USING GAP-FILLED MAIAC AOD AND WRF-CHEM TO ESTIMATE DAILY $PM_{2.5}$ CONCENTRATIONS AT 1 KM RESOLUTION IN THE EASTERN UNITED STATES

Daniel L. Goldberg*[,1,2], Pawan Gupta[3,4], Kai Wang[5], Chinmay Jena[5], Yang Zhang[5], Zifeng Lu [1,2], and David G. Streets [1,2]

[1]Energy Systems Division, Argonne National Laboratory, Argonne, IL 60439, USA

[2]Consortium for Advanced Science and Engineering, University of Chicago, Chicago, IL 60637, USA

[3]University Space Research Association, Greenbelt, MD 20770, USA

[4]NASA Goddard Space Flight Center, Greenbelt, MD 20770, USA

[5]Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695

Revised Paper submitted to

Atmospheric Environment

October 2018

Reference Number: ATMENV-D-18-01145

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

*Corresponding author. Phone: (202)488-2417; Fax: (630) 252-8007; Email: dgoldberg@anl.gov.

Using gap-filled MAIAC AOD and WRF-Chem to estimate daily $PM_{2.5}$ concentrations at 1 km resolution in the Eastern United States

Daniel L. Goldberg*[,1,2], Pawan Gupta[3,4], Kai Wang[5], Chinmay Jena[5], Yang Zhang[5], Zifeng Lu[1,2], and David G. Streets[1,2]

[1]Energy Systems Division, Argonne National Laboratory, Argonne, IL 60439, USA
[2]Computation Institute, University of Chicago, Chicago, IL 60637, USA
[3]University Space Research Association, Greenbelt, MD 20770, USA
[4]NASA Goddard Space Flight Center, Greenbelt, MD 20770, USA
[5]Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695

Correspondence to: Daniel L. Goldberg (dgoldberg@anl.gov)

Abstract

To link short-term exposures of air pollutants to health outcomes, scientists must use high temporal and spatial resolution estimates of $PM_{2.5}$ concentrations. In this work, we develop a daily $PM_{2.5}$ product at 1 × 1 km^2 spatial resolution across the eastern United States (east of 90° W) with the aid of 1 × 1 km^2 MAIAC aerosol optical depth (AOD) data, 36 × 36 km^2 WRF-Chem output, 1 × 1 km^2 land-use type from the National Land Cover Database, and 0.125° × 0.125° ERA-Interim re-analysis meteorology. A gap-filling technique is applied to MAIAC AOD data to construct robust daily estimates of AOD when the satellite data are missing (e.g., areas obstructed by clouds or snow cover). The input data are incorporated into a multiple-linear regression model trained to surface observations of $PM_{2.5}$ from the EPA Air Quality System (AQS) monitoring network. The model generates a high-fidelity estimate (r^2 = 0.75 using a 10-fold random cross-validation) of daily $PM_{2.5}$ throughout the eastern United States. Of the inputs to the statistical model, WRF-Chem output (r^2 = 0.66) is the most important contributor to the skill of the model. MAIAC AOD is also a strong contributor (r^2 = 0.52). Daily $PM_{2.5}$ output from our statistical model can be easily integrated into county-level epidemiological studies. The novelty of this project is that we are able to simulate $PM_{2.5}$ in a computationally efficient manner that is constrained to ground monitors, satellite data, and chemical transport model output at high spatial resolution (1 × 1 km^2) without sacrificing the temporal resolution (daily) or spatial coverage (>2,000,000 km^2).

Introduction

Inhaling fine particulate matter with an aerodynamic diameter of less than 2.5 μm ($PM_{2.5}$) is a serious health hazard.[1-3] Health studies demonstrate that $PM_{2.5}$ has substantially greater toxicity than larger particles.[4] Furthermore, some constituents of $PM_{2.5}$, such as organic carbon matter and elemental carbon, may be more toxic than others.[5,6] Due to the difficulties of measuring particles with such small size and the emerging science demonstrating that the smallest particles pose the greatest risk, the U.S. EPA did not require measurement of $PM_{2.5}$ until 1997.[7,8]

EPA-approved $PM_{2.5}$ monitors are often located in urban areas or areas of known concern, leaving residents in many areas, especially rural areas, with a limited idea of their exposure to air pollution. Furthermore, epidemiological studies often rely on these sparse ground networks to link air pollution to health outcomes. This leaves a substantial knowledge gap between what people are actually exposed to and what we "think" people are exposed to. This suggests that any alternative observations or robust models may be powerful in providing constraints on the range of surface concentrations.

Satellite observations of AOD can be helpful in providing a first-order estimate of $PM_{2.5}$ at the surface.[9-14] Satellite instruments can be advantageous because they capture greater spatial coverage than any other observational technique. Satellite instruments measure AOD by quantifying how much light at certain wavelength (e.g., 550 nm; yellow-green light) is attenuated between the Earth's surface and the instrument's detector.[15] Therefore, AOD measurements are indicative of the total atmospheric column content and not necessarily the surface concentration. There are three different algorithms used to retrieve AOD from MODerate-resolution Imaging Spectroradiometer (MODIS) spectral reflectance: Dark Target (DT)[16] , Deep Blue (DB)[17], and Multi-Angle Implementation of Atmospheric Correction (MAIAC)[18]. No algorithm has shown the ability to retrieve AOD over ice/snow or cloudy scenes.[19] The DT algorithm has the longest legacy and works best over dark surfaces (i.e., vegetated land or ocean). The DB algorithm has been developed more recently and was specifically refined to acquire AOD over bright surfaces (i.e., deserts or dry brush). Both the DT and DB products are both retrieved at 10 km, with DT having an additional product at 3 km.[20] The MAIAC product became operational in June 2018 and reports AOD at 1 km spatial resolution. The MAIAC algorithm is fundamentally different from

previous AOD algorithms because it remembers the measurements over time and corrects for atmospheric and reflectance parameters that are specific to each pixel.[18] Because of this, the MAIAC algorithm is more computationally demanding and the MAIAC product is only available for land areas.

Satellite AOD measurements can be merged with other variables known to be strongly correlated with surface air quality, such as meteorology[21] and land-use type[22], to better estimate $PM_{2.5}$ concentrations. These models must be trained to surface observations. Initial studies showed how AOD can be joined with meteorological predictors to estimate $PM_{2.5}$ with reasonable skill in the eastern United States.[23-30] Some of the newest approaches use neural networks[31,32] and a random forest approach[33] to estimate $PM_{2.5}$.

A fundamentally different approach to estimate surface $PM_{2.5}$ is to apply the AOD-$PM_{2.5}$ relationship from a chemical transport model to satellite data.[34-37] A distinct advantage of this technique is the ability to estimate $PM_{2.5}$ at high spatial coverage in areas with no ground monitors. This technique can be further improved by using geographic weighted regression, which uses land-cover type, elevation, and surface observations to further refine the $PM_{2.5}$ estimate.[38,39] However, this approach is computationally expensive, relies on the model's capability to provide an accurate estimate of $PM_{2.5}$-AOD ratios, and is proven more accurate when estimates are averaged over longer time periods.

A known limitation with the use of AOD to predict $PM_{2.5}$ is the satellite's inability to provide a valid retrieval over snowy and cloudy scenes. This is particularly problematic because some polluted days occur when clouds or snow obstruct the retrieval, especially in the mid-latitudes. For example, in the winter, snow cover or warm-air advection aloft associated with a cloudy warm front can cause strong temperature inversions to develop near the surface layer trapping air pollution close to the ground level. Therefore, some $PM_{2.5}$ episodes cannot be captured by passive satellite data. A novel solution to this is by gap-filling AOD when daily satellite data are missing.[40,41] The "artificial" AOD is an observed seasonal mean adjusted based on the daily $PM_{2.5}$ measurements by ground stations. This allows for the use of daily AOD covering an entire domain. Lv et al.[40,41] applied this technique to the Beijing metropolitan area and was able to predict $PM_{2.5}$ with a cross-validated skill of r^2=0.78.

Marshall et al.[42] discussed how using a chemical transport model output as an independent variable in a land-use regression model could be an optimal solution. De Hoogh et al.[43] demonstrate the incremental utility of adding satellite data and a chemical transport model as independent inputs to a typical land-use regression model. They found that satellite data adds the most skill increasing the "hold-out" validation r^2 from 0.27 to 0.56 for monitors in Europe. Chemical transport model data further increases the skill to $r^2 = 0.58$.

In this study, we employ a statistical regression technique with meteorological data, land-use parameters, satellite data, and chemical transport model output as independent inputs to estimate daily 24-hour averaged $PM_{2.5}$ at 1×1 km^2 spatial resolution throughout the eastern United States. Our approach most closely matches the methodology followed by de Hoogh et al.[43], but instead we estimate daily $PM_{2.5}$. To provide daily AOD, we utilize a gap-filling AOD methodology[40] to fill-in missing AOD pixels. The novelty in this project lies in the fact that we are able to simulate high spatial resolution data (1×1 km^2) without sacrificing the temporal resolution (daily) or spatial coverage (>2,000,000 km^2).

Methods

Regression model

We calculate daily $PM_{2.5}$ in the eastern United States by fitting a temporally varying multiple least-squares linear regression model to daily 24-hour averaged $PM_{2.5}$ from the EPA ground monitor observations. The model is derived from eleven input variables: MAIAC AOD, WRF-Chem SO_4^{-2}, WRF-Chem NO_3^-, WRF-Chem NH_4^+, WRF-Chem secondary organic aerosol (SOA), 2-m air temperature, total water column, forest percentage, high-density urban percentage, medium-density urban percentage, and open-area urban percentage. Variables were chosen based on their individual cross-validated r^2 skill greater than 0.4, as shown in Table 1.

Multiple least-squares linear regression determines the best-fitting predictor coefficients (i.e., α_i, $\beta_{1,i}$, $\beta_{2,i}$, etc.) by minimizing the sum of the squares between observed values and the resulting estimates, as shown in Equation 1.

$$PM2.5_{i,j} = \alpha_i + \beta_{1,i} x_{1,i,j} + \beta_{2,i} x_{2,i,j} + \ldots + \beta_{n,i} x_{n,i,j} \quad (1)$$

$PM_{2.5}$ for day *i* and grid cell *j* is calculated by adding the constant α and multiplicative sum of the regression parameters β_n and the spatial value x_n at each day *i* and grid cell *j*. AOD, meteorological, and WRF-Chem variables vary spatially and temporally, while land-use parameters vary only spatially. The 'regress' function in the IDL software package is used to determine the coefficients for each individual day. Once the coefficients are determined, the model can be applied to all land-based locations within the eastern United States. Since the model is only fit to monitors located on the land in the eastern United States, we do not attempt to apply the model to areas in eastern Canada or over water bodies, such as the Atlantic Ocean, Great Lakes, and other large land-locked lakes. The model, as currently developed, is not speciated by chemical constituents of $PM_{2.5}$, but this is under our consideration for implementation in future model versions. The model is evaluated using a 10-fold cross-validation, which is further described in the Results section. A block flow diagram of the methodology can be found in Figure 1.

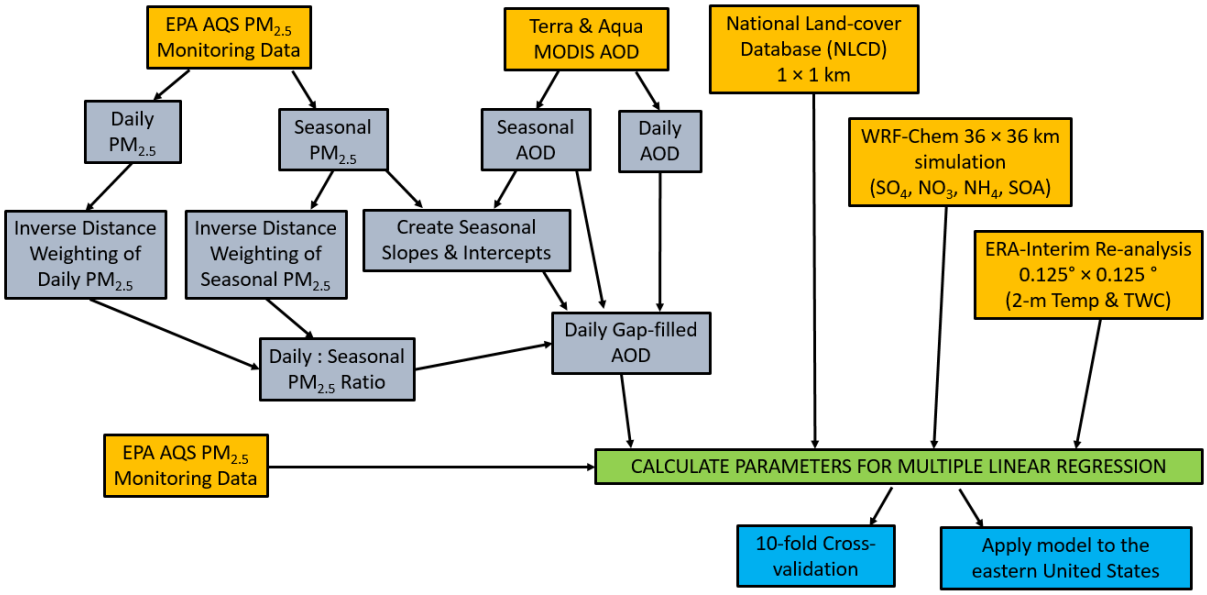

Figure 1. Methodology used to estimate daily $PM_{2.5}$ concentrations. Model inputs are shown in orange, the gap-filling procedure is shown in gray, the regression model is shown in green, and the model outputs are shown in blue.

MAIAC AOD

Two MODIS instruments are onboard NASA low-earth-orbiting satellites: Terra and Aqua. Both Terra and Aqua provide near daily global coverage with local equatorial overpass times of 10:30 and 13:30 respectively. MODIS has a broad swath – 2330 km in width – allowing for a complete

picture of the Earth once per day with minimal gaps in tropical regions. MODIS collects information within thirty-six spectral bands spanning the visible and infrared spectrum (0.41 to 14.5 μm), with individual pixel sizes ranging from 250 to 1000 m; the aerosol retrievals make use of seven of these channels (0.47 to 2.13 μm).[15]

For this study, AOD was retrieved from the NASA MAIAC Collection 6 Level 2G product, which utilizes the MODIS sensors on both the Terra and Aqua satellites.[18] We use all reported MAIAC data, which includes pixels with partly cloudy scenes and pixels adjacent to cloudy scenes. Applying a "clear-only" filter erroneously eliminates many pixels in urban areas and smoke plumes,[18] which are important for our study. From the daily product, we develop three "seasonal" mean products used for the gap-filling procedure: fall and winter (October – March), spring (April – May), and summer (June – September). The gap-filling procedure is described in detail later. In Figure 2, we show the seasonal means of MAIAC AOD in the eastern United States for 2008. On average, AOD in the eastern United States is lower in the cold season than in the warm season. Fall and winter values generally range between 0.05 and 0.1 (mean: 0.076, standard deviation (sd): 0.023), while , spring values range between 0.12 and 0.22 (mean: 0.17, sd: 0.048) and summer values range between 0.15 and 0.3 (mean: 0.23, sd: 0.066). The larger AOD during the summer months in the eastern United States is due to secondary aerosol formation processes,[45] which rely on sunlight, warm temperatures, humidity, and biogenic volatile organic compound emissions. Winter values are low because most residential areas in the northern United States use natural gas in lieu of coal or biomass to heat[46]: natural gas combustion emits almost no sulfate or black carbon, which are important components of particulate matter. In the winter, despite AOD being relatively low, local maximums occur near the agricultural areas of the Midwest U.S.; the lowest values occur in the rural forested areas. In the warm season, AOD is the largest over major cities and the forested areas of the southeastern U.S.

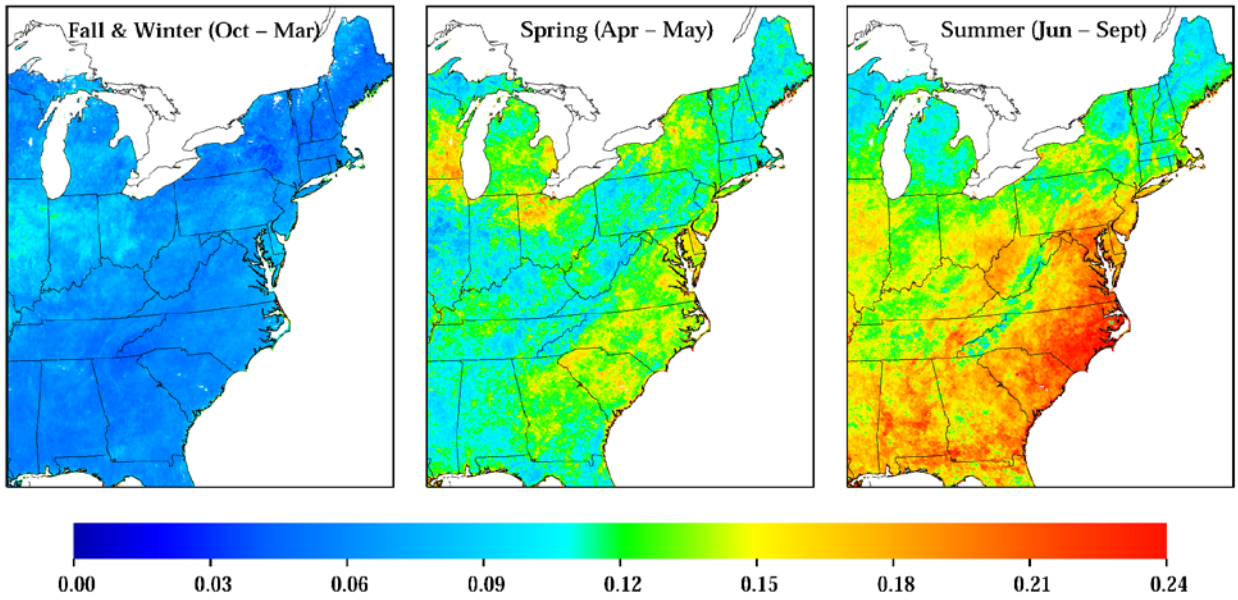

Figure 2. MAIAC AOD Collection 6 in the eastern United States for 2008. Both Terra MODIS and Aqua MODIS data are included; 1 km grid cells with fewer than 5 valid AOD retrievals have no data. Values over water and in Canada are masked.

EPA Air Quality System (AQS) Monitors

Daily 24-hour averaged $PM_{2.5}$ concentrations were acquired from the EPA AQS pre-generated data files (https://aqs.epa.gov/aqsweb/airdata/download_files.html). Most EPA monitoring stations measure 24-hr averaged $PM_{2.5}$ every third day (~120 samples per year) using a Federal Reference Method (FRM): a procedure involving gravimetric sample collection on filter paper.[8] Starting in 2008, EPA began to install continuous Federal Equivalence Method (FEM) $PM_{2.5}$ monitors, which acquire data on an hourly basis.[47] In our eastern United States domain, approximately 90 monitors measure $PM_{2.5}$ every day using continuous the FEM, while an additional ~400 monitors measure daily $PM_{2.5}$ every third day using the FRM. We used over 80,000 observations of 24-hour averaged $PM_{2.5}$ in the eastern United States during 2008. Calibration and quality control of the monitors are conducted by the organizations operating the monitors, which are often state environmental agencies. On any given day, a few monitors, usually less than 10, register values that are inconsistent with the overall trends in the region. These monitors are outliers and are not representative of the local region perhaps because they are affected by a hyperlocal influence, such as a plume from a small factory, a small wildfire, or a roadway. For this reason, we filter out the top 2.5% of observed values when training and evaluating our regression model.

WRF-Chem

The WRF-Chem simulation was conducted by North Carolina State University using an improved version of WRF-Chem v3.7.1 that incorporates the advanced Volatility Basis Set (VBS) secondary organic aerosol (SOA) treatment[48] over a domain that covers the continental U.S. (CONUS) in 2008. The simulation was performed at a horizontal grid resolution of 36-km with 148 × 112 grid cells and a vertical resolution of 34 layers from surface to 100 hPa. This improved version of WRF-Chem uses the 2005 Carbon Bond (CB05) gas-phase mechanism with additional chloride chemistry coupled with the aerosol model Modal for Aerosol Dynamics in Europe (MADE) available in the standard released version as described by Wang et al.[49] but with an improved VBS SOA module as described in Yahya et al.[48]. The anthropogenic emissions used for the simulation are based on the 2008 National Emission Inventory (NEI). The chemical initial and boundary conditions (ICONs/BCONs) are from the global simulations using the modified CESMv1.2.2/CAMv5.3.[50,51] More details on the WRF-Chem model configuration can be found in Yahya et al.[48]

Other Spatiotemporal Predictors

Meteorological data were obtained from the European Center for Medium-Range Forecast (ECMWF) Re-Analysis Interim (ERA-Interim) dataset at a downscaled spatial resolution of 0.125° × 0.125°.[52] We utilized mid-afternoon values (18:00 UTC; 14:00 local time) of 2-m temperature and total water column content from the re-analysis; temperature and humidity are the two strongest meteorological predictors of $PM_{2.5}$ in the eastern United States.[53] Land-cover data were obtained from the Biogenic Emission Land-use Database (BELD) version 4.1 (ftp://ftp.epa.gov/EmisInventory/2011v6/v3platform/beld4.1/tiles/), which compiles land-cover data at 1 × 1 km^2 spatial resolution across the United States. BELD is derived from the 2011 National Land-Cover Database (NLCD). For this study, we utilized the following variables from BELD: high-density urban, medium-density urban, open-area urban, and total forest percentages.

Gap-filling MAIAC AOD

Missing AOD was gap-filled on a daily basis using a technique first described by Lv et al.[40] and integrated in this study in a similar manner. Gap-filled AOD (AODgf) is calculated using Equation 2.

$$AODgf_{i,j} = a_{s,j}\left(\frac{PM2.5_{i,j}}{PM2.5_{s,j}}\right)AOD_{s,j} + b_{s,j} \quad (2)$$

AODgf for day *i* at grid cell *j* is a function of a slope ($a_{s,j}$) during season *s*, the ratio of daily observed 24-hour averaged $PM_{2.5}$ ($PM_{2.5i,j}$) to seasonal $PM_{2.5}$ ($PM_{2.5s,j}$), the observed seasonal mean AOD ($AOD_{s,j}$), and a seasonal constant ($b_{s,j}$). The coefficients, $a_{s,j}$ and $b_{s,j}$, are derived using a linear fit to the observed daily/seasonal AOD and $PM_{2.5}$ at each ground monitor. The daily mean $PM_{2.5}$ anomaly ($PM_{2.5i,j}$/ $PM_{2.5s,j}$) is assumed to be analogous to the anomaly at the MODIS overpass time (10:30 or 13:30 local time). To obtain estimates of $PM_{2.5i,j}$, $PM_{2.5s,j}$, $a_{s,j}$, $b_{s,j}$ within grid cells with no ground monitoring data we use an inverse-distance weighted method with a 50 km smoothing distance using the IDL function "GridData". We define the "cold" season as October – March and the "warm" season as April – September. The warm season is further segregated into two portions, April – May and June – September, for a total of three "seasons". The warm season was separated into two parts due to the varying aerosol characteristics within the warm season. In the cold season and the early part of the warm season, aerosols in the eastern United States are dominated by a nitrate component due to the longer photochemical lifetime of NO_2; in the mid-summer, the nitrate component is overtaken by a secondary organic aerosol (SOA) component.[54]

Results

In this study, we first demonstrate the skill of the gap-filling technique in the eastern United States by comparing to the AERONET ground-based spectrometer network and MAIAC AOD under clear skies. We then develop daily $PM_{2.5}$ estimates at 1×1 km^2 spatial resolution for the eastern

United States. Finally, we compare $PM_{2.5}$ concentrations output by our statistical model to the $PM_{2.5}$ concentrations output from the WRF-Chem simulation.

Gap-filling AOD

The amount of valid MAIAC AOD spatial coverage within our domain during 2008 ranges from a minimum of 0.2% (February 22, 2008) to 92.3% (October 31, 2008), with the median value being 28.8%. As seen in Figure 3, the lowest spatial coverage values occur in the mid-winter due to snow cover, while the largest spatial coverage values occur in the autumn; this is consistent with results shown in Lv et al.[40] Overall, there are more days with valid pixels in the southern states than northern states, but it varies by season. In the southeastern United States, more valid AOD pixels occur during the winter months than the summer months, while the opposite is true in the northeastern United States. Similar results were found by Gupta et al.[55,56]

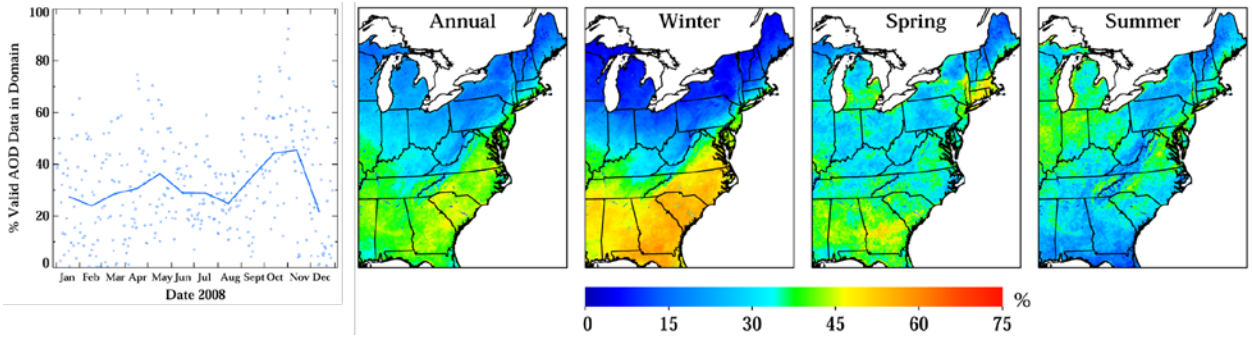

Figure 3. The percentage of valid MAIAC AOD pixels within our domain (eastern U.S.) (a) temporally by day with the monthly mean denoted by a line, (b) spatially for an annual average, (c) spatially for the winter months, (d) spatially for the spring months, (e) spatially for the summer months

To demonstrate the importance of AOD gap-filling, a day representing the average amount of valid MAIAC AOD pixels amongst all days in 2008 is shown in Figure 4. On August 22, 2008, clear skies exist throughout much of the Northeast United States yielding a valid AOD retrieval in these areas. However, cloud cover associated with a frontal boundary in the Midwest and Tropical Storm Fay in the Southeast obstructed the AOD retrieval in a large fraction of our domain: only 25.9% of the region had a valid AOD retrieval. Using equation 1, the area with no retrieval is re-populated with an artificial, best-guess estimate of daily AOD. The gap-filled AOD matches the general spatial pattern of $PM_{2.5}$ observed by the ground EPA AQS sites: higher AOD in the Midwest, lower

AOD in the Southeast. This example serves as a reminder that the surface $PM_{2.5}$ can be both higher or lower than average in the presence of cloud cover.

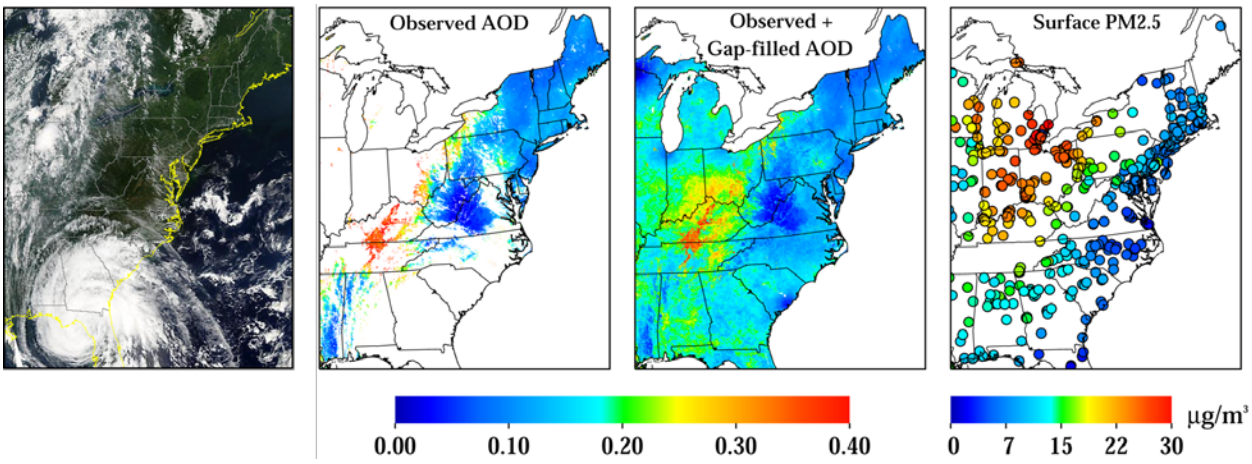

Figure 4. Procedure showing AOD "gap-filling" on August 22, 2008: a) True color visible imagery from Aqua MODIS showing clouds obstructing the AOD measurement in many areas, b) Observed MAIAC AOD from Terra & Aqua, c) Observed + "Gap-filled" AOD, d) Surface observations of $PM_{2.5}$ from the EPA AQS (note: there are very few observations in Tennessee for 2008).

Daily MAIAC AOD and gap-filled daily MAIAC AOD are compared to AOD from the ground-based AERONET observing network. MAIAC AOD is calculated at 550 nm, while the AERONET spectrometers record AOD at a spectrum of wavelengths between 340 nm and 1640 nm. An interpolation of AERONET AOD at 550 nm was created from a log-log best fit of AOD within the AERONET spectrum. For AERONET, we use all valid data acquired between a timeframe of 10:00 – 15:00 local time and average it into a single value. For MAIAC AOD, we use an average of all valid Terra and Aqua pixels within a 5 km radius of the monitoring site; the area encompassing the entire 5 km radius must have complete data in order to be considered valid. In Figure 5, we show AERONET vs. MAIAC AOD and gap-filled MAIAC AOD during days in which there are valid collocated measurements (in our domain, there are 436 valid collocated 5-hour daily averages). The r^2 between observed MAIAC and AERONET AOD is 0.90. A slope near one indicates excellent performance of the MAIAC product and is consistent with other studies validating the MAIAC Collection 6 product. The r^2 between gap-filled MAIAC and AERONET AOD is 0.52, which represents worse but still appropriate correlation, while the slope of the best-fit line is similar to the observed MAIAC. This demonstrates that urban bias in the original retrieval will be inherent in a gap-filled retrieval. The r^2 is lower in the gap-filled product

because it is not a true observation, but instead an estimate. We show that when using a seasonal AOD mean to gap-fill, the performance is degraded even further (r^2=0.31).

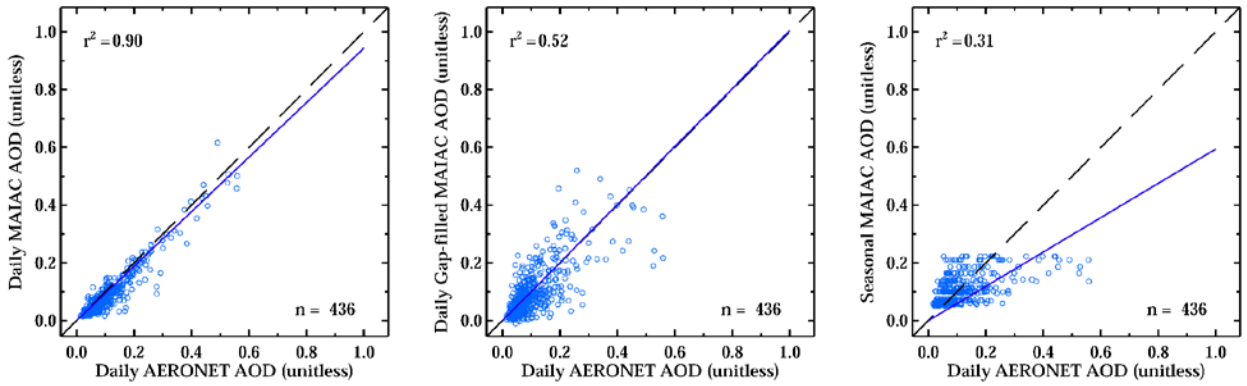

Figure 5. MAIAC AOD vs. AERONET AOD at 550 nm for all locations within the eastern United States during 2008. (left) Observed daily MAIAC AOD, (center) Gap-filled-only daily MAIAC AOD, (right) Observed seasonal MAIAC AOD

In Figure 6, we directly compare the MAIAC gap-filled product to the observed MAIAC AOD, and find good agreement at lower values (AOD < 0.6; 98.5% of all pixels), but poor agreement at the highest observed AOD values (AOD > 0.6; 1.5% of all pixels). This is to be expected as the gap-filled product cannot account for anomalous events such as small wildfires or aerosol plumes that do not reach the surface; some of high observed AOD values may be artificially due to cloud contamination.

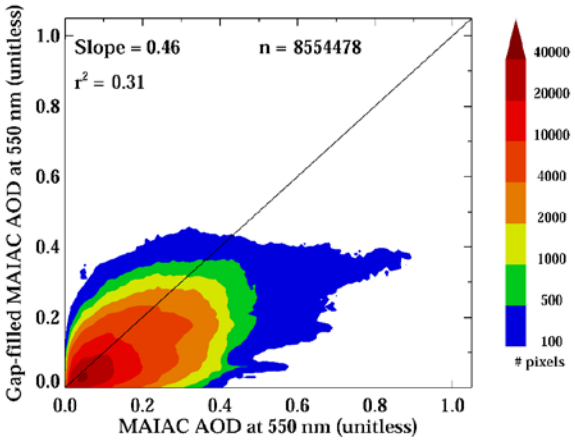

Figure 6. Gap-filled MAIAC AOD vs. observed MAIAC AOD for all valid MAIAC pixels (Terra and Aqua) during 2008.

Daily $PM_{2.5}$

Figure 7 depicts daily $PM_{2.5}$ concentrations averaged over the entire year in the eastern United States; insets of four metropolitan areas are also shown: Chicago, New York City, Baltimore/Washington D.C., and Atlanta. The model is tested using a random 10-fold cross-validation. A random 10-fold cross-validation is a technique in which the model is trained using 90% of the sites, and then is used to predict the values at the remaining 10% of the sites; this process is repeated ten times until there are predictions at all ground sites. To further test the robustness of the model, a spatial 10-fold cross-validation is performed. A spatial cross-validation is identical to a random cross-validation, but instead of randomly removing sites, all sites in a particular region are removed. For example, the model is trained on all sites except the northeastern United States, and then is used to predict data in the northeastern United States. All of the sites and their respective location bins for the spatial cross-validation are shown in Figure 8.

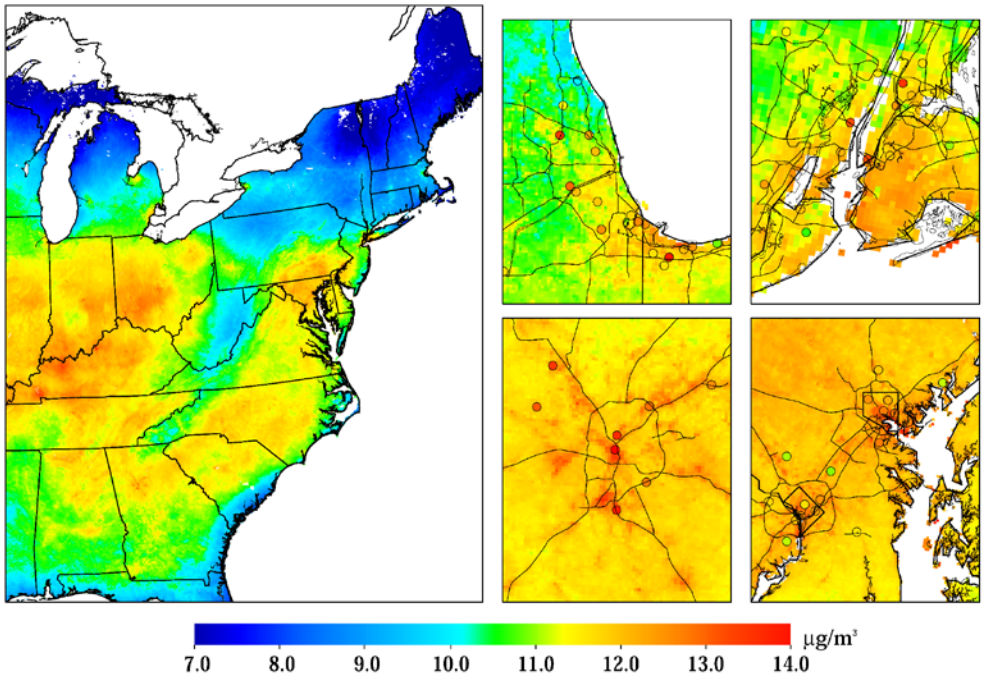

Figure 7. Predicted 2008 Annual $PM_{2.5}$ for the eastern United States and 4 cities, clockwise from top left: a) Chicago, IL, b) New York City, NY, c) Baltimore, MD / Washington, DC, and d) Atlanta, GA. Observations of annual $PM_{2.5}$ are overlaid.

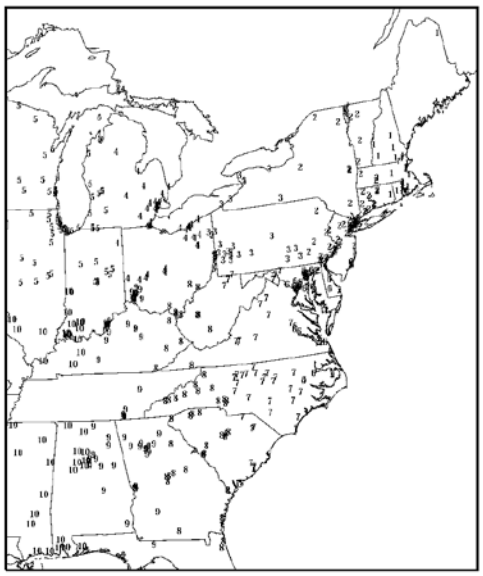

Figure 8. The bins of each monitoring location used in the spatial cross-validation.

The model predicts daily $PM_{2.5}$ with high skill (r^2 = 0.75 using a random cross validation and r^2 = 0.73 using a spatial cross-validation) as shown in Figure 9 and Table 1. The slope of the 10-fold cross-validation best-fit line is close to one (m = 0.89), indicating that the model can reasonably capture low and high observations with similar skill. The median observed daily $PM_{2.5}$ concentration in the eastern United States is 10.1 μg/m^3, while the mean is 11.2 μg/m^3. Table 2 lists the mean and range of the daily coefficients derived by the multiple linear regression model.

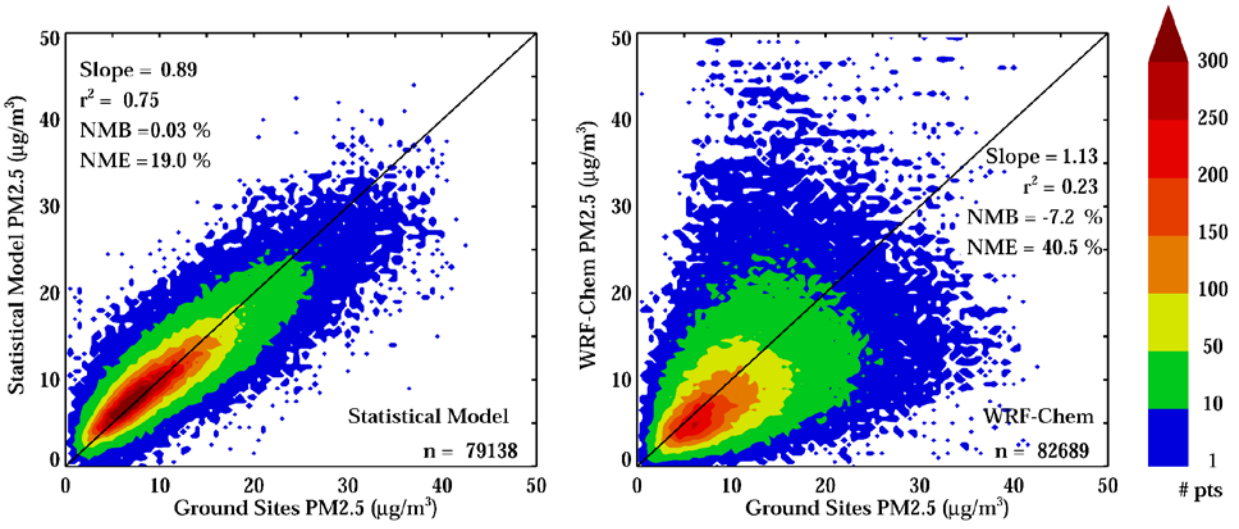

Figure 9. 10-fold random cross-validation of a) the statistical model, and b) the 36 km WRF-Chem simulation matched to closest $PM_{2.5}$ monitor.

Table 1. A 10-fold cross-validation of the statistical model with various inputs used. The italicized model is the one used to generate Figures 7, 9a, and 10.

	Random CV				Spatial CV			
Name of run	**R^2**	**Slope**	**Const**	**NME**	**R^2**	**Slope**	**Const**	**NME**
Met	0.46	0.68	3.5	29.2%	0.43	0.70	3.3	29.6%
LUT	0.41	0.66	3.8	30.5%	0.37	0.68	3.5	31.0%
WRF-Chem	0.66	0.82	2.0	22.3%	0.64	0.83	1.9	23.0%
Met, LUT & WRF-Chem	0.72	0.87	1.4	20.0%	0.70	0.87	1.4	20.8%
AOD	0.52	0.77	3.0	27.3%	0.51	0.73	3.0	27.6%
AOD, Met, LUT & WRF-Chem	*0.75*	*0.89*	*1.3*	19.0%	*0.73*	*0.89*	*1.2*	19.8%
AODallgf, Met, LUT & WRF-Chem	0.77	0.90	1.2	18.4%	0.76	0.90	1.2	18.9%
WRF-Chem-PM2.5*	0.23	1.13	-2.2	40.5%	0.23	1.13	-2.20	40.5%

Met= Meteorological variables (2-m afternoon temperature & total water column from ERA-Reanalysis)

LUT= Land-use Type; forest and urban percentages from the 2011 National Land Cover Database

WRF-Chem = Concentrations of NH4, NO3, SO4, and Black Carbon from WRF-Chem simulation

AOD = MAIAC AOD from Terra & Aqua; for the daily fit, the gap-filling technique is used

AODallgf = Daily observed Aerosol Optical Depth is overwritten with gap-filled AOD

WRF-Chem-PM2.5 = PM2.5 from the WRF-Chem simulation

*Not a cross-validation

Table 2. The percentiles (10^{th}, 50^{th}, and 90^{th}) of the daily coefficients derived by the multiple-linear regression model. The column on the right represents the 50^{th} percentile coefficient multiplied by the mean independent variable to which the coefficient corresponds.

Variable name	**10 %ile**	**50 %ile**	**90 %ile**	**50 %ile × $\bar{x}_n$ (μg/m^3)**
α	-4.86	3.46	13.52	3.46
β_1: AOD	0.43	12.90	33.95	0.98
β_2: 2-m Temperature	-0.26	0.099	0.54	2.10
β_3: Precipitation	-0.22	0.016	0.27	0.10
β_4: High Urban %	-0.019	0.013	0.042	0.21
β_5: Medium Urban %	-0.019	0.0068	0.032	0.16
β_6: Open Urban %	-0.044	-0.0031	0.028	-0.04
β_7: Forest %	-0.047	-0.0008	0.034	-0.02
β_8: WRF-Chem SO_4	-7.71	-1.63	2.48	-1.65
β_9: WRF-Chem NO_3	-7.24	-2.40	0.76	-2.08
β_{10}: WRF-Chem NH_4	-2.71	8.76	24.68	5.68
β_{11}: WRF-Chem SOA	-0.94	0.30	2.04	0.77

When replacing all of the daily AOD with gap-filled AOD – previously we only replaced missing AOD with gap-filled AOD – the skill improves further (r^2 = 0.77 using a random cross validation

and $r^2 = 0.75$ using a spatial cross-validation). This demonstrates the utility of using an artificial daily retrieval MAIAC constrained by a seasonal mean.To determine which independent variables are most important to the regression, we run the model for each set of input variables: meteorology, land-use, WRF-Chem, and AOD; 10-fold cross-validation results (random and spatial) are shown in Table 1. The strongest predictors of daily $PM_{2.5}$ are WRF-Chem ($r^2 = 0.65$) and MAIAC gap-filled AOD ($r^2 = 0.52$); the land-use type parameters ($r^2 = 0.41$) and meteorological data ($r^2 = 0.46$) are also useful inputs. When all independent variables are used, the model performs better ($r^2 = 0.75$). Because the model is trained to the observations, the normalized mean bias (NMB) for all model set-ups is under 0.1%. The more relevant statistic, in this case, is the mean normalized error (NME), which represents the average percentage error of each prediction. The NME is 19.0% for the full regression model. In Figure 10, we depict daily examples of output from the regression model.

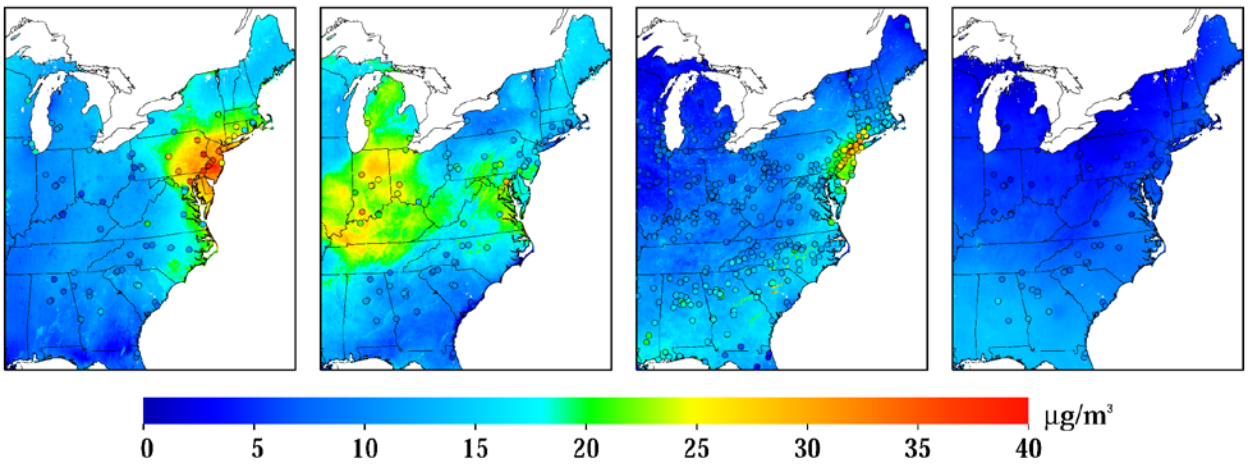

Figure 10. Predicted $PM_{2.5}$ by our statistical model for 4 days: a) January 29, 2008: $PM_{2.5}$ episode in the northeastern US, b) February 23, 2008: $PM_{2.5}$ episode in the Midwest US, c) June 23, 2008: biomass burning plume in North & South Carolina, d) October 2, 2008: a clean day with somewhat elevated $PM_{2.5}$ in the southeastern US.

Comparison to WRF-Chem

Chemical transport models, such as WRF-Chem, are often used to determine how changes in emissions yield changes in secondary pollutants. It is a superior tool to better under the complex chemical relationships in the atmosphere when compared to statistical models. However in this study, our goal is different and narrower: to create a high spatial and temporal resolution re-analysis of $PM_{2.5}$. To do so, running a high spatial resolution chemical transport model is

403 economically unfeasible. It is important to know how much improvement, if any, can be gained
404 by using a high-resolution statistical model to estimate $PM_{2.5}$ concentrations while also using a
405 chemical transport model at coarser spatial resolution as an input.

406 The WRF-Chem model performance is excellent when compared to other simulations of similar
407 spatial resolution. The $PM_{2.5}$ normalized mean bias (NMB) of the WRF-Chem simulation is -7.2%
408 indicating that the model has a minimal low bias. The statistical model generally performs better
409 than WRF-Chem primarily because the statistical model is trained by the observations and is at a
410 higher spatial resolution. The r^2 for estimating daily $PM_{2.5}$ is meaningfully higher for the statistical
411 model ($r^2 = 0.75$) than WRF-Chem ($r^2 = 0.23$) (Figure 9). When we train the individual $PM_{2.5}$
412 chemical constituents from WRF-Chem to $PM_{2.5}$ observations, the r^2 rises dramatically ($r^2 = 0.66$),
413 indicating that information from the WRF-Chem simulation is an essential contributor to the
414 performance of our statistical model.

415 In Figure 11, we demonstrate the seasonal skill of each model. WRF-Chem's skill in simulating
416 $PM_{2.5}$ is highest in autumn and worst in the late spring and early summer. The "AOD-only" model
417 produces the best skill in the summer months. Correspondingly, the skill of the full regression
418 model peaks in the late summer and early fall since WRF-Chem and AOD are most helpful during
419 this timeframe.

420

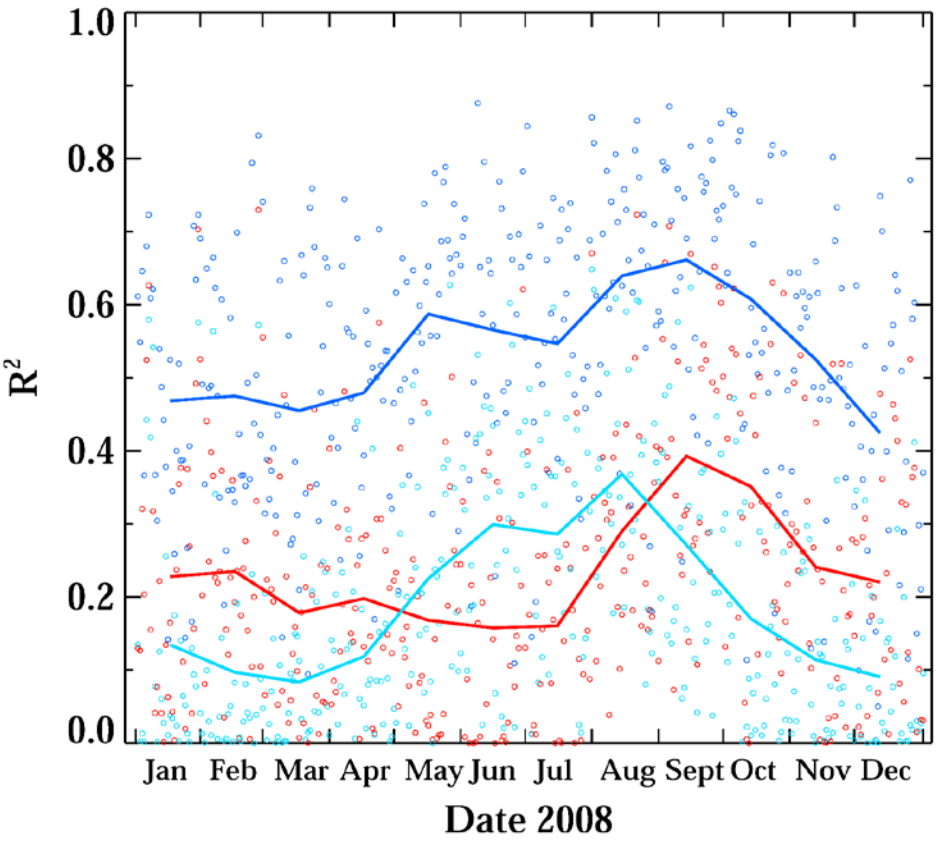

Figure 11. Daily performance of the WRF-Chem simulation (red), the AOD-only statistical model (light blue), and the full statistical model (dark blue). WRF-Chem has highest skill in Aug – Oct, AOD statistical model has best skill Jun – Sept, while full statistical model always has best skill, which peaks in Sept.

Discussion

In this study, ground observations, land-use data, meteorological re-analyses, chemical transport model output, and satellite data are efficiently combined in a statistical model to create a best-estimate of daily $PM_{2.5}$ in the eastern United States in 2008. The model developed herein generates a high-fidelity estimate ($r^2 = 0.75$ using a 10-fold cross-validation) of daily $PM_{2.5}$ over a large area at a high spatial resolution. Information from WRF-Chem is a substantial contributor to the skill of the model. Gap-filled satellite AOD is also an important contributor to the performance of the model. Meteorological information and land-use data are secondary contributors to model skill. The statistical model performs better than WRF-Chem as a stand-alone product.

This study is novel in several ways. First, we show the benefits of a gap-filled AOD product for the eastern United States. Less than half of the eastern United States returns a valid AOD pixel on any given day due to clouds and snow cover, which obstruct the retrieval – more obstructed retrievals occur in the northern US and during winter. Days with obstructed retrievals may also have poor surface air quality, which goes undetected by the MODIS instrument. We demonstrate that the gap-filled AOD product is a reasonable estimate of AOD in areas when actual satellite AOD is missing. This allows us to generate an estimate of daily $PM_{2.5}$ in areas that have no other indicators of air quality. The error characterization of the gap-filled AOD remains challenging due to limited AERONET measurements under cloudy conditions.

The regional chemical transport model is a critical contributor to the overall skill of our statistical model. When speciated components of particulate matter regional chemical transport models are combined in a statistical manner with other variables correlated with air quality, it proves to be a useful contributor, especially during the late summer and autumn.

The domain size includes many metropolitan areas at high spatial resolution. Many previous studies focus on smaller regions which encompass only a few cities, or larger regions at coarse spatial resolution. Here, we develop a daily 1×1 km^2 product for most major cities in the United States east of the Mississippi River including New York City, Baltimore/Washington D.C., Atlanta, and Chicago among others. Despite the large domain, the model generates a high-fidelity estimate ($r^2 = 0.76$ using a 10-fold cross-validation) of daily $PM_{2.5}$ throughout the eastern United States. By enhancing the satellite data to be an important contributor in estimating daily $PM_{2.5}$, we can better fit the needs of those linking $PM_{2.5}$ to health outcomes, which are often conducted at the county level. This 1×1 km^2 $PM_{2.5}$ product could be very helpful for epidemiological studies.

Our study yields comparable statistics to other studies producing similar $PM_{2.5}$ products. Lv et al.[40,41] applied a similar gap-filling AOD technique to the Beijing metropolitan area and was able to predict $PM_{2.5}$ with a cross-validated skill of r^2=0.78. Cross-validation from similar studies in the varying regions of the United States for the year 2008 are between r^2=0.67 and r^2=0.88[27-33]. However, in some cases, the areas of focus are smaller[27-31] and in all cases many more covariates (>>11) are used to train the model. A benefit of this study is that we only use 11 covariates that

are most correlated to $PM_{2.5}$, which allows us to be more computationally efficient and minimizes overfitting.

While this study focuses only on 2008, future work will extend this forward to the current year using the same techniques described herein. Additional work will also focus on developing a statistical model that can simulate speciated $PM_{2.5}$. We also intend to update this model as future AOD products are publicly released. In particular, AOD data from a geostationary satellite such as GOES-16 could be especially helpful in improving daily $PM_{2.5}$ estimation.

Acknowledgments

This publication was developed under Assistance Agreement No. RD835871 awarded by the U.S. Environmental Protection Agency to Yale University. It has not been formally reviewed by EPA. The views expressed in this document are solely those of the SEARCH Center and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication. This publication was also developed using support from the Office of Biological and Environmental Research (BER) in the U.S. Department of Energy (DOE). The WRF-Chem simulation used in this work was performed on Stampede, provided as an Extreme Science and Engineering Discovery Environment (XSEDE) digital service by the Texas Advanced Computing Center (TACC) (http://www.tacc.utexas.edu), which is supported by National Science Foundation grant number ACI-1053575 and on Yellowstone (ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation and Information Systems Laboratory. The authors are grateful to Rob Levy of NASA Goddard Space Flight Center (GSFC) for his constructive comments on this manuscript. EPA $PM_{2.5}$ data were downloaded from the AQS Data Mart, and can be freely retrieved from: https://aqs.epa.gov/aqsweb/documents/data_mart_welcome.html. We acknowledge use of MAIAC data (http://dx.doi.org/10.5067/MODIS/MCD19A2) from the Terra and Aqua MODIS sensors available at: https://earthdata.nasa.gov/ and processed by Alexei Lyapustin and Yujie Wang. We would also like to thank the AERONET team – Brent Holben, David Giles among others – for maintaining the AERONET measurement database and for making the data freely available at: https://aeronet.gsfc.nasa.gov/. Argonne National Laboratory is operated by UChicago Argonne, LLC, under contract no. DE-AC2-06CH11357 with the U.S. Department of Energy.

References

(1) Franklin, M.; Zeka, A.; Schwartz, J. Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. J. Exposure Sci. Environ. Epidemiol. 2007, 17 (3), 279-287.

(2) Anenberg, S. C.; Horowitz, L. W.; Tong, D. Q.; West, J. J. An Estimate of the Global Burden of Anthropogenic Ozone and Fine Particulate Matter on Premature Human Mortality Using Atmospheric Modeling. Environ. Health Perspect. 2010, 118 (9), 1189-1195.

(3) Lelieveld, J.; Evans, J. S.; Fnais, M.; Giannadaki, D.; Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature. 2015, 525 (7569), 367-371.

(4) Schwartz, J.; Dockery, D. W.; Neas, L. M. Is daily mortality associated specifically with fine particles? J. Air Waste Manage. 1996, 46 (10), 927-939.

(5) Peng, R. D.; Bell, M. L.; Geyh, A. S.; McDermott, A.; Zeger, S. L.; Samet, J. M.; Dominici, F. Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. Environ. Health Perspect. 2009, 117 (6), 957.

(6) Krall, J. R.; Anderson, G. B.; Dominici, F.; Bell, M. L.; Peng, R. D. Short-term exposure to particulate matter constituents and mortality in a national study of US urban communities. Environ. Health Perspect. 2013, 121 (10), 1148.

(7) U.S. EPA. Revised requirements for designation of reference and equivalent methods for PM2.5 and ambient air quality surveillance for particulate matter – final rule. 40 CFR part 53. 1997, Federal Register, 62(138), 38763-38830.

(8) Noble, C. A.; Vanderpool, R. W.; Peters, T. M.; McElroy, F. F.; Gemmill, D. B.; Wiener, R. W. Federal reference and equivalent methods for measuring fine particulate matter. Aerosol Sci. Technol. 2001, 34 (5), 457-464.

(9) Wang, J.; Christopher, S. A. Intercomparison between satellite derived aerosol optical thickness and PM2.5 mass: implications for air quality studies. Geophys. Res. Lett. 2003, 30(21).

(10) Chu, D. A.; Kaufman, Y. J.; Zibordi, G.; Chern, J. D.; Mao, J.; Li, C.; Holben, B. N. Global monitoring of air pollution over land from the Earth Observing System-Terra Moderate Resolution Imaging Spectroradiometer (MODIS). J. Geophys. Res.-Atmos. 2003, 108 (4661), D21.

(11) 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 (16), 2495-2509.

(12) Al-Saadi, J.; Szykman, J.; Pierce, R. B.; Kittaka, C.; Neil, D.; Chu, D. A.; Remer, L.; Gumley, L.; Prins, E.; Weinstock, L. Improving national air quality forecasts with satellite aerosol observations. Bull. Amer. Met. Soc. 2005, 86 (9), 1249-1261.

(13) Gupta, P.; Christopher, S. A.; Wang, J.; Gehrig, R.; Lee, Y. C.; Kumar, N. Satellite remote sensing of particulate matter and air quality assessment over global cities. Atmos. Environ. 2006, 40 (30), 5880-5892.

(14) Zhang, H.; Hoff, R. M.; Engel-Cox, J. A. The relation between Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth and PM2.5 over the United States: a geographical comparison by US Environmental Protection Agency regions. J. Air Waste Manage. 2009, 59 (11), 1358-1369.

(15) Remer, L. A.; Kaufman, Y. J.; Tanré, D.; Mattoo, S.; Chu, D. A.; Martins, J. V.; Li, R.-R.; Ichoku, C.; Levy, R. C.; Kleidman, R. G. The MODIS aerosol algorithm, products, and validation. J. Atmos. Sci. 2005, 62 (4), 947-973.
(16) Levy, R. C.; Mattoo, S.; Munchak, L. A.; Remer, L. A.; Sayer, A. M.; Patadia, F.; Hsu, N. C. The Collection 6 MODIS aerosol products over land and ocean. Atmos. Meas. Tech. 2013, 6 (11), 2989.
(17) Hsu, N. C.; Jeong, M.; Bettenhausen, C.; Sayer, A. M.; Hansell, R.; Seftor, C. S.; Huang, J.; Tsay, S. Enhanced Deep Blue aerosol retrieval algorithm: The second generation. J. Geophys. Res-Atmos. 2013, 118 (16), 9296-9315.
(18) Lyapustin, A.;Wang, Y.; Korkin, S.; Huang, D. MODIS Collection 6 MAIAC algorithm. Atmos. Meas. Tech. 2018, 11, 5741-5765.
(19) Sayer, A. M.; Munchak, L. A.; Hsu, N. C.; Levy, R. C.; Bettenhausen, C.; Jeong, M. MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and "merged" data sets, and usage recommendations. J. Geophys. Res-Atmos. 2014, 119, 13965-13989.
(20) Remer, L. A.; Mattoo, S.; Levy, R. C.; Munchak, L. A. MODIS 3 km aerosol product: algorithm and global perspective. Atmos. Meas. Tech. 2013, 6 (7), 1829-1844.
(21) Gupta, P.; Christopher, S. A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. J. Geophys. Res-Atmos. 2009, 114, D14205.
(22) Hoek, G.; Beelen, R.; De Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 2008, 42 (33), 7561-7578.
(23) Liu, Y.; Sarnat, J. A.; Kilaru, V.; Jacob, D. J.; Koutrakis, P. Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing. Environ. Sci. Technol. 2005, 39 (9), 3269-3278.
(24) Liu, Y.; Franklin, M.; Kahn, R.; Koutrakis, P. Using aerosol optical thickness to predict ground-level PM2.5 concentrations in the St. Louis area: a comparison between MISR and MODIS. Remote Sensing Environ. 2007, 107 (1), 33-44.
(25) Liu, Y.; Paciorek, C. J.; Koutrakis, P. Estimating Regional Spatial and Temporal Variability of PM2.5 Concentrations Using Satellite Data, Meteorology, and Land Use Information. Environ. Health Persepct. 2009, 117 (6), 886.
(26) Lee, H. J.; Liu, Y.; Coull, B. A.; Schwartz, J.; Koutrakis, P. A novel calibration approach of MODIS AOD data to predict PM2.5 concentrations. Atmos. Chem. Phys. 2011, 11 (3), 9769-9795.
(27) Kloog, I.; Koutrakis, P.; Coull, B. A.; Lee, H. J.; Schwartz, J. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmos. Environ. 2011, 45 (35), 6267-6275.
(28) Kloog, I.; Nordio, F.; Coull, B. A.; Schwartz, J. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the Mid-Atlantic states. Environ. Sci. Technol. 2012, 46 (21), 11913-11921.
(29) Kloog, I.; Chudnovsky, A.; Just, A.; Nordio, F.; Koutrakis, P.;Coull, B. A.; Lyapustin, A.; Wang, Y.; Schwartz, J. A new hybrid spatio-tempoarl model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth. Atmos. Environ. 2014, 95, 581-590.

(30) Hu, X.; Waller, L. A.; Al-Hamdan, M. Z.; Crosson, W. L.; Estes, M. G.; Estes, S. M.: Quattrochi, D. A.; Sarnet, J. A.; Liu, Y. Estimating ground-level PM2.5 concentrations in the southeastern US using geographically weighted regression. Environ. Res. 2013, 121, 1-10.
(31) Gupta, P.; Christopher, S. A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. J. Geophys. Res-Atmos. 2009, 114, D20205.
(32) Di, Q.; Kloog, I.; Koutrakis, P.; Lyapustin, A.; Wang, Y.; Schwartz, J. Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. Environ. Sci. Technol. 2016, 50 (9), 4712-4721.
(33) Hu, X.; Belle, J. H.; Meng, X.; Wildani, A.; Waller, L.; Strickland, M.; Liu, Y. Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach. Environ. Sci. Technol. 2017,
(34) Liu, Y.; Park, R. J.; Jacob, D. J.; Li, Q.; Kilaru, V.; Sarnat, J. A. Mapping annual mean ground-level PM2.5 concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. J. Geophys. Res-Atmos. 2004, 109, D22206.
(35) van Donkelaar, A.; Martin, R. V.; Park, R. J. Estimating ground level PM2.5 using aerosol optical depth determined from satellite remote sensing. J. Geophys. Res-Atmos. 2006, 111, D21201.
(36) van Donkelaar, A.; Martin, R. V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P. J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. Environ. Health Persepct. 2010, 118 (6), 847.
(37) van Donkelaar, A.; Martin, R. V.; Spurr, R. J. D.; Drury, E.; Remer, L. A.; Levy, R. C.; Wang, J. Optimal estimation for global ground level fine particulate matter concentrations. J. Geophys. Res-Atmos. 2013, 118 (11), 5621-5636.
(38) van Donkelaar, A.; Martin, R. V.; Spurr, R. J. D.; Burnett, R. T. High-resolution satellite-derived PM2.5 from optimal estimation and geographically weighted regression over North America. Environ. Sci. Technol. 2015, 49 (17), 10482-10491.
(39) van Donkelaar, A.; Martin, R. V.; Brauer, M.; Hsu, N. C.; Kahn, R. A.; Levy, R. C.; Lyapustin, A.; Sayer, A. M.; Winker, D. M. Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. Environ. Sci. Technol. 2016, 50 (7), 3762–3772.
(40) Lv, B.; Hu, Y.; Chang, H. H.; Russell, A. G.; Bai, Y. Improving the accuracy of daily PM2. 5 distributions derived from the fusion of ground-level measurements with aerosol optical depth observations, a case study in north China. Environ. Sci. Technol. 2016, 50 (9), 4752-4759.
(41) Lv, B.; Hu, Y.; Chang, H. H.; Russell, A. G.; Cai, J.; Xu, B.; Bai, Y. Daily estimation of ground-level PM2.5 concentrations at 4km resolution over Beijing-Tianjin-Hebei by fusing MODIS AOD and ground observations. Sci. Total Environ. 2017, 580 235-244.
(42) Marshall, J. D.; Nethery, E.; Brauer, M. Within-urban variability in ambient air pollution: comparison of estimation methods. Atmos. Environ. 2008, 42 (6), 1359-1369.
(43) de Hoogh, K.; Gulliver, J.; van Donkelaar, A.; Martin, R. V.; Marshall, J. D.; Bechle, M. J.; Cesaroni, G.; Pradas, M. C.; Dedele, A.; Eeftens, M. Development of West-European PM2.5 and NO2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 2016, 151 1-10.

(44) Di, Q.; Koutrakis, P.; Schwartz, J. A hybrid prediction model for PM2.5 mas and componenets using a chemical transport model and land use regression. Atmos. Environ. 2016, 131, 390-399
(45) Hand, J. L.; Schichtel, B. A.; Pitchford, M.; Malm, W. C.; Frank, N. H. Seasonal composition of remote and urban fine particulate matter in the United States. J. Geophys. Res-Atmos. 2012, 117, D05209.
(46) Energy Information Administration (EIA). Household heating fuels vary across the country. 2011, https://www.eia.gov/todayinenergy/detail.php?id=3690
(47) Gobeli, D; Meyer, M. B.; Schloesser, H.; Pottberg, T. Finally, a continuous FEM for PM2.5. Environ. Manage. 2009, 2, 6-9.
(48) Yahya, K.; Glotfelty, T.; Wang, K.; Zhang, Y.; Nenes, A. Modeling regional air quality and climate: improving organic aerosol and aerosol activation processes in WRF/Chem version 3.7.1. Geosci. Model Dev. 2017, 10 (6), 2333.
(49) Wang, K.; Zhang, Y.; Yahya, K.; Wu, S.-Y.; Grell, G. Implementation and initial application of new chemistry-aerosol options in WRF/Chem for simulating secondary organic aerosols and aerosol indirect effects for regional air quality. Atmos. Environ. 2015, 115, 716-732.
(50) Glotfelty, T.; He, J.; Zhang, Y. Impact of future climate policy scenarios on air quality and aerosol-cloud interactions using an advanced version of CESM/CAM5: Part I. model evaluation for the current decadal simulations. Atmos. Environ. 2017, 152, 222-239.
(51) Glotfelty, T.; Zhang, Y. Impact of future climate policy scenarios on air quality and aerosol-cloud interactions using an advanced version of CESM/CAM5: Part II. Future trend analysis and impacts of projected anthropogenic emissions. Atmos. Environ. 2017, 152, 531-552.
(52) Dee, D. P.; Uppala, S. M.; Simmons, A. J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M. A.; Balsamo, G.; Bauer, P. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly J. Royal Met. Soc. 2011, 137 (656), 553-597.
(53) Zhang, H.; Wang, Y.; Park, T.-W.; Deng, Y. Quantifying the relationship between extreme air pollution events and extreme weather events. Atmos. Res. 2017, 188 64-79.
(54) Bell, M. L.; Dominici, F.; Ebisu, K.; Zeger, S. L.; Samet, J. M. Spatial and temporal variation in PM2.5 chemical composition in the United States for health effects studies. Environ. Health Perspect. 2007, 115 (7), 989.
(55) Gupta, P., Christopher, S.A., 2008. An evaluation of Terra-MODIS sampling for monthly and annual particulate matter air quality assessment over the Southeastern United States. Atmos. Environ. 42, 6465–6471.
(56) Christopher, S. A.; Gupta, P. Satellite remote sensing of particulate matter air quality: The cloud-cover problem. J. Air Waste Manage. Assoc., 2010, 60(5), 596-602.