Environmental Research 121 (2013) 1–10

Contents lists available at SciVerse ScienceDirect

Environmental Research

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

Estimating ground-level $PM_{2.5}$ concentrations in the southeastern U.S. using geographically weighted regression ☆

Xuefei Hu [a], Lance A. Waller [b], Mohammad Z. Al-Hamdan [c], William L. Crosson [c], Maurice G. Estes Jr. [c], Sue M. Estes [c], Dale A. Quattrochi [d], Jeremy A. Sarnat [a], Yang Liu [a,*]

[a] Department of Environmental Health, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322, USA
[b] Department of Biostatistics & Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA
[c] Universities Space Research Association at the National Space Science and Technology Center, NASA Marshall Space Flight Center, Huntsville, AL 35805, USA
[d] Earth Science Office at NASA Marshall Space Flight Center, National Space Science and Technology Center, Huntsville, AL 35805, USA

ARTICLE INFO

Article history:
Received 31 August 2011
Received in revised form
21 June 2012
Accepted 13 July 2012
Available online 6 December 2012

Keywords:
$PM_{2.5}$
MODIS
Remote sensing
Aerosol optical depth
Geographically weighted regression

ABSTRACT

Most of currently reported models for predicting $PM_{2.5}$ concentrations from satellite retrievals of aerosol optical depth are global methods without considering local variations, which might introduce significant biases into prediction results. In this paper, a geographically weighted regression model was developed to examine the relationship among $PM_{2.5}$, aerosol optical depth, meteorological parameters, and land use information. Additionally, two meteorological datasets, North American Regional Reanalysis and North American Land Data Assimilation System, were fitted into the model separately to compare their performances. The study area is centered at the Atlanta Metro area, and data were collected from various sources for the year 2003. The results showed that the mean local R^2 of the models using North American Regional Reanalysis was 0.60 and those using North American Land Data Assimilation System reached 0.61. The root mean squared prediction error showed that the prediction accuracy was 82.7% and 83.0% for North American Regional Reanalysis and North American Land Data Assimilation System in model fitting, respectively, and 69.7% and 72.1% in cross validation. The results indicated that geographically weighted regression combined with aerosol optical depth, meteorological parameters, and land use information as the predictor variables could generate a better fit and achieve high accuracy in $PM_{2.5}$ exposure estimation, and North American Land Data Assimilation System could be used as an alternative of North American Regional Reanalysis to provide some of the meteorological fields.

1. Introduction

Many recent epidemiological studies have shown that fine particles ($PM_{2.5}$; particles having aerodynamic diameter less than 2.5 μm), which in populated regions are emitted primarily from anthropogenic and biogenic sources, are associated with various health outcomes, including increased risk of cardiovascular and respiratory diseases (Dominici et al., 2006), myocardial infarction (Peters et al., 2001), and significantly reduced heart rate variability (Gold et al., 2000). Furthermore, $PM_{2.5}$ exposure can cause respiratory problems and deficits in lung development in children (Gauderman et al., 2004; Schwartz and Neas, 2000). Long-term $PM_{2.5}$ concentration monitoring and accurate $PM_{2.5}$ exposure prediction are crucial to air quality assessment and to address public health concerns.

Although many epidemiological studies have used measurements from stationary ambient monitoring sites as surrogates for personal exposure of $PM_{2.5}$ (Ito et al., 2001; Pope et al., 2002), the number of such sites is limited and its distribution is often sparse and unbalanced, which makes continuous spatial monitoring difficult. Remote sensing technology provides a new way to generate an extensive coverage of $PM_{2.5}$ monitoring within a study area at various scales by using satellite-retrieved aerosol optical depth. Aerosol optical depth is a measure of the degree to which aerosols prevent light from penetrating the atmosphere. In addition, aerosol optical depth retrieved using visible channels is most sensitive to particles with

Abbreviations: NARR, North American Regional Reanalysis; NLDAS, North American Land Data Assimilation System; MPE, Mean Prediction Error; RMSPE, Root Mean Squared Prediction Error; CMAQ, Community Multiscale Air Quality; OLS, Ordinary Least Squares; NASA, National Aeronautics and Space Administration; MODIS, Moderate Resolution Imaging Spectroradiometer; MISR, Multiangle Imaging SpectroRadiometer; GASP, Geostationary Operational Environmental Satellite Aerosol/Smoke Product; SEARCH, the Southeastern Aerosol Research and Characterization Study Experiment

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

☆ This project is supported and funded by NASA Applied Sciences Public Health Program managed by John Haynes (grant number NNX09AT52G).

0013-9351/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.envres.2012.11.003

sizes from 0.1 to 2 μm (Kahn et al., 1998) and can be used to measure loadings of fine particles. To date, many studies have examined the linkage between ground-level $PM_{2.5}$ concentrations and aerosol optical depth retrieved from various satellite sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Liu et al., 2007a; Zhang et al., 2009), the Multiangle Imaging SpectroRadiometer (MISR) (Liu et al., 2007a, 2007b, 2007c), and the Geostationary Operational Environmental Satellite Aerosol/Smoke Product (GASP) (Liu et al., 2009; Paciorek et al., 2008).

Some previous studies predicted surface $PM_{2.5}$ concentrations by establishing the direct relationship between $PM_{2.5}$ and aerosol optical depth with aerosol optical depth being the only independent variable (Hu, 2009; Schaap et al., 2009), while others estimated ground-level $PM_{2.5}$ concentrations using satellite aerosol optical depth in conjunction with meteorological fields (e.g., boundary layer height, relative humidity, air temperature, and wind speed) (Paciorek et al., 2008) and land use information (e.g., road length, land use types, and population density) (Liu et al., 2009). Those meteorological and land use variables have been recognized as effective predictors of $PM_{2.5}$ and can significantly improve the model predictability (Liu et al., 2005, 2007a, 2009; Tian and Chen, 2010).

Numerous studies used linear regression models to examine the association between $PM_{2.5}$ measured at fixed-site monitors and aerosol optical depth (Kumar et al., 2007; Liu et al., 2005; Schaap et al., 2009; Schafer et al., 2008; Wallace et al., 2007). Previous studies showed that the correlations between $PM_{2.5}$ and aerosol optical depth are spatially non-stationary (Engel-Cox et al., 2004; Hu, 2009). The spatial nonstationarity denotes that the relationships between dependent and independent variables are not constant across space and change with spatial context. The spatial nonstationarity led to a poorer performance of models that used globally fixed parameters, which are based on an assumption that the relationship between $PM_{2.5}$ and aerosol optical depth does not vary spatially, than local models such as the geographically weighted regression (Wang et al., 2005; Zhao et al., 2010). Geographically weighted regression is a technique that can examine the spatial variability and nonstationarity by producing local regression results (Fotheringham et al., 1998). To date, few studies have developed geographically weighted regression models to estimate $PM_{2.5}$ concentrations. Hu (2009) used geographically weighted regression to generate a $PM_{2.5}$ surface by establishing a local relationship between $PM_{2.5}$ and aerosol optical depth. However, a drawback of this study was that aerosol optical depth was the only predictor variable in the model without accounting for the meteorological parameters and land use information. As discussed above, many meteorological variables can substantially affect the relationship between $PM_{2.5}$ and aerosol optical depth and are significant predictors of $PM_{2.5}$ concentrations. Neglecting those parameters might reduce the prediction accuracy and increase the bias, and thus should be avoided.

The first objective of this paper is to establish a quantitative local relationship between ground-level $PM_{2.5}$ and aerosol optical depth using geographically weighted regression together with meteorological parameters and land use information. The local regression model is cross-validated and used to predict ground-level $PM_{2.5}$ concentrations within the study domain. A $PM_{2.5}$ surface is then derived from the predictions to illustrate the distribution of ground-level $PM_{2.5}$ concentrations within the study area. The second objective is to compare the impacts of meteorological fields from two different datasets, the North American Regional Reanalysis (NARR) and the North American Land Data Assimilation System (NLDAS), on the $PM_{2.5}$ concentration predictions to examine the performance of NLDAS, given its higher temporal and spatial resolution than NARR.

2. Materials and methods

2.1. Study area

The study area for this analysis is approximately 750 km by 750 km, covering most of Georgia, Tennessee, and Alabama and small portions of Kentucky, West Virginia, North and South Carolina, and Florida, centered at the Atlanta metro area (Fig. 1).

2.2. Materials

2.2.1. EPA $PM_{2.5}$ measurements

The 24-h average EPA $PM_{2.5}$ concentrations in the study area for year 2003 were downloaded from the EPA's Air Quality System Technology Transfer Network. The data were collected from the federal reference monitors within the study region. The measured $PM_{2.5}$ concentrations were used as the dependent

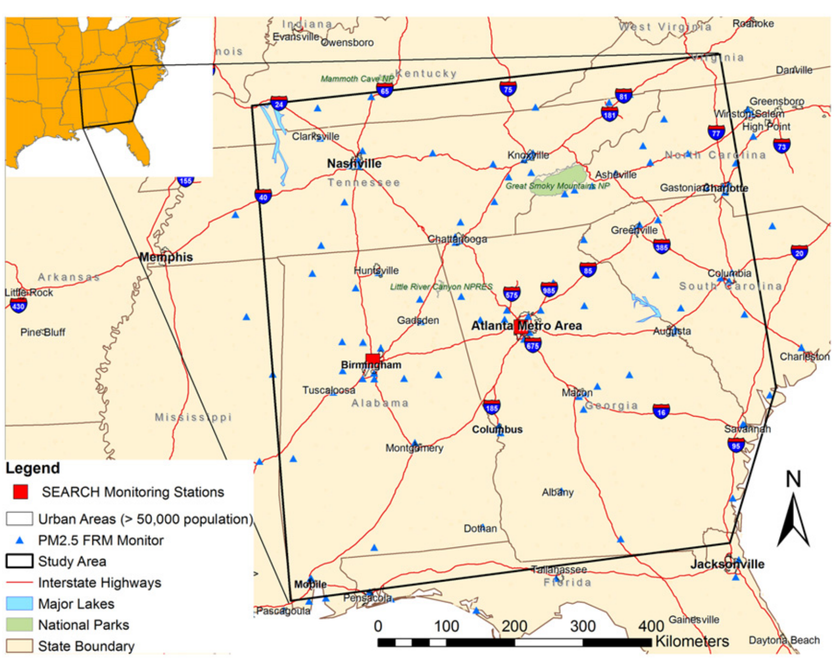

Fig. 1. Study area.

variable in the model, and the $PM_{2.5}$ concentrations less than 2 μg/m^3 (3.6% of total records) were discarded as they are below the established limit of detection (EPA, 2008) to avoid introducing errors into the model. There are a total of 119 $PM_{2.5}$ monitoring sites within the study domain in 2003. Although some of the monitoring sites are close to each other, we did not combine them because the number of monitoring sites is very limited, and combining them will reduce the number of records. Moreover, geographically weighted regression modeling is based on distance between points. Thus, combining original sites might alter the relationship between $PM_{2.5}$ and the predictors for each regression point. Only 7.5% of the total data points fall inside the same grid cell on the same day, and spatially averaging these sites does not change our findings in any significant way. In addition, forest cover is one of the predictors and is calculated within 1 km radius of each station (900 m by 900 m). As a result, merging nearby stations will reduce the resolution of this parameter.

2.2.2. Aerosol optical depth data

MODIS is an instrument aboard the Terra and Aqua satellites operated by National Aeronautics and Space Administration (NASA) (Remer et al., 2005). MODIS measures the abundance of atmospheric particles on the global scale at a moderate spatial resolution (10 km). The 2003 MODIS aerosol data (collection 5) at 550 nm wavelength were downloaded from the Earth Observing System Data Gateway at the Goddard Space Flight Center (http://delenn.gsfc.nasa.gov/-imswww/pub/imswelcome). Because MODIS aerosol optical depth pixels shift in space between orbits, it is difficult to compare model predictions with ground $PM_{2.5}$ measurements. Therefore, a base grid is needed for prediction. We re-sampled these data to the 12 km Community Multiscale Air Quality (CMAQ) grid using a nearest neighbor approach. CAMQ is a commonly used grid that results in little sacrifice of resolution. The re-sampling might introduce additional variability. However, both CMAQ and MODIS have similar spatial resolutions (12 km and 10 km, respectively), and the study domain is large (750 km by 750 km). Thus, the variability should be relatively small. In case a grid cell had aerosol optical depth observations from both Terra and Aqua, their average value was assigned for that grid cell. In 2003, that case only happened for 10.2% of the days on average. On the other hand, grid cells had either only Terra observations or only Aqua observations in 14.1% and 13.1% of the days, respectively. On average, grid cells had no observations at all for 62.6% of the days, due to cloud coverage.

2.2.3. Meteorological fields

The National Centers for Environmental Prediction has released the NARR dataset, a long term, consistent, high-resolution climate dataset for North America (Mesinger et al., 2006). The NARR project is an extension of the National Centers for Environmental Prediction Global Reanalysis which is run over the North American Region (http://www.emc.ncep.noaa.gov/mmb/rreanl/). The NARR model uses the high resolution Eta Model (32 km/45 layer) (Black, 1988; Janjic, 1994; Mesinger et al., 1988) together with the Regional Data Assimilation System which assimilates precipitation along with other variables. The Eta model is a state-of-the-art atmospheric model initially developed in the 1970s. It was formally adopted for operation by the National Centers for Environmental Prediction in 1993. The improvements in the model/assimilation have resulted in a dataset with substantial improvements in the accuracy of temperature, winds and precipitation compared to the earlier National Centers for Environmental Prediction—Department of Energy Global Reanalysis 2 (Kanamitsu et al., 2002). The spatial resolution of the NARR meteorological data is approximately 32 km. The data set contains parameters such as boundary layer height, relative humidity, air temperature, and wind speed at 3-h intervals. The daily meteorological data were obtained by averaging the 3-h NARR measurements sampled from 10 a.m. to 4 p.m. local time to cover the satellites overpass time.

The NLDAS (Phase 2) meteorological data were downloaded from the NLDAS website (http://ldas.gsfc.nasa.gov/nldas/). The NLDAS provided quality controlled, spatially and temporally consistent, real-time, and retrospective forcing datasets (Cosgrove et al., 2003). The non-precipitation land-surface forcing fields for NLDAS (Phase 2) are derived from the analysis fields of the NARR. Those NARR fields that are utilized to generate NLDAS (Phase 2) forcing fields are spatially interpolated to the finer resolution of the NLDAS 1/8th-degree (~13 km) grid and then temporally disaggregated to the NLDAS hourly frequency. Also, the fields of surface pressure, surface downward longwave radiation, near-surface air temperature, and near-surface specific humidity are adjusted vertically to account for the vertical difference between the NARR and NLDAS fields of terrain height. This vertical adjustment applies the traditional vertical lapse rate of 6.5 K/km for air temperature. The details of the spatial interpolation, temporal disaggregation, and vertical adjustment are presented by Cosgrove et al. (2003). The hourly NLDAS measurements sampled from 10 a.m. to 4 p.m. local time were averaged to generate the daily meteorological fields.

2.2.4. Land use information

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 (http://www.epa.gov/mrlc/nlcd-2001). A forest map was generated by converting the values of the forest pixels into one and others into zero.

2.2.5. The southeastern aerosol research and characterization study experiment (SEARCH) data

The SEARCH network is designed to establish detailed aerosol climatology for the Southeast United States (http://www.atmospheric-research.com/studies/SEARCH/index.html). The 2003 SEARCH data from two monitor stations within our study region including North Birmingham, Alabama and Jefferson St. Atlanta, Georgia were downloaded from the website. The predictions at these two monitor sites were compared to the observed $PM_{2.5}$ mass as independent validation of our model.

2.2.6. Data integration

Since the original projections and spatial resolutions of the datasets varied, all the datasets were re-projected to the USA Contiguous Albers Equal Area Conic USGS coordinate system before combining them. For meteorological parameters and the aerosol optical depth data, a nearest neighbor approach was applied. That is, the meteorological and aerosol optical depth values acquired from the nearest center point of the pixel which is either in NARR, NLDAS, or CMAQ grid were assigned to the $PM_{2.5}$ monitor site. For the forest cover, a 900 m by 900 m square buffer centered at each $PM_{2.5}$ monitor site was created, with each buffer containing 900 forest cover pixels. We chose 900 m because the smallest distance between any two monitor sites is around 1 km. For each buffer, the percentage of forest cover was calculated and assigned to the corresponding $PM_{2.5}$ monitor site. The geographically weighted regression modeling was conducted for each day by selecting 10 as a threshold for number of records. For year 2003, there were 137 days that met the criterion, and the maximum number of observations in one day was 101.

2.3. Methods

2.3.1. Geographically weighted regression

Instead of estimating global parameters, geographically weighted regression can generate a continuous surface of parameter values by taking measurements of the parameters at each local observation to denote the spatial variations of the surface. In this study, the adaptive bandwidths were used due to the uneven distribution of the data points, and the bandwidths were obtained by minimizing the corrected Akaike Information Criterion value. There are other bandwidth selection criteria such as Akaike Information Criterion and Bayesian Information Criterion. Compared to Akaike Information Criterion, the corrected Akaike Information Criterion is less biased and can avoid the large variability and tendency to undersmooth (Hurvich et al., 1998). Unlike Akaike Information Criterion, Bayesian Information Criterion is not an estimator of Kullback–Leibler information distance that can be used to decide which model is closest to reality. How Bayesian Information Criterion could be extended to variable bandwidth non-parametric models with effective degrees of freedom is also not clear (Fotheringham et al., 2002). In general, the best model should have the lowest corrected Akaike Information Criterion value (Fotheringham et al., 2002).

2.3.2. Model structure

In this study, a separate geographically weighted regression model was established for each selected day (137 days in this case). The model structure could be expressed as:

$$PM_{2.5} \sim HPBL + RH + TEMP + WIND\_SPEED + FOREST\_COVER + MODIS\_AOD \quad (1)$$

where $PM_{2.5}$ refers to the daily ground-level $PM_{2.5}$ concentrations (μg/m^3), *HPBL* is the boundary layer height (*m*), *RH* denotes the relative humidity (%), *TEMP* is the air temperature (*K*), *WIND_SPEED* refers to the surface wind speed (m/sec), *FOREST_COVER* denotes the percentage of the forest cover (unitless), and *MODIS_AOD* is the MODIS aerosol optical depth value (unitless). we also tested the u-wind and v-wind as separate parameters instead of wind speed. The prediction accuracy drops significantly, and the model becomes more complex due to one more predictor included. This phenomenon might be due to the limited number of records on some days, and too many predictors might lead to model over-fitting. Hence, wind speed is used in our model because it can not only improve the prediction accuracy, but keep the model simpler. Two combinations of meteorological parameters were applied. One was completely derived from NARR. The other contained boundary layer height from NARR and other parameters from NLDAS, since NLDAS does not provide the boundary layer height. Both models were then used to predict the ground-level $PM_{2.5}$ concentrations in the study domain and generate a continuous $PM_{2.5}$ surface for each day. The annual mean $PM_{2.5}$ surfaces for the two models were derived from the daily surfaces and compared visually to examine the impacts of the two different meteorological datasets on the $PM_{2.5}$ concentration prediction. In addition, relative difference values were calculated to quantitatively denote the difference between the two annual mean $PM_{2.5}$ surfaces. The relative difference formula is defined as

$$Diff = \left| \frac{\beta_{NARR} - \beta_{NLDAS}}{0.5(\beta_{NARR} + \beta_{NLDAS})} \right| \times 100\% \quad (2)$$

where β_{NARR} and β_{NLDAS} denote the two annual mean predicted $PM_{2.5}$ surfaces derived from the geographically weighted regression model using NARR and NLDAS, respectively.

2.3.3. Model fitting and residual spatial autocorrelation

The geographically weighted regression model generates a local R^2 for each regression point which indicates how well a local model can replicate the data in the vicinity of the regression point (Fotheringham et al., 2002).

In this study, a mean local R^2 for each day was determined since the geographically weighted regression model was implemented on a daily basis, and a mean local R^2 for all the regression points and days was also calculated in order to quantitatively compare the overall performance of NARR and NLDAS on the geographically weighted regression model calibration.

Spatial autocorrelation measures the similarity between samples for a given variable as a function of spatial distance (Legendre, 1993). A commonly used index to measure spatial autocorrelation is the Moran's I statistic. Moran's I values larger than zero indicate positive spatial autocorrelation, values smaller than zero denote negative spatial autocorrelation, and values near zero means no spatial autocorrelation. Moran's I is defined as

$$I = \frac{n}{\sum_i^n \sum_j^n w_{ij}} \times \frac{\sum_i^n \sum_j^n w_{ij}(r_i - \bar{r})(r_j - \bar{r})}{\sum_i^n (r_i - \bar{r})^2} \tag{3}$$

where n is the sample size. w_{ij} is the spatial weight between point i and point j. $\bar{r}$ is the mean of the residuals. r_i and r_j are the geographically weighted regression residuals for the ith and jth points, respectively.

Wang et al. (2005) and Zhao et al. (2010) found that geographically weighted regression can solve the problems of spatially autocorrelated error terms in global Ordinary Least Squares (OLS) models, and ideally the residuals of the geographically weighted regression model have no significant spatial autocorrelation. Spatially autocorrelated error terms indicate that some variations cannot be explained by the current covariates, and statistical inference is affected, yielding biased results. In this study, the global Moran's I index of the geographically weighted regression residuals was calculated for each day, and a significance test was conducted to examine if there was any significant spatial autocorrelation in the error terms.

2.3.4. Model validation for prediction

To validate model performance, a prediction of the ground-level $PM_{2.5}$ concentration was made at each regression point for each day using the geographically weighted regression model. The predicted $PM_{2.5}$ concentrations were fitted against the observed values using a linear regression model with zero intercept, and the slope indicates the overall prediction bias. Furthermore, the correlation coefficient was calculated to denote the degree of goodness of fit between predicted and observed $PM_{2.5}$ concentrations. Mean prediction error (MPE) and root mean squared prediction error (RMSPE) (Yanosky et al., 2008) were adopted to evaluate the model prediction accuracy. All the statistics were calculated for all days. In addition, predictions were made at the two monitor stations (North Birmingham, Alabama and Jefferson St. Atlanta, Georgia) from the SEARCH network to compare to the observed $PM_{2.5}$ concentrations for independent validation. The data collected from SEARCH network were independent from those used in the modeling. The data used in the modeling were collected from the federal reference monitors. The correlation coefficient, MPE, and RMSPE for all days were calculated to quantitatively denote the goodness of fit between the predicted and observed values.

10-fold cross validation was applied for each daily geographically weighted regression model to test for potential model over-fitting. The dataset was first split into ten folds with approximately 10% of the total data points in each fold. In each round of the cross validation, one fold (10% of the total data points) was used as the testing data, the remaining nine folds (90% of the total data points) were used to fit the model, and predictions were made on the held-out testing fold (10% of the total data points). In next round, another fold (another 10% of the total data points) was used for testing, and the remaining nine folds (another 90% of the total data points) were used for training. The process is repeated ten times until every fold is tested. To incorporating different combinations of random selections, we repeated cross validation for one hundred times. The same statistics (e.g., the slope, the correlation coefficient, MPE, and RMSPE) for all days were calculated for the cross validation results. Furthermore, a comparison was conducted between the cross-validation and the model-fitting statistics to assess the degree of potential model over-fitting.

The geographically weighted regression results were also compared to those derived from an Ordinary Least Squares (OLS) model using the same aerosol optical depth, land use, and meteorological parameters to examine if the geographically weighted regression model has better predictability of $PM_{2.5}$ concentrations than a global model.

2.3.5. Sensitivity analysis for the number of records

A sensitivity analysis was conducted on the R^2 values calculated from regressing the predicted against the observed $PM_{2.5}$ concentrations as a function of a daily minimum number of observations. Our model is built for each day. For some days, the number of records is limited, which might impact the model significance. The sensitivity analysis is to remove the days with too few valid records in order to improve the overall prediction accuracy. By examining the fluctuation of the R^2 variations, an optimum minimum number of records for each day which yielded relatively stable R^2 values was identified and adopted as the threshold. The days with a number of records smaller than the threshold were considered insufficient and excluded from the analysis. After filtering, there were 137 out of 365 days that met the criteria and therefore were included in the geographically weighted regression modeling.

3. Results

3.1. Descriptive statistics

The histograms of variables for all days are illustrated in Fig. 2, which shows that all the variables are roughly unimodal and log-normally distributed. The geometric mean, standard deviation, maximum, and minimum for all the variables for all days are also presented (Table 1). The annual mean $PM_{2.5}$ concentration for all the monitor sites is 13.66 μg/m^3, and the overall mean of MODIS aerosol optical depth is 0.13.

A Pearson's correlation was performed among all the independent variables to avoid potential multicollinearity problems. The results showed that the correlation coefficients among all

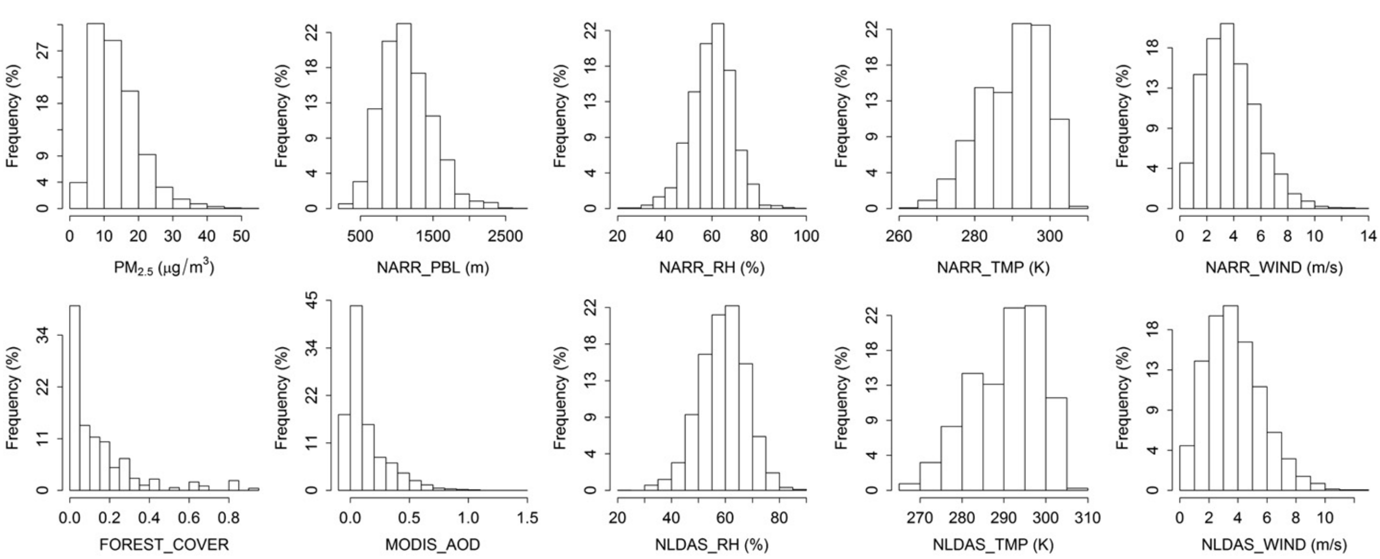

Fig. 2. Histograms of dependent and independent variables.

the independent variables were relatively low, and thus all of them were kept in the model calibration.

The agreement between NARR and NLDAS meteorological parameters was examined by conducting a linear regression analysis with zero intercept. The results indicate that all the meteorological fields are highly comparable between NARR and NLDAS with slopes of 1.02, 0.999, and 1.003, and the correlation coefficients of 0.98, 0.998, and 0.98 for relative humidity, air temperature, and wind speed, respectively.

Fig. 3 shows the results of the sensitivity analysis for the number of daily observations. When the number of matched data records reaches ten, model prediction becomes relatively stable. Hence, the minimum number of ten observations for each day was accepted as the threshold. When the number of records reaches seventy, the R^2 values become instable, because the number of days and records starts to drop significantly.

3.2. Results of model fitting and residual spatial autocorrelation

As shown in Fig. 4(a), the mean local R^2 values of daily geographically weighted regression models range from 0.22 to 0.99 for the models using NARR and from 0.20 to 0.99 for the models using NLDAS. The overall mean local R^2 values were 0.60 and 0.61 for the models using NARR and NLDAS, respectively. These two values indicate that despite poor performance of the geographically weighted regression model for some days, the overall performance was relatively high. Furthermore, we also fitted a geographically weighted regression model with aerosol optical depth as the only predictor variable, and the overall mean local R^2 was 0.38 which was less than the models incorporating NARR and NLDAS, indicating that meteorological fields can significantly improve the model performance. The mean adjust R^2 of the OLS model is 0.44 and 0.47 for NARR and NLDAS, respectively, suggesting that the geographically weighted regression model has better performance than the OLS model.

Fig. 4(b) illustrates the results of Moran's I calculation for the error terms. The results show that the Moran's I values of model residuals ranged from -0.75 to 0.33 and from -0.72 to 0.31 for the models using NARR and NLDAS, respectively. The mean Moran's I value was -0.10 for the models with NARR and -0.14 for the models with NLDAS. In addition, a two-tailed significance test ($\alpha=0.05$) with the null hypothesis of no significant spatial autocorrelation was conducted, and the results showed that there were no significant spatial autocorrelation within the geographically weighted regression model residuals of 107 days (78.1% of days) for both NARR and NLDAS, while there were only 95 days (69.3% of days) for NARR and 91 days (66.4% of days) for NLDAS in which the OLS model residuals did not show significant spatial autocorrelation. The result suggested that geographically weighted regression sufficiently captures any spatially autocorrelated error terms in the global parameter OLS models.

Table 1
Descriptive statistics for dependent and independent variables ($N=4477$).

		Mean	Standard deviation	Minimum	Maximum
NARR	$PM_{2.5}$ ($\mu g/m^3$)	13.66	6.85	2.10	53.30
	Boundary layer Height (m)	1134.12	348.40	236.19	2694.47
	Relative humidity (%)	60.00	8.87	23.38	95.51
	Air temperature (*K*)	290.27	8.43	264.00	305.41
	Wind speed (m/s)	3.75	1.97	0.03	13.01
NLDAS	Relative humidity (%)	59.01	8.63	23.70	89.11
	Air temperature (*K*)	290.56	8.39	265.07	305.58
	Wind speed (m/s)	3.75	1.92	0.04	12.77
Forest cover		0.14	0.18	0	0.91
MODIS Aerosol optical depth		0.13	0.18	-0.05	1.47

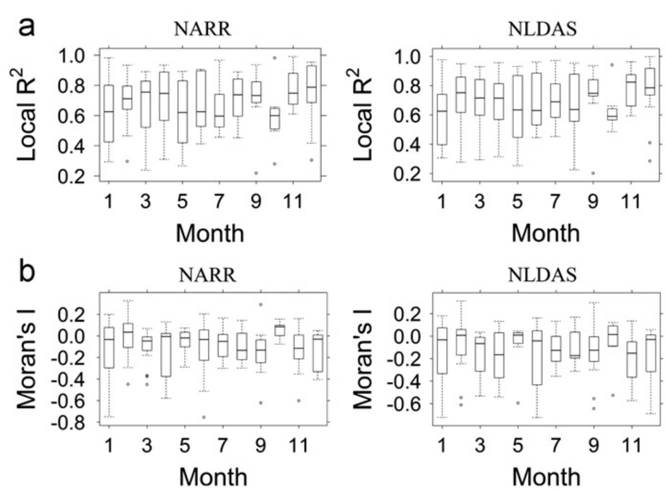

Fig. 4. Model validation. (a) Daily mean local R^2; (b) global Moran's I index. The box gives 25–75% percentile, and the line in box denotes the median. The whisker indicates the minimum and maximum data values, and if there are outliers, the whisker extends to a maximum of 1.5 times the inter-quartile range. The points are outliers.

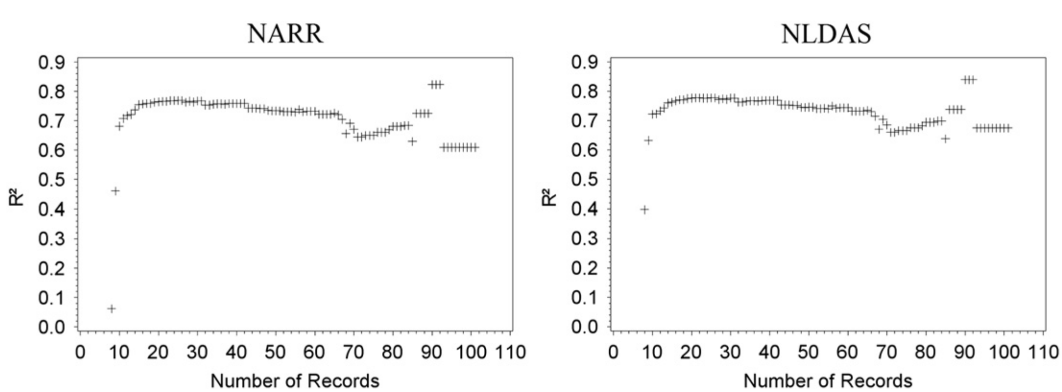

Fig. 3. Sensitivity analysis for record numbers.

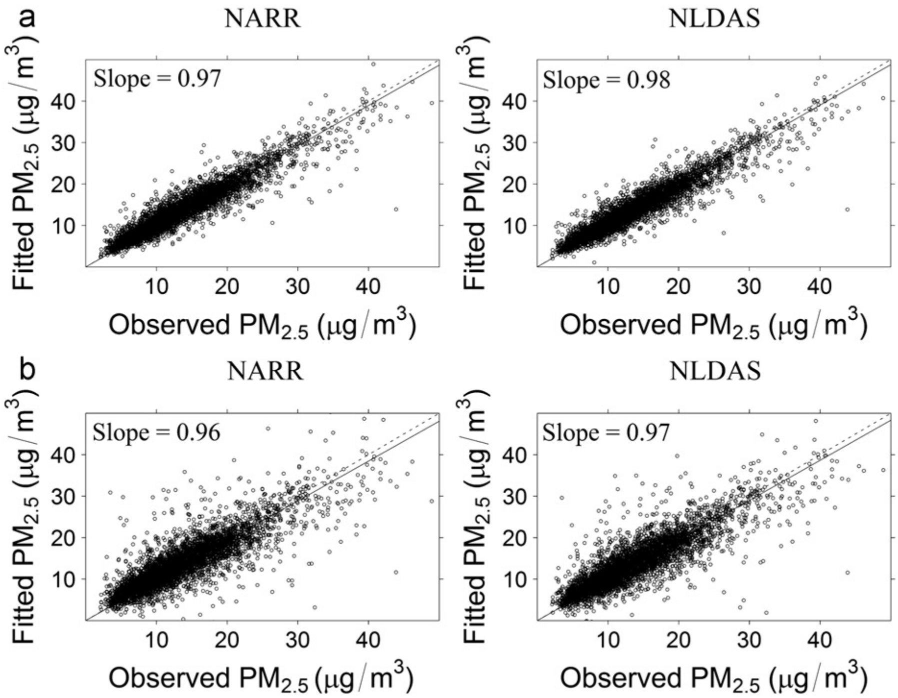

Fig. 5. Predicted vs. Observed $PM_{2.5}$ concentrations. (a) Model fitting and (b) cross validation.

Table 2
Model validation.

			Correlation coefficient *r*	Slope	Mean prediction error	Root mean squared prediction error	Relative accuracy (%)[a]
Model fitting	Geographically weighted regression	NARR	0.94	0.97	1.60	2.37	82.7
		NLDAS	0.94	0.98	1.55	2.32	83.0
	Ordinary least squares	NARR	0.91	0.97	1.91	2.78	79.6
		NLDAS	0.92	0.97	1.85	2.70	80.2
Cross validation	Geographically weighted regression	NARR	0.82	0.96	2.59	4.14	69.7
		NLDAS	0.84	0.97	2.46	3.81	72.1
	ordinary least squares	NARR	0.80	0.96	2.74	4.32	68.4
		NLDAS	0.83	0.96	2.60	4.00	70.7

[a] Relative accuracy is defined as 100%-Root Mean Squared Prediction Error/the mean $PM_{2.5}$ concentration. The mean $PM_{2.5}$ concentration is 13.66 μg/m^3.

3.3. Results of model validation

A regression with zero intercept (Fig. 5) was performed to fit the predicted against the observed values, and the correlation coefficient, the slope, MPE, and RMSPE were calculated to evaluate the predictive power of the geographically weighted regression model (Table 2). The results showed that all the correlation coefficients and the slopes of NARR and NLDAS were close to unity, which indicated that the predictions made from both model fitting and cross validation agreed well with the observed values. However, the correlation coefficients and the slopes of the predicted against the observed values, derived from the cross validation, were smaller than those generated from the model fitting. On the other hand, MPE and RMSPE obtained from the cross validation were larger than those derived from the model fitting, which indicated a slight model over-fitting. The results also showed that the geographically weighted regression model outperformed the OLS model. The geographically weighted regression model generated higher correlation coefficient and lower RMSPE and MPE than the OLS model for both model fitting and cross validation. We also tested the model with aerosol optical depth removed. The RMSPE values from model fitting for NARR and NLDAS are 2.46 and 2.42 μg/m^3, respectively, and those from cross validation are 4.09 and 3.79 μg/m^3. The results indicate that the prediction accuracy does not change significantly with aerosol optical depth removed from the model, which is in line with the findings of Paciorek and Liu (2009) that aerosol optical depth might be helpful for estimating temporal heterogeneity rather than spatial heterogeneity due to the weak spatial correlation between $PM_{2.5}$ and aerosol optical depth. In addition, there are many powerful predictors in the model, and the predicting power of a single predictor might be reduced. Without aerosol optical depth, the model over-fitting is also reduced, which is due to the small sample sizes on some days. Fig. 5 shows that there were underestimations primarily at high concentration levels, the model fitting under-predicted the $PM_{2.5}$ concentrations by 2–3% (e.g., fitted $PM_{2.5}$=0.98 or 0.97 observed $PM_{2.5}$), and the cross validation under-predicted the $PM_{2.5}$ concentrations by 3–4%. The mean differences of observed and predicted $PM_{2.5}$ concentrations for model fitting at each monitor site ranged from -2.64 to 2.35 μg/m^3 for NARR and from -2.47 to 2.30 μg/m^3 for NLDAS. The mean differences for cross validation at each monitor site ranged from -4.43 to 3.67 μg/m^3 for NARR and from -4.14 to 3.72 μg/m^3 for NLDAS. Fig. 6 illustrates the distribution of the mean differences of observed and predicted $PM_{2.5}$ concentrations, which shows that the negative and positive differences are largely intertwined and randomly distributed, and no systematic spatial pattern

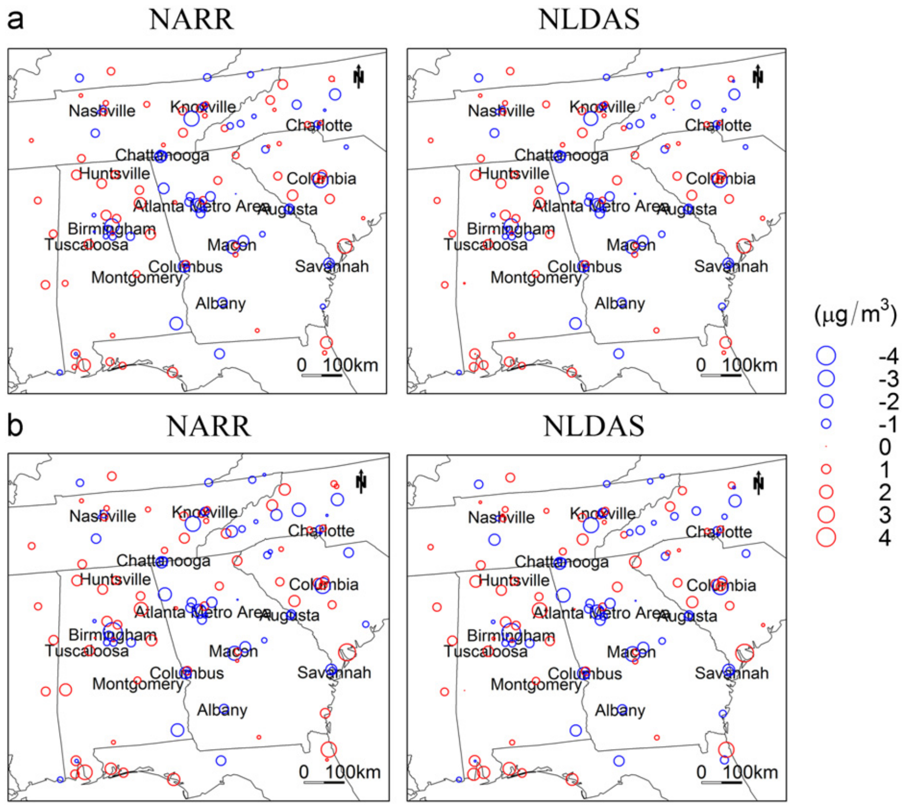

Fig. 6. The mean difference between predicted and observed $PM_{2.5}$ concentrations at each monitor site. (a) Model fitting and (b) cross validation.

Table 3
Independent validation.

			Correlation coefficient *r*	Mean prediction error	Root mean squared prediction error	Relative accuracy (%)[a]
North Birmingham, Alabama	Geographically weighted regression	NARR	0.90	3.07	4.37	77.1
		NLDAS	0.90	2.95	4.21	77.9
	Ordinary least squares	NARR	0.87	3.44	4.96	74.0
		NLDAS	0.88	3.28	4.73	75.2
Jefferson St. Atlanta, Georgia	Geographically weighted regression	NARR	0.82	2.90	4.39	71.2
		NLDAS	0.80	2.94	4.58	70.0
	Ordinary least squares	NARR	0.81	2.90	4.41	71.1
		NLDAS	0.80	2.98	4.63	69.7

[a] Relative accuracy was defined as 100%-root mean squared prediction error/the mean $PM_{2.5}$ concentration. The mean $PM_{2.5}$ concentrations of 19.07 μg/m^3 for the North Birmingham, Alabama site and 15.26 μg/m^3 for the Jefferson St. Atlanta, Georgia site.

is observed. Large negative differences occurring in large urban areas are in line with the fact that the model tends to under-estimate in areas with high $PM_{2.5}$ concentrations. However, large negative differences are likely to occur in but not limited to large urban areas.

The independent validation for the geographically weighted regression model also showed a good fit between the predicted and observed values, and the geographically weighted regression model outperformed the OLS model for both North Birmingham, Alabama and Jefferson St. Atlanta, Georgia sites (Table 3).

3.4. Predictions of $PM_{2.5}$ concentrations

The NARR and NLDAS annual mean (based on 137 days) $PM_{2.5}$ surfaces on the CMAQ grid (12 km by 12 km) are shown in Fig. 7. The results show that NARR predicted the $PM_{2.5}$ concentrations within the range from 4.93 to 17.84 μg/m^3 with a mean of 12.16 μg/m^3, while NLDAS predicted $PM_{2.5}$ concentrations from 4.69 to 17.76 μg/m^3 with a mean of 12.56 μg/m^3. The prediction of NLDAS was slightly higher than NARR. The relative differences between the predictions of the NARR and NLDAS are around 0–10% in most of the domain, except for some rural areas and the mountainous areas in the northeast, where differences exceed 20%. The $PM_{2.5}$ surfaces predicted by the geographically weighted regression model with NARR and NLDAS are similar and consistent with land use and transportation corridors, as well as the patterns showed in previous research (Al-Hamdan et al., 2009), i.e., high values of $PM_{2.5}$ exposure primarily appeared in large urban areas and along major highways, while low values occurred in rural or mountainous areas.

Table 4 lists the parameters estimated from the geographically weighted regression model, which were averaged at each CMAQ pixel for all the daily geographically weighted regression models to illustrate the spatial variation. The results show that there is spatial nonstationarity in the data, and the parameters vary spatially. In most of the study region, boundary layer height, relative humidity, wind speed, and forest cover have generally negative relationships with $PM_{2.5}$ concentrations (median beta values were negative), while the air temperature and aerosol optical depth values show a positive

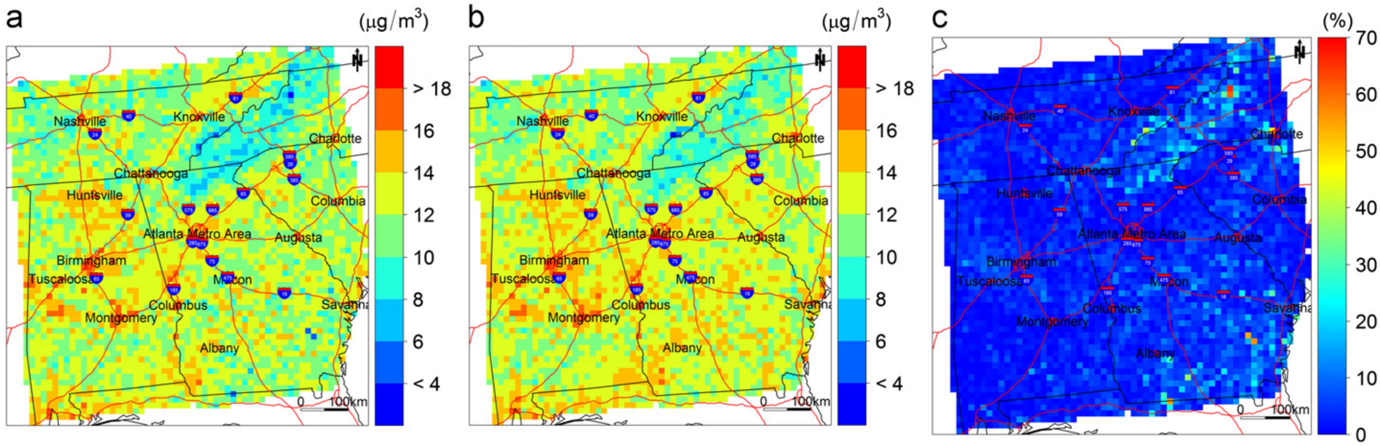

Fig. 7. Prediction results. (a) North American Regional Reanalysis and (b) North American Land Data Assimilation System and (c) relative differences between North American Regional Reanalysis and North American Land Data Assimilation System.

Table 4
Parameter estimates for the geographically weighted regression model.

		$Beta_{PBL}$	$Beta_{RH}$	$Beta_{TMP}$	$Beta_{Wind}$	$Beta_{Forest}$	$Beta_{MODIS}$
NARR	Mean	−0.003	−0.06	0.25	−0.39	−4.33	12.78
	Minimum	−0.007	−0.24	−0.09	−0.998	−6.72	5.31
	25% Quantile	−0.004	−0.09	0.18	−0.50	−4.76	11.81
	Median	−0.003	−0.05	0.27	−0.39	−4.34	12.76
	75% Quantile	−0.0028	−0.03	0.32	−0.28	−3.95	13.76
	Maximum	−0.0002	0.10	0.54	0.18	−1.82	19.49
NLDAS	Mean	−0.003	−0.08	0.35	−0.32	−3.70	14.18
	Minimum	−0.006	−0.32	−0.11	−0.99	−6.56	6.90
	25% Quantile	−0.004	−0.11	0.26	−0.47	−4.23	12.85
	Median	−0.003	−0.08	0.36	−0.33	−3.65	14.18
	75% Quantile	−0.0026	−0.05	0.44	−0.15	−3.18	15.64
	Maximum	0.001	0.59	1.43	0.53	−0.26	22.27

association with $PM_{2.5}$ exposure (median beta values were positive). The phenomenon can be explained by several factors. First, lower boundary layer height could increase the ground-level $PM_{2.5}$ concentrations by reducing vertical mixing. Second, high relative humidity can increase the size and light extinction efficiencies of some particles such as ammonium sulfate and ammonium nitrate, while $PM_{2.5}$ measurements only account for dry particle mass under controlled relative humidity conditions (relative humidity $= \sim 40\%$). Thus, the same aerosol optical depth value at high relative humidity levels represents lower particle dry mass (lower $PM_{2.5}$ concentrations) than those at low relative humidity conditions (Liu et al., 2005). Third, high wind speed can increase horizontal mixing, therefore diluting $PM_{2.5}$ concentrations. Finally, higher forest cover implies fewer particle emission sources such as industries, resulting in lower $PM_{2.5}$ concentrations. The air temperature is positively related to $PM_{2.5}$ because the high air temperature accelerates the generation of secondary particles near the surface (Liu et al., 2007a). Aerosol optical depth values are directly related to the number of particles in the air, and thus showed a strongly positive relationship with $PM_{2.5}$ concentrations.

4. Discussion

The predictions derived from the models using NARR and NLDAS shared a similar pattern and the relative differences were fairly small in most of the domain, except for some rural or mountainous areas, because most of the NLDAS (Phase 2) land surface forcing fields are derived from the NARR analysis fields after applying some spatial interpolation and temporal disaggregation techniques. Also, some meteorological fields of the NLDAS (Phase 2) data such as surface pressure, near-surface air temperature, and near-surface specific humidity were adjusted vertically to account for the vertical difference between the NARR and NLDAS fields of terrain height (Cosgrove et al., 2003), which might be the reason for differences of $PM_{2.5}$ predictions in high elevation areas. There was strong agreement between the NARR and NLDAS meteorological fields in the annual comparison, and the advantage of NLDAS was that its spatial resolution was higher than NARR. However, NLDAS did not provide boundary layer height, which must be obtained from NARR or another meteorological dataset. Due to the higher spatial resolution, predictions from NLDAS have the potential to generate finer structure than those from NARR. Nevertheless, in this study, both data sets were re-sampled to a 12 km grid, and thus the results of NLDAS did not show much difference from the NARR results. It is expected that NLDAS will illustrate more details of the predicted surface than NARR when the spatial resolution of the surface increases. Another explanation is that our domain is relatively flat and has some distance from the ocean. As a result, the temperature and humidity fields are rather smooth, so the advantage of higher spatial resolution of NLDAS cannot be fully demonstrated.

The number of $PM_{2.5}$ sites is very limited in the study domain. For different days, the model significance varies due to different number of records that is used to fit the model. However, our objective is to generate an annual mean prediction. We assess the prediction accuracy for all days rather than for each day. The results showed that the overall accuracy is satisfying. In addition, the number of observations on some days was insufficient compared to the number of variables. In general, the number of independent observations must be equal to or larger than the number of predictor variables (n) plus one because there are $n+1$ unknown parameters. However, on many days, the number of observations does not meet this criterion. In addition, even if this criterion were met, too few observations might lead to model over-fitting and therefore reduce the model prediction accuracy. At the same time, the annual mean $PM_{2.5}$ prediction required as many days as possible to be accounted for. Thus, a trade-off between the number of days and the number of observations per day needs to be established, and it is important to identify an optimal minimum of observation numbers for each day in order to include the maximum number of days to generate an accurate annual mean $PM_{2.5}$ exposure prediction. In this study, we conducted a sensitivity analysis for the daily number of records. The results showed that the predictions of the models became relatively stable when the minimum number of records reached ten. As a result, we accepted ten observations as the threshold in the model calibration.

Both negative and positive differences between predicted and observed $PM_{2.5}$ concentrations occurred in some urban areas, because the measured $PM_{2.5}$ concentrations recorded by numerous monitor sites which are close to each other geographically can vary substantially. Some monitor sites might be close to heavy emission sources such as industrial areas, and therefore record a much higher value of $PM_{2.5}$ concentrations than their neighbors. Those monitor sites in the urban areas cluster together within a range of one to five kilometers which are much smaller than the spatial resolution of our datasets (e.g., aerosol optical depth in CMAQ:12 km, NARR: 32 km, and NLDAS: ~13 km), and thus our model is unable to capture the difference. The results are expected to be improved when data with higher spatial resolution (e.g., 1 km) become available.

5. Conclusions

This study moved a step forward by including other significant parameters such as meteorological fields (e.g., boundary layer height, relative humidity, air temperature, and wind speed) and land use information (e.g., forest cover) as the independent variables in the geographically weighted regression modeling to predict the ground-level $PM_{2.5}$ concentrations. The results showed that meteorological fields and land use information can significantly improve the model performance. Moreover, there was spatial nonstationarity within the data, and the model parameters varied over space. In addition, the predicted $PM_{2.5}$ surfaces illustrated a reasonable spatial pattern with high concentrations of $PM_{2.5}$ in large urban areas and along major highways and low values of $PM_{2.5}$ concentrations in rural or mountainous areas. The results also indicated that the geographically weighted regression model combined with aerosol optical depth, meteorological fields, and land use information as the predictor variables had a strong predictive power for the ground-level $PM_{2.5}$ concentrations.

We also examined the possibility of using the NLDAS as an alternative to the NARR for providing the meteorological data in the geographically weighted regression modeling. Annual comparison between the two meteorological datasets showed good agreement between them. In addition, the prediction results from the NARR and NLDAS shared a similar pattern, and the relative differences between them were fairly small in most of the domain. Although the NLDAS lacked some valuable parameters such as the boundary layer height, given the high spatial and temporal resolution, NLDAS has the potential to be used as an alternative to NARR for providing some of meteorological fields in the $PM_{2.5}$ exposure prediction, and NLDAS is expected to predict surfaces with finer structures when the spatial resolution of the surfaces increases.

Acknowledgments

We thank Dr. Judith Qualters, Helen Flowers, and Ambarish Vaidyanathan at the Environmental Health Tracking Branch, Centers for Disease Control and Prevention; and Sarah Hemmings at the Universities Space Research Association, National Space Science and Technology Center for their technical support.

References

Al-Hamdan, M.Z., et al., 2009. Methods for characterizing fine particulate matter using ground observations and remotely sensed data: potential use for environmental public health surveillance. J. Air Waste Manage. Assoc. 59, 865–881.

Black, T.L., 1988. The Step-Mountain Eta Coordinate Regional Model: A Documentation. NOAA/NWS National Meteorological Center, 47 pp. [Available from NOAA Environmental Modeling Center, Room 207, 5200 Auth Road, Camp Springs, MD 20746.].

Cosgrove, B.A., et al., 2003. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. J. Geophys. Res. Atmos., 108.

Dominici, F., et al., 2006. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA 295, 1127–1134.

Engel-Cox, J.A., et al., 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. Atmos. Environ. 38, 2495–2509.

EPA, Quality assurance handbook for air pollution measurement systems. Volume II. Ambient Air Quality Monitoring Program. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, RTP, NC 27711. Vol. II, 2008.

Fotheringham, A., et al., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons Inc..

Fotheringham, A.S., et al., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environ. Planning A 30, 1905–1927.

Gauderman, W.J., et al., 2004. The effect of air pollution on lung development from 10 to 18 years of age. N. Engl. J. Med. 351, 1057–1067.

Gold, D.R., et al., 2000. Ambient pollution and heart rate variability. Circulation 101, 1267–1273.

Hu, Z.Y., 2009. Spatial analysis of MODIS aerosol optical depth, $PM_{2.5}$, and chronic coronary heart disease. Int. J. Health Geographics 8.

Hurvich, C.M., et al., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. J. R. Stat. Soc. B 60, 271–293.

Ito, K., et al., 2001. Monitor-to-monitor temporal correlation of air pollution and weather variables in the North-Central US. J. Exposure Anal. Environ. Epidemiol. 11, 21–32.

Janjic, Z.I., 1994. The step-mountain eta coordinate model—further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon. Weather Rev. 122, 927–945.

Kahn, R., et al., 1998. Sensitivity of multiangle imaging to aerosol optical depth and to pure-particle size distribution and composition over ocean. J. Geophys. Res. Atmos. 103, 32195–32213.

Kanamitsu, M., et al., 2002. NCEP-DOE AMIP-II reanalysis (R-2). Bull. Am. Meteorol. Soc. 83, 1631–1643.

Kumar, N., et al., 2007. An empirical relationship between $PM_{2.5}$ and aerosol optical depth in Delhi Metropolitan. Atmos. Environ. 41, 4492–4503.

Legendre, P., 1993. Spatial autocorrelation—trouble or new paradigm. Ecology 74, 1659–1673.

Liu, Y., et al., 2007a. Using aerosol optical thickness to predict ground-level $PM_{2.5}$ concentrations in the St. Louis area: a comparison between MISR and MODIS. Remote Sens. Environ. 107, 33–44.

Liu, Y., et al., 2007b. Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: Part 1—Method development. J. Air Waste Manage. Assoc. 57, 1351–1359.

Liu, Y., et al., 2007c. Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: Part 2—A case study. J. Air Waste Manage. Assoc. 57, 1360–1369.

Liu, Y., et al., 2009. Estimating regional spatial and temporal variability of $PM_{2.5}$ concentrations using satellite data, meteorology, and land use information. Environ. Health Perspect. 117, 886–892.

Liu, Y., et al., 2005. Estimating ground-level $PM_{2.5}$ in the eastern united states using satellite remote sensing. Environ. Sci. Technol. 39, 3269–3278.

Mesinger, F., et al., 2006. North American regional reanalysis. Bull. Am. Meteorol. Soc. 87, 343–360.

Mesinger, F., et al., 1988. The step-mountain coordinate—model description and performance for cases of alpine lee cyclogenesis and for a case of an appalachian redevelopment. Mon. Weather Rev. 116, 1493–1518.

Paciorek, C.J., Liu, Y., 2009. Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter. Environ. Health Perspect. 117, 904–909.

Paciorek, C.J., et al., 2008. Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level $PM_{2.5}$. Environ. Sci. Technol. 42, 5800–5806.

Peters, A., et al., 2001. Increased particulate air pollution and the triggering of myocardial infarction. Circulation 103, 2810–2815.

Pope, C.A., et al., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287, 1132–1141.

Remer, L.A., et al., 2005. The MODIS aerosol algorithm, products, and validation. J. Atmos. Sci. 62, 947–973.

Schaap, M., et al., 2009. Exploring the relation between aerosol optical depth and $PM_{2.5}$ at Cabauw, the Netherlands. Atmos. Chem. Phys. 9, 909–925.

Schafer, K., et al., 2008. Correlation of aerosol mass near the ground with aerosol optical depth during two seasons in Munich. Atmos. Environ. 42, 4036–4046.

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.

Tian, J., Chen, D., 2010. A semi-empirical model for predicting hourly ground-level fine particulate matter ($PM_{2.5}$) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. Remote Sens. Environ. 114, 221–229.

Wallace, J., et al., An investigation of air pollution in southern Ontario, Canada, with MODIS and MISR aerosol data. In: Igarss: 2007 IEEE International Geoscience and Remote Sensing Symposium, Vols. 1–12—Sensing and Understanding Our Planet. IEEE, New York, 2007, pp. 4311–4314.
Wang, Q., et al., 2005. Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. Global Ecol. Biogeogr. 14, 379–393.
Yanosky, J.D., et al., 2008. Spatio-temporal modeling of chronic PM_{10} exposure for the nurses' health study. Atmos. Environ. 42, 4047–4062.
Zhang, H., et al., 2009. The relation between moderate resolution imaging spectroradiometer (MODIS) aerosol optical depth and $PM_{2.5}$ over the United States: a geographical comparison by US environmental protection agency regions. J. Air Waste Manage. Assoc. 59, 1358–1369.
Zhao, N., et al., 2010. Application of geographically weighted regression in estimating the effect of climate and site conditions on vegetation distribution in Haihe Catchment, China. Plant Ecol. 209, 349–359.