**Fine particulate matter predictions using high resolution aerosol optical depth (AOD) retrievals**

Alexandra A. Chudnovsky1\*,2, Petros Koutrakis1, Itai Kloog1,3, Steven Melly1, Francesco Nordio1, Alexei Lyapustin4, Yujie Wang5, and Joel Schwartz1

1Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA

[achudnov@hsph.harvard.edu](mailto:achudnov@hsph.harvard.edu), phone: +1-617-3848846, 972-3-546996221

2Department of Geography and Human Environment, Tel-Aviv University, Israel

3Department of Geography and Environmental Development, Ben-Gurion University, Israel

4GEST / UMBC, NASA Goddard Space Flight Center, Baltimore, Maryland, USA

5University of Maryland Baltimore County, Baltimore, MD, USA

**Abstract**

To date, spatial-temporal patterns of particulate matter (PM) within urban areas have primarily been examined using models. On the other hand, satellites extend spatial coverage but their spatial resolution is too coarse. In order to address this issue, here we report on spatial variability in PM levels derived from high 1km resolution AOD product of Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm developed for MODIS satellite images. We apply day-specific calibrations of AOD data to predict PM2.5 concentrations within the New England area of the United States. To improve the accuracy of our model, land use and meteorological variables were incorporated. We used inverse probability weighting (IPW) to account for nonrandom missingness of AOD and nested regions within days to capture spatial variation in the daily calibration across New England. With this approach we can control for the inherent day-to-day variability in the AOD-PM2.5 relationship, which depends on time-varying parameters such as particle optical properties, vertical and diurnal concentration profiles and ground surface reflectance among others. Out-of-sample "ten-fold" cross-validation was used to quantify the accuracy of model predictions. Our results show that the model-predicted PM2.5 mass concentrations are highly correlated with the actual observations, with out-of-sample R2of 0.89. Furthermore, our study clearly shows that the model captures the pollution levels along highways and many urban locations thereby extending our ability to investigate the spatial patterns of urban air quality, such as examining exposures in areas with high traffic. Importantly, our results show high accuracy within the cities of Boston and New Haven thereby indicating that MAIAC data can be used to examine intra-urban exposure contrasts in PM2.5 levels.

**Keywords**: Particulate matter, PM2.5, Aerosol Optical Depth (AOD), High resolution aerosol retrieval, MAIAC, intra-urban pollution, variability in PM2.5 levels, scales of pollution

**Introduction**

Numerous studies have reported associations between mortality and morbidity outcomes and particulate matter with aerodynamic diameter≤2.5 μm (PM2.5) (Dockery et al. 1993; Schwartz 1994; Schwartz et al., 1995; Pope et al. 2002; U.S. Environmental Protection Agency (EPA) 2004). Although ground-level PM2.5 monitoring sites provide accurate measurements, their spatial coverage is limited and thus often insufficient to capture the PM2.5 spatial variability for exposure and epidemiological studies.

Satellite imagery adds another important tool to air quality and pollution monitoring due to its extensive spatial coverage and repeated observations of the earth surface and atmosphere. The most common parameter derived from satellite observations is the aerosol optical depth (AOD), which is a measure of the extinction of electromagnetic radiation at a given wavelength due to the presence of aerosols in an atmospheric column. However, the satellite-derived AOD is a measure of light attenuation in the column which is affected by ambient conditions (e.g., humidity, vertical profile, chemical composition etc.), while PM2.5 is a measure of dry particle mass near the surface; therefore, these two parameters are not expected to be strongly correlated.

To date, spatial patterns of particle exposure within populated areas have been examined by models, such as land use regression (Jerrett et al., 2005; Hoek et al., 2008), line dispersion models such as CAlifornia LINE Source Dispersion Model (CALINE, Benson 1984; Benson 1992), or atmospheric chemistry and transport models such as GEOS-Chemical transport model (CTM) (<http://geos-chem.org>). Indeed, it is challenging to deploy hundreds of monitors, needed to measure spatial variability in PM levels. On the other hand, satellites extend spatial coverage but their spatial resolution is insufficient. The Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua satellites provide a daily global coverage but the conventional resolution of its aerosol product (10 km Dark Target (DT)) is often too coarse for suitable exposure estimates in urban areas. The widely anticipated 3 km MODIS AOD product is expected to be generated as a part of Collection 6 re-processing (Levy et al., 2013; Remer et al., 2013). Recently, a new Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm was developed for MODIS which provides aerosol information at 1 km resolution (Lyapustin et al., 2011a,b). Several studies published in the last 3 years have shown that high spatial resolution is essential to detect spatial variability in PM levels (Kumar et al., 2011) and in aerosol loadings at regional and at a sub-10 km scale (e.g. intra-urban domain) (Emili et al., 2011, Chudnovsky et al., 2013a). In another study Chudnovsky et al., (2013b) showed that MAIAC provides a factor of 1.5-1.8 more data than the DT 10 km over New England; the improvement being over brighter urban surfaces and partly cloudy days. However authors conducted straightforward analyses of AOD vs PM2.5 association as a simple metric to compare between 10 km and 1 km retrievals and its seasonal frequency without any goal to provide PM2.5 estimates at the urban scale

Recently, several studies proposed that the effects of the time-varying parameters influencing the AOD-PM2.5 relationship can be taken into account with daily adjustments, resulting in much higher R2 than previously reported in the literature, ranging from 0.83 to 0.92 (Lee et al., 2011; Kloog et al., 2011; 2012; Chudnovsky et al., 2012). Furthermore, Kloog et al., 2011 and Kloog et al., 2012 extended the previously established model on day-specific calibrations of AOD data using ground PM2.5 measurements by incorporating commonly used Land Use (LU) and meteorological variables. The goal of the current study is to investigate if the improved 1 km spatial resolution would better explain the variations in PM2.5. We use MAIAC 1 km resolution AOD data from MODIS Aqua for the New England area to obtain daily PM2.5 estimates for AOD retrieval days and explore its applicability at the intra-urban scale. By further developing the approach described in Kloog et al. (2011), we investigate the extent of spatial variability of the AOD-PM2.5 relationship on a daily basis and show how this variability can be captured by a mixed effects model approach during the period of 2003. Finally, we present the modeled spatial pattern of PM2.5 levels within the study domain for selected days.

**2 Material and methods**

**2.1 Ground-level PM2.5 data**

Twenty-four hour PM2.5 concentrations were calculated at 62 U.S. Environmental Protection Agency (EPA) PM2.5 monitoring sites during 2003 (Figure 1). These include 14 sites in Maine (ME), 10 sites in New Hampshire (NH), 6 sites in Vermont (VT), 16 sites in Massachusetts (MA), 16 sites in Connecticut (CT) and 2 sites in Rhode Island (RI). Sampling frequency differed by site and included samples collected every day, every third day, and every sixth day. Additionally, we used 24 hour PM2.5 concentrations from the Harvard School of Public Health (HSPH) supersite located near downtown Boston, MA.

**2.2 Satellite data**

A new algorithm MAIAC (Lyapustin et al. 2011a,b; 2012a) has been developed to process MODIS data. MAIAC retrieves aerosol parameters over land at 1 km resolution simultaneously with parameters of a surface bidirectional reflectance distribution function (BRDF). This is accomplished by using a time series of MODIS measurements and simultaneous processing of groups of pixels. The MAIAC algorithm ensures that the number of measurements exceeds the number of unknowns, a necessary condition for solving an inverse problem without empirical assumptions typically used by current operational algorithms. The MODIS time series accumulation also provides multi-angle coverage for every surface grid cell, which is required for the BRDF retrievals from MODIS data. The aerosol parameters include optical depth, Angstrom exponent from 0.47 and 0.67 μm, and aerosol type including background, smoke and dust models (Lyapustin et al., 2012b). The background models are specified regionally based on the climatology of the Aerosol Robotic Network (AERONET) (Holben et al. 1998) sun-photometer data for relatively low AOD days (<0.5). The smoke/dust types are identified as aerosols with increased shortwave absorption (Lyapustin et al., 2012b) and dominant fine/coarse mode particles, respectively. AERONET validation over the continental USA showed that the MAIAC and MODIS Dark Target (DT) algorithms have a similar accuracy over dark and vegetated surfaces, but also showed that MAIAC generally improves accuracy over brighter surfaces, including most urban areas (Lyapustin et al. 2011b). The improved accuracy of MAIAC results from using the explicit surface characterization method in contrast to the empirical surface parameterization approach, which is utilized in the DT algorithm. Further, MAIAC incorporates a cloud mask (CM) algorithm based on spatio-temporal analysis which augments traditional pixel-level cloud detection techniques (Lyapustin et al., 2008). In this work, the residual contamination by clouds and cloud shadows was additionally reduced by discarding 2 pixels adjacent to detected clouds. Importantly, the MAIAC approach becomes indispensable in heterogeneous aerosol environments, *e.g*. with local sources such as fire smoke plumes or in urban/industrial areas.

**2.3 Spatial (LU) and temporal (meteorology) predictors of PM2.5**

Different land use (LU, spatial predictors) and meteorological (MET, temporal predictors) variables were examined to improve the predictions of PM2.5 at the 1 km scale. Whereas land use parameters enable us to capture traffic and point sources, meteorological conditions can influence the AOD-PM relationship and AOD retrieval accuracy. For example, AOD generally increases with relative humidity for hygroscopic particles due to hygroscopic growth (Bergin et al., 2000; Altaratz et al., 2013).

Multiple LU and MET potential predictors of PM measured at EPA sites within the grid cell were examined. Different sets of predictors were combined and the model for each set was run. The selection of a best model was based on the AIC score (Akaike information criterion, a measure of the relative goodness of fit that is asymptotically related to out of sample prediction R2) to further backward-select the variables in each combination. Finally, the combination of predictors with the minimum AIC was selected. As a result of this work, the model was constructed with the following spatial and temporal predictors: percent of the grid cell that is urban, elevation, Normalized Difference Vegetative Index (NDVI), major road density within a grid, distance to PM2.5 point emissions, relative humidity, height of the planetary boundary layer and wind speed.

2.3.1 Percent of urban areas

We used 30 m resolution 2001 national land cover data (NLCD). Data were obtained as raster files with 30 m cell size from mrlc.gov. The ArcGIS Spatial Analyst Neighborhood Focal Statistics tool was used to calculate count of urban cells in 33 X 33 cell rectangle, where the category “developed land” was counted as urban. Then, percent of urban spaces data in 33 X 33 30 m cell rectangle surrounding each EPA monitor were calculated.

2.3.2 Elevation

Elevation data above sea level in meters were obtained through the national elevation dataset (NED) (Maune, 2007). NED is the distributed by the U.S.Geological Survey (USGS) and provides seamless raster elevation of the conterminous United States. A 30 m resolution raster was created from 1/3 arc second data using ArcGIS.

2.3.3 Normalized Difference Vegetative Index

Satellite-derived normalized difference vegetation index (NDVI) based on red and near infrared (NIR) reflectances a temporal indicator of the vegetation cover and its phenological state (Tucker et al., 1979; Tucker and Sellers, 1986):

NDVI = (ρ NIR − ρ RED)/(ρ NIR + ρ RED) (1)

 We used the 1km MODIS monthly vegetation indices product (MOD13A3).Except for periods of spring green-up and fall senescence associated with seasonal surface change, the NDVI can generally be considered relatively constant on a monthly basis. Many other studies suggested that NDVI was found to be good predictors for local means of pollutant concentrations (e.g. Su et al., 2009).

2.3.4 Distance to PM2.5 point source emissions

PM2.5 point source emissions were obtained through the 2005 US EPA National Emissions Inventory (NEI) facility emissions report (EPA, 2010). The distance (in km) from the AOD grid centroid to the nearest point source emission (from the EPA emissions dataset, tons per year) was calculated in ArcGIS. Out of the 62 monitors used in the analysis, 6 were within 150 m, 19 were within 550 m and 26 were within 2-10 km of an industrial source, respectively.

2.3.5 Road density

Major roads were selected from ESRI Street Map data provided with ArcGIS (version 10) using census Feature Classification Code (FCC) (e.g. A1: primary highway with limited access, A2: primary road without limited access, A3: secondary and connecting road). Road density was calculated using ArcGIS to create a line density raster at a 1x1 km resolution to match the AOD grid cells. We used all A1, A2 and A3 roads in our study region. Thus each AOD gird cell was assigned the density (e.g. road density) in each grid based on the density of intersecting roads in each grid. Because the distributions of major roads were highly right-skewed, we used a logarithmic transformation.

2.3.6 Temporal predictors: meteorological data

All meteorological variables used in the analysis (wind speed (WS), relative humidity (RH)) were obtained through the national climatic data center (NCDC) (NCDC, 2010). There were 44 active MET stations across New-England during the study period. Daily meteorological (MET) data was used in the analysis. Using ArcGIS we assigned to each AOD grid the closest MET station in our dataset (based on multiple MET data sources including EPA and The National Climatic Data Center (NCDC)). Height of the planetary boundary layer (PBL) was obtained from the NOAA Reanalysis Data.

**2.4 Statistical model**

We apply a mixed effects model approach to MAIAC AOD retrievals and other meteorological and LU variables to predict PM2.5 concentrations in each grid cell. This model allows for the regression intercepts and slopes to vary daily in order to account for the inherent day-to-day variability in the AOD-PM2.5 relationship. Furthermore, since New England is a relatively large area and PM-AOD relationship can vary spatially, we partitioned our study area into three sub-regions (Figure 1) and allowed for the daily AOD-PM2.5 slopes to vary by region. Region 1 included ME, VT and NH states, region 2 included MA while CT and RI formed region 3 in our analyses. Although there are some variations among the three regions in topography and climate conditions mostly via the usual north-south snow cover gradient in winter, the main difference appears at the level of urbanization and land use affecting surface brightness and thus the AOD vs PM2.5 relationships (Chudnovsky et al., 2013b) and the quality of aerosol product. For instance, validation analysis of MODIS 3km product (Munchak et al., 2013) showed a strong correlation between percent of retrievals above expected error and percent of the urban land cover. A similar investigation is ongoing for MAIAC. Of the three, region 1 is least urbanized with high fraction of forest cover and region 2 is most urbanized. Thus, by dividing the study area into regions we can evaluate the role of environmental conditions (e.g. snow coverage) and different land use settings on AOD-PM2.5 relationship. We used the following mixed effects model with random intercepts and slopes that was applied on AOD retrieval days/pixels (Eq.2):

*PMij = (α + uj+gj(reg)) + [ (β1 + vj+ hj(reg)) × AODij] + β2Elevation +β3NDVI +β4WS + β5Urban + β6Log(RoadDensity) + β7PBL + β8Humidity + β9 DistanceEmission + β10WS\*PBL +β11Log(RoadDensity)\*PBL+β12 AOD2+εij* (2) *(ujvj ) ~ [(o o), Σβ ]*

*(gj (reg)hj (reg) ) ~ [(o o), Σreg ]*

where *PMij* is the PM2.5 concentration at a spatial site *i* on day *j*; *AODij* is the AOD value in the grid cell corresponding to site *i* on day *j*; *α* and *uj* are the fixed and random intercepts, respectively; *β1*and*vj* are the fixed and random slopes, respectively; *Wind speed (WS), Humidity, PBL (Planetary Boundary Layer)* are the values in the grid cells corresponding to site i on a day j. *Elevation*, *road density (RoadDensity), distance to emission sources (DistanceEmission), percent of urban space (Urban*) are the values in grid cells corresponding to site i. *gj(reg) and hj(reg)* are the daily random intercepts and AOD slopes specific to each study area region. *εij ~N(0, σ2)* is the error term at site i on a day j and *Σβ* is the variance-covariance matrix for the random effects. The AOD fixed effect in the model (Eq.2) accounts for the effect of AOD on PM2.5, which was the same for all study days. The AOD random effects explain the daily variability in the PM2.5-AOD relationship. Since height of the boundary layer may vary with wind speed (Oke 1987), influencing the concentration and vertical profile of pollutants, both terms were included as interaction terms. For example, boundary layer not only controls transport and location of pollutants and aerosols but also their concentrations would be different in variable boundary layer structures (Angevine et al., 2013). The solution of the mixed model equations is maximum likelihood, a form of estimation that accounts for the parameters in the fixed-effects structure of the model to reduce the bias in the covariance parameter estimates (Lindstrom and Bates 1988; Laird et al. 1982). Currently, this method is implemented in the SAS (Statistical Analysis System) statistical software package version 9.3 (proc mixed).

In addition, we incorporated inverse probability weighting (IPW) to potentially avoid bias in the regression coefficient estimates and thus in the resulting predictions. This approach effectively up-weights dates and grid cells which are under-represented due to missing data, as described in Kloog et al., 2012.Finally, PM2.5 concentrations for each grid cell on a day j were estimated using the corresponding AOD values where the fixed and random intercepts, the fixed and random slopes for each study day and for each region were derived from Eq. (2).

Importantly, in this study we compared the full model above with reduced models to ascertain the benefits of increasing complexity. The models examined were: 1) AOD only model;2) AOD+MET model; 3) AOD + LU model; and 4) AOD+LU+MET.

**2.6 Model validation**

We use a cross-validation (CV) approach to evaluate the ability of the model to predict PM2.5 concentrations for each pixel in the study area. Thus, the dataset was repeatedly randomly divided into 90% (calibration) and 10% (held-out test) splits. We applied the fitted calibration model to estimate PM2.5 for the held-out test set. This “out-of-sample” process was repeated ten times to calculate the cross-validated (CV) R2 values. Subsequently, the predicted PM2.5 concentrations were compared to those measured at each site. Overall temporal R2 was calculated by regressing Δ PM against Δ predicted where: Δ PM is the difference between the observed PM2.5 at a given site on a given day and the annual mean PM2.5 at that location, and Δ predicted is defined similarly for the predicted values generated from the model. Overall spatial R2 was calculated by regressing the annual mean PM2.5 at a given site against the annual mean predicted PM2.5 at that location.

**2.7 Estimation of PM2.5 levels in the study domain and their spatial variability**

We examined spatial PM2.5 patterns in New England using the mixed effects models described above. We focus here on five days based on the MAIAC AOD retrieval from Aqua in 2003: (i) a medium to high pollution event (regional source of pollution): June 25, and (ii) low pollution events (local sources): April 25, May 19, July 31 and November 15. We focus in more detail on the June 25th and November 15th data to analyze the consistency and quality of the high-resolution AOD retrievals. In particular, it is known that the MODIS dark target algorithm (Levy et al., 2007) has a bias over brighter urban surfaces (e.g., Munchaket al., 2013), and this comparison has been performed to assure that MAIAC algorithm has reduced this error.

**3 Results**

**3.1 PM2.5 prediction based on a mixed-effects model**

Using the entire data set, estimates of PM2.5 concentrations were obtained for 196 days (e.g. days with available AOD vs PM2.5 pairs) during 2003. The retrieval rate is lower for higher numbers of AOD vs PM2.5 pairs: there were 95 days with at least five pairs and 44 days with at least 10 pairs. Note that during the same period of observations but using the conventional 10 km AOD retrieval (DT) for the entire data set there were 177 days available (Kloog et al., 2011).

The fixed effects of the AOD intercept and slope were statistically significant: α=8.91 (p<0.0001); and β1=16.20 (p<0.0001), respectively. The fixed effects of spatial and temporal predictors were also significant. In addition, the random slopes for AOD by day, and by day and region were both significant (p<0.0001). Figure 2 shows the daily variation of random AOD intercepts and slopes. Note that these results (significant random effects) support the hypothesis that because the parameters influencing the relationship between PM2.5 and AOD vary from day to day within a given domain, it is necessary to adjust for this daily variability. Table 1 presents modeled and cross-validated R2 for 2003 for four different models: the AOD model, AOD+MET model, AOD + LU model and our final model: AOD+LU+MET. As can be seen, adding the LU terms to the AOD model significantly improves the spatial R2, and the full model, incorporating LU and MET, improves it further. In contrast, the AOD model is sufficient to well characterize the temporal variation. The final CV R2 resulted in an R2 value of 0.88 with a spatial R2 of 0.80. In addition, if we regress the out of sample measured PM2.5 against the predicted we get a measure of bias in the relationship. We obtained a slope of 0.99, and intercept of 0.01, indicating a very low (or negligible) bias in the prediction model. Furthermore, the increase in spatial R2 (from R2=0.58 when the main explanatory variable was AOD to R2=0.80 when all parameters were incorporated in a model) is of special importance since spatial variation is essential for chronic exposures studies. We can contrast this to previous results of Kloog et al (2011), with overall cross-validated R2 of 0.83 in New England, with a spatial R2 of 0.78 and a temporal R2 of 0.84 when using 10 km AOD. Hence use of the MAIAC data improves overall, temporal, and spatial R2.

Figure 3 shows the seasonal mean residuals per EPA site. As can be seen, there is a high prediction accuracy for most of sites and for all seasons, with the mean yearly residual value of 1.71 μg/m3 (standard deviation of the mean=1.20 μg/m3). The fall season shows the lowest mean residual value of 1.27 μg/m3 (stdev=1.12 μg/m3) whereas the winter has the highest mean of 2.48μg/m3 (stdev=2.21 μg/m3). This is due to the higher AOD retrieval error during winter (undetected residual snow) and lower AOD retrieval rate. Note that one site (EPA code 33-007-4002) located in the forest on a state highway was excluded from our analyses (denoted as 1, Figure 3), since it had the highest average residual compared to other sites. There were only a limited number of PM2.5 measurements which differed significantly from those observed at the nearby sites. Located on the east of Mount Washington, we also assume that the nearby upwind peaks affect this site when the wind is from the prevailing wind direction.

**3.2 Spatial variability in PM2.5 levels during moderate and low pollution events.**

We compare the spatial variability in PM2.5 levels using the AOD model (panel A, Fig. 4) and AOD+LU+MET model (panel B, Fig. 4) during the moderate pollution day on June 25. Generally, both models show similar spatial pattern. Importantly, this pattern of pollution transport can only be captured by using satellite AOD retrievals and cannot be estimated solely based on LU and MET parameters. Furthermore, both models exhibit variability in PM2.5 concentrations across the domain. However, this variability is much larger when using the AOD+LU+MET model. For example, consider the variability in PM2.5concentrations inside of areas marked 1 and 2. The AOD+LU+MET model shows not only higher variations per unit area but also higher predicted concentrations of PM2.5 along the highway (Route 495, area 2). This area corresponds to the Merrimack River Valley, so elevation may be a factor as well as industrial development along the river. The lower left part of it corresponds to the historically industrial city of Lawrence whereas to the west, elevated PM2.5 concentrations correspond to the industrial city of Lowell. Importantly, the area marked 1 in the figures shows uniformly high concentrations in the AOD model, but a more mixed pattern, with pockets of lower concentrations and a few places with even higher concentrations in the AOD+LU+MET model.

What governs spatial variability in PM2.5 levels during low pollution day? To answer this question in Figure 5 we show several days generated by AOD+LU+MET model: April 25, May, 19 and July, 31. For all, PM2.5 concentrations differ by date and location. As expected, highly populated areas such as Bridgeport, New Haven, Hartford, Boston, Springfield and Providence exhibited higher PM2.5 levels, compared to rural areas of Vermont and southwestern New Hampshire. Furthermore, grid cells along major highways (e.g., Interstate Highways 91 and 95) tend to have higher PM2.5 concentrations, perhaps because these cells are more impacted by traffic and are also densely populated. On May 19, urban areas (Providence, Boston, and New Haven), roads, and the area along the coast line (CT) exhibited higher PM2.5 concentrations. Furthermore, there is a higher variability in PM2.5 concentrations between urban and rural locations. Note however, that for all days central Boston appears to have variability in PM2.5 levels.

To rule out algorithm biases with surface brightness, we present an additional AOD map of a clear day acquired on November 15, 2003 (Fig. 6, panel A). As can be seen, the MAIAC retrieval is free from this artifact of processing above urban areas. This day also served as an example of the most extreme difference in PM2.5 spatial pattern using AOD (Fig. 6, panel B) models and AOD+LU+MET (Fig. 6, panel C) on low pollution day. Not surprisingly, using AOD model, the spatial pattern of PM2.5 concentrations follows AOD retrievals. Furthermore, the variability in PM2.5 concentrations between different locations on this day is 1.5 μg/m3. Importantly, as can be seen from Figure 6 (B), the AOD model is prone to overestimate PM2.5concentrations in areas adjacent to clouds. This may represent a residual noise from cloud-contaminated pixels as well as a real physical signal, for example, presence of elevated hydrated aerosols in the vicinity of cloud which is not taken into account by LU and MET ancillary information. When a AOD+LU+MET model was run, urban areas and roads appears to be more polluted than surrounding non-urban areas, with the highest difference in PM2.5 concentration levels of 4.5 μg/m3for the entire domain and between different urban settings, or three times the variation of the AOD model. The results of the AOD+LU+MET model were supported by ground EPA measurements: the seven sites that were available on that day measured PM2.5 concentrations ranging from 2.3-8.9 mg/m3, with a difference of 6.6 μg/m3.

Finally, in Figure 7 we compared annual mean measured and predicted PM2.5 concentrations using AOD and AOD+LU+METmodels for all sites located along major highways (I91, I93 and I95).Generally, the AOD+LU+MET model generates better accuracy (R2=0.82) for those sites than the AOD model (R2=0.61). Using the AOD model we tend to underestimate the PM2.5 levels from road pollution.

**3.3 Model accuracy at the city scale**

High resolution AOD may provide information about local conditions and intra-urban variability, at scales below 10km. With this in mind, we analyzed the accuracy of our model inside the greater Boston area and New Haven (each city contains five ground monitoring sites) to examine performance within cities. Figure 8 shows that using the AOD+LU+MET model there is a good agreement between measured and predicted PM2.5 concentrations for both locations. As can be seen, the model and cross-validated R2 are high for both cities indicating a good agreement between the measured and predicted concentrations. Furthermore, cross-sectional comparison between the mean measured and mean predicted PM2.5 concentrations for each site over the study period for Boston and New Haven R2 are 0.80 and 0.87 respectively (data not shown). This comparison is especially important for determining whether model predictions are suitable assessments for epidemiological studies, which require accurate estimation of spatial patterns. This improvement can be related to the improved MAIAC accuracy over bright/urban areas that has a direct impact on the model accuracy.

**4. Discussion**

In this paper we use the new high-resolution (1 km) AOD retrieval from MODIS data based on MAIAC algorithm to predict PM2.5 concentrations within the New England area of the United States. The main goal was to study if the high resolution AOD can improve our ability to distinguish qualitatively and quantitatively spatial patterns of PM2.5 levels. Toward this end we developed mixed effects model, similar to our previous studies, to explore the advantages of high resolution data set. Importantly, we have shown that PM2.5 prediction accuracy improves further by adding meteorological and land use parameters. We have demonstrated that high resolution remote sensing data can supply critical information in our efforts to obtain accurate exposure estimates for future health effects and risk assessment studies. Moreover, we have shown that using the MAIAC data we obtain better predictive power than with the DT data, temporally, spatially, and overall. While an overall improvement in R2 of 3% explained may seem modest, we were starting from a high baseline (83% explained) and this represents 17% of the remaining unexplained variance in concentration. Moreover we have shown that this model specifically better captures the effects of urban highways, and high frequency spatial variability.

From the epidemiological and exposure assessment point of view, it is of high importance to have information about the spatial variability of the exposures in the city. In a previous study Kumar et al., 2011 employed the same DT algorithm to retrieve the 2- and 5-km AOD resolutions, which were used to predict PM2.5 concentrations. The authors reported that their model was less accurate in urban locations than the suburban ones. In contrast, our study showed high accuracy in selected urban locations (Boston and New Haven) thereby indicating that our model based on MAIAC data can be used to investigate the intra-urban exposure contrasts in PM2.5 levels. Furthermore, Kloog and colleagues 2013 employed a similar model for the Mid-Atlantic region but using the coarser DT retrieval. While overall model fit was good (cross validated R2 of 0.81) the authors reported occasional low PM2.5 concentrations around the major Mid-Atlantic Highway (the I-95) presumably because the relatively coarse 10 ×10 km2 grid cell for AOD that cannot always capture the concentration gradients near line or point sources. Our study clearly shows that the final constructed model capture the pollution levels along highways and many other urban locations. Importantly, these high concentrations are not artifacts of retrieval, thereby extending our ability to investigate the spatial patterns of urban particulate pollution, such as examining exposures in areas with high traffic. The direct implementation of our results will result in more accurate accounting for the magnitude of the association between PM2.5 and health outcomes. Finally, additional parameter that should be considered in the future modeling of PM2.5 is traffic counts from National Transportation Atlas 2011. Unfortunately, this data was unavailable for 2003.

Although the 1 km resolution is still far from optimal, it offers a clear advantage over the 10 km and even 3 km AOD data in urban studies. First, the improved resolution is expected to not only reduce the exposure error but also generally result in larger health effects estimates. For example, fine-scale variations in PM2.5 have been shown to associate with larger health effects than those that vary regionally (Jerrett et al. 2005; 2009), suggesting the potential importance of refining exposure predictions. However, it should be also noted that recently-developed statistical approaches use the land use information to get within grid spatial variation at finer than 1 km scales (Beckerman et al., 2013; Vienneau et al., 2013; Sampson et al., 2013) which is potentially a complimentary approach to what has been done in our study.

Despite promising results, more data need to be pre-processed and analyzed. First, our model was developed for AOD retrieval days/pixels and the next study should expand previously developed methodology described in Kloog et al., 2011 aimed to assess PM2.5 concentrations on non-retrieval days. Furthermore, to further investigate the strengths and limitations of high resolution AOD data for modeling PM2.5 concentrations we are planning a comprehensive multi-year study based on the full set of MODIS measurements. Next, further improve­ments to the MAIAC AOD retrieval algorithm would improve accuracy in PM2.5 estimation. For example, lack of vertical information highlights the importance of combining the satellite image with vertical profiles, like LIDARs. It should be noted also that this approach requires a large amount of daily PM2.5 stations, which are not always available in any given region. Therefore, the developed model would not be directly transferable for areas without sufficient PM2.5 monitors, such as Africa or Latin America.

**Acknowledgments:**

This work was made possible by USEPA grant RD 83479801. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication. The support for A. Lyapustin and Y. Wang is provided by the NASA Terra and Aqua Science Program.

**References**

Altaratz, O., Bar-Or, R., Wollner, U., and Koren, I. [Relative humidity and its effect on aerosol optical depth in the vicinity of convective clouds](http://iopscience.iop.org/1748-9326/8/3/034025), Environ. Res. Lett. 8 (2013), 034025 [doi:10.1088/1748-9326/8/3/034025](http://dx.doi.org/10.1088/1748-9326/8/3/034025),

Beckerman, B.S., Jerrett, M., Serre, M., Martin, R.V., Lee, S.-J., Van Donkelaar, A., Ross, Z., Su, J., Burnett, R.T. [A hybrid approach to estimating national scale spatiotemporal variability of PM2.5 in the contiguous United States](http://www.scopus.com.scopeesprx.elsevier.com/record/display.url?eid=2-s2.0-84880112020&origin=reflist&sort=plf-f&cite=2-s2.0-84880112020&src=s&imp=t&sid=AD128FF83D3D881D9A0277A224E01D38.Vdktg6RVtMfaQJ4pNTCQ%3a1080&sot=cite&sdt=a&sl=0), Environ Science and Technol, 47 (13),  (2013), pp. 7233-7241.doi: 10.1021/es400039u

[Benson, P.E. 1984](http://www.sciencedirect.com.ezp-prod1.hul.harvard.edu/science/article/pii/S1352231007006735#bib3). CALINE4-A Dispersion Model for Prediction Air Pollutant Concentrations Near Roadways. California Department of Transportation, Sacramento, CA (1984).

[Benson, P.E. 1992](http://www.sciencedirect.com.ezp-prod1.hul.harvard.edu/science/article/pii/S1352231007006735#bib4). A Review of the Development and Application of the CALINE3 and CALINE4 Models. Atmospheric Environment, 26B (3) (1992), p. 379–390.

Bergin, M. H., et al., 2000. Comparison of aerosol optical depth inferred from surface measurements with that determined by sunphotometer for cloud-free conditions at a continental U.S. site,J. Geophys. Res.,105, (2000), p.6807–6816.

Chudnovsky, A., Lee, H-J., Kostinski, A., Kotlov, T., and Koutrakis, P.: Prediction of daily fine particulate matter concentrations using aerosol optical depth retrievals from the Geostationary Operational Environmental Satellite. J. Air Waste Manage., 62(9), (2012), p.1022-1031,doi: [10.1080/10962247.2012.695321](http://sfx.hul.harvard.edu/sfx_local?__char_set=utf8&id=doi:10.1080/10962247.2012.695321&sid=libx%3Ahul.harvard&genre=article).

Chudnovsky, A., Kostinski, A., Lyapustin, A., and Koutrakis, P.: Spatial scales of pollution from variable resolution satellite imaging. Environ. Pollut., 172, (2013a), p.131-138.

Chudnovsky, A., Tang, C., Lyapustin, A., Wang, Y., Schwartz, J., and Koutrakis, P.: A critical assessment of high-resolution aerosol optical depth retrievals for fine particulate matter predictions, Atmos. Chem. Phys., 13, (2013b), p.10907-10917, doi:10.5194/acp-13-10907-2013.

Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG,Speizer FE. An association between air pollution and mortality in six UScities. N Engl J Med, 29 (1993), p. 1753–1759.

Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger S, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA 295, (2006), p.1127–1134.

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. 42 ,2008, p.7561 −7578.

Hoff RM, and Christopher SA. Remote Sensing of Particulate Pollution from Space: Have We Reached the Promised Land?J Air Waste Manage Assoc 59(6), (2009), p.645-675.

Jensen, John R., 2006, RemoteSensing of theEnvironment: AnEarthResourcePerspective(2nd edition), Prentice Hall: UpperSaddleRiver, NJ, 608 pages.

Jerrett, M.; Arain, A.; Kanaroglou, P.; Beckerman, B.; Potoglou, D.;Sahsuvaroglu, T.; Morrison, J.; Giovis, C. A review and evaluation of intraurban air pollution exposure models. J. Expo. Anal.Environ.Epidemiol, 15, (2005), p.185 −204.

Jerrett M, Burnett RT, Pope CA 3rd, Ito K, Thurston G, Krewski D, Shi Y, Calle E, Thun M.

[Long-term ozone exposure and mortality.](http://www.ncbi.nlm.nih.gov/pubmed/19279340) N Engl J Med. 2009 Mar 12;360(11):1085-95. doi: 10.1056/NEJMoa0803894.

Kloog, I., Koutrakis, P., Coull, B., Lee, H.,and Schwartz, J.: Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmos. Environ., 45(35), (2011), p.6267-6275.

Kloog I, Nordio F, Coull BA, Schwartz J. 2012. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM(2.5) exposures in theMid-Atlanticstates. Environ Sci Technol. 2012 Nov 6;46(21):11913-21. doi: 10.1021/es302673e.

Kumar, N., A. Foster, A. Chu, T. Peters, and Willis, R.: SatelliteRemoteSensingforDeveloping Time-Space Resolved Estimates of AmbientParticulate in Cleveland, OH. AerosSciTechnol., 45(9), 1090-1108, DOI: [10.1080/02786826.2011.581256](http://sfx.hul.harvard.edu/hvd?__char_set=utf8&id=doi:10.1080/02786826.2011.581256&sid=libx%3Ahul.harvard&genre=article), 2011.

Laird, Nan M.; Ware, James H. 1982.Random-EffectsModelsfor Longitudinal Data. Biometrics (International BiometricSociety) 38 (4): 963–974. [doi](http://en.wikipedia.org/wiki/Digital_object_identifier):[10.2307/2529876](http://dx.doi.org/10.2307%2F2529876). [JSTOR](http://en.wikipedia.org/wiki/JSTOR) [2529876](http://www.jstor.org/stable/2529876).

Lee, H.J., Liu, Y., Coull, B.A., Schwartz, J.,andKoutrakis, P.:A novel calibration approach of MODIS AOD data to predict PM2.5 concentrations, Atmos. Chem. Phys., 11, 7991-8002, 2011.

[Levy, R.,](mailto:Robert.C.Levy@nasa.gov) [S. Mattoo,](mailto:Shana.Mattoo-1@nasa.gov) [L.A. Munchak,](mailto:Leigh.Munchak@gmail.com) [L.A. Remer,](mailto:laremer@hotmail.com) [A.M. Sayer,](mailto:Andrew.Sayer@nasa.gov) and [N. Hsu](mailto:Christina.Hsu@nasa.gov).The Collection 6 MODIS Aerosol Products over Land and Ocean.Atmos. Meas. Tech. Disc., (Submitted) (6), 159-259, 2013.

Lindstrom, ML; Bates, DM. 1988. "Newton-Raphson and EM algorithmsfor linear mixed-effectsmodelsforrepeated-measures data". JASA 83 (404): 1014-1021.

Liu, Y.; Paciorek, C. J.; Koutrakis, P. Estimating regional spatialand temporal variability of PM2.5 concentrations using satellite data,meteorology, and land use information. Environ Health Perspect, 117, (2009), p.886−892.

Lyapustin A, Martonchik J. Wang Y, LaszloI, Korkin S. 2011a. Multi-Angle Implementation of Atmospheric Correction (MAIAC): Part 1. Radiative Transfer Basis and Look-Up Tables, J. Geophys. Res., 116, D03210, doi:10.1029/2010JD014985.

LyapustinA, Wang Y, LaszloI, KahnR, KorkinS, RemerL, LevyR, and Reid JS. 2011b. Multi-Angle Implementation of Atmospheric Correction (MAIAC): Part 2. Aerosol Algorithm, J. Geophys. Res., 116, D03211, doi:10.1029/2010JD014986.

Maune, D. F. (Ed.) (2007). Digital elevation model technologies and applications: The DEM users manual, 2nd edition. Bethesda, MD: American Society for Photogrammetric Engineering and Remote Sensing, p.550.

Monks P, Simpson D., et al. 2009. Atmospheric composition change - global and regional air quality, Atmos Environ 43, (2009), p.5268-5350.

Munchak, L. A., Levy, R. C., Mattoo, S., Remer, L. A., Holben, B. N., Schafer, J. S., Hostetler, C. A., and Ferrare, R. A.: MODIS 3 km aerosol product: applications over land in an urban/suburban region, Atmos. Meas. Tech., 6, p.1747-1759, doi:10.5194/amt-61747-2013, 2013.

Oke T.R. Boundary Layer Climates, 2nd edn., 435 p. London: Methuen (1987).

Pope A III, Burnett R,Thun M, Calle E, Krewski D, Ito K, Thurston G. 2002. Lung Cancer, Cardiopulmonary Mortality and Long-term Exposure to Fine Particulate Air Pollution.J Am Med Assoc287, p.1132-1141.

[Remer, L.,](mailto:laremer@hotmail.com) [S. Mattoo,](mailto:Shana.Mattoo-1@nasa.gov) [R.C. Levy,](mailto:Robert.C.Levy@nasa.gov) and [L. Munchak](mailto:Leigh.Munchak@gmail.com). MODIS 3km Aerosol Product: Algorithm and Global Perspective. Atmos. Meas. Tech. Disc., (In Press)(6), p.69-112.  
[doi:10.5194/amtd-6-69-2013](http://qb3vf9cp4a.search.serialssolutions.com/?SS_LibHash=qb3vf9cp4a&genre=article&sid=sersol%3AuniqueIDQuery&id=doi:10.5194/amtd-6-69-2013), 2013.

Samet JM, Zeger SL, Dominici F, Curriero F, Coursac I, Dockery DW, et al. 2000. The national morbidity, mortality, and air pollution study, part II: morbidity and mortality from air pollution in the United States. Cambridge: Health Effects Institute, June.

Schwartz, J.(1994) What are people dying of on high air pollution days? Environ. Res. 64, p.26-35.

Schwartz, J., Dockery, D.W., Neas, LM. (1996). Is daily mortality associated specifically with fine particles? J. Air Waste Manage. Assoc. 46, p.2-14.

Su GJ, Jerrett M, Beckerman B. A distance-decay variable selection strategy for land use regression modeling of ambient air pollution exposures. Science of the Total Environment. 2009a;407:3890–3898.

Tucker,C.J.,1979:Redandphotographicinfraredlinearcombinationsformonitoringvegetation. RemoteSens.Environ., 8, p.127–150.

Tucker, C.J., andP.J.Sellers,1986:Satelliteremotesensingofprimaryproduction. Int.J.RemoteSens., 7, p.1395–1416.

U.S. EPA 2004.Air Quality Criteria for Particulate Matter.EPA/600/P-99/002aF-bF. Research Triangle Park, NC:U.S. Environmental Protection Agency Office of Research and Development.

vanDonkelaar A, Martin RV, Brauer M, Kahn R, Levy R, Verduzco C, Villeneuve P. Global Estimates of Exposure to Fine Particulate Matter from Satellite-based Aerosol Optical Depth. [Environmental Health Perspectives. 2010;118(6), p.847-55.](http://ehp03.niehs.nih.gov/article/info:doi/10.1289/ehp.0901623)

[Vienneau, D.](http://www.scopus.com.scopeesprx.elsevier.com/authid/detail.url?authorId=16302849000&amp;eid=2-s2.0-84889853426) , [De Hoogh, K.](http://www.scopus.com.scopeesprx.elsevier.com/authid/detail.url?authorId=16205025600&amp;eid=2-s2.0-84889853426), [Bechle, M.J.](http://www.scopus.com.scopeesprx.elsevier.com/authid/detail.url?authorId=37561110700&amp;eid=2-s2.0-84889853426), [Beelen, R.](http://www.scopus.com.scopeesprx.elsevier.com/authid/detail.url?authorId=15829316600&amp;eid=2-s2.0-84889853426), [Van Donkelaar, A.](http://www.scopus.com.scopeesprx.elsevier.com/authid/detail.url?authorId=15926418500&amp;eid=2-s2.0-84889853426), [Martin, .V.](http://www.scopus.com.scopeesprx.elsevier.com/authid/detail.url?authorId=7501381728&amp;eid=2-s2.0-84889853426), [Millet, D.B.](http://www.scopus.com.scopeesprx.elsevier.com/authid/detail.url?authorId=7003644552&amp;eid=2-s2.0-84889853426), [Hoek, G.](http://www.scopus.com.scopeesprx.elsevier.com/authid/detail.url?authorId=7004818581&amp;eid=2-s2.0-84889853426), [Marshall, J.D.](http://www.scopus.com.scopeesprx.elsevier.com/authid/detail.url?authorId=7404860039&amp;eid=2-s2.0-84889853426)  Western European land use regression incorporating satellite- and ground-based measurements of NO2 and PM10. [Environmental Science and Technology](http://www.scopus.com.scopeesprx.elsevier.com/source/sourceInfo.url?sourceId=21537&origin=recordpage) 47 (23), (2013), pp. 13555-13564.