

Estimating Ground-Level PM_{2.5} in the Eastern United States Using Satellite Remote Sensing

YANG LIU,^{*,†} JEREMY A. SARNAT,[‡]
VASU KILARU,[§] DANIEL J. JACOB,^{||} AND
PETROS KOUTRAKIS[‡]

Division of Engineering and Applied Sciences (DEAS) and Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts 02215, and National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27709

An empirical model based on the regression between daily PM_{2.5} (particles with aerodynamic diameters of less than 2.5 μm) concentrations and aerosol optical thickness (AOT) measurements from the multiangle imaging spectroradiometer (MISR) was developed and tested using data from the eastern United States during the period of 2001. Overall, the empirical model explained 48% of the variability in PM_{2.5} concentrations. The root-mean-square error of the model was 6.2 μg/m³ with a corresponding average PM_{2.5} concentration of 13.8 μg/m³. When PM_{2.5} concentrations greater than 40 μg/m³ were removed, model results were shown to be unbiased estimators of observations. Several factors, such as planetary boundary layer height, relative humidity, season, and other geographical attributes of monitoring sites, were found to influence the association between PM_{2.5} and AOT. The findings of this study illustrate the strong potential of satellite remote sensing in regional ambient air quality monitoring as an extension to ground networks. With the continual advancement of remote sensing technology and global data assimilation systems, AOT measurements derived from satellite remote sensors may provide a cost-effective approach as a supplemental source of information for determining ground-level particle concentrations.

Introduction

Epidemiological studies around the world have found strong and consistent correlations between adverse health effects and levels of fine particles (PM_{2.5}, particles with aerodynamic diameters of less than 2.5 μm) measured at central monitoring stations, which serve as a major surrogate to actual population exposure level (1–3). In addition, health effects associated with particle exposure have shown no apparent threshold at lower concentrations (4). To date, assessments of chronic population exposures over a large geographical region have

been limited since they generally require long-term monitoring data from a comprehensive network such as the United States Environmental Protection Agency's (EPA) compliance network. Operating and maintaining such networks are very costly, especially for many developing countries. Where monitoring networks do not exist, air quality models can be used to estimate PM_{2.5} concentrations. However, daily PM_{2.5} concentrations predicted by these models may be biased for various reasons such as a lack of background information of certain particle species and simplified model assumptions. Another major hurdle with air quality modeling is that they rely heavily on detailed emission inventories which are often difficult to accurately estimate and maintain.

Polar orbiting satellites can provide information on aerosol optical properties for almost complete global coverage at a moderate spatial resolution over multiple years, which have emerged as another potential method of estimating ground-level air quality. In December 1999, the National Aeronautics and Space Administration (NASA) launched its Terra Earth Observing Satellite (5). The multiangle imaging spectroradiometer (MISR) aboard Terra employs nine cameras pointed at different fixed angles to simultaneously observe reflected and scattered sunlight in four wavelength bands. This unique design enables MISR to retrieve columnar aerosol optical thickness (AOT) at 17.6 km resolution over ocean and most land surfaces (6–8). In a previous study, we showed that MISR AOT values agreed well with ground-level standard AOT measurements from the Aerosol Robotic Network (AERONET) (9). It has also been shown that MISR AOT measurements are sensitive to particles with diameters ranging from 0.05 to 2.0 μm (10), which roughly corresponds to the definition of PM_{2.5}.

Earlier studies have shown that earth-observing satellites can detect and track the transport of particles as well as severe pollution episodes on a regional scale (11–14). In this analysis, we examine the relationship between ground-level PM_{2.5} measurements and MISR AOT measurements in the eastern United States using a generalized linear regression model. To account for the variation in particle vertical profiles, composition, and optical properties, planetary boundary layer height and relative humidity data from the Goddard Earth Observing System (GEOS-3) have been included in the model (15, 16). Model validation using an independent dataset and a graphical display of the model results are also presented. The objective of this study is to explore the efficacy and accuracy of satellite remote sensing data as a cost-effective approach for predicting ground-level PM_{2.5}, thus providing an independent and supplemental data source to in situ monitoring and computational modeling.

Method

(a) Data Collection and Processing. (1) PM_{2.5} Measurement Collection and Processing. A total of 2505 gravimetrically based daily average PM_{2.5} measurements were collected from 346 sites within the EPA's compliance network in the eastern United States from the year 2001 (Figure 1). The study area was, subsequently, divided into three subregions to examine geographic variability among the observed results. The "New England" region included Maine, New Hampshire, Massachusetts, Connecticut, Vermont, and Rhode Island. The "mid-Atlantic" region included New York, New Jersey, Pennsylvania, Delaware, District of Columbia, Maryland, West Virginia, and Virginia. The "south Atlantic" region included North Carolina, South Carolina, Georgia, and Florida. In addition to PM_{2.5} mass concentrations (reported in μg/m³), site geographic location (latitude and longitude), land use,

* Corresponding author phone: (703) 516-2366; e-mail: yliu@enviromcorp.com. Present address: ENVIRON International Corp., 4350 N. Fairfax Dr., Suite 300, Arlington, VA 22203.

[†] DEAS, Harvard University.

[‡] Harvard School of Public Health.

[§] U.S. Environmental Protection Agency.

^{||} Department of Earth and Planetary Sciences, Harvard University.

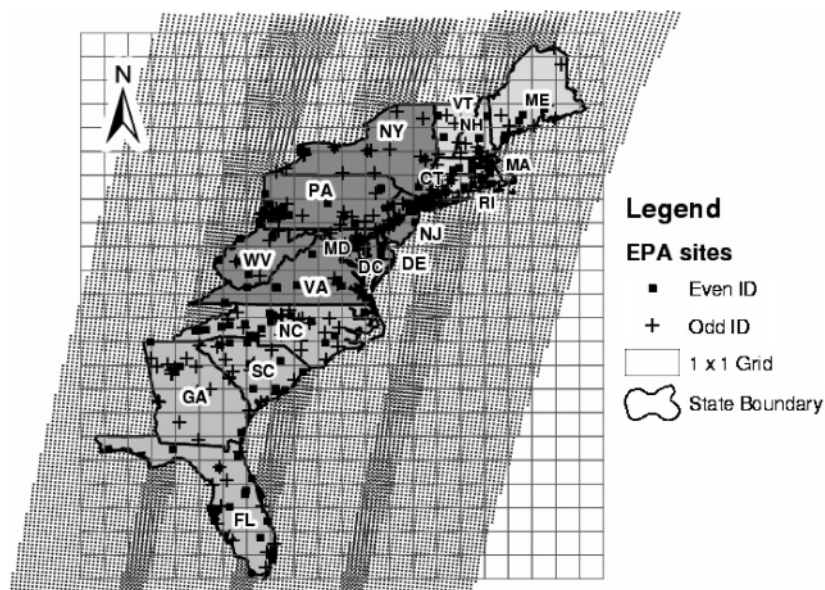


FIGURE 1. MISR spatial coverage of the study region. The ground tracks of MISR paths are shown as dotted strips in the map. Only the ground tracks of MISR paths 11, 14, 17, and 20 are shown for clarity of the map. The three subregions (New England, mid-Atlantic, and south Atlantic) are marked in different gray colors. The study area is also divided into $1^\circ \times 1^\circ$ grids. A grid is counted in if its center coordinates fall into the region.

and other site attribute information were also obtained. The sites were also classified according to their distance to the coast (i.e., ≤ 100 km away or > 100 km away), since a previous study indicated that mixing height growth is heavily influenced by the land–ocean interaction of the atmosphere in areas located within 100 km from the coast (17).

(2) MISR Level 2 Aerosol Data Collection and Processing.

For the current analysis, MISR AOT data covering the east coast was downloaded from the Atmospheric Sciences Data Center at NASA Langley Research Center (<http://edg.larc.nasa.gov/>). The mean and standard deviation of the AOT measurements from each 3×3 MISR region (a 17.6×17.6 km² MISR pixel is called an MISR region) centered at a given EPA site were calculated and matched with the PM_{2.5} measurement taken at that site on the same day. Studies have shown a relatively high degree of spatial homogeneity of PM_{2.5} concentrations over a 24 h period (18). Extreme variability of AOT values within the 3×3 MISR regions may indicate that cloud screening prior to AOT retrieval in some of the MISR regions was insufficient or that the MISR retrieval algorithm could not identify the particle composition in certain regions. To limit the impact of erroneous spatial variation in AOT measurements, we require each 3×3 region to have at least three AOT measurements. In addition, the upper limit of the coefficient of variation (standard deviation divided by mean AOT) calculated from those valid 3×3 regions was set to be less than 0.5 to further reduce the likelihood of data contamination.

(3) GEOS-3 Assimilated Meteorological Field Processing.

The GEOS-3 meteorological fields were given at 1° latitude \times 1° longitude resolution (integrated to $2^\circ \times 2.5^\circ$ in this study due to data storage limitations) and 30 sigma vertical layers. The bottom 10 layers were centered at approximately 10, 50, 100, 200, 400, 600, 900, 1200, 1700, and 2300 m above ground. The mean relative humidity (referred to as RH, %) within the lower troposphere (roughly the lower 3 km of atmosphere) and planetary boundary layer height (referred to as PBL, km) were used in this analysis. These parameters were first interpolated to 10–11 a.m. values, and subsequently matched to EPA MISR measurements.

(4) Data Integration and Randomization for Model Development and Testing. To preserve a subset of the monitoring data for statistical model validation, the moni-

toring dataset was divided into two subsets according to site ID numbers that are randomly assigned by EPA. Observations from the sites with odd ID numbers ($N = 1315$) were used to fit an empirical model between PM_{2.5} and MISR AOT and other factors (called “the model dataset”), while those with even ID numbers ($N = 1190$) were used for model validation (called “the validation dataset”). As seen in Figure 1, site ID numbers do not exhibit any notable pattern regarding their geographic locations. In addition, no autocorrelation was found between consecutive MISR EPA observations in each site since observations were spaced, on average, 30 days apart. Finally, since the repetition of MISR AOT measurements varies between 2 and 9 days, the matching process between MISR regions and ground PM_{2.5} monitoring sites which also have various measurement schedules (daily, every third day, every sixth day) serves as a random sampling of PM_{2.5} concentrations over a given area on the ground. As a result, all the data points regardless of their temporal and spatial distribution may be treated as independent observations. Thus, we believe that the data points finally collected are randomly divided into the model dataset and validation dataset. To confirm this, we also used a complete randomization approach in data division, which generated highly comparable results. Therefore, it is not further discussed.

(5) Model Result Visualization. To display the model results in a geographical information system (GIS), all available MISR data (~40000 measurements), EPA data (~10000 measurements), and GEOS meteorological fields were integrated into seasonal means in $1^\circ \times 1^\circ$ grids that cover the entire study area (Figure 1). The empirical model developed in the current analysis was applied to this gridded dataset, and the geographical pattern of the predicted surface PM_{2.5} concentrations were compared with seasonal mean EPA observations.

(b) Model Development. As an indicator of the abundance of particles in the vertical air column, AOT is defined as the integral of aerosol extinction coefficients (σ_{ext}) along the vertical atmospheric column from the ground to the top of the atmosphere:

$$\text{AOT} = \int_0^\infty \sigma_{\text{ext}}(z) dz \quad (1)$$

TABLE 1. Definition of All Categorical Variables Used in Estimating PM_{2.5} Concentration with MISR AOT and Mixing Height

variable	level
region	New England mid-Atlantic south Atlantic
season	winter (December to February) spring (March to May) summer (June to August) fall (September to November)
site location	rural suburban urban
distance from the coast	within 100 km beyond 100 km

where σ_{ext} can be calculated using the extinction cross-sectional area, C_{ext} , which is a function of particle size and complex refractive index m , and the particle size distribution, $n(r)$

$$\sigma_{\text{ext}}(z) = \int_0^{2.5 \mu\text{m}} C_{\text{ext}}(r, m) n(r) dr \quad (2)$$

Equations 1 and 2 show that the particle composition, size distribution, and vertical profile are key factors that link MISR AOT measurements with ground-level PM_{2.5} concentrations. These factors were modeled by using several variables in the regression model. Airborne and ground-level measurements have shown that the majority of particle mass loading resides in the lower troposphere and the particle mass distribution below the planetary boundary layer tends to be more homogeneous due to the active mixing (19–22). Therefore, GEOS-3 PBL data were included in the regression model as a continuous variable to represent particle vertical distribution. The impact of aerosol composition and size distribution on the association between ground-level PM_{2.5} concentration and MISR AOT was expressed by a series of categorical variables such as geographical location, season, etc.

The empirical regression model used in the current analysis can be expressed as

$$[\text{PM}_{2.5}] = (e^{\beta_0 + \beta_1(\text{variable}_1) + \beta_2(\text{variable}_2) + \dots + \beta_n(\text{variable}_n)}) \times (e^{\beta_{\text{RH}}(\text{RH})}) (\text{AOT})^{\beta_{\text{AOT}}} (\text{PBL})^{\beta_{\text{PBL}}} \quad (3)$$

The dependent variable on the left-hand side, [PM_{2.5}], is the 24 h average ground-level PM_{2.5} concentration measured at various monitoring sites in 2001. The independent variables on the right-hand side include RH, AOT, and PBL, which are geographically matched to each PM_{2.5} measurement, as well as various categorical variables (variable₁ through variable_n) listed in Table 1. The parameters β_0 through β_n are regression coefficients for variable₁ through variable_n. Likewise, β_{RH} , β_{AOT} , and β_{PBL} are regression coefficients for RH, AOT, and PBL, respectively. An exponential function of RH was used to account for the superlinear growth of particle size with increasing relative humidity (23, 24). It should be noted that, for eq 3 to be valid, we assumed that the particle vertical profile is smooth and particle concentrations at different altitudes are correlated to the surface concentration.

Equation 3 was linearized by log-transforming both sides, which resulted in the following model form:

$$\ln([\text{PM}_{2.5}]) = \beta_0 + \beta_1(\text{variable}_1) + \dots + \beta_n(\text{variable}_n) + \beta_{\text{RH}}(\text{RH}) + \beta_{\text{AOT}} \ln(\text{AOT}) + \beta_{\text{PBL}} \ln(\text{PBL}) \quad (4)$$

This model form provides a stronger physical background and more flexibility and predicting power as compared to a

previous study, where a simple linear regression model was fitted between AOT and PM_{2.5} concentrations (14). All statistical analyses were conducted using the SAS system (SAS Institute Inc., Cary, NC). Model parameter estimates were presented with three decimal places or at least one significant digit. The statistical significance of parameter estimates was reported at the $\alpha = 0.05$ level.

Results and Discussion

(a) Descriptive Statistics. Histograms of the various parameter distributions showed that, for both the model and the validation datasets, AOT, PM_{2.5}, PBL, and RH data were unimodal and log-normally distributed (Figure 2). As a result, geometric means and standard deviations are also reported in the summary statistics. Since the summary statistics for the modeling and validation datasets were highly comparable, only the dynamic range and seasonal pattern of the modeling dataset are presented (Table 2). The annual mean PM_{2.5} concentration for all sites of 14.0 $\mu\text{g}/\text{m}^3$ was slightly lower than the ambient annual air quality standard of 15 $\mu\text{g}/\text{m}^3$ for the United States. The overall mean AOT was 0.12.

There was strong seasonal variability for all of the variables. Mean AOT values were 50–100% larger in the spring (0.16 \pm 0.10) and summer (0.19 \pm 0.12) as compared to the fall (0.10 \pm 0.07) and winter (0.08 \pm 0.04). Likewise, AOT values exhibited a broader dynamic range in the summer and the fall than the other two seasons. The PM_{2.5} mass concentration was approximately 20% higher during the winter (14.19 \pm 9.53) and the summer (15.72 \pm 9.25) as compared to the spring (11.89 \pm 5.94) and the fall (13.84 \pm 8.36). The largest and the most variable mixing height values were found in the summer with the lowest and least variable values in the winter, due to the seasonal variation of solar radiation levels. With the exception of slightly lower springtime RH values, the average RH values were comparable in all seasons, except during the spring, when RH values were lower than in the other seasons.

(b) Regression Analysis. The empirical model described in eq 4 was fitted using the model dataset. Overall, the model results were highly significant ($p < 0.0001$), explaining 43% of the variability in corresponding ground-level PM_{2.5} concentrations. MISR AOT, PBL, RH, and all the categorical variables listed in Table 1 were found to be highly significant predictors of PM_{2.5} ($p < 0.0001$) (Table 3). Concentration impact factors (CI factors) for the categorical variables were calculated as the exponentials of the parameter estimates. A CI factor can be interpreted as the impact of a categorical variable at a certain level on the association between MISR AOT and PM_{2.5} as compared to the reference level of this factor.

The estimated power of AOT (0.447 \pm 0.022) was positive and less than 1, indicating surface PM_{2.5} concentrations varied sublinearly with MISR AOT measurements. The greater variability of the observed MISR AOT values is likely due to the fact that AOT measures particle abundance within the entire atmospheric column. Both photochemical reactions, which occur mainly within the boundary layer, and the long-range transport of particles, which occurs in the free troposphere, can have substantial impacts on AOT values. As a result, MISR AOT measurements exhibited a greater variability as compared to ground-level PM_{2.5} concentrations, which are less influenced by long-range transport of particles.

The significance of PBL in predicting surface PM_{2.5} concentrations reflects the difference in the particle vertical profile within and above the boundary layer. Fine particles tend to be more homogeneous within the boundary layer due to convective mixing as compared to particles in the free troposphere. Fine particles emitted from the surface are diluted within the boundary layer as PBL increases, resulting in a lower PM_{2.5} concentration. This may explain the negative

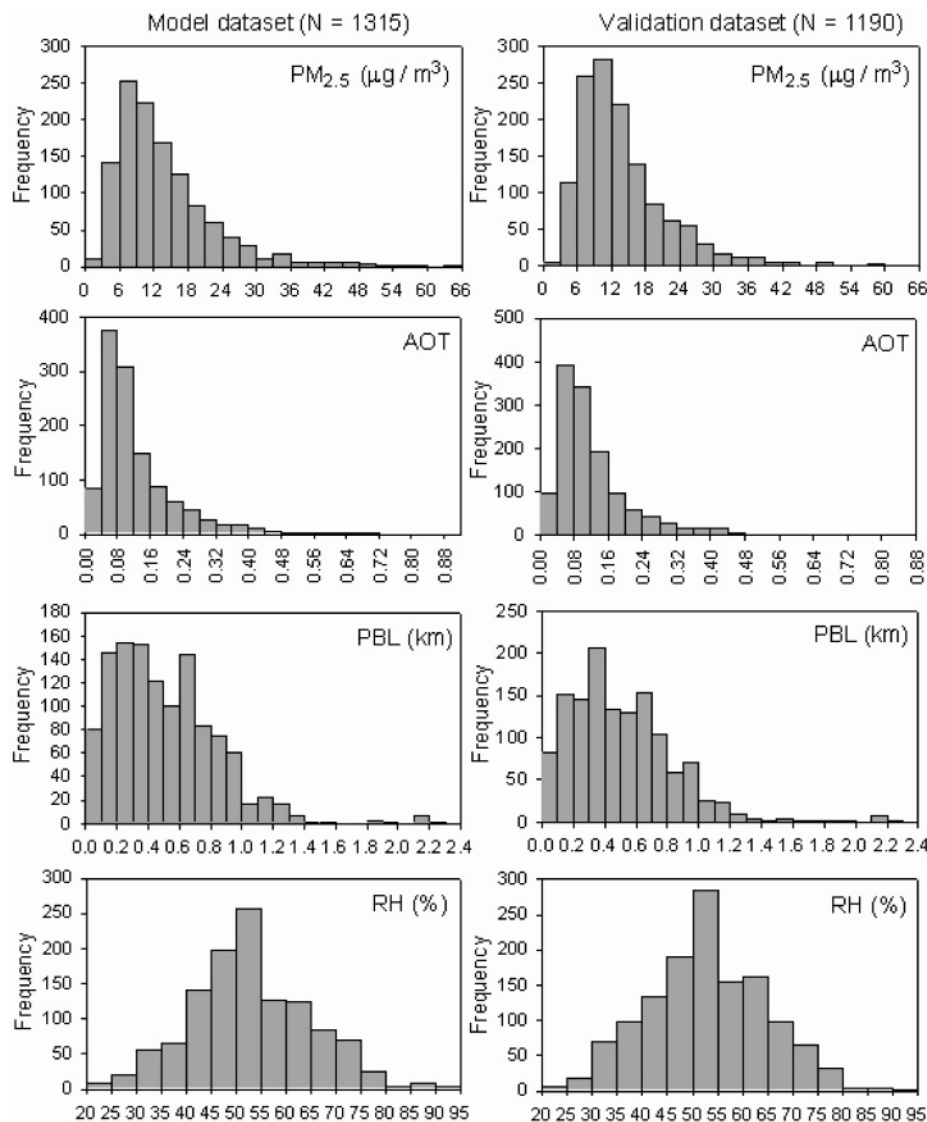


FIGURE 2. Histograms of primary variables in modeling (left panel, $N = 1315$) and validation (right panel, $N = 1190$) datasets, respectively.

sign of the estimated power on the boundary layer height (-0.361 ± 0.023). In addition, the observation that the magnitude of the parameter estimate is substantially smaller than 1 indicates that surface $PM_{2.5}$ concentrations vary at a slower rate as compared to the boundary layer height.

The negative parameter estimate of the exponential of RH (-0.634 ± 0.115) indicated that the same AOT values would correspond to lower $PM_{2.5}$ concentrations as RH increases. This result shows that the model provides a correction for the humidification effect on particle light extinction. MISR measures AOT at ambient meteorological conditions. Under high relative humidity (RH > 70%), which was observed frequently in this study, hygroscopic particles such as ammonium sulfate and ammonium nitrate can grow 2–10 times in size, resulting in a dramatic increase of their light extinction efficiencies (23, 24). In contrast, $PM_{2.5}$ measurements correspond to dry particle mass, since filter storage and weighing are conducted under controlled RH conditions (at 40% RH). As a result, the same AOT values at high RH levels will correspond to smaller particle dry mass as compared to those at low RH conditions. It should be noted that the actual particle size does not grow strictly exponentially and depends on the particle composition. Therefore, the exponential form of RH is a simplified representation of the particle growth effect based on the regression statistics.

The impact of particle size, composition, and vertical distribution on the association between MISR AOT and $PM_{2.5}$ concentrations is also reflected by the categorical variables. Possible interpretations of the impact of these variables are given below. The effect of season was highly significant ($p < 0.0001$), with the association between $PM_{2.5}$ and AOT found to be significantly weaker in the spring as compared to the other seasons. The CI factor of 0.74 for the spring indicates that the predicted $PM_{2.5}$ was approximately 26% lower in the spring than in the fall, with all other parameters being equal. This could be because the particle vertical distribution in the spring is generally different from those in the other seasons. During the spring, for example, long-range transport of Asian dust can significantly increase the particle concentrations in the free troposphere (25), resulting in a larger proportion of particles above the boundary layer as compared to those of the other seasons. Because the particle mass loading below the boundary layer determines the surface $PM_{2.5}$ concentration, similar AOT levels will predict a lower surface $PM_{2.5}$ concentration in the spring as compared to the other seasons.

The CI factor of 0.85 for New England region suggests that MISR predicts lower $PM_{2.5}$ levels in the New England region as compared to the other regions. It has been shown that $PM_{2.5}$ concentrations in the New England region are heavily impacted by transported pollutants from distant industrial and urban sources in the mid-Atlantic and mid-

TABLE 2. Annual and Seasonal Summary Statistics for the Average of the MISR Regional Mean AOT, Ambient 24 h Average PM_{2.5} Mass Concentration, Atmospheric Boundary Layer Height, and Relative Humidity for the Modeling Dataset

	variable	units	mean	SD ^a	min	max	geometric mean	geometric SD
annual N = 1315	PM _{2.5} ^b	μg/m ³	14.0	8.1	2.1	58.8	12.1	1.7
	AOT ^c	unitless	0.13	0.10	0.02	0.88	0.10	1.93
	PBL ^d	km	0.61	0.33	0.05	2.39	0.52	1.75
	RH ^e	%	53	12	23	93	52	1.26
winter N = 438	PM _{2.5}	μg/m ³	14.8	9.1	2.7	58.8	12.7	1.7
	AOT	unitless	0.08	0.05	0.03	0.38	0.07	1.69
	PBL	km	0.42	0.21	0.05	1.21	0.38	1.65
	RH	%	55	10	23	84	54	1.23
spring N = 286	PM _{2.5}	μg/m ³	12.1	5.9	2.7	48.1	10.8	1.6
	AOT	unitless	0.15	0.08	0.05	0.55	0.14	1.64
	PBL	km	0.65	0.27	0.05	1.71	0.59	1.57
	RH	%	47	11	27	77	46	1.25
summer N = 259	PM _{2.5}	μg/m ³	15.6	8.6	2.1	57.4	13.6	1.7
	AOT	unitless	0.20	0.14	0.04	0.80	0.16	1.85
	PBL	km	0.82	0.40	0.17	2.39	0.74	1.60
	RH	%	58	10	36	82	57	1.19
fall N = 336	PM _{2.5}	μg/m ³	13.5	7.6	3.4	38.5	11.6	1.8
	AOT	unitless	0.10	0.08	0.02	0.88	0.08	1.85
	PBL	km	0.64	0.32	0.11	1.66	0.55	1.75
	RH	%	51	12	24	93	50	1.28

^a Standard deviation. ^b Daily PM_{2.5} mass concentration. ^c Mean MISR AOT over the 3 × 3 MISR regions. ^d Mixing height between 10 and 11 a.m. local time. ^e Mean relative humidity of the lower troposphere.

TABLE 3. Estimated Regression Coefficients of the Model Presented in Equation 4 Using the Modeling Dataset (N = 1315, R² = 0.43)

model variable	estimate ^a	std error ^b	p ^c	CI factor ^d
intercept	3.891	0.102	<0.0001	48.97
season				
winter	0.048	0.031	0.12	1.05
spring	-0.296	0.036	<0.0001	0.74
summer	0.009	0.038	0.80	1.01
fall ^e	0.000			1.00
region				
New England	-0.157	0.038	<0.0001	0.85
mid-Atlantic	0.005	0.027	0.84	1.01
south Atlantic ^e	0.000			1.00
distance from coast				
≤ 100 km	-0.193	0.028	<0.0001	0.82
> 100 kme	0.000			1.00
site location				
rural	-0.296	0.047	<0.0001	0.74
suburban	-0.083	0.025	0.001	0.92
urbane	0.000			1.00
relative humidity	-0.634	0.115	<0.0001	e ^{-0.634(RH)}
ln(AOT)	0.447	0.022	<0.0001	AOT ^{0.447}
ln(PBL)	-0.361	0.023	<0.0001	h ^{-0.361}

^a Parameter estimate (standard error) and p value. ^b Standard error of the parameter estimate. ^c Probability that an estimated regression coefficient would be equal to zero. ^d Concentration impact factor. ^e Reference level in each categorical factor.

western U.S. as well as southern Canada (26, 27). As a result, more sulfate particles from transported precursors such as SO₂ and less carbonaceous particles may be found in particle mixtures throughout New England. Sulfate particles generally have higher light extinction efficiencies than carbonaceous particles, especially under high relative humidity conditions (24). Consequently, under the same meteorological conditions, lower particle concentrations in the New England region will be needed to achieve the same AOT level in the other two regions.

Other variables being equal, MISR predicts lower PM_{2.5} concentrations for rural sites (CI factor 0.74) as compared to those at suburban (CI factor 0.92) and urban sites (reference

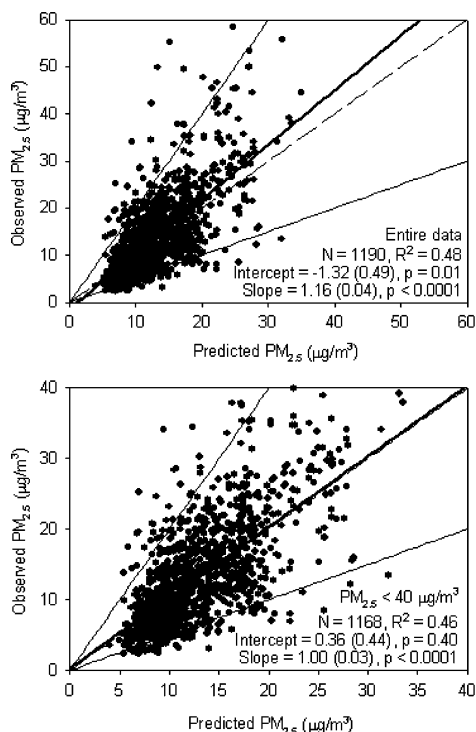


FIGURE 3. Scatter plots of predicted vs observed PM_{2.5} mass concentration for the entire validation dataset (upper panel) and for observations of less than 40 μg/m³. The adjusted R², parameter estimates, and p values of the estimates are shown in each plot. The 1:1 line (dashed) is shown as a reference. The regression line is shown as a thick solid line, and the upper and lower bounds (factor of 2) are shown as thin solid lines.

state, CI factor 1.0). Urban sites are generally characterized by greater anthropogenic PM_{2.5} emission sources than suburban or rural sites. Therefore, a larger proportion of particle mass at urban sites is nitrate and carbonaceous particles, which are typically generated from mobile source emissions, as compared to that at rural sites (28). As previously mentioned, sulfate particles have higher light extinction efficiencies than carbonaceous particles. Consequently, a

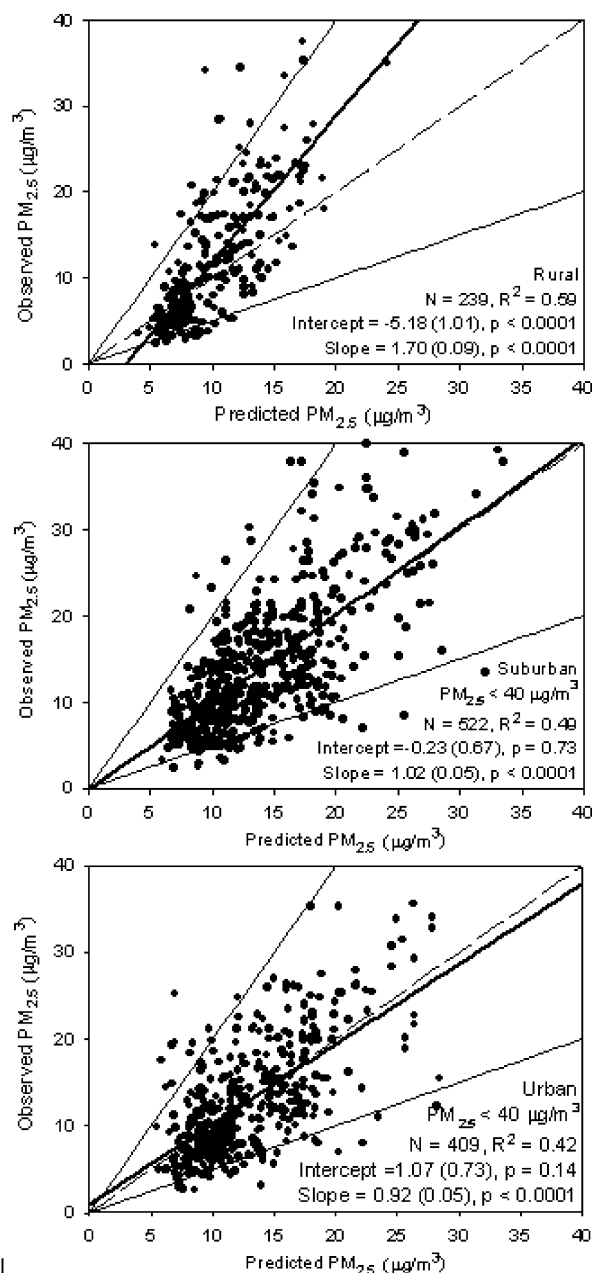


FIGURE 4. Scatter plots of predicted vs observed $PM_{2.5}$ mass concentrations at rural sites (upper panel), suburban sites (middle panel), and urban sites (lower panel). The adjusted R^2 , parameter estimates, and p values of the estimates were calculated when $PM_{2.5}$ concentrations greater than $40 \mu g/m^3$ were excluded. The 1:1 line (dashed) is shown as a reference. The regression line is shown as a thick solid line, and the upper and lower bounds (factor of 2) are shown as thin solid lines.

larger amount of particle mass is needed to achieve the same AOT level in urban areas as in rural areas. The particle composition in suburban sites may be influenced by both long-range transport and local emissions. Therefore, a slight correction effect is noted.

Other variables being equal, MISR also predicts lower $PM_{2.5}$ concentrations at coastal sites (CI factor 0.82) as compared to inland sites. Previous research has shown that the warm conveyor belts (i.e., moist air streams that rise ahead of surface cold fronts), which can lift ground-level pollutants to the upper troposphere and then transport them over the continents, most frequently originate in the boundary layer of the eastern seaboard of North America and Asia, close

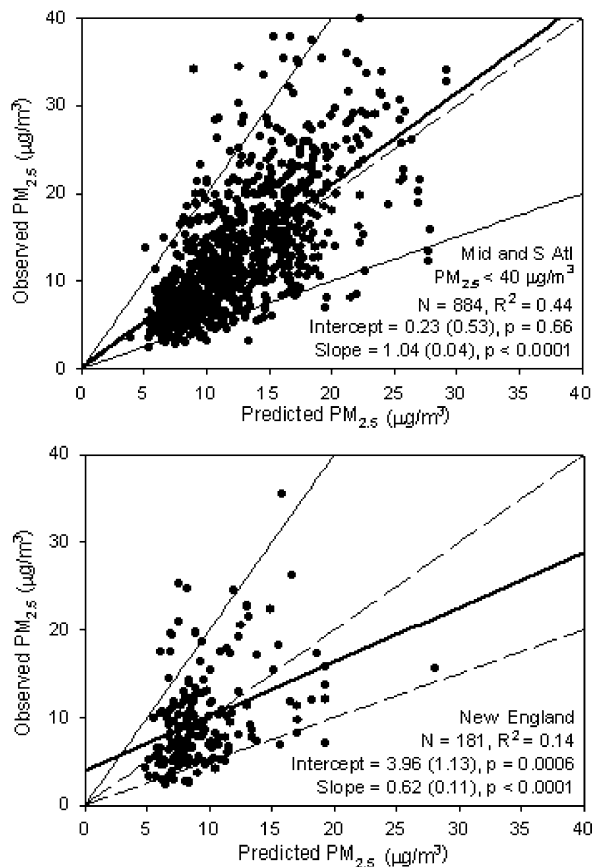


FIGURE 5. Scatter plots of predicted vs observed $PM_{2.5}$ mass concentrations in the mid-Atlantic and south Atlantic regions (upper panel) and in the New England region (lower panel). The adjusted R^2 , parameter estimates, and p values of the estimates were calculated when $PM_{2.5}$ concentrations greater than $40 \mu g/m^3$ were excluded. The 1:1 line (dashed) is shown as a reference. The regression line is shown as a thick solid line, and the upper and lower bounds (factor of 2) are shown as thin solid lines.

to the heavy anthropogenic emissions (29). Therefore, it is possible that a larger proportion of particles reside at higher altitude in the coastal region as compared to the inland region in this study. Similar levels of AOT will correspond to lower surface $PM_{2.5}$ concentrations at coastal sites, therefore, than at inland sites given the fact that the particle mass loading below the PBL determines the surface $PM_{2.5}$ concentration.

Partial F tests (30) indicated that the independent variables MISR AOT and PBL, which measure the total particle abundance and its distribution, are the two most important predictors of $PM_{2.5}$ concentrations. RH and each of the categorical variables make comparable contributions to the predictability of the regression model. However, they each contribute significantly less than MISR AOT or PBL. Further analysis revealed that MISR AOT and PBL are able to explain approximately 18% and 15% of the variation in $PM_{2.5}$ concentration, respectively, measured by the accumulative change of R^2 . The rest of the variables together help explain approximately 10% of the variation.

(c) Model Validation. Bootstrapping was used to test the stability of the regression coefficients. The population of the 2505 data points was randomly sampled 1000 times with a sample size of 1300. Each random sample was used to fit the regression model presented in eq 4, and the corresponding regression coefficients were recorded. The mean and standard deviation of the regression coefficient for each variable in eq 4 were calculated from this population. None of the regression coefficients presented in Table 3 are significantly different

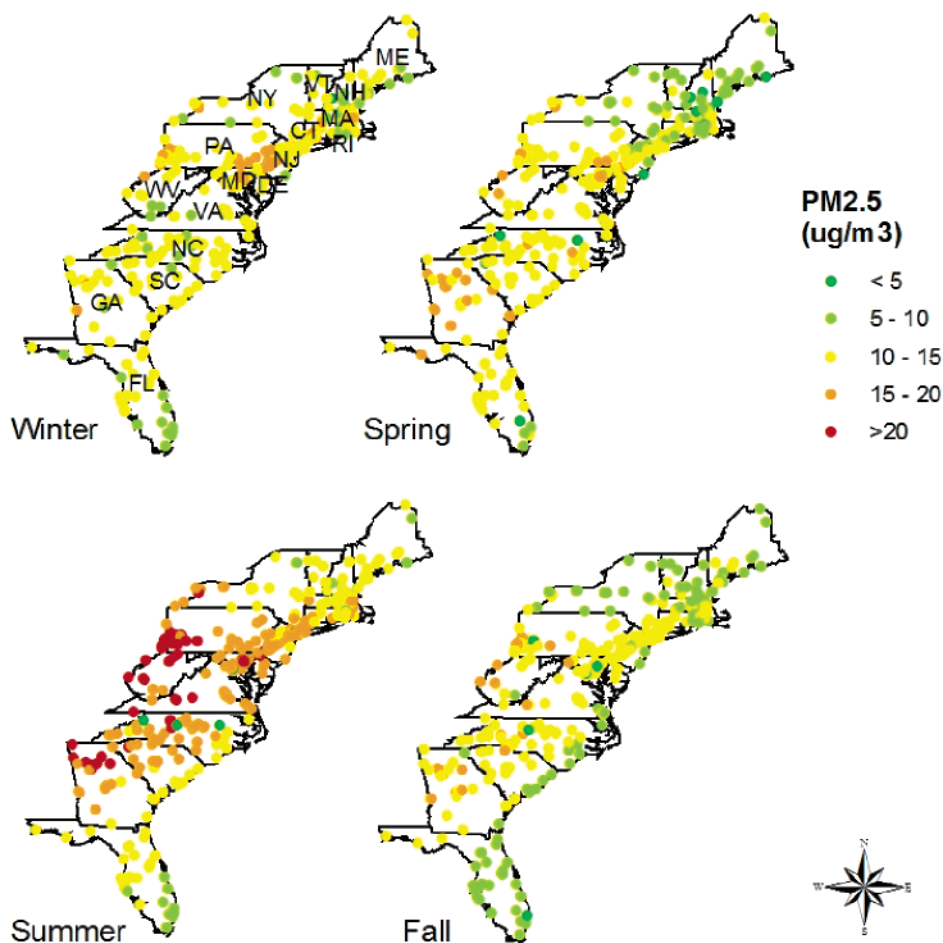


FIGURE 6. Seasonal mean PM_{2.5} concentration at each EPA site. Only sites with gravimetric measurements are included.

from the bootstrapping estimates at the $\alpha = 0.05$ level except the regression coefficient of $\ln(\text{PBL})$ (-0.361), which is slightly beyond the 95% confidence interval ($[-0.365, -0.461]$). The robustness of the regression coefficients is likely due to the relatively long residence time of fine particles in the atmosphere, which results in smooth spatial and temporal variation of PM_{2.5} concentrations over large geographic regions. In addition, a significant proportion of PM_{2.5} is generated by photochemical reactions. Emission patterns and the meteorological conditions that control these photochemical reactions also vary slowly and smoothly over time and region. As a result, different datasets randomly sampled from the same region will likely produce consistent and stable regression coefficients.

To evaluate the model performance, the regression model developed in the previous section was applied to the validation dataset to generate predicted PM_{2.5} concentrations. On average, predicted PM_{2.5} concentrations were $1.2 \mu\text{g}/\text{m}^3$ lower than the observations. Differences decreased to $0.4 \mu\text{g}/\text{m}^3$ when PM_{2.5} concentrations greater than $40 \mu\text{g}/\text{m}^3$ were removed. A linear regression between predicted and observed PM_{2.5} concentrations yielded an R^2 of 0.48 (Figure 3). The model RMSE was $\pm 6.2 \mu\text{g}/\text{m}^3$ for a mean PM_{2.5} concentration of $13.8 \mu\text{g}/\text{m}^3$. Overall, model predictions were within a factor of 2 of the monitored values, when PM_{2.5} concentrations were less than $40 \mu\text{g}/\text{m}^3$. The model substantially underestimated PM_{2.5} concentrations at higher concentrations ($>40 \mu\text{g}/\text{m}^3$). This could be because over 98% of the PM_{2.5} concentrations are below $40 \mu\text{g}/\text{m}^3$ in the modeling dataset. Therefore, current parameter estimates do not sufficiently represent the association between PM_{2.5} and the independent variables at higher PM_{2.5} concentrations. In addition, higher daily

average PM_{2.5} concentrations were often strongly influenced by pollution episodes that occur during a short period of the day. The impact of these episodes may not be captured within the MISR measurement time window (10–11 a.m. local time). As a result, MISR AOT measurements cannot sufficiently represent the daily average PM_{2.5} concentrations under such circumstances. The predicted vs observed regression slope approached 1.00 and intercepts were insignificant when those high observations were excluded (again, less than 2% of the total data). Additionally, the model RMSE was reduced to $5.3 \mu\text{g}/\text{m}^3$ for a mean PM_{2.5} concentration of $13.2 \mu\text{g}/\text{m}^3$ when these high PM_{2.5} concentrations were excluded. Since current air quality models, including Eulerian box models (31), Lagrangian plume models (32), and 3-D Eulerian models (33–35), have been shown to agree within 17–46% of ground-based measurements, the results from our regression model are comparable with these models.

Model predictions and observations were not significantly different in urban sites. The largest discrepancy between model predictions and observations existed in rural sites where observations were $2.1 \mu\text{g}/\text{m}^3$ (20%) greater than model predictions. When observed PM_{2.5} concentrations greater than $40 \mu\text{g}/\text{m}^3$ were removed, regressions at both suburban and urban sites had slopes close to 1 with insignificant intercepts (Figure 4). In addition, model predictions and observations did not significantly differ in New England. However, the agreement between the model and observations was weaker in New England than elsewhere (Figure 5). After PM_{2.5} concentrations greater than $40 \mu\text{g}/\text{m}^3$ were removed, the model agreed well with observations in the mid-Atlantic and south Atlantic regions with an insignificant intercept and a slope very close to 1.

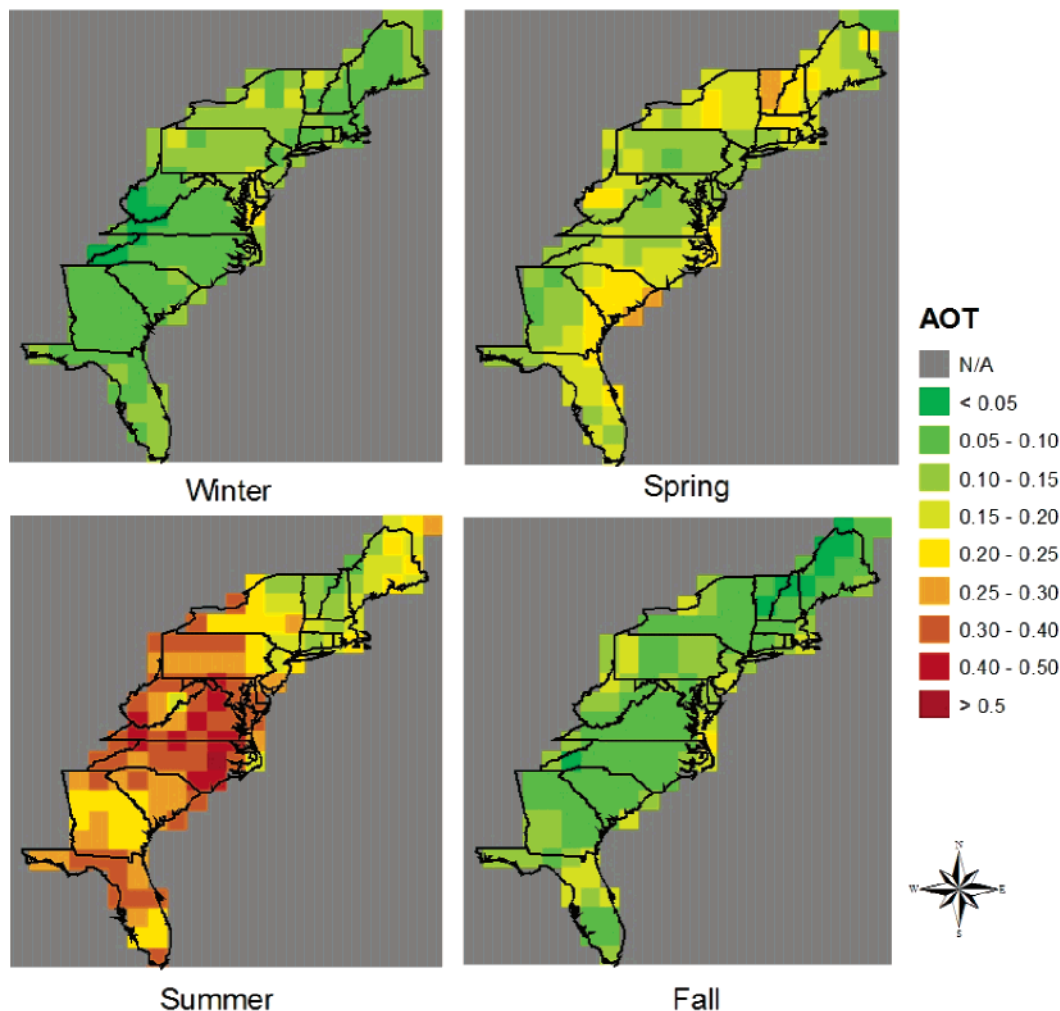


FIGURE 7. Seasonal mean MISR AOT in $1^\circ \times 1^\circ$ grids over the eastern United States in 2001.

(d) Visualization of Model Results. As seen in Figure 6, $PM_{2.5}$ concentrations exhibited strong seasonal and spatial variation, with the levels generally decreasing toward the coastal region in all seasons except in the winter. During the winter, $PM_{2.5}$ concentrations ranged between 10 and 15 $\mu g/m^3$ in the majority of sites across the study area, with the exception of eastern Pennsylvania and New Jersey, where $PM_{2.5}$ concentrations ranged between 15 and 20 $\mu g/m^3$. During the spring, the areas of higher $PM_{2.5}$ concentrations shifted southward, with the concentrations in New England generally below 10 $\mu g/m^3$, between 10 and 15 $\mu g/m^3$ throughout the mid-Atlantic and south Atlantic regions, and between 15 and 20 $\mu g/m^3$ in Georgia. Overall pollution levels were the highest during the summer, with the $PM_{2.5}$ concentrations between 15 and 20 $\mu g/m^3$ in the entire mid-Atlantic region and northern south Atlantic regions. $PM_{2.5}$ levels above 20 $\mu g/m^3$ were found in many sites near Pittsburgh and Atlanta. The spatial pattern in the fall was similar to that in the spring except that its overall $PM_{2.5}$ concentration was slightly lower.

MISR AOT measurements exhibited strong seasonal and spatial variations which were often different from ground-level $PM_{2.5}$ observations (Figure 7). During the winter, most of the study area had AOT levels below 0.10 except in western New York and Pennsylvania, where AOT levels were between 0.1 and 0.15. The AOT level in the spring was generally higher than in the winter. As previously analyzed, the higher AOT values in southern New England were probably caused by Asian dust events. AOT levels along the coast were generally

higher than in the inland region. In addition, AOT measurements did not show higher concentrations in New Jersey and Maryland in the winter as $PM_{2.5}$ concentrations did. High AOT values (0.25–0.50) were observed across the mid-Atlantic region and along the coast of the south Atlantic region. This is probably due to the influence of long-range transport of African dust above the mixing height (36). The overall AOT level was the lowest and uniform in the fall (0.05–0.10).

Surface $PM_{2.5}$ concentrations were predicted from gridded seasonal mean AOT measurements and GEOS meteorological fields using the empirical model discussed in the previous sections (regression coefficients listed in Table 3). The predicted $PM_{2.5}$ concentrations showed spatial characteristics similar to the observations in the winter, summer, and fall, with a slight overestimation in coastal North Carolina in the summer (Figure 8). In the spring, the model substantially underestimated observations over the mid-Atlantic region and Georgia by approximately 5–10 $\mu g/m^3$. Overall, predicted $PM_{2.5}$ concentrations tended to be slightly lower than the observations. This is probably because EPA sites are primarily clustered in populated and more polluted areas while seasonal AOT values were calculated from all available MISR measurements that provided a complete coverage of the entire study area. These results suggest that mixing height and relative humidity information as well as the categorical variables is crucial in the association between AOT values and ground-level $PM_{2.5}$ concentrations.

In conclusion, this study shows the promising potential of air quality models driven by satellite remote sensing and

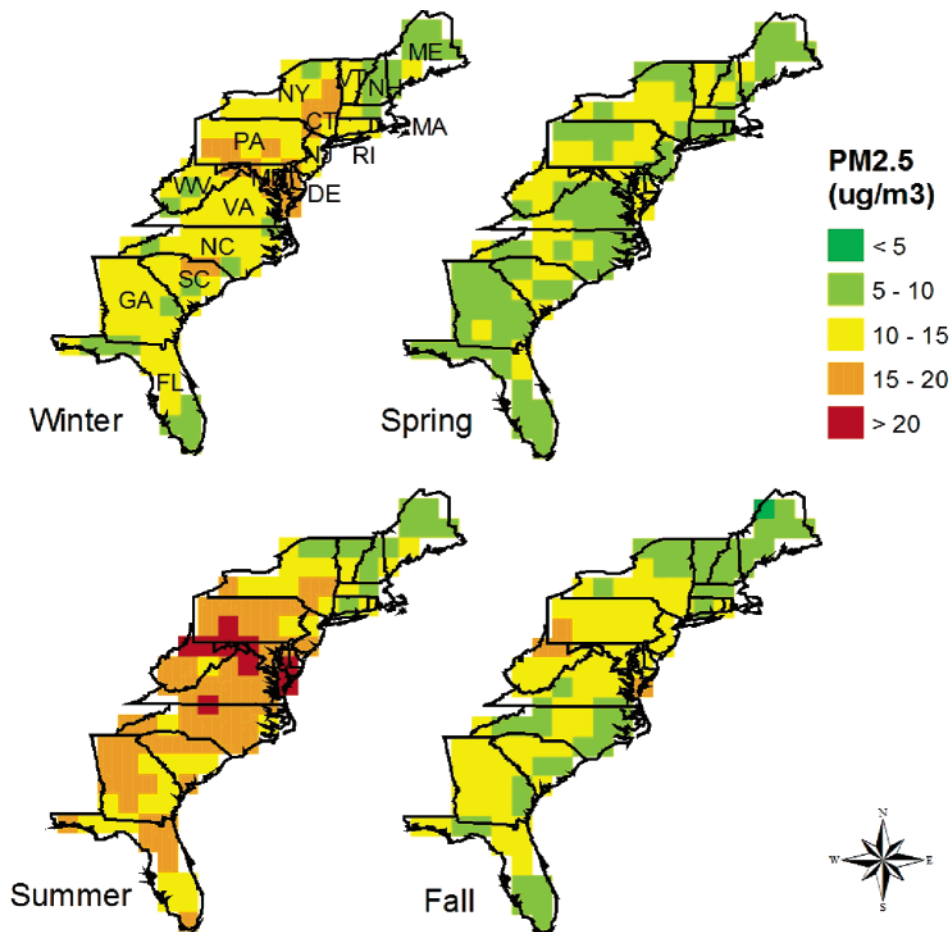


FIGURE 8. Estimated seasonal mean surface PM_{2.5} mass concentrations in 2001 using the empirical model.

global assimilated meteorological fields. There are several advantages to models utilizing this approach. First, these models do not require the extensive resources to develop source emission inventories necessary for conventional computational air quality models. In many of the most polluted areas in the world, emissions inventories are often not available. Second, since satellite remote sensing data and assimilated meteorological fields are internally consistent and available globally, models using these datasets may be applied to national or even global scale studies with relatively simple calibration. Finally, the advancement of remote sensing technology and enhancement in our understanding of the chemistry and transport of atmospheric particles will quickly improve model performance. Clearly, the empirical model cannot be used to replace the compliance monitoring network since coverage is not guaranteed due mainly to cloud cover and a narrow observation time window. However, it could serve as an important extension of ground measurement networks providing supplemental PM_{2.5} concentration estimates for use in population health effect studies in which an unbiased long-term dataset with complete spatial coverage is extremely important. This capability is particularly useful in areas without existing ground-level monitoring capabilities. Furthermore, the use of satellite-based air pollution information may also help in reconfiguring and refining existing ground monitoring networks.

The overall R^2 of 0.43–0.48 suggests that a substantial amount of variability in PM_{2.5} concentrations is not explained by this model. This limited model predictability can be attributed to the lack of accurate information on particle size distribution, composition, and vertical profile, as well as the accuracy and resolution of input data. The performance

of this empirical model can be enhanced by using higher quality versions of MISR data as well as higher resolution meteorological data. The latest generation of GEOS meteorological fields with spatial resolution as high as $0.25^\circ \times 0.25^\circ$, which will be made available soon, can also help improve the model performance. Finally, information about the vertical distribution of particle mass will make it possible to further improve the developed model.

Acknowledgments

This research is funded by the Herbert Winokur Jr. Fund, Harvard University Center for the Environment (HUCE) Research Project Award, and Harvard-EPA Center on Particle Health Effects (Grant R827353-01-0). We thank Peter Rogers at Harvard DEAS for his thoughtful comments on methodology development. We also acknowledge Dr. John Evans, Jim Sullivan, and Steve Melly at the Harvard-EPA Center, Bob Yantosca at the Harvard Department of Earth and Planetary Sciences, and Tesh Rao and David Mintz of the EPA's Office of Air Quality Planning and Standards for their assistance. Finally, we thank the MISR science team for their technical support on MISR data. The United States Environmental Protection Agency through its Office of Research and Development collaborated in the research described here. It has been subjected to Agency review and approved for publication.

Literature Cited

- (1) Samet, J. M.; Dominici, F.; Curriero, F. C.; Coursac, I.; Zeger, S. L. Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. *N. Engl. J. Med.* **2000**, *343*, 1742–1749.

- (2) Pope, C. A., III. Review: epidemiological basis for particulate air pollution health standards. *Aerosol Sci. Technol.* **2000**, *32*, 4–14.
- (3) Wallace, L. Correlations of personal exposure to particles with outdoor air measurements: A review of recent studies. *Aerosol Sci. Technol.* **2000**, *32*, 15–25.
- (4) Smith, K. R.; Jantunen, M. Why particles? *Chemosphere* **2002**, *49*, 867–871.
- (5) Kaufman, Y. J.; Herring, D.; Ranson, K.; Collatz, G. Earth Observing System AM 1 Mission to Earth. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1045–1055.
- (6) Diner, D.; Beckert, J.; Reilly, T. H.; Bruegge, C.; Conel, J. E.; Kahn, R. A.; Martonchik, J.; Ackerman, T. P.; Davies, R.; Gerstl, S. A. W.; Gordon, H.; Muller, J.-P.; Myneni, R. B.; Sellers, P. J.; Pinty, B.; Verstraete, M. M. Multi-angle Imaging SpectroRadiometer (MISR) Instrument Description and experiment overview. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1072–1087.
- (7) Martonchik, J. V.; Diner, D. J.; Kahn, R. A.; Ackerman, T. P.; Verstraete, M. M.; Pinty, B.; Gordon, H. Techniques for the Retrieval of Aerosol properties Over Land and Ocean Using Multiangle Imaging. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1212–1227.
- (8) Martonchik, J.; Diner, D. J.; Crean, K. A.; Bull, M. A. Regional aerosol retrieval results from MISR. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1520–1531.
- (9) Liu, Y.; Sarnat, J.; Coull, B. A.; Koutrakis, P.; Jacob, D. J. Validation of Multiangle Imaging Spectroradiometer (MISR) aerosol optical thickness measurements using Aerosol Robotic Network (AERONET) observations over the contiguous United States. *J. Geophys. Res.* **2004**, *109*, doi: 10.1029/2003JD003981.
- (10) Kahn, R.; Banerjee, P.; McDonald, D.; Diner, D. Sensitivity of multiangle imaging to aerosol optical depth and to pure-particle size distribution and composition over ocean. *J. Geophys. Res.* **1998**, *103*, 32195–32213.
- (11) Hutchison, K. D. Applications of MODIS satellite data and products for monitoring air quality in the state of Texas. *Atmos. Environ.* **2003**, *37*, 2403–2412.
- (12) Falke, S.; Husar, R.; Schichtel, B. Fusion of SeaWiFS and TOMS satellite data with surface observations and topographic data during extreme aerosol events. *J. Air Waste Manage. Assoc.* **2001**, *51*, 1579–1585.
- (13) Wang, J.; Christopher, S. A. Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies. *Geophys. Res. Lett.* **2003**, *30*, doi:10.1029/2003GL018174.
- (14) Engel-Cox, J. A.; Holloman, C. H.; Coutant, B. W.; Hoff, R. M. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmos. Environ.* **2004**, *38*, 2495–2509.
- (15) Schubert, S. D.; Rood, R. B.; Pfendtner, J. An assimilated data set for earth science applications. *Bull. Am. Meteorol. Soc.* **1993**, *74*, 2331–332342.
- (16) Allen, D. J.; Rood, R. B.; Thompson, A. M.; Hudson, R. D. Three-dimensional radon 222 calculations using assimilated meteorological data and a convective mixing algorithm. *J. Geophys. Res.* **1996**, *101*, 6871–6881.
- (17) Berman, S.; Ku, J.-Y.; Rao, S. Spatial and Temporal Variation in the Mixing Depth over the Northeastern United States during the Summer of 1995. *J. Appl. Met.* **1999**, *38*, 1661–1673.
- (18) Ito, K.; Thurston, G.; Nádas, A.; Lippmann, M. Monitor-to-monitor temporal correlation of air pollution and weather variables in the North-Central U.S. *J. Exposure Anal. Environ. Epidemiol.* **2001**, *11*, 21–32.
- (19) Clarke, A.; Porter, J.; Valero, F.; Pilewskie, P. Vertical profiles, aerosol microphysics, and optical closure during the Atlantic Stratocumulus Transition Experiment: Measured and modeled column optical properties. *J. Geophys. Res.* **1996**, *101*, 4443–4453.
- (20) Sheridan, P.; Ogren, J. Observations of the vertical and regional variability of aerosol optical properties over central and eastern North America. *J. Geophys. Res.* **1999**, *104*, 16793–16805.
- (21) Hegg, D.; Livingston, J.; Hobbs, P.; Novakov, T.; Russell, P. Chemical apportionment of aerosol column optical depth off the mid-Atlantic coast of the United States. *J. Geophys. Res.* **1997**, *102*, 25293–25303.
- (22) Schmid, B.; Redemann, J.; Russell, P.; Hobbs, P.; Hlavka, D.; McGill, M.; Holben, B.; Welton, E.; Campbell, J.; Torres, O.; Kahn, R.; Diner, D.; Helmlinger, M.; Chu, D.; Robles-Gonzalez, C.; de Leeuw, G. Coordinated airborne, spaceborne, and ground-based measurements of massive thick aerosol layers during the dry season in southern Africa. *J. Geophys. Res.* **2003**, *108*, Art. No. 8496.
- (23) Malm, W. C.; Day, D. E.; SM, K. Light scattering characteristics for aerosols as a function of relative humidity: part I: A comparison of measured scattering and aerosol concentrations using the theoretical models. *J. Air Waste Manage. Assoc.* **2000**, *50*, 686–700.
- (24) Chin, M.; Ginoux, P.; Kinne, S.; Torres, O.; Holben, B. N.; Duncan, B. N.; Martin, R. V.; Logan, J. A.; Higurashi, A.; Nakajima, T. Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements. *J. Atmos. Sci.* **2002**, *59*, 461–483.
- (25) Thulasiraman, S.; O'Neill, N. T.; Royer, A.; Holben, B.; Westphal, D.; McArthur, L. J. B. Sunphotometric observations of the 2001 Asian dust storm over Canada and the US. *Geophys. Res. Lett.* **2002**, *29*, Art. No. 1255.
- (26) Slater, J.; Dibb, J.; Keim, B.; Talbot, R. Light extinction by fine atmospheric particles in the White Mountains region of New Hampshire and its relationship to air mass transport. *Sci. Total Environ.* **2002**, *287*, 221–239.
- (27) Jordan, C.; Talbot, R.; Keim, B. Water-soluble nitrogen at the New Hampshire sea coast: HNO₃, aerosols, precipitation, and fog. *J. Geophys. Res.* **2000**, *105*, 26403–26431.
- (28) Kleeman, M.; Schauer, J.; Cass, G. Size and composition distribution of fine particulate matter emitted from motor vehicles. *Environ. Sci. Technol.* **2000**, *34*, 1132–1142.
- (29) Stohl, A. A 1-year Lagrangian “climatology” of airstreams in the Northern Hemisphere troposphere and lowermost stratosphere. *J. Geophys. Res.* **2001**, *106*, 7263–7279.
- (30) Wesolowsky, G. O. *Multiple Regression and Analysis of Variance*; John Wiley & Sons: New York, 1976; pp 65–70.
- (31) Pun, B.; Seigneur, C. Sensitivity of particulate matter nitrate formation to precursor emissions in the California San Joaquin Valley. *Environ. Sci. Technol.* **2001**, *35*, 2979–2987.
- (32) Hudischewskyj, A. B.; Seigneur, C. Mathematical Modeling of the Chemistry and Physics of Aerosols in Plumes. *Environ. Sci. Technol.* **1989**, *23*, 413–421.
- (33) Jacobson, M. Z. Development and application of a new air pollution modeling system—Part III. Aerosol-phase simulations. *Atmos. Environ.* **1997**, *31*, 587–608.
- (34) Seigneur, C. Current status of air quality models for particulate matter. *J. Air Waste Manage. Assoc.* **2000**, *51*, 1508–1521.
- (35) Seigneur, C.; Pai, P.; Hopke, P.; Grosjean, D. Modeling atmospheric particulate matter. *Environ. Sci. Technol.* **1999**, *33*, 80A–86A.
- (36) Prospero, J. Long-term measurements of the transport of African mineral dust to the southeastern United States: Implications for regional air quality. *J. Geophys. Res.* **1999**, *104*, 15917–15927.

Received for review April 28, 2004. Revised manuscript received February 1, 2005. Accepted February 2, 2005.

ES049352M