially and used the spatial varying relationship to predict $PM_{2.5}$ concentrations. Failure to account for spatial variability in the relationship may lead to poor model performance.

The objective of this analysis is to evaluate the performance of 1 km MAIAC AOD as a major predictor of ground level $PM_{2.5}$ concentrations in the setting of a two-stage spatial statistical model using MAIAC AOD as the primary predictor, and meteorological and land use information as ancillary parameters. The two-stage model is expected to account for both temporal and spatial variability in the $PM_{2.5}$–AOD relationship. The accuracy and spatial patterns of estimated $PM_{2.5}$ concentrations were examined by various 2-D and 3-D maps, standard model fitting, and cross validation statistics. As a reference, this model was also applied to the MODIS AOD data with a 10 km spatial resolution. The results derived from MODIS and MAIAC models were then compared in order to examine the impact of spatial resolution on $PM_{2.5}$ concentration estimates.

2. Materials and methods

2.1. Study area

The study area is approximately 800 × 1200 km^2 in the southeastern U.S., covering Georgia, Alabama, Tennessee, and Mississippi, most of

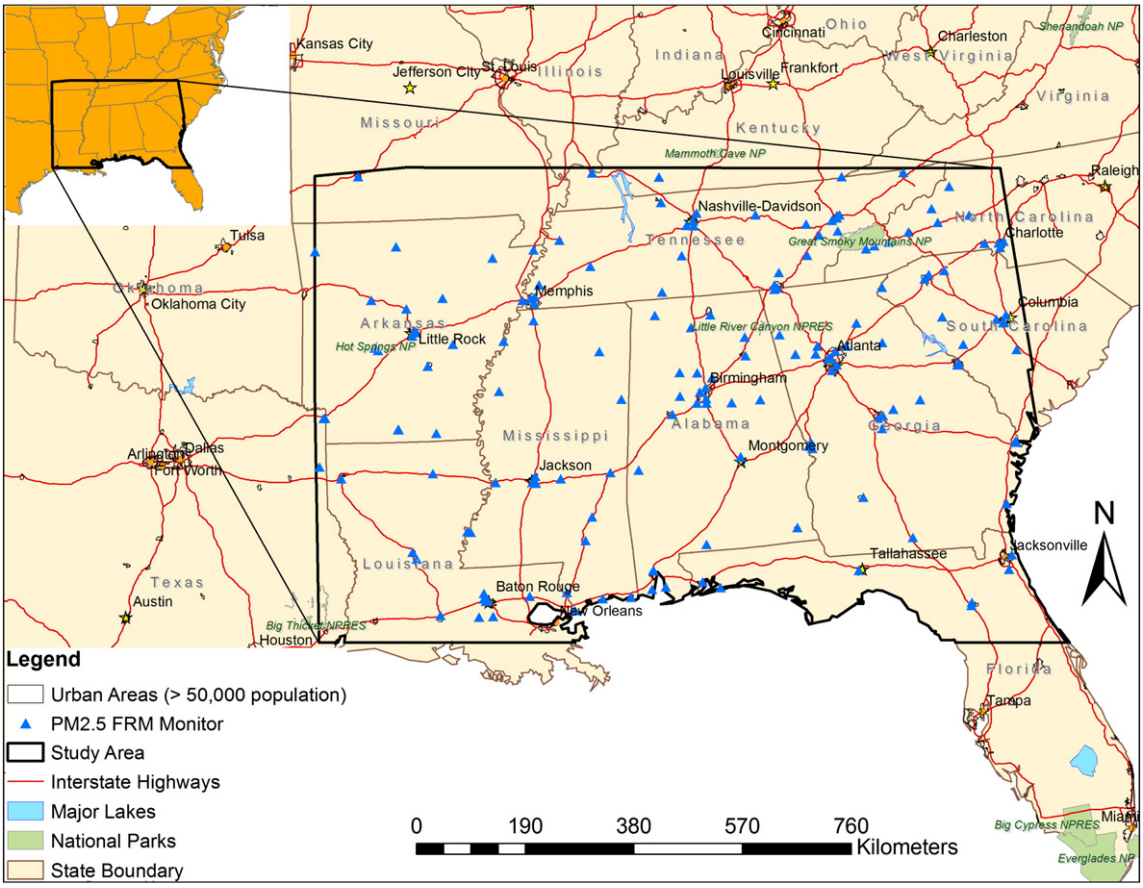

Fig. 1. Study area.

Louisiana and Arkansas, and parts of Florida, Missouri, Kentucky, North Carolina, and South Carolina (Fig. 1). This domain includes various terrains, numerous large urban centers, medium to small cities, and suburban and rural areas. In addition, this region also suffers from active prescribed burns, especially in the spring.

2.2. EPA $PM_{2.5}$ measurements

The 24-h average $PM_{2.5}$ concentrations for 2003 collected from 166 U.S. Environmental Protection Agency (EPA) federal reference monitors (FRM) were downloaded from the EPA's Air Quality System Technology Transfer Network (http://www.epa.gov/ttn/airs/airsaqs/). $PM_{2.5}$ concentrations less than 2 μg/m^3 (~2% of total records) were discarded as they are below the established limit of detection (EPA, 2008).

2.3. Remote sensing data

2.3.1. MAIAC AOD

MAIAC processing includes cloud masking, deriving column water vapor, and retrieval of aerosol parameters over land at 1 km resolution simultaneously with parameters of a surface bidirectional reflectance distribution function (BRDF). This is accomplished by using the time series of MODIS measurements and simultaneous processing of a group of pixels in fixed 25 × 25 km^2 blocks (Lyapustin, Martonchik, et al., 2011; Lyapustin, Wang, et al., 2011; Lyapustin, Wang, et al., 2012). MAIAC uses a sliding window approach to accumulate 5 (over poles)–16 (over equator) days of MODIS radiance observations over the same area. MODIS data are initially gridded to a 1 km resolution in a selected projection, and the algorithm is applied to Terra and Aqua data separately. The surface BRDF retrievals are conducted for conditions with relatively low AOD (e.g., less than 0.5 globally) using the regional background aerosol model with fixed size distribution and refractive index. The BRDF is retrieved when the surface remains stable during the 5–16-day accumulation period, which is established with the internal surface change detection algorithm (Lyapustin, Wang, et al., 2012). Over the dark and moderately bright surfaces, the aerosol and surface reflectance retrieval problems are decoupled through the use of the 2.1 μm channel. As this band is generally transparent, its BRDF model is derived first. The aerosol retrieval (e.g., at 0.47 μm) requires knowledge of the spectral regression coefficient (SRC), which relates reflectance at 0.47 and 2.1 μm. The SRC is obtained using four or more low AOD days by inverting all available measurements in 25 × 25 km^2 blocks. The assumptions (such as constant AOD in the block on a given day and stable surface during the selected period) are verified by the algorithm internally as discussed in Lyapustin, Wang, et al. (2011). Once SRC is obtained, the AOD is retrieved from the last MODIS measurement. In clear conditions, aerosol retrievals are performed with the regional background aerosol model tuned to the AERONET measurements. In hazy conditions with sufficient sensitivity to aerosol type (smoke/mineral dust), MAIAC's knowledge of spectral surface BRDF from the previous retrievals is used for the aerosol type classification and retrievals (Lyapustin, Korkin, et al., 2012). The AOD retrieval error is characterized internally based on the uncertainty of the surface spectral BRDF, although it is not currently reported. Validation over the continental USA, based on the Aerosol Robotic Network (AERONET) (Holben et al., 1998) data, showed that the MAIAC and operational Collection 5 MODIS Dark Target AOD have a similar accuracy over dark and vegetated surfaces, but also showed that MAIAC generally improves accuracy over brighter surfaces, including most urban areas (Lyapustin, Wang, et al., 2011).

In this study, Aqua (overpass at ~1:30 pm local time) and Terra (overpass at ~10:30 am local time) MAIAC AOD values were first combined to improve spatial coverage. Wang and Christopher (2003) built simple empirical linear relationships between MODIS AOD and 24-h $PM_{2.5}$. In addition, Zhang et al. (2012) found that Terra and Aqua may provide a good estimate of the daily average of AOD, thus the average of these two measurements should be able to be used to predict 24-h $PM_{2.5}$ concentrations, which has been successfully applied in previous research (Lee et al., 2011). Changing cloud cover causes the temporal and spatial coverage of MAIAC-Aqua and MAIAC-Terra AOD values to differ. Therefore when combining the two MAIAC products on its common pixel grid, we came across two scenarios, one where a given grid cell has one of the MAIAC products and the other where both are present. In the grid cells that have both MAIAC-Terra and MAIAC-Aqua AOD, the averaged value represents the mean of the AOD distribution from 10 am to 2 pm local time. In the other scenario, the averaged AOD at the grid cell is biased towards either morning condition or afternoon condition. To overcome this bias, Lee et al. (2011) used the average Terra AOD/Aqua AOD ratio to estimate the missing AOD values. In this study, we fitted a simple linear regression to define the relationship between daily mean AOD values of MAIAC-Terra and MAIAC-Aqua. By using the MAIAC data present on a given day, we predicted the missing AOD value and averaged them together. As a result, each MAIAC grid cell contains a mean value that better represents the average conditions from 10 am to 2 pm local time. Although the relationship between Aqua and Terra AOD may vary by season, we found that the variation is relatively small in our case. Therefore, the regression model was built using the annual data in this paper. The regression equation was provided as follows

$$\begin{aligned}\hat{\tau}_{AQUA} &= 0.78761\tau_{TERRA} + 0.11542\\ \hat{\tau}_{TERRA} &= 0.92194\tau_{AQUA} + 0.06444\end{aligned} \quad (1)$$

where τ is the AOD, and R^2 of 0.73 was obtained for both regression models. Finally, a simple filter with an upper bound of 2.0 was used for combined MAIAC AOD to reduce potential cloud contamination (~0.1% of total data records were filtered).

2.3.2. MODIS AOD and fire product

As a reference, the 2003 Terra and Aqua MODIS aerosol data (Collection 5) were downloaded from the Earth Observing System Data Gateway at the Goddard Space Flight Center (http://delenn.gsfc.nasa.gov/~imswww/pub/imswelcome). We re-sampled these data to the 12 km Community Multiscale Air Quality (CMAQ) grid using a nearest neighbor approach. Because standard MODIS algorithm provides swath data with pixels shifting in space and changing footprint size from 10 × 10 km^2 at nadir to 20 × 40 km^2 at the edge of scan, a base grid is needed for prediction. The CMAQ grid is a commonly used grid in air quality modeling, which can facilitate future intercomparison between CMAQ simulation results with satellite predictions. Both CMAQ and MODIS have similar spatial resolutions (12 km and 10 km, respectively), and the study domain is large (800 × 1200 km^2). Thus, the variability due to MODIS re-sampling should be relatively small. The same procedure used to combine MAIAC Aqua and Terra AOD was applied to combine MODIS Aqua and Terra AOD. The fire detections of 2003 in the study region were obtained from the MODIS data processing system (MODAPS), the definitive version of collection 5 (version 5.1). The fire data were used for analyzing the potential cause of abnormally high $PM_{2.5}$ predictions.

2.4. Meteorological fields

The meteorological fields provided by the North American Land Data Assimilation System (NLDAS) Phase 2 were obtained from the NLDAS website (http://ldas.gsfc.nasa.gov/nldas/). NLDAS provides quality controlled, spatially and temporally consistent, real-time, and retrospective forcing datasets (Cosgrove et al., 2003). The spatial resolution of NLDAS meteorological data is 1/8th-degree (~13 km). The non-precipitation land-surface forcing fields for NLDAS (Phase 2) are spatially interpolated and temporally disaggregated from the North American Regional Reanalysis (NARR) dataset. The spatial resolution of NARR is ~32 km,

and its temporal resolution is 3-hourly. In this paper, we used NLDAS meteorological fields to take advantage of its higher spatial resolution. Hourly NLDAS measurements for the period from 10 am to 4 pm local standard time, which correspond to NARR measurements at 10 am, 1 pm, and 4 pm local standard time, were averaged to generate daytime meteorological fields corresponding to the MODIS overpass times.

2.5. Land use variables

Elevation data were obtained from the national elevation dataset (NED) (http://ned.usgs.gov). NED is the seamless elevation dataset covering the conterminous United States and is distributed by the U.S. Geological Survey (USGS). The elevation data are at a spatial resolution of 1 arc sec (~30 m). The road data were obtained from ESRI StreetMap USA (Environmental Systems Research Institute, Inc., Redland, CA). The road data at level A1 (limited access highway) were extracted, and the sum of the road segment lengths was determined for each 1×1 km^2 MAIAC grid cell. Grid cells with no roads were assigned a value of zero. A 2001 Landsat-derived land cover map covering the study area with a spatial resolution of 30 m was downloaded from the National Land Cover Database (NLCD) (http://www.epa.gov/mrlc/nlcd-2001). A forest cover map was generated by assigning a value of one to the forest pixels and zero to others. The 2001 Percent Developed Imperviousness map was also downloaded from the NLCD to examine the relationship between $PM_{2.5}$ concentrations and built-up areas in the Atlanta metro area. Primary $PM_{2.5}$ emissions (tons per year) were obtained from the 2002 EPA National Emissions Inventory (NEI) facility emissions report. Grid cells with multiple emission sources were assigned the summed value, and grid cells with no emissions were assigned a value of zero.

2.6. Data integration

All the data were first re-projected to the USA Contiguous Albers Equal Area Conic USGS coordinate system. For model fitting, the meteorological and AOD values acquired from the nearest centroid of the pixel were assigned to the $PM_{2.5}$ monitoring site, i.e. the nearest neighbor approach was applied using ArcGIS 9.3. Forest cover and elevation values were averaged, and road lengths and point emissions were summed over a 1×1 km^2 square buffer centered at each $PM_{2.5}$ monitoring site. For $PM_{2.5}$ prediction, forest cover and elevation values were averaged, while road lengths and point emissions were summed in each 1×1 km^2 MAIAC grid cell. Meteorological fields were assigned to each grid cell using the nearest neighbor approach. To maintain consistency between the two statistical models, the MODIS AOD model also used parameters at 1 km spatial resolution by creating a 1×1 km^2 square buffer around the centroid of each 12×12 km^2 CMAQ grid cell. In this study, the days with fewer than three matched data records in the study domain were discarded (~0.6% of total data records for MAIAC, and ~0.8% for MODIS). After filtering, there were 8033 data records in 309 days for MAIAC (~85% temporal coverage) and 6556 data records in 279 days for MODIS (~76% temporal coverage).

2.7. Model structure and validation

We developed a two-stage modeling framework to calibrate the $PM_{2.5}$–AOD relationship varying in both space and time. The first stage is a linear mixed effects model with day-specific random intercepts and slopes for AOD and wind speed (both are time-varying variables) to account for the temporally varying relationship between $PM_{2.5}$ and AOD (Kloog et al., 2011; Lee et al., 2011). A linear mixed effects model incorporates both fixed-effects terms and random-effects terms. Fixed effects affect the population mean, while random effects are associated with a sampling procedure and contribute to the covariance structure of the data. Unlike many previous studies that used log-transformed independent variables to deal with the skewed data, we used the original scale to simplify the modeling as we found no significant impact on the overall model fit. Additional predictors were considered, including surface temperature, relative humidity, wind direction, and boundary layer height. Only statistically significant predictors were included in the final model, which can be expressed as:

$$\begin{aligned} PM_{2.5,st} = &\left(b_0 + b_{0,t}\right) + \left(b_1 + b_{1,t}\right) AOD_{st} + \left(b_2 + b_{2,t}\right) Wind\,Speed_{st} \\ &+ b_3 Elevation_s + b_4 Major\,Roads_s + b_5 Forest\,Cover_s \\ &+ b_6 Point\,Emissions_s + \varepsilon_{st}\left(b_{0,t} b_{1,t} b_{2,t}\right) \sim N[(0,0,0), \Psi] \end{aligned} \quad (2)$$

where $PM_{2.5,st}$ is the measured ground level $PM_{2.5}$ concentration ($\mu g/m^3$) at site s in day t; b_0 and $b_{0,t}$ (day-specific) are the fixed and random intercepts, respectively; AOD_{st} is the MAIAC AOD value (unitless) at site s in day t; b_1 and $b_{1,t}$ (day-specific) are the fixed and random slopes for AOD, respectively; $Wind\ Speed_{st}$ is the 2-m wind speed (m/s) at site s in day t; b_2 and $b_{2,t}$ (day-specific) are the fixed and random slopes for wind speed, respectively; $Elevation_s$ is elevation values (m) at site s; $Major\ Roads_s$ is road length values (m) at site s; $Forest\ Covers$ is forest cover values (unitless) at site s; $Point\ Emissions_s$ is point emissions (tons per year) at site s; $b_{0,t}$, $b_{1,t}$, and $b_{2,t}$ are multivariate normally distributed; and Ψ is an unstructured variance–covariance matrix for the random effects. The fixed effects of AOD and wind speed represent the average effects on $PM_{2.5}$ concentrations for the entire study period, while random effects account for the daily variability in the relationship between dependent and independent variables. This equation was applied to the entire fitting dataset to generate fixed-effects intercept and slopes for all the days and random-effects intercept and slopes for each individual day. The first stage linear mixed effects model can account for the day-to-day variability in $PM_{2.5}$–AOD relationship by generating a daily AOD slope for all the sites for each day.

While the first stage incorporates temporal variation in the $PM_{2.5}$–AOD association, we expect that there may be additional spatial variations in the association as well. A significance test ($\alpha = 0.05$) (Fotheringham, Brunsdon, & Charlton, 2002) was conducted to examine the spatial non-stationary for each day (Table 1). The results indicate that there are a number of days showing significant spatial non-stationary after the stage one model, and with the increase of the minimum number of records per day, the percentage also increases. Although there might be potential factors to cause spatial variation after including land use and meteorological variables in the stage one model, we did not expect it to be large. In fact, spatial variation was slightly reduced after the stage one model. Our significance test showed that after the stage one model, the percentage of days showing significant spatial non-stationary drops 3.6%. However, there are still 15.2% of days showing significant spatial non-stationary in the relationship after the stage one model. To accommodate for this, we consider a second stage to our model using a geographically weighted regression (GWR), which generates a continuous surface of estimates for each parameter at each location instead of a universal value for all observations. In order to describe these spatially varying associations, we fitted GWR by adopting an adaptive bandwidth selection

Table 1
A significance test for spatial non-stationary ($\alpha = 0.05$).

N[a]	Percentage of days[b]
N > 2	15.2%
N > 3	16.6%
N > 4	18.2%
N > 5	19.4%
N > 6	20.3%
N > 7	21.9%

[a] Denotes minimum number of records per day.
[b] Denotes percentage of days showing significant spatial non-stationary in the relationship.

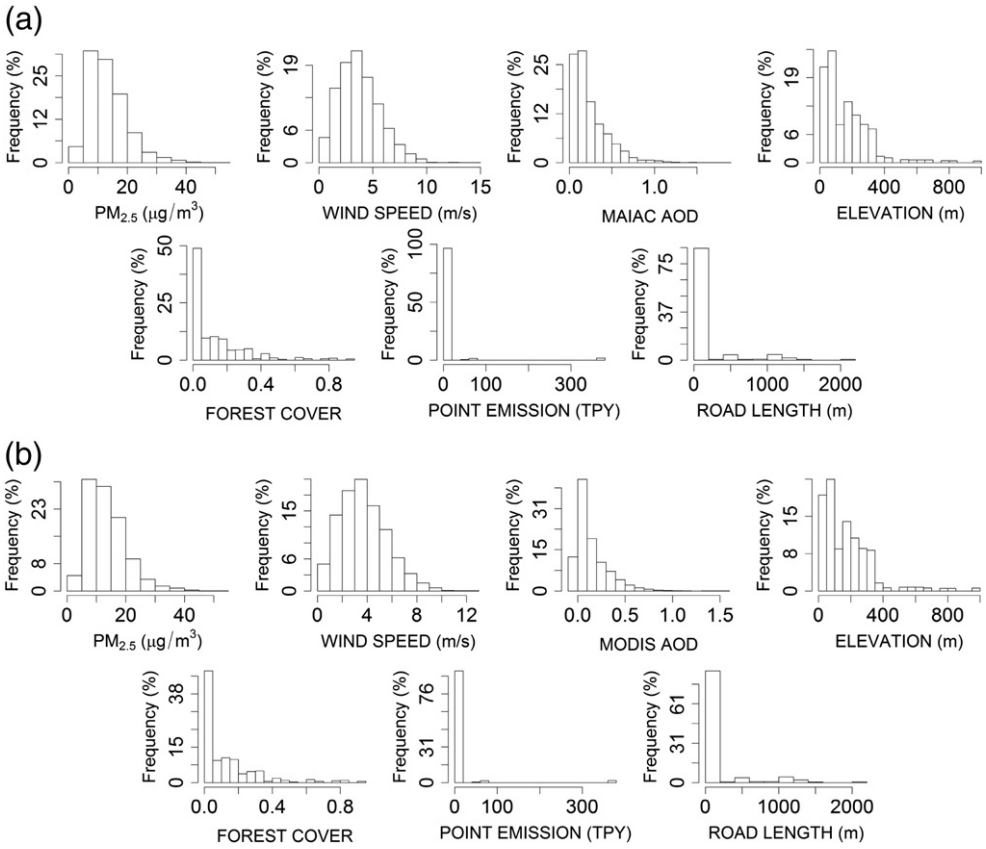

Fig. 2. Histograms of dependent and independent variables for MAIAC (a) and MODIS (b).

method to minimize the corrected Akaike Information Criterion (AIC_c) value. The best model should have the lowest AIC_c value (Fotheringham et al., 2002). It should be noted that GWR can be fitted using averaged dependent and independent variables for all days or be fitted for each day separately. We tested both and found that the separate GWR models for each day typically generate better results and were adopted in this analysis. The model structure can be expressed as

$$PM_{2.5}\_resi_{st} = \beta_{0,s} + \beta_{1,s} AOD_{st} + \varepsilon_{st} \quad (3)$$

where $PM_{2.5}\_resi_{st}$ denotes the residuals from the stage one model at site s in day t, AOD_{st} is the MAIAC AOD value (unitless) at site s in day t, and $\beta_{0,s}$ and $\beta_{1,s}$ denote the location-specific intercept and slope, respectively. β is calculated based on the geographical weighting (e.g., generally a Gaussian distance-decay weighting function) of each observation (e.g., a $PM_{2.5}$ monitoring site) relative to the location of the regression point (e.g., a $PM_{2.5}$ monitoring site or the centroid of a gird cell). The weighting of each observation for the regression point will decrease according to a Gaussian curve as the distance between them increases. For the second stage model, a threshold for minimum number of daily records must be established. The absolute minimum number of matched observations required is two in order to fit an intercept and a slope. We required a minimum of three observations to improve overall model performance, while covering as many days as possible in the analysis. The second stage GWR model can account for the spatial variability in $PM_{2.5}$–AOD relationship by generating a local AOD slope for each site.

To assess the goodness of fit of the model, various statistical indicators such as the coefficient of determination (R^2), mean prediction error (MPE) and square root of the mean squared prediction errors (RMSPE) were calculated between the predicted $PM_{2.5}$ concentrations from the fitted model and the observations. A 10-fold cross validation (CV) method was adopted to test for potential model over-fitting; i.e., the model could perform better on the data used to fit the model than on unobserved data. The entire model-fitting dataset was first randomly split into ten subsets with approximately 10% of the total data records in each subset. In each round of cross validation, we select one subset (10% of the data) as testing samples and use the remaining nine subsets (90% of the data) to fit the model. Predictions of the held-out subset (10% of the data) were made from the fitted model. In the next round, another subset was used for testing, and the remaining nine subsets were used for training. The

Table 2
Descriptive statistics for dependent and independent variables.

	MAIAC (N = 8033) (days = 309)				MODIS (N = 6556) (days = 279)			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
$PM_{2.5}$ (μg/m^3)	13.31	6.58	2.00	53.30	13.54	6.68	2.00	53.30
Wind speed (m/s)	3.75	1.91	0.04	14.67	3.71	1.89	0.04	12.77
Elevation (m)	160.76	149.21	1.90	981.26	170.77	157.90	1.90	981.26
Point emission (tons/year)	8.25	49.79	0.00	364.42	8.24	49.14	0.00	364.42
Limited access highway length (m)	125.90	341.05	0.00	2072.0	128.30	346.57	0.00	2072.0
Forest cover	0.13	0.17	0.00	0.94	0.13	0.18	0.00	0.94
AOD	0.24	0.21	0.002	1.83	0.14	0.17	−0.05	1.55

Table 3
Fixed effect of the linear mixed effects model (stage 1).

	MAIAC		MODIS	
	b	P-value	*b*	P-value
Intercept	13.05	<0.0001	13.75	<0.0001
AOD	10.33	<0.0001	12.67	<0.0001
Wind speed (m/s)	−0.68	<0.0001	−0.65	<0.0001
Elevation (m)	−0.0007	<0.05	−0.0003	0.3181[a]
Major roads (m)	0.0005	<0.001	0.0005	<0.001
Forest cover	−2.20	<0.0001	−2.06	<0.0001
Point emission (tons/year)	0.01	<0.0001	0.01	<0.0001

[a] Elevation is not significant in the MODIS model, and we kept it for comparison purpose.

process was repeated ten times until every subset was tested. The agreement between the predicted and observed values was evaluated using the slope, R^2, MPE, and RMSPE. A comparison was conducted between the CV and the model-fitting statistics to assess the degree of potential model over-fitting. A similar two-stage model was also developed using MODIS AOD as the primary predictor. Both two-stage models were used to estimate ground-level $PM_{2.5}$ concentrations in the study domain where there are no $PM_{2.5}$ observations and to generate a continuous $PM_{2.5}$ surface for each day. The annual and seasonal mean predicted $PM_{2.5}$ surfaces were derived from the daily surfaces and compared visually. In addition, annual mean $PM_{2.5}$ surfaces for the Atlanta metro area were generated for MAIAC and MODIS to examine the effect of spatial resolution on the $PM_{2.5}$ concentration estimation. All modeling was done using the R statistical software version 2.15.2.

3. Results

3.1. Descriptive statistics

The histograms of variables are illustrated in Fig. 2, which shows that all the variables are approximately unimodal and log-normally distributed. The mean, standard deviation, maximum, and minimum for all the variables are presented in Table 2. The annual mean $PM_{2.5}$ concentration for all the monitoring sites is 13.31 μg/m^3 and 13.54 μg/m^3 for MAIAC and MODIS matched data, respectively. The overall mean of AOD is 0.24 and 0.14 for MAIAC and MODIS, respectively. The difference in AOD reporting wavelengths (MAIAC at 470 nm vs. MODIS at 550 nm) to a large extent leads to the difference in their mean AOD values. Despite the difference, MAIAC and MODIS AOD are highly correlated. The correlation coefficient between MAIAC and MODIS AOD is 0.91 for matched pairs in our study domain.

3.2. Results of model fitting

The fixed effects of model fitting are shown in Table 3. The intercept and all the independent variables in the MAIAC model are statistically significant at $\alpha = 0.05$ level. The fixed slopes of the independent variables indicate that AOD, point emission, and road length have a positive relationship with $PM_{2.5}$ concentrations (positive *b* values), while wind speed, elevation, and forest cover show a negative association with $PM_{2.5}$ exposure (negative *b* values). This is attributed to several factors. AOD values are related to the number of particles in the air, point emissions indicate the amount of near-surface particle emissions, and thus they show a positive relationship with ground-level $PM_{2.5}$ concentrations. Road length has a positive association with $PM_{2.5}$ exposure because it is related to the amount of vehicle emissions. Elevation is negatively related to $PM_{2.5}$. In general, locations at higher altitude are less populated and the higher altitude makes pollution dispersion easier due to relatively higher wind speed, $PM_{2.5}$, however, tends to concentrate in valleys as a result of the relatively closed structure and reduced horizontal mixing. A higher percentage of forest cover implies that there are fewer emission sources such as industries, traffic, and population, which lowers $PM_{2.5}$ concentrations. In addition, a high wind speed can increase horizontal and vertical mixing, therefore diluting $PM_{2.5}$ concentrations (Chudnovsky et al., 2012; Liu, Franklin, et al., 2007).

3.3. Results of model validation

The coefficient of determination (R^2), MPE, and RMSPE of our model are given in Table 4. The results show that R^2 is relatively high, and MPE and RMSPE remain low, indicating that the estimates made from both model fitting and cross validation agree well with the observed values. The results also show that model over-fitting is present; that is, in the first stage from model fitting to cross validation, R^2 decreased 0.07 for both MAIAC and MODIS; MPE increased 0.23 μg/m^3 for MAIAC and 0.25 μg/m^3 for MODIS; and RMSPE increased 0.36 μg/m^3 for MAIAC and 0.38 μg/m^3 for MODIS. Model over-fitting became more severe when the second stage GWR model was incorporated because of the limited number of matched data records per day. From model fitting to cross validation, R^2 decreased 0.16 for MAIAC and 0.14 for MODIS; MPE increased 0.65 μg/m^3 for MAIAC and 0.63 μg/m^3 for MODIS; and RMSPE increased 1.15 μg/m^3 for MAIAC and 1.07 μg/m^3 for MODIS. However, the overall prediction accuracy was improved when the second stage GWR model was incorporated. From the first stage to the second stage, CV R^2 increased 0.03 for both MAIAC and MODIS; CV MPE decreased 0.27 μg/m^3 for MAIAC and 0.28 μg/m^3 for MODIS; and CV RMSPE decreased 0.05 μg/m^3 for MAIAC and 0.09 μg/m^3 for MODIS, indicating that the GWR model captures the spatial variability in the $PM_{2.5}$–AOD relationship. In addition, Fig. 3 shows that when the minimum number of matched data records per day increased from four to eight (we used three as the minimum number in this analysis), overall CV RMSPE decreased 0.17 μg/m^3 for MAIAC and 0.23 μg/m^3 for MODIS; 0.29 μg/m^3 for MAIAC and 0.28 μg/m^3 for MODIS; 0.28 μg/m^3 for MAIAC and 0.30 μg/m^3 for MODIS; 0.35 μg/m^3 for MAIAC and 0.33 μg/m^3 for MODIS; 0.34 μg/m^3 for MAIAC and 0.36 μg/m^3 for MODIS, respectively. Overall CV R^2 increased 0.02 for both MAIAC and MODIS; 0.04 for MAIAC and 0.03 for MODIS; 0.04 for both MAIAC and MODIS; 0.05 for MAIAC and 0.04 for MODIS; 0.05 for both MAIAC and MODIS, respectively. The results showed that when the minimum number of matched data records per day increased, model over-fitting was reduced, and performance improved. This indicates that with a sufficiently high number of matched data records per day, the second stage GWR model can significantly improve prediction

Table 4
Model validation.

		MAIAC (N > 2; days = 309)[a]			MODIS (N > 2; days = 279)[a]		
		R^2	MPE (μg/m^3)	RMSPE (μg/m^3)	R^2	MPE (μg/m^3)	RMSPE (μg/m^3)
Model fitting	Stage 1	0.71	2.58	3.57	0.73	2.52	3.50
	Stage 2	0.83	1.89	2.73	0.83	1.86	2.72
Cross validation	Stage 1	0.64	2.81	3.93	0.66	2.77	3.88
	Stage 2	0.67	2.54	3.88	0.69	2.49	3.79

[a] N denotes the minimum number of records per day, and stage 2 denotes the overall accuracy, including both stage 1 and stage 2.

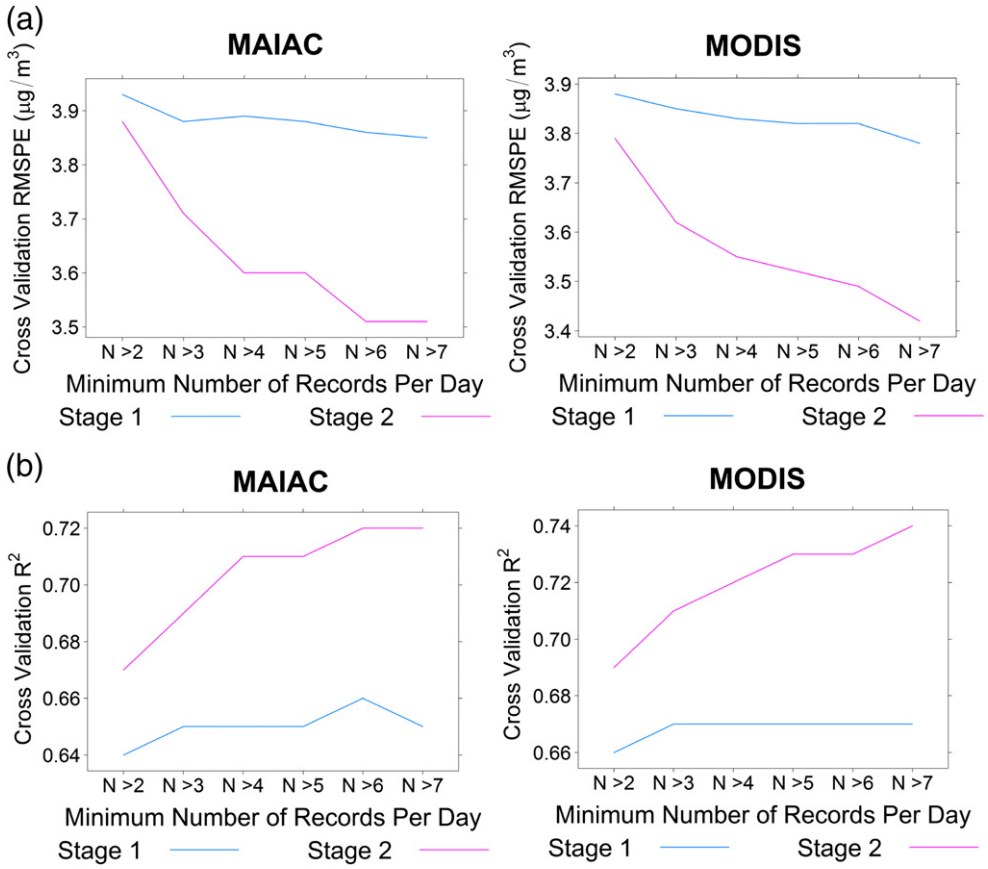

Fig. 3. The impact of minimum number of matched data records per day on model performance assessed using RMSPE (a) and R^2 (b). Stage 2 denotes the overall accuracy, including both stage 1 and stage 2.

accuracy. A regression with zero intercept (Fig. 4) was performed to fit the predicted against the observed values. The figure shows that at high concentration levels, both model fitting and cross validation under-predicted the $PM_{2.5}$ concentrations by 3–4% (e.g. fitted $PM_{2.5}$ = 96% to 97% observed $PM_{2.5}$).

3.4. Estimation of $PM_{2.5}$ concentrations

The annual mean $PM_{2.5}$ surfaces on MAIAC grid (1×1 km^2) and CMAQ grid (12×12 km^2) are shown in Fig. 5. The mean $PM_{2.5}$ concentrations estimated by MAIAC and MODIS in the study domain

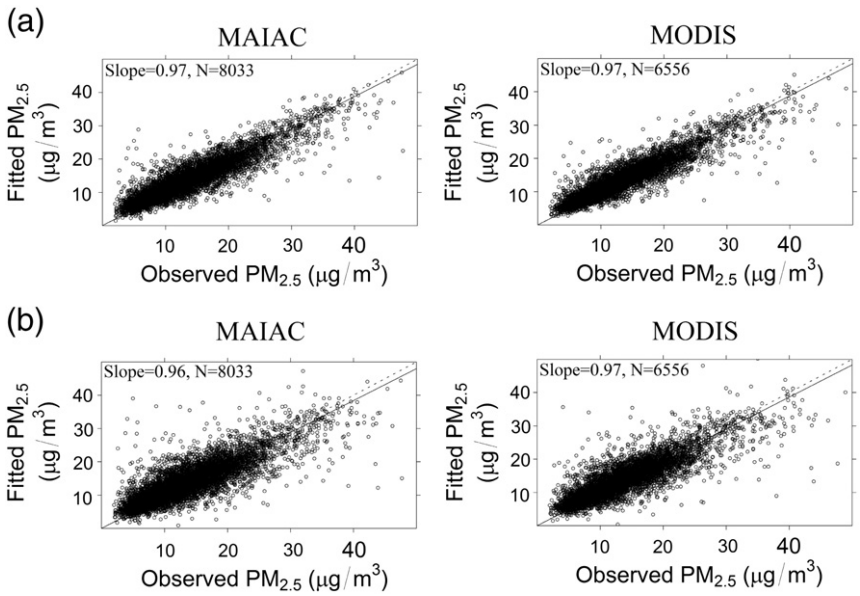

Fig. 4. Estimated vs. observed $PM_{2.5}$ concentrations for Model Fitting (a) and Cross Validation (b).

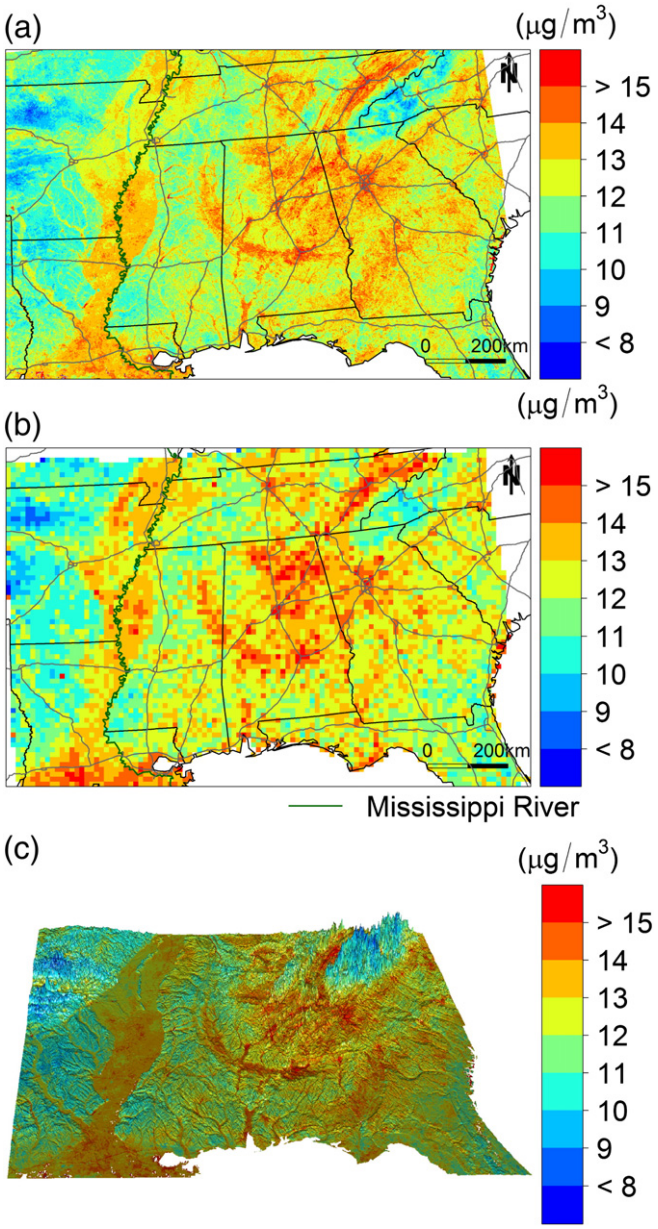

Fig. 5. Annual mean $PM_{2.5}$ estimated using MAIAC (a) and MODIS (b). 3-D $PM_{2.5}$ surface generated using MAIAC estimates (elevation values are projected as Z) (c).

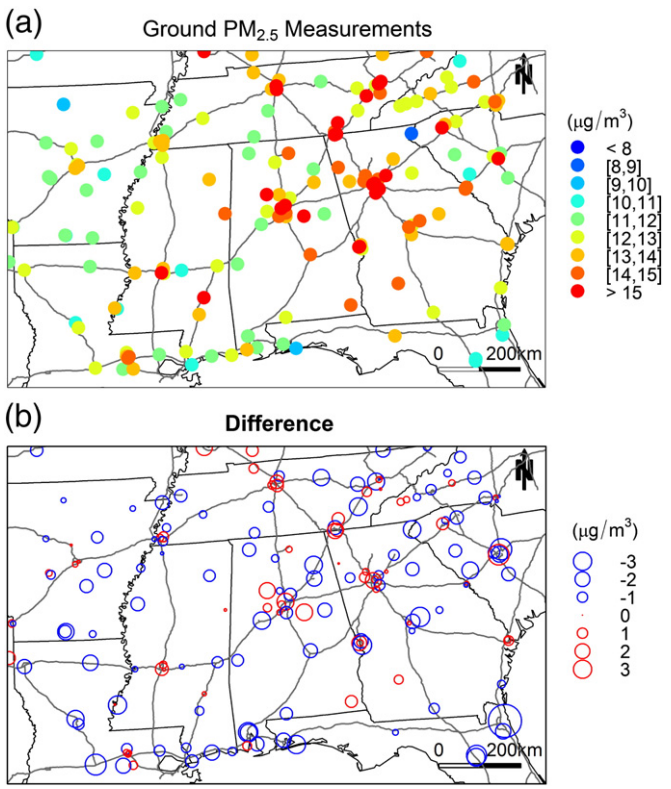

Fig. 6. (a) Annual mean ground $PM_{2.5}$ measurements at each FRM monitor; (b) The difference between observed and estimated $PM_{2.5}$ concentrations at each FRM monitor.

are 12.48 μg/m^3 and 12.30 μg/m^3, respectively. The patterns of $PM_{2.5}$ surfaces predicted by the two-stage model with MAIAC and MODIS are very similar. For example, high levels of $PM_{2.5}$ concentrations primarily appear in large urban areas and along major highways and valleys (e.g. the Mississippi river valley), while low levels occur in rural or mountainous areas. The results correspond well with land cover patterns, indicating an association between $PM_{2.5}$ levels and land cover types, which agrees with previous studies (Mao, Qiu, Kusano, & Xu, 2012). However, the 1 km MAIAC predictions can provide much finer details than the MODIS predictions. Fig. 6 illustrates the annual mean ground $PM_{2.5}$ measurements and the difference between observed and estimated $PM_{2.5}$ concentrations at each monitoring site. The results show that the pattern of ground $PM_{2.5}$ measurements corresponds well with that of our estimated concentrations, and the differences at 95% of the monitoring sites are within ± 3 μg/m^3, indicating a good agreement between observed and estimated values. Additionally, Fig. 6 shows that the FRM monitors observed high $PM_{2.5}$ concentrations in the south of our domain (e.g., southern Georgia and Alabama). The annual mean $PM_{2.5}$ measurements from five FRM monitors located in that region were 12.81, 12.46, 13.06, 14.52, and 14.95 μg/m^3, respectively. The differences between observed and estimated concentrations for those five sites are relatively small, which are −1.43, −1.02, −0.32, 0.78, and 1.32 μg/m^3, respectively. In addition, CV RMSPE, MPE, and R^2 for those five sites are 3.84 μg/m^3, 2.57 μg/m^3, and 0.66, respectively, which are similar to domain-wide accuracy.

Fig. 7a illustrates the MAIAC predictions in the Atlanta Metro area. Compared with the Urban Impervious Surface map (Fig. 7c), the MAIAC predictions show that high $PM_{2.5}$ concentrations appear in the areas with a high percentage of urban land use and along major highways, while the low concentrations appear in parks and forested areas. The MODIS predictions (Fig. 7b) cannot show this trend due to its coarser spatial resolution. Moreover, MAIAC predictions within a 12×12 km^2 CMAQ grid cell (Fig. 7d) can provide much more details (e.g., high $PM_{2.5}$ concentrations along highways) than MODIS predictions (Fig. 7e), while MAIAC can reach a similar accuracy to MODIS in $PM_{2.5}$ concentration estimation.

3.5. Seasonal patterns of $PM_{2.5}$ concentrations

Figs. 8 and 9 illustrate the seasonal mean $PM_{2.5}$ surfaces. MAIAC predicted $PM_{2.5}$ concentrations with a mean of 9.27 μg/m^3 in winter, 12.63 μg/m^3 in spring, 15.53 μg/m^3 in summer, and 12.48 μg/m^3 in fall, while MODIS estimated $PM_{2.5}$ concentrations with a mean of 8.81 μg/m^3 in winter, 12.71 μg/m^3 in spring, 16.17 μg/m^3 in summer, and 12.73 μg/m^3 in fall. The results show that $PM_{2.5}$ concentrations are the highest in summer and lowest in winter. Spring and fall $PM_{2.5}$ levels are in the intermediate range as cooler temperatures reduce the secondary $PM_{2.5}$ production. Although we expect high $PM_{2.5}$ concentrations in urban areas and along major highways, abnormally high $PM_{2.5}$ concentrations occur in southern Georgia and Alabama where there are no large urban areas or major highways. These high $PM_{2.5}$ concentrations might be caused by the fire incidents

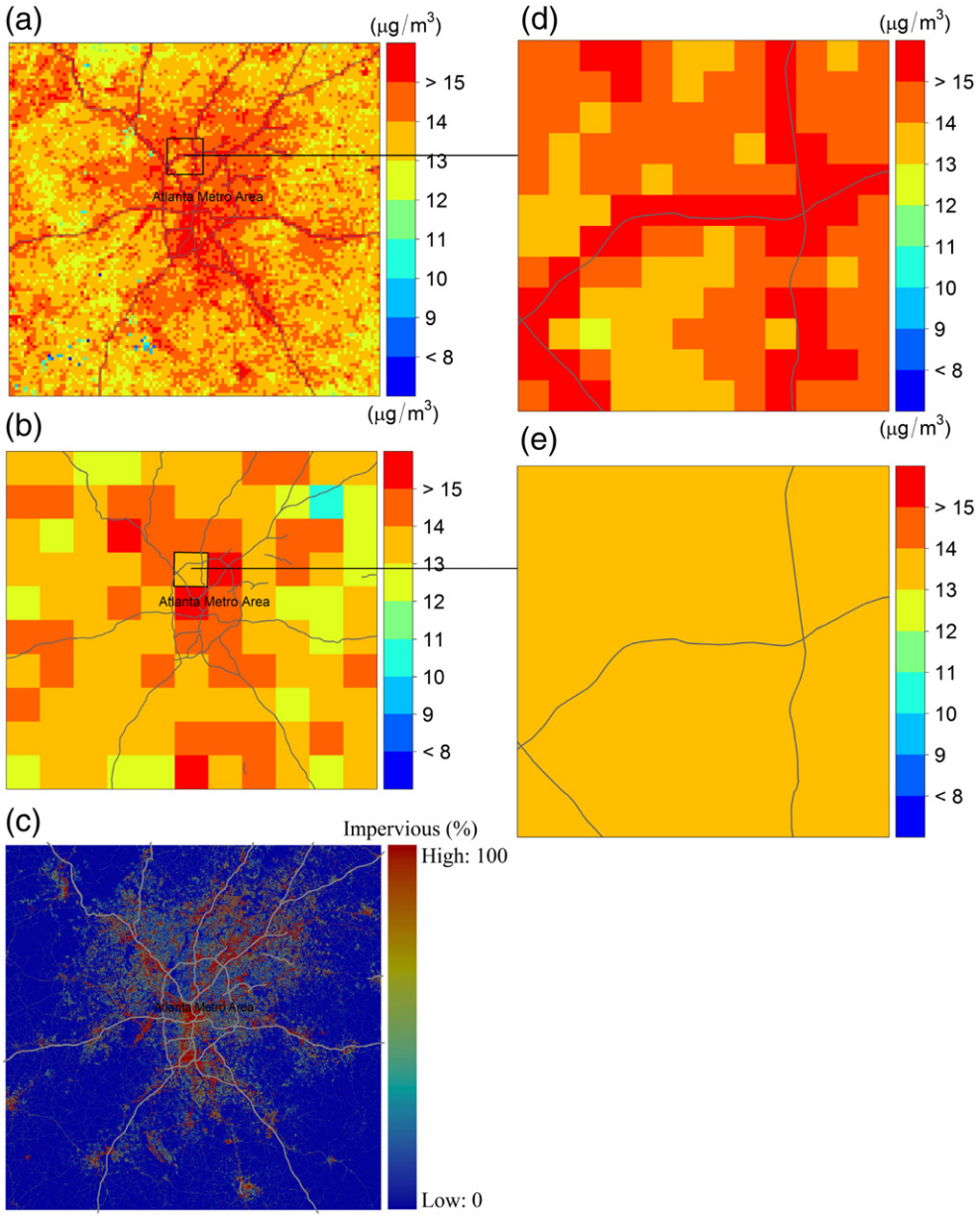

Fig. 7. Annual mean $PM_{2.5}$ for the Atlanta Metro area estimated using MAIAC (a) and MODIS (b), compared to urban built-up area (c). MAIAC estimation of $PM_{2.5}$ concentrations within a CMAQ (12 × 12 km) grid cell (d), compared to MODIS estimation in the CMAQ grid cell (e).

that occurred in the region in 2003. Zeng et al. (2008) suggested that prescribed fire emissions can result in a daily increase of $PM_{2.5}$ mass up to 25 μg/m^3, indicating that fire might have a significant impact on $PM_{2.5}$ levels. Fig. 10 was generated using the MODIS fire product, and it shows that in the spring and fall of 2003, fire incidents occurred much more frequently in the south than in the north. Correspondingly, abnormally high $PM_{2.5}$ concentrations in the south also occur in these two seasons. Meanwhile, $PM_{2.5}$ concentrations are high in most of the area in the summer, which is caused by more active generation of secondary particles near the surface due to strong solar radiation, higher temperature, and more abundant water vapor (Liu, Franklin, et al., 2007; Zheng, Cass, Schauer, & Edgerton, 2002). High $PM_{2.5}$ concentrations along the Gulf of Mexico coast in Louisiana are also observed, which are likely linked to emissions from a large number of oil refineries in Texas and Louisiana (Jarrell & Ozymy, 2010). Emissions from this area might be partly responsible for the high $PM_{2.5}$ concentrations in the southern part of our study region. Model simulations with emission sources toggled on and off are necessary to test this hypothesis, which is beyond the scope of this work. Another major emission source of fine particles in the region is agricultural emission. As reported by previous studies, ammonia (NH_3) and nitrogen oxides (NO_x) generated by agricultural activities, such as farm vehicles, domestic and farm animals, and fertilizer applications, can significantly increase the number of suspended particles (Kurvits & Marta, 1998). According to the NLCD map, cropland and pasture/hay are widely distributed in our domain such as along the Mississippi river valley, from northern Mississippi to central Alabama, and in southern Georgia and Alabama. As a result, agricultural emissions might be another critical factor responsible for elevated $PM_{2.5}$ levels in those regions. However, some of the high estimates might be due to bias coming from AOD retrieval algorithms.

3.6. The impact of AOD on model fitting

In order to test if AOD helps improve predictions relative to just using the other variables, we fitted the two-stage model without AOD.

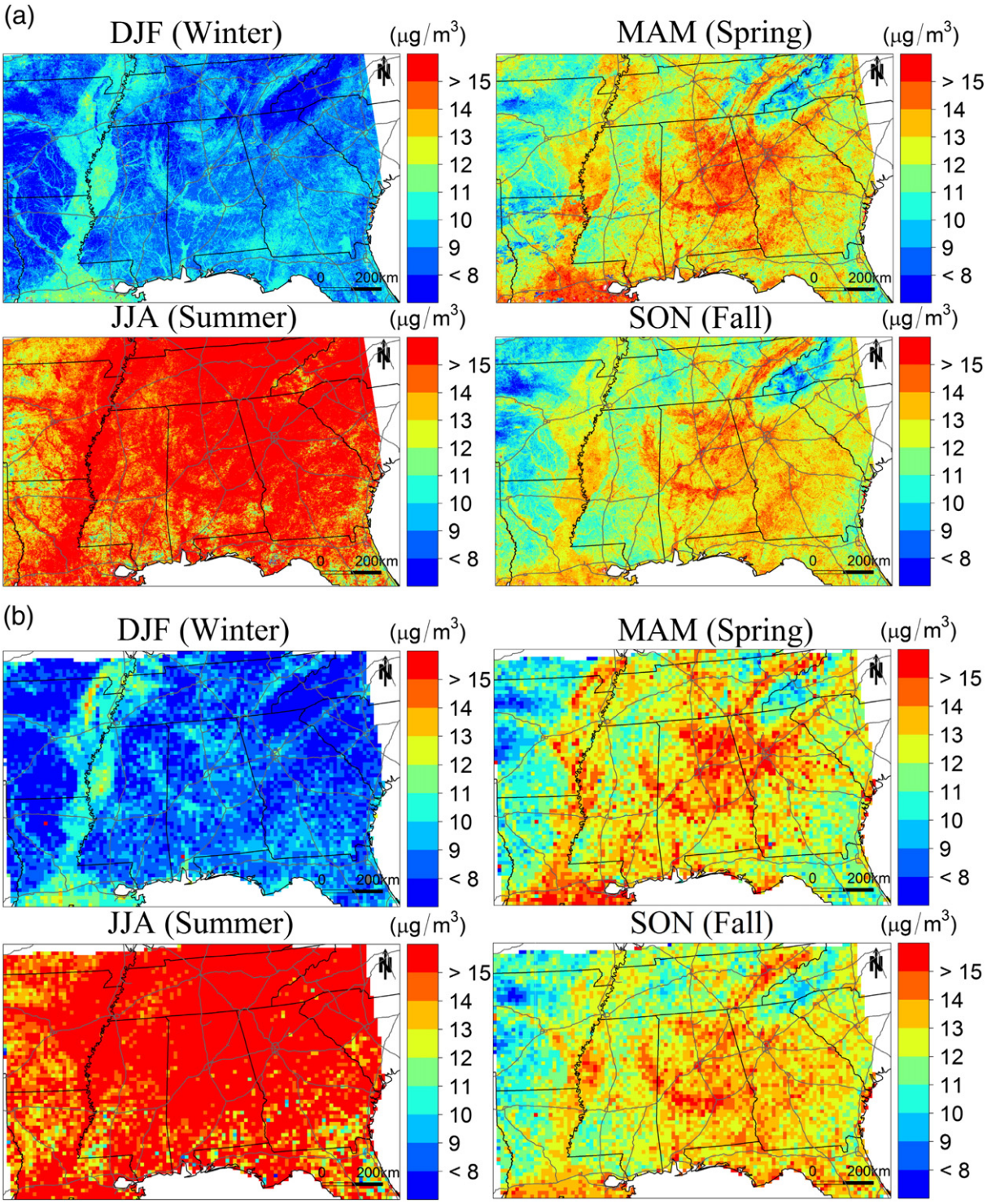

Fig. 8. Seasonal mean $PM_{2.5}$ estimated using MAIAC (a) and MODIS (b).

For the first stage, all the predictor variables remained in the model except AOD. For the second stage GWR model, wind speed, forest cover, major road, elevation, and point emissions were individually used to replace AOD in model fitting. The results (Table 5) showed that overall CV RMSPE increased 0.97 μg/m^3 from the MAIAC model and 1.06 μg/m^3 from the MODIS model, 0.38 μg/m^3 from the MAIAC model and 0.47 μg/m^3 from the MODIS model, 0.89 μg/m^3 from the MAIAC model and 0.98 μg/m^3 from the MODIS model, and 1.14 μg/m^3 from the MAIAC model and 1.23 μg/m^3 from the MODIS model for elevation, forest cover, wind speed, and major road, respectively, while overall CV R^2 decreased 0.13 from the MAIAC model and 0.15 from the MODIS model, 0.07 from the MAIAC model and 0.09 from the MODIS model, 0.12 from the MAIAC model and 0.14 from the MODIS model, 0.16 from the MAIAC model and 0.18 from the MODIS model for elevation, forest cover, wind speed, and major road, respectively. In addition, without AOD, point emissions generated extreme outliers in the distribution of the predictions and led to a significant decrease in prediction accuracy. These results suggest that AOD is essential for improving the prediction accuracy of our two-stage modeling framework.

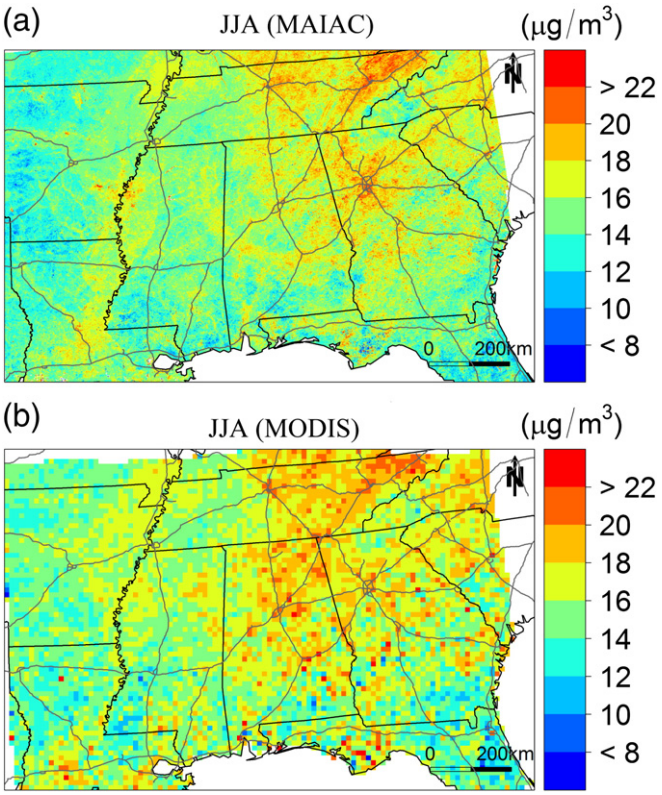

Fig. 9. Summer mean $PM_{2.5}$ estimated using MAIAC (a) and MODIS (b).

4. Discussion

The method developed in this analysis has several benefits over conventional methods such as linear regression. First, we used high spatial resolution (1 km) MAIAC AOD to estimate $PM_{2.5}$ concentrations. High spatial resolution AOD data can make accurate predictions in small grid cells, providing exposure information linked more precisely to the microenvironments of population exposure (e.g., business, industrial, and residential areas). Therefore they may be more suitable for spatially-resolved environmental health research since many epidemiological studies use health records based on small geographical regions (e.g., zip code or census block group), many of which are much smaller than the spatial resolutions of MODIS and MISR. In addition, compared to the typical size of an urban area, the spatial resolutions of MODIS and MISR are too coarse to be used for urban air pollution studies, which demand the fine scale satellite aerosol data. Our comparison between MAIAC-based and MODIS-based $PM_{2.5}$ predictions showed that MODIS estimated $PM_{2.5}$ concentrations are slightly more correlated with ground observations. The difference, however, is small. On the other hand, MAIAC provides a considerably greater spatial coverage and a larger number of AOD retrievals than MODIS. For example, the study of Chudnovsky et al. (2013) conducted in the New England region showed that MAIAC has a factor of 1.52 higher coverage of EPA sites with available PM measurements than MYD04, and the factor grows to 1.77 when only considering the coverage of EPA locations regardless of available PM data. The coverage increases because (1) MAIAC is not limited to dark surfaces, providing retrievals over brighter regions including many urban areas; (2) while MAIAC has an improved and robust detection of both cloudy and clear-sky conditions (Hilker et al., 2012), its approach to data filtering is less conservative than that of MOD04 algorithm. For instance, the study of Chudnovsky et al. (2013) revealed that on a large number

Table 5
Cross validation for models without AOD.

[a]Independent Variable	RMSPE (μg/m^3)	R^2
Elevation (m)	4.85	0.54
Forest cover	4.26	0.60
Wind speed (m/s)	4.77	0.55
Major road (m)	5.02	0.51

[a] The independent variable individually fitted in the second stage GWR model to replace AOD, and the first stage model was conducted using all the independent variables except AOD.

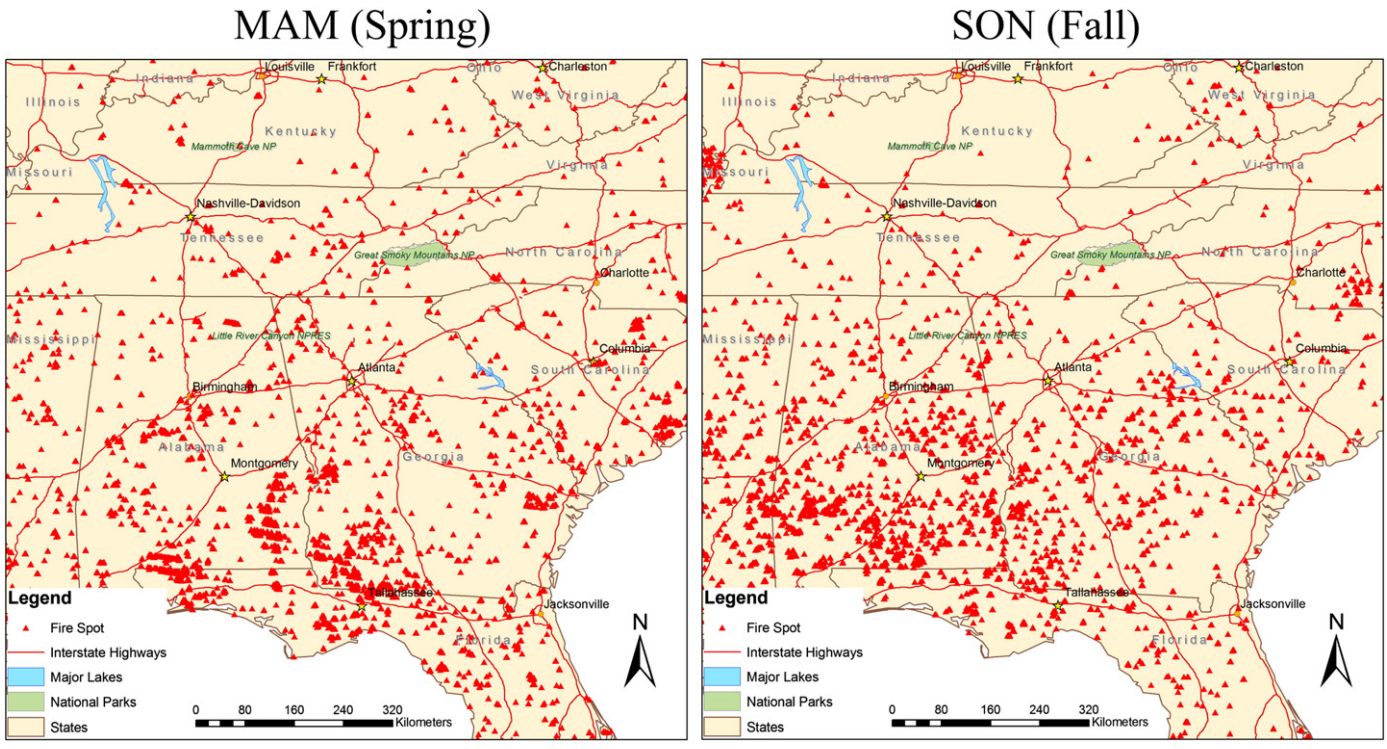

Fig. 10. Seasonal fire incidents.

(e.g., 344) of cloudy days during 2002 to 2008, standard MODIS Aqua AOD product was not available (less than two collocations with EPA sites) whereas MAIAC would provide on average eight collocations; and (3) in the MOD04 algorithm, AOD is not reported if there are fewer than twelve dark 500 m pixels in a 20 × 20 pixel box, which becomes restrictive in partly cloudy conditions or over brighter surfaces. It should be noted that the 20 × 20 pixel box corresponding to a 10 × 10 km^2 area at nadir expands to ~20 × 40 km^2 at the edge of MODIS scan due to pixel's footprint growth by approximately a factor of 2 × 4 (Wolfe, Roy, & Vermote, 1998). At the same time, MAIAC grid resolution of 1 km remains the same regardless of the MODIS scan angle. Since the resolution of the original MODIS land bands is 500 m at nadir, MAIAC "under-samples" AOD by a factor of 4 at nadir as compared to "potential" information which could be derived from 500 m measurements. At the edge of scan, MAIAC 1 km product "over-samples" AOD by a factor of 2. Our analysis shows that this "over-sampling" does not create a problem as aerosol retrievals are robust at the edge of scan due to high air mass producing generally smooth AOD distributions with least artifacts from spatially variable surface. In addition, our experience with MAIAC does not show any noticeable increase in the AOD retrieval error at high view zenith angle (VZA) due to cloud contamination. Second, we used a two-stage model incorporating both a linear mixed effects model and a GWR model to account for temporal as well as spatial variability in the $PM_{2.5}$–AOD relationship. The linear mixed effects model allows for day-to-day variability in the relationship by incorporating daily variation as a random effect, while the GWR model can effectively capture the spatial variability.

A limitation of the developed approach is the lack of a method to fill the gaps in areas where AOD is not retrieved. The lack of AOD data in the operational products is usually caused by the presence of clouds or high surface reflectance and is a generic feature of all AOD products. Several empirical gap-filling methods have been developed to alleviate this problem (Kloog et al., 2011). However, in this paper, our main objectives were to develop a two-stage model that can account for both temporal and spatial variability in the $PM_{2.5}$–AOD relationship and demonstrate the ability of the 1 km MAIAC AOD product as the primary estimator of $PM_{2.5}$ concentrations. Filling the missing data gaps using statistical approaches will inevitably introduce additional measurement errors and complicate result interpretation. Hence it was not pursued in the analysis. Another limitation comes from the number of records per day, since our second stage GWR model was implemented on a daily basis. Too few observations may lead to model over-fitting and reduce prediction accuracy. In the meantime, we attempt to account for as many days as possible in the analysis to calculate an annual prediction. Thus, a trade-off between number of days and minimum number of records per day needs to be made. In this paper, a minimum number of three records per day was selected as the threshold in order to both include a sufficient number of days and maintain prediction accuracy. Although the increase is limited due to model over-fitting when the threshold is three, our results further show that as the minimum number of records per day increases, the prediction accuracy also increases. This indicates that as long as there are a sufficiently high number of observations, our second stage GWR model can improve the prediction accuracy.

5. Conclusions

This paper demonstrates the feasibility of using 1 km spatial resolution MAIAC AOD data to estimate ground-level $PM_{2.5}$ concentrations using a two-stage model. The results show that the overall accuracy of MAIAC predicted $PM_{2.5}$ concentrations at 1 km resolution is comparable with MODIS predicted $PM_{2.5}$ concentrations at 12 km resolution. Both satellite-driven models point out interesting features of the $PM_{2.5}$ spatial distribution in the southeastern U.S. and their possible causes, which warrant further analysis in conjunction with an air quality model simulation. In a smaller area, the high spatial resolution of MAIAC AOD product has substantial advantages over MODIS by offering more spatially refined contrasts of $PM_{2.5}$ levels that track fine-scale land use patterns closely. As MAIAC AOD data go back to 2000 and are available almost twice a day, it has the great potential to serve $PM_{2.5}$ health effects studies nationwide related to both chronic and acute exposures.

In future studies, we will focus on four aspects. First, we will develop new statistical models and introduce additional estimators to fill the gaps in areas where AOD is not retrieved. For example, this can be implemented based on prior knowledge of AOD distribution in background conditions from the time series of MAIAC data. Hierarchical Bayesian models offer an attractive analytic framework for addressing both our temporal random effects and spatially-varying coefficients but at a higher computational cost for data sets such as ours. We will investigate the inferential and implementation constraints in both approaches. A second focus will involve conducting a time series analysis of $PM_{2.5}$ concentrations estimated at 1 km spatial resolution to facilitate epidemiological studies about the impact of air pollution on public health issues. Third, we will examine the impact of aerosol vertical profiles on $PM_{2.5}$ concentration estimation by including model simulated vertical profiles in our statistical models. Finally, since our goal is to demonstrate the performance of MAIAC AOD and the benefit of its high spatial resolution, we did not consider the non-random missingness in AOD values, which might bias the regression coefficient estimates in the first stage model. We will address this problem in future studies.

Acknowledgements

This work was partially supported by NASA Applied Sciences Program (grant no. NNX09AT52G). In addition, this publication was made possible by USEPA grant R834799. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication.

References

Chudnovsky, A. A., Kostinski, A., Lyapustin, A., & Koutrakis, P. (2012). Spatial scales of pollution from variable resolution satellite imaging. *Environmental Pollution, 172*, 131–138.

Chudnovsky, A. A., Tang, C., Lyapustin, A., Wang, Y., Schwartz, J., & Koutrakis, P. (2013). A critical assessment of high resolution aerosol optical depth (AOD) retrievals for fine particulate matter (PM) predictions. *Atmospheric Chemistry and Physics Discussions, 13*, 14581–14611.

Cosgrove, B.A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., et al. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research-Atmospheres, 108*.

Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L., et al. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA: The Journal of the American Medical Association, 295*, 1127–1134.

Drury, E., Jacob, D. J., Spurr, R. J.D., Wang, J., Shinozuka, Y., Anderson, B. E., et al. (2010). Synthesis of satellite (MODIS), aircraft (ICARTT), and surface (IMPROVE, EPA-AQS, AERONET) aerosol observations over eastern North America to improve MODIS aerosol retrievals and constrain surface aerosol concentrations and sources. *Journal of Geophysical Research-Atmospheres, 115*.

Engel-Cox, J. A., Holloman, C. H., Coutant, B. W., & Hoff, R. M. (2004). Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment, 38*, 2495–2509.

EPA (2008). Quality assurance handbook for air pollution measurement systems. *Ambient air quality monitoring program, vol. II*, : U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division (RTP, NC 27711. In).

Fotheringham, A., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships.* : John Wiley & Sons Inc.

Gauderman, W. J., Avol, E., Gilliland, F., Vora, H., Thomas, D., Berhane, K., et al. (2004). The effect of air pollution on lung development from 10 to 18 years of age. *The New England Journal of Medicine, 351*, 1057–1067.

Gold, D. R., Litonjua, A., Schwartz, J., Lovett, E., Larson, A., Nearing, B., et al. (2000). Ambient pollution and heart rate variability. *Circulation, 101*, 1267–1273.

Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *Journal of Geophysical Research-Atmospheres, 114*.

Hilker, T., Lyapustin, A. I., Tucker, C. J., Sellers, P. J., Hall, F. G., & Wang, Y. (2012). Remote sensing of tropical ecosystems: Atmospheric correction and cloud masking matter. *Remote Sensing of Environment*, *127*, 370–384.

Hoff, R. M., & Christopher, S. A. (2009). Remote sensing of particulate pollution from space: Have we reached the promised land? *Journal of the Air & Waste Management Association*, *59*, 645–675.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., et al. (1998). AERONET — A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*, 1–16.

Hu, Z. Y. (2009). Spatial analysis of MODIS aerosol optical depth, PM2.5, and chronic coronary heart disease. *International Journal of Health Geographics*, *8*.

Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G., Jr., Estes, S. M., et al. (2013). Estimating ground-level PM2.5 concentrations in the southeastern U.S. using geographically weighted regression. *Environmental Research*, *121*, 1–10.

Jarrell, M. L., & Ozymy, J. (2010). Excessive air pollution and the oil industry: Fighting for our right to breathe clean air. *Environmental Justice*, *3*, 111–115.

Kloog, I., Koutrakis, P., Coull, B.A., Lee, H. J., & Schwartz, J. (2011). Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment*, *45*, 6267–6275.

Kurvits, T., & Marta, T. (1998). Agricultural NH3 and NOx emissions in Canada. *Environmental Pollution*, *102*, 187–194.

Laden, F., Schwartz, J., Speizer, F., & Dockery, D. (2006). Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities study. *American Journal of Respiratory and Critical Care Medicine*, *173*, 667–672.

Lee, H. J., Liu, Y., Coull, B.A., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict PM2.5 concentrations. *Atmospheric Chemistry and Physics*, *11*, 7991–8002.

Liu, Y., Franklin, M., Kahn, R., & Koutrakis, P. (2007a). Using aerosol optical thickness to predict ground-level PM2.5 concentrations in the St. Louis area: A comparison between MISR and MODIS. *Remote Sensing of Environment*, *107*, 33–44.

Liu, Y., Koutrakis, P., & Kahn, R. (2007). Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: Part 1 — Method development. *Journal of the Air & Waste Management Association*, *57*, 1351–1359.

Liu, Y., Koutrakis, P., Kahn, R., Turquety, S., & Yantosca, R. M. (2007). Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: Part 2 — A case study. *Journal of the Air & Waste Management Association*, *57*, 1360–1369.

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. *Environmental Health Perspectives*, *117*, 886–892.

Liu, Y., Sarnat, J. A., Kilaru, A., Jacob, D. J., & Koutrakis, P. (2005). Estimating ground-level PM2.5 in the eastern united states using satellite remote sensing. *Environmental Science & Technology*, *39*, 3269–3278.

Lyapustin, A., Korkin, S., Wang, Y., Quayle, B., & Laszlo, I. (2012). Discrimination of biomass burning smoke and clouds in MAIAC algorithm. *Atmospheric Chemistry & Physics Discussions*, *12*, 18651–18670.

Lyapustin, A., Martonchik, J., Wang, Y. J., Laszlo, I., & Korkin, S. (2011). Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *Journal of Geophysical Research-Atmospheres*, *116*.

Lyapustin, A. I., Wang, Y., Laszlo, I., Hilker, T., Hall, F. G., Sellers, P. J., et al. (2012b). Multi-angle implementation of atmospheric correction for MODIS (MAIAC): 3. Atmospheric correction. *Remote Sensing of Environment*, *127*, 385–393.

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., et al. (2011). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research-Atmospheres*, *116*.

Mao, L., Qiu, Y. L., Kusano, C., & Xu, X. H. (2012). Predicting regional space-time variation of PM2.5 with land-use regression model and MODIS data. *Environmental Science and Pollution Research*, *19*, 128–138.

Paciorek, C. J., Liu, Y., Moreno-Macias, H., & Kondragunta, S. (2008). Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level PM2.5. *Environmental Science & Technology*, *42*, 5800–5806.

Peters, A., Dockery, D. W., Muller, J. E., & Mittleman, M.A. (2001). Increased particulate air pollution and the triggering of myocardial infarction. *Circulation*, *103*, 2810–2815.

Samet, J. M., Dominici, F., Curriero, F. C., Coursac, I., & Zeger, S. L. (2000). Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. *The New England Journal of Medicine*, *343*, 1742–1749.

Schwartz, J., & Neas, L. M. (2000). Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren. *Epidemiology*, *11*, 6–10.

van Donkelaar, A., Martin, R. V., Spurr, R. J.D., Drury, E., Remer, L. A., Levy, R. C., et al. (2013). Optimal estimation for global ground-level fine particulate matter concentrations. *Journal of Geophysical Research-Atmospheres*, *118*, 5621–5636.

Wallace, J., Kanaroglou, P., & Ieee (2007). An investigation of air pollution in southern Ontario, Canada, with MODIS and MISR aerosol data. *IGARSS: 2007 IEEE International Geoscience and Remote Sensing Symposium. Sensing and understanding our planet, vol. 1–12*. (pp. 4311–4314). New York: Ieee.

Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies. *Geophysical Research Letters*, *30*, 2095.

Wang, J., Xu, X., Spurr, R., Wang, Y., & Drury, E. (2010). Improved algorithm for MODIS satellite retrievals of aerosol optical thickness over land in dusty atmosphere: Implications for air quality monitoring in China. *Remote Sensing of Environment*, *114*, 2575–2583.

Wolfe, R. E., Roy, D. P., & Vermote, E. (1998). MODIS land data storage, gridding, and compositing methodology: Level 2 grid. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1324–1338.

Zeng, T., Wang, Y., Yoshida, Y., Tian, D., Russell, A. G., & Barnard, W. R. (2008). Impacts of prescribed fires on air quality over the Southeastern United States in spring based on modeling and ground/satellite measurements. *Environmental Science & Technology*, *42*, 8401–8406.

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. *Journal of the Air & Waste Management Association*, *59*, 1358–1369.

Zhang, Y., Yu, H., Eck, T. F., Smirnov, A., Chin, M., Remer, L. A., et al. (2012). Aerosol daytime variations over North and South America derived from multiyear AERONET measurements. *Journal of Geophysical Research-Atmospheres*, *117*, D05211.

Zheng, M., Cass, G. R., Schauer, J. J., & Edgerton, E. S. (2002). Source apportionment of PM2.5 in the Southeastern United States using solvent-extractable organic compounds as tracers. *Environmental Science & Technology*, *36*, 2361–2371.