Harvard University
Harvard University Biostatistics Working Paper Series

Year 2007 *Paper* 74

Spatio-temporal Associations Between GOES Aerosol Optical Depth Retrievals and Ground-Level PM2.5

Christopher J. Paciorek* Yang Liu†
Hortensia Moreno-Macias‡ Shobha Kondragunta**

*Harvard School of Public Health, paciorek@hsph.harvard.edu
†Harvard School of Public Health, yangliu@hsph.harvard.edu
‡Harvard School of Public Health
**National Oceanic and Atmospheric Administration

Spatio-temporal Associations Between GOES Aerosol Optical Depth Retrievals and Ground-Level PM2.5

Christopher J. Paciorek, Yang Liu, Hortensia Moreno-Macias, and Shobha Kondragunta

Abstract

We assess the strength of association between aerosol optical depth (AOD) retrievals from the GOES Aerosol/Smoke Product (GASP) and ground-level fine particulate matter (PM2.5) to assess AOD as a proxy for PM2.5 in the United States. GASP AOD is retrieved from a geostationary platform and therefore provides dense temporal coverage with half-hourly observations every day, in contrast to once per day snapshots from polar-orbiting satellites. However, GASP AOD is based on a less-sophisticated instrument and retrieval algorithm. We find that correlations between GASP AOD and PM2.5 over time at fixed locations are reasonably high, except in the winter and in the western U.S. Correlations over space at fixed times are lower. Simple averaging over time actually reduces correlations over space dramatically, but statistical calibration allows averaging over time that produces strong correlations. These results and the data density of GASP AOD highlight its potential to help improve exposure estimates for epidemiological analyses. On average 40% of days in a month have a GASP AOD retrieval compared to 14% for MODIS and 4% for MISR. Furthermore, GASP AOD has been retrieved since November 1994, providing the possibility of a long-term record that pre-dates the availability of most PM2.5 monitoring data and other satellite instruments.

Spatio-temporal associations between GOES aerosol optical depth retrievals and ground-level $PM_{2.5}$

December 18, 2007

Christopher J. Paciorek

Department of Biostatistics, Harvard School of Public Health

Yang Liu

Department of Environmental Health, Harvard School of Public Health

Hortensia Moreno-Macias

Department of Biostatistics, Harvard School of Public Health

Shobha Kondragunta

National Oceanic and Atmospheric Administration

Abstract

We assess the strength of association between aerosol optical depth (AOD) retrievals from the GOES Aerosol/Smoke Product (GASP) and ground-level fine particulate matter ($PM_{2.5}$) to assess AOD as a proxy for $PM_{2.5}$ in the United States. GASP AOD is retrieved from a geostationary platform and therefore provides dense temporal coverage with half-hourly observations every day, in contrast to once per day snapshots from polar-orbiting satellites. However, GASP AOD is based on a less-sophisticated instrument and retrieval algorithm. We find that correlations between GASP AOD and $PM_{2.5}$ over time at fixed locations are reasonably high, except in the winter and in the western U.S. Correlations over space at fixed times are lower. Simple averaging over time actually reduces correlations over space dramatically, but statistical calibration allows averaging over time that produces strong correlations. These results and the data density of GASP AOD highlight its potential to help improve exposure estimates for epidemiological analyses. On average 40% of days in a month have a GASP AOD retrieval compared to 14% for MODIS and 4% for MISR. Furthermore, GASP AOD has been retrieved since November 1994, providing the possibility of a long-term record that pre-dates the availability of most $PM_{2.5}$ monitoring data and other satellite instruments.

1 Introduction

Epidemiological studies provide strong evidence that chronic exposure to particulate matter (PM) is related to increased mortality, as well as outcomes such as ischemic heart disease, dysrhythmias, heart failure, cardiac arrest, and lung cancer (*1–5*). Studies of the chronic health effects of PM rely on spatial heterogeneity in PM concentrations to estimate the effects, but most studies have characterized concentrations based on city- or county-wide averages of ambient measurements, relying on spatial heterogeneity at relatively large scales to estimate health effects. A combination of spatial modeling and land use regression in an additive modeling statistical framework can help to estimate concentrations at a finer scale (*6*). While this approach borrows strength from covariate information to help estimate concentrations at locations without monitors, it still suffers from the sparse spatial representation in the monitoring network. Evidence of health effects of acute exposure to PM, e.g., (*7*, *8*), relies on temporal heterogeneity in PM, but the fact that many monitors operate only once every three or six days reduces statistical power in these time series studies.

Remote sensing holds promise for adding information, particularly spatial information in suburban and rural areas far from monitors, and temporal information on days without monitoring. Recent work suggests that satellite-derived aerosol optical depth (AOD) is correlated with ground level PM (*9–14*), specifically particles with diameter ranging from 0.05-2 microns (*15*), which is roughly the definition of $PM_{2.5}$ (particles with aerodynamic diameter less than 2.5 microns), the size fraction on which current EPA regulatory efforts focus. These correlations occur despite the mismatch in vertical detail between total column aerosol, as measured by AOD, and ground-level $PM_{2.5}$, the level of interest for health studies. In general, if the atmosphere is well-mixed, total column aerosol is expected to be a good proxy for ground-level $PM_{2.5}$. One approach to help account for the mismatch is calibration via a regression model based on season, spatial location, and meteorological infomation reflected in planetary boundary layer (PBL) and relative humidity (RH) (*12*).

Efforts to use AOD as a proxy for $PM_{2.5}$ have to this point concentrated on the Multiangle Imaging Spectroradiometer (MISR) and Moderate Resolution Imaging Spectroradiometer (MODIS) instruments. These are on polar-orbiting satellites, which causes relatively sparse coverage over time at any given location, with individual locations in the eastern United States sampled via a single snapshot every 4-7 days by MISR and every 1-2 days by MODIS. Intensifying the problem, retrievals are often missing because of cloud cover. The result is sparse coverage in space and time that can impede use in health studies. Geostationary satellites provide much more complete data; the GOES Aerosol/Smoke Product (GASP) AOD provides observations every 30 minutes on a nominal 4 km grid. However, the GASP AOD retrievals are less precise than those from the polar-orbiting instruments because the GOES instrument is a broadband sensor with a single angle of view (*16*). To determine the GASP AOD retrieval, surface reflectivity is calculated based on generating a composite background image using images taken from the past 28 days at the same time. The composite background may be contaminated by possible aerosol extinction, residual cloud contamination, cloud shadows, and temporal surface variations. Since AOD retrievals use only the visible channel (520 - 720 nm) signals, all atmospheric and aerosol properties (e.g., size distribution, composition, and scattering phase function) must be assumed and only AOD is allowed to vary in the radiative transfer model. Overall, GOES AOD retrieval uncertainty is ± 18 - 34%, higher than MODIS, or particularly MISR (*17*). Despite these limitations, GASP retrievals are reasonably well-correlated with AERONET ground measurements of total column aerosol and MODIS AOD retrievals in the northeastern/mid-Atlantic United States and eastern Canada (*16*). To date, no studies have been done to understand the relationship between GASP AOD and ground-level $PM_{2.5}$.

Here we assess the potential of GASP AOD to act as a proxy for ground-level $PM_{2.5}$ at the daily, monthly, and yearly time scales. First, we assess the basic strength of association between GASP AOD and $PM_{2.5}$ in space and time. We build flexible regression-style models, relating AOD to $PM_{2.5}$ in a way that allows us to calibrate daily AOD based on meteorological, spatial, and temporal effects. We compare the predictive ability of calibrated AOD for daily, monthly, and yearly average $PM_{2.5}$. Finally, we assess whether the presence or absence of an AOD retrieval is associated with the $PM_{2.5}$ level, allowing us to determine if bias is induced by ignoring the pattern of missingness and simply using the available retrievals. Our goal is to understand the association of GASP AOD with $PM_{2.5}$ and show how to calibrate GASP AOD to increase its utility, not to physically interpret our statistical modeling of AOD.

2 Data

We make use of GASP AOD from GOES-12 (East) imager data, provided by the U.S. National Oceanic and Atmospheric Administration (NOAA), using all the retrievals from 2004. (*16*) describe the GOES-12 imager data and GASP AOD algorithm in detail; in brief, AOD is calculated from a single visible channel (520-720 nm) based on a set of assumptions about surface reflectivity and atmospheric and aerosol properties, while the cloud mask is determined from infrared channels 2 (3.9 um) and 4 (10.7 um) and the visible channel.

GASP AOD retrievals are available during daylight, from the time period 10:45-23:45 UTC. However, our analyses make use of the data from times and locations with a solar zenith angle less than 70 degrees, as retrievals are generally less accurate at high zenith angle because of limitations of the radiative transfer model, which ignores the earth's curvature. Our sensitivity analysis considers using data from higher angles, finding that these additional data may be useful (see supplementary material). The pixel centroids of the GASP AOD retrievals are nominally on a 4 km grid, but the distance between centroids is not generally 4 km. Retrievals are attempted every half-hour, but cloud cover and high surface reflectivity lead to many missing observations.

In our core analysis, we follow NOAA's criteria for screening valid AOD observations, described in the supplementary material. Negative retrievals occur due to errors in the estimation of surface reflectivity when AOD is low. Unlike (*16*), we make use of negative retrievals in the hope that they indicate low AOD. In the supplementary material, we assess this choice in a sensitivity analysis, showing that including these retrievals provides useful information.

Defining potential retrievals as those occurring at times with solar zenith angle less than 70 degrees, Fig. 1 shows the spatial pattern of available retrievals for the eastern U.S. (we show later that correlations of

AOD and $PM_{2.5}$ are low in the western U.S.). There are few retrievals satisfying the criteria in the northern US during fall and winter, due to high levels of cloudiness and surface reflectivity. During summer and spring, the spatial differences in availability occur at small spatial scales.

To assess the relationship between ground-level $PM_{2.5}$ and AOD, we matched monitoring data from the US EPA Air Quality System (AQS) to the nearest GASP AOD pixel, omitting a small number of monitors for which the nearest AOD pixel centroid is closer to another monitor. Since we use AOD as the dependent variable in our regression modeling, this avoids having duplicate AOD values that would be induced by using all the $PM_{2.5}$ monitors. We then selected days for which the EPA monitor reported a $PM_{2.5}$ concentration. Our interest is in fine resolution estimation of $PM_{2.5}$, so unlike other analyses that aggregate AOD across adjoining pixels, we consider only individual pixels.

We use data with EPA parameter 88101, the primary identifier for $PM_{2.5}$ data. This excludes most $PM_{2.5}$ data from IMPROVE sites, which are generally in very rural areas, avoiding issues of comparability between AQS and IMPROVE observations and focusing our calibration efforts on populated areas where people live. We included all observations regardless of any quality flags in the data record, at the suggestion of EPA personnel who indicate that all data reported to AQS should be valid data. For simplicity, we used only data with parameter occurrence code (POC) equal to one, thereby including only the primary monitor at a site. The AQS data and measurements collected by similar methods are the primary source of data for estimating exposure in epidemiological studies, so we consider them as the gold standard here, while acknowledging that the ground measurements are not error-free.

For meteorological information, we follow (*12*) and concentrate on planetary boundary layer (PBL, i.e., mixing height) and relative humidity (RH) as key meteorological variables that affect the relationship between $PM_{2.5}$ and AOD. PBL is used to represent the vertical distribution of $PM_{2.5}$; most particle mass loading resides in the lower troposphere, and the PBL gives an indication of how much of the column is more actively mixed and relatively homogeneous. Higher PBL is expected to be associated with a larger ratio of AOD to $PM_{2.5}$ because aerosol emitted from the surface is distributed over a larger volume of air. The size of hygroscopic particles such as sulfates and organic carbonaceous species grows with increasing relative humidity, resulting in greatly increased light extinction efficiency. Since $PM_{2.5}$ is measured as dry particle mass (measured at controlled RH of approximately 40%) we expect higher RH to be associated with a larger ratio of AOD to $PM_{2.5}$. We use the North American Regional Reanalysis (NARR) meteorological fields; the NARR assimilates available data with a state-of-the-art meteorological model to estimate meteorological

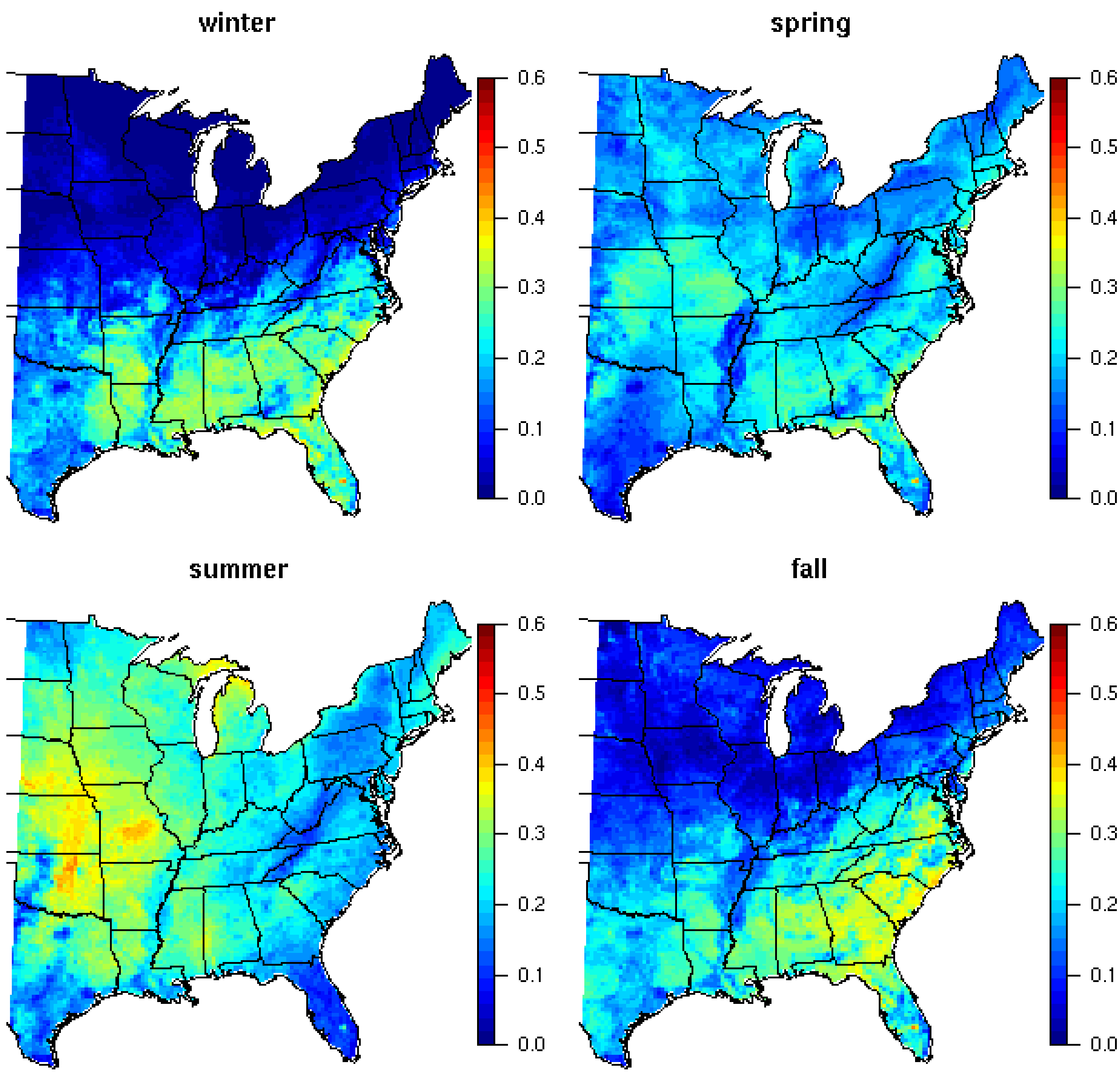

Figure 1: Proportion of potential retrievals (defined as those with solar zenith angle less than 70 degrees) that satisfy the GASP AOD screening criteria, by season.

parameters every 3 hours on a 32 km grid covering North America. (*18*) reports that NARR PBL is highly correlated with LIDAR measurements, although in urban areas the correlation decreases, possibly because of small-scale heat island effects. We use data from 12:00, 15:00, 18:00, and 21:00 UTC to match the time range of the GASP AOD retrievals, and we take the inverse-squared distance weighted average of the values from the four closest grid points to each EPA monitor.

To characterize the human environment at the locations of the matched monitor-pixel pairs we estimate road and population density nearby. In the larger project of which this is a part, we have divided the eastern U.S. into 4 km square grid cells and estimated the population density in each cell from the 2000 U.S. Census, as well as the density of roads in each cell based on the ESRI StreetMap 9.2 product. Using the cell whose centroid was closest to the AOD pixel centroid, we assigned road and population density estimates to each matched pair. We also have information from EPA on the local land use and monitoring objective for many of the monitors.

In principle, since GASP AOD is at the half-hourly resolution, one might want to calibrate to hourly $PM_{2.5}$ measurements, also available through AQS. However, the number of hourly monitors is much smaller than daily monitors, limiting our ability to calibrate AOD to hourly $PM_{2.5}$, and there is no Federal Reference Method (FRM) for hourly PM2.5. The relationship of hourly $PM_{2.5}$ (averaged to the day) to daily FRM measurements can vary by location and season of year, as is evident in the AQS data, suggesting the need to calibrate hourly $PM_{2.5}$ to daily $PM_{2.5}$. This occurs in part because of the loss of semi-volatile compounds in continuous $PM_{2.5}$ measurements due to pre-heating the air sample. For simplicity and because our interest is in relationships between $PM_{2.5}$ and AOD at time scales longer than hourly (note that the EPA air quality standard is a 24-hour average), we restrict our analysis to daily associations after averaging the GASP AOD retrievals available for each day for a given pixel, using only $PM_{2.5}$ data from daily FRM monitors.

3 Analyses

3.1 Spatio-temporal associations between daily AOD and $PM_{2.5}$

We first investigate the associations between daily AOD and $PM_{2.5}$ in space and time to understand the relative strength of the association. For simplicity, we calculate a daily estimate of AOD as the simple average of the available retrievals.

In Figure 2 we see correlations calculated over time for each pixel-monitor match. The correlations are much higher in the eastern U.S. than the western U.S., as has been found for both MODIS (*11*) and MISR (*12*), and as expected based on higher surface reflectivity in the less-vegetated western part of the United States. Also, a higher proportion of the aerosol in the western U.S. is in the free troposphere, rather than the boundary layer, with less local anthropogenic pollution than the eastern U.S., with the exception of California (*19*). Correlations vary by season, with lower correlations during winter (and few retrievals, particularly outside of the southeast), while summer, spring, and fall (not shown) have similar correlations to the all season results. There were no clear and substantial relationships between the correlations and local information about the site, such as land use, population density, monitoring objective or local emissions. The results are robust with respect to various thresholds in terms of the number of AOD retrievals required to calculate the daily average AOD and the number of days required to calculate the correlation at a site, although the correlations are lower when fewer retrievals are required in a day. Note that we ignore temporal correlation in the AOD and $PM_{2.5}$ measurements themselves in calculating these correlations, treating the days as independent, as is done in other published analyses.

(*11*) report correlations between MODIS AOD retrievals and $PM_{2.5}$ at nearby sites (less than 40-50 km) for selected major cities. Here we compare correlations between GASP AOD retrievals and $PM_{2.5}$ at the same sites with their results. Figure 3 suggests that when using days with any number of GASP AOD retrievals, there is a lower correlation than found by (*11*), but that with more retrievals, the GASP AOD correlations become as strong as the MODIS correlations. This highlights the strength of the GASP AOD retrievals, their high temporal coverage, while indicating that individual retrievals are not as highly correlated with $PM_{2.5}$ as MODIS AOD, even though GASP AOD pixels are much closer (within about 6 km) to the monitoring sites than in (*11*).

Because of the low correlations in the western United States, we henceforth restrict our analyses to locations east of 100°W. Note that most of the counties violating the EPA air quality standard for $PM_{2.5}$ are east of 100°W, with the exception of California counties.

Next we consider associations across space, with Fig. 4 showing correlations for individual days. These correlations over space are less strong than those over time, suggesting that GASP AOD can better distinguish high from low $PM_{2.5}$ over time at fixed locations than over space at fixed times. This may be related to spatially-varying factors such as average reflectivity, aerosol type, or climate over time that may obscure, and potentially confound, the relationship between AOD and $PM_{2.5}$. Our calibration work (Section 3.2) suggests that the relationship between AOD and $PM_{2.5}$ varies by location, which helps to explain the low

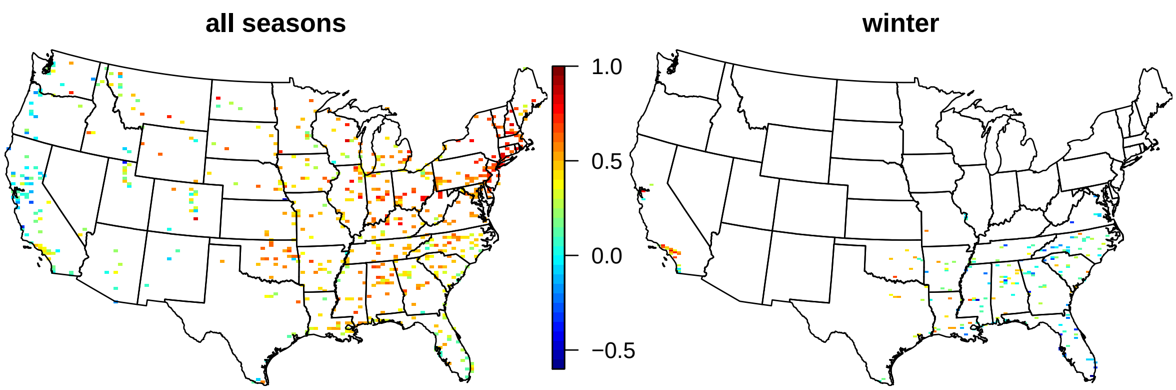

Figure 2: Correlations at individual sites between daily average $PM_{2.5}$ and the average of half-hourly AOD retrievals for all seasons (left) and winter only (right). Plots are based on site-days with at least three AOD retrievals, and only locations with at least 10 days of matched pairs are shown.

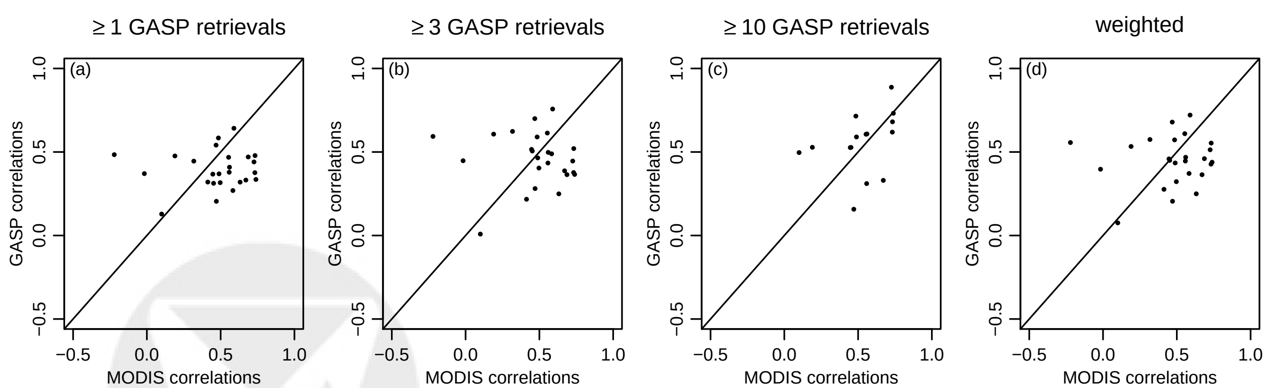

Figure 3: Scatterplots of correlations (calculated at individual sites over time, with each point representing a site) of GASP AOD with $PM_{2.5}$ against correlations of MODIS AOD with $PM_{2.5}$ for sites in Table 2 of (*11*): (a) GASP AOD correlations calculated using all days with at least one retrieval (b) GASP AOD correlations calculated using all days with at least three retrievals (c) GASP AOD correlations calculated using all days with at least ten retrievals (d) GASP AOD correlations weighted by number of retrievals in each day. Some sites are excluded because there were not at least five matched pairs (following (*11*)) for GASP AOD or because the site selected in (*11*) was not the nearest monitoring site to that GASP AOD pixel.

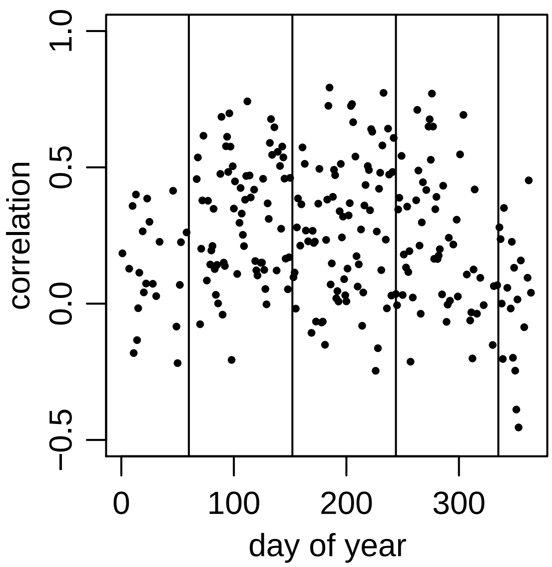

Figure 4: Correlations of GASP AOD with $PM_{2.5}$ by day of year, requiring at least three retrievals per site-day and including only days with at least 30 such sites. Vertical lines divide the seasons.

cross-sectional correlations seen here; after calibration based on location, the associations improve. The correlations tend to be lower in winter, as found for the correlations over time. The results are again robust with respect to various thresholds. The lower associations in winter do not appear to be primarily caused by any marked differences in the variability of either $PM_{2.5}$ or AOD between winter and other seasons. Another possibility is that reflectance changes with vegetation loss in winter, and this may not be accurately captured by GOES imager.

3.2 Statistical calibration of daily AOD and $PM_{2.5}$

3.2.1 Basic model

Our goal here is to understand the factors that modify the relationship between GASP AOD (henceforth referred to as AOD) and $PM_{2.5}$ and to build a regression model to provide a calibrated AOD variable that is more strongly associated with $PM_{2.5}$. Ultimately, as part of the larger project of which this work is a part, the calibrated AOD will be used in a statistical prediction model for $PM_{2.5}$.

To address potential overfitting in our regression models, we divided the data into 10 random sets, each set containing all the observations over time from approximately one-tenth of the locations. We left the tenth set in reserve for final testing and used the other nine sets in a cross-validation approach. That is, for each regression model under consideration, we sequentially left out one of the nine sets, fit the model to the remaining eight sets, and calculated calibrated AOD for the observations in the held-out set. Aggregating over the nine sets, this gives us cross-validated values of calibrated AOD for the nine sets that we can compare to the held-out $PM_{2.5}$ observations to assess how well the calibrated AOD correlates with $PM_{2.5}$. We have 99,159 matched daily observations, of which 46,684 have at least one valid AOD retrieval during the day. Of the matched observations with valid AOD retrievals, 6,558 are in winter, with 13,361, 15,454 and 11,311 in spring (March-May), summer (June-August), and fall (September-November), respectively.

The calibration model (*12*) treats $PM_{2.5}$ as the dependent (response) variable in regression models, using the log transform of both AOD and $PM_{2.5}$ to create an additive model on the log scale. Here we consider log AOD as the dependent variable and regress on $PM_{2.5}$ and other factors, treating AOD as observed data. We believe this makes sense because of the high variability in AOD, reflecting its noisiness as a proxy for $PM_{2.5}$, and the varying number of retrievals contributing to average daily AOD. These are difficult to account for if AOD is considered to be the independent variable. In our models the observed $PM_{2.5}$ values at the monitors stand in for true $PM_{2.5}$, ignoring any monitor instrument error. Once we model AOD as a function of $PM_{2.5}$, we can use the fitted model to calibrate AOD as described below.

The basic model structure we employ builds on (*12*) but uses smooth regression functions (*20*) in place of linear functions and indicator variables for region and season. The model is

$$\log \bar{a}_{it} \sim \mathcal{N}(\mu + g(s_i) + f_t(t) + f_{\mathrm{PBL}}(\mathrm{PBL}_{it}) + f_{\mathrm{RH}}(\mathrm{RH}_{it}) + \beta \mathrm{PM}_{it}, \tau^2). \tag{1}$$

Here μ is an overall mean (simple additive bias). $g(s_i)$, $f_t(t)$, $f_{\mathrm{PBL}}(\mathrm{PBL}_{it})$, and $f_{\mathrm{RH}}(\mathrm{RH}_{it})$ are smoothly-varying regression functions that account for additive bias due to spatial location, s_i (represented in the Albers equal-area projection), time (day of year), PBL, and RH, respectively. β is a multiplicative bias coefficient that scales from units of $PM_{2.5}$ to unitless AOD, and PM_{it} is the matched $PM_{2.5}$ measurement. We use the simple average of the available AOD retrievals in each day, $\bar{a}_{it}$, but below we consider a more sophisticated approach. We considered using both a log transformation of the average AOD values and staying on the original scale; the two approaches performed very similarly. Because the log transformation gives residuals that are slightly less skewed, we used the log transformation in our models. Since there are negative retrievals for GASP, we added 0.6 to each observation (the minimum value is -0.5) and then log transformed. Fitting separate models of the form (1) for each season, we estimated β to be 0.0018

(95% confidence interval of (0.0010, 0.0027)) in winter, an order of magnitude smaller than 0.0164 (0.0157, 0.0170) in spring, 0.0164 (0.0158, 0.0169) in summer, and 0.0129 (0.0123, 0.0134) in fall, showing that the coefficient for winter is close to zero compared to the other seasons and that the other seasons are fairly comparable. As a result we chose to fit models only for spring, summer, and fall. We fit separate models (1) for each season, to facilitate computations with such a large dataset and to allow the relationships to vary by season.

The model (1) can be fit in the statistical software, R, using the gam() function, designed for fitting generalized additive models (*20*). The software uses penalized splines to parameterize the smooth functions of time, space, and covariates, with penalty terms to help avoid overfitting, thereby ensuring that the functions are sufficiently smooth to allow for generalizability while allowing estimation of non-linear relationships. Having fit the model and estimated the smooth functions, we can create a calibrated AOD variable, a^*_{it}, by subtracting off the values of all the fitted functions from the observed value, $\log \bar{a}_{it}$, except for the value of $\mathrm{PM}_{2.5}$:

$$a^*_{it} = \log \bar{a}_{it} - \hat{\mu} - \hat{g}(s_i) - \hat{f}_t(t) - \hat{f}_{\mathrm{PBL}}(\mathrm{PBL}_{it}) - \hat{f}_{\mathrm{RH}}(\mathrm{RH}_{it}). \qquad (2)$$

Our hope is that by adjusting for factors that modify the relationship of AOD and $\mathrm{PM}_{2.5}$, the calibrated AOD, which we note is on a different scale than raw AOD, is more strongly associated with $\mathrm{PM}_{2.5}$ than raw AOD and has a reasonably linear relationship. If linearity holds, it will allow averaging to longer time scales, to produce more robust proxy estimates of $\mathrm{PM}_{2.5}$ that average over short-term fluctuations. For example linearity allows us to calculate a monthly average proxy,

$$\frac{1}{T}\sum_{t=1}^{T} a^*_{it} \approx \beta_0 + \beta_1 \frac{1}{T}\sum_{t=1}^{T} \mathrm{PM}_{it} \qquad (3)$$

where β_0 and β_1 would be estimated within the statistical prediction model used in the larger project.

To investigate whether linearity in the relationship of $\mathrm{PM}_{2.5}$ to AOD was a reasonable assumption and to consider whether using $\mathrm{PM}_{2.5}$ or $\log\mathrm{PM}_{2.5}$ in (1) is preferable, we compared models of the form (1) but using a smooth regression function of pollution, either $f_{\mathrm{PM}}(\mathrm{PM}_{it})$ or $f_{\log \mathrm{PM}}(\log \mathrm{PM}_{it})$. We found a reasonably linear relationship of $\log \bar{a}_{it}$ with $\mathrm{PM}_{2.5}$ on the original scale while the association of $\log \bar{a}_{it}$ with $\log\mathrm{PM}_{2.5}$ was not linear, which would complicate the construction of the calibration model (2,3). Further justifying the linearity of $\mathrm{PM}_{2.5}$ in (1), the model using $f_{\mathrm{PM}}(\mathrm{PM}_{it})$ explained only slightly more of the variability in

$\log \bar{a}_{it}$ than when using the linear term.

Results In Section 3.2.2, we compare a number of alternative model specifications. Here we focus on a set of key results with respect to the importance of calibration and the relationships between AOD and $PM_{2.5}$ at different temporal resolutions. For a given proxy, either raw AOD or calibrated AOD from a particular regression model, we can calculate correlations of the proxy with the matched $PM_{2.5}$ values. We calculated correlations at the daily scale as well as after averaging across available matched pairs within a month and within a year at each site. Note that correlations are calculated only based on days for which both the AOD proxy and $PM_{2.5}$ were available, so we overstate the predictive ability of AOD for true monthly and yearly average $PM_{2.5}$, since there will be days with no AOD retrievals. Our correlation results measure the ability of the calibrated AOD values to mirror heterogeneity in $PM_{2.5}$ over space and time.

First we report that on the daily scale calibrated AOD (2) from the final model provides stronger correlations with $PM_{2.5}$ than using the raw daily log average, $\log \bar{a}_{it}$ (Table 1). More importantly, without calibration, we cannot average over time and achieve more robust relationships; we discuss this surprising result below. Requiring a threshold of AOD retrievals in a day (e.g., five in Table 1) improves associations over the shorter time periods. However, over the yearly period, by reducing the number of days with matched pairs, the resulting year-long averages are less robust and correlations decrease compared to using all days with at least one retrieval. Based on this we suggest that analyses that average to monthly or yearly resolution include all available AOD retrievals. Fig. 5 shows the associations between $PM_{2.5}$ and either raw AOD or calibrated AOD for the different averaging periods, graphically illustrating the results in Table 1. Note that our final model is of the form (1), but with a restriction on the flexibility of the function of time, forcing the dimension of the basis functions to be less than five; the resulting estimated degrees of freedom are between 3 and 4 for all seasons. This restriction was chosen based on model comparisons in Section 3.2.2.

The reduction in correlations between AOD and $PM_{2.5}$ when averaging over time (Table 1, column (a)) mirrors the fact that correlations over time, holding space fixed, tend to be stronger than correlations over space, holding time fixed (Section 3.1). Somehow the within-site relationships between AOD and $PM_{2.5}$ are positive, but across sites and most noticeably at the yearly resolution, AOD is only weakly associated with $PM_{2.5}$. The most likely explanation for this is that there are spatially-varying confounders that tend to obscure the long-term average relationship between $PM_{2.5}$ and AOD, driving long-term average AOD

Table 1: Correlations between various AOD proxies and $PM_{2.5}$ at different temporal resolutions, excluding winter. The three AOD proxies are: (a) raw AOD, calculated using the log average daily AOD; (b) calibrated AOD (2) based on $\bar{a}_{it}$; (c) calibrated AOD (4) based on $\hat{a}_{it}$ from a time series model (8); and (d) calibrated AOD (5) based on $\bar{a}_{it}$ from the simplified model without time, PBL and RH . Correlations are shown both when using matched pairs for days with any number of AOD retrievals and restricting to days with at least five retrievals.

temporal resolution of correlations	(a) Raw AOD ($\log \bar{a}_{it}$)	(b) Calibrated AOD (a^*_{it}) using $\log \bar{a}_{it}$	(c) Calibrated AOD (a^*_{it}) using $\log \hat{a}_{it}$	(d) Calibrated AOD (a^*_{it}) based on (5)
	any number of AOD retrievals in a day			
daily	0.41	0.50	0.51	0.50
monthly averages (at least 3 matched days for each site-month)	0.34	0.62	0.63	0.63
yearly averages (at least 10 matched days for each site)	0.17	0.75	0.76	0.74
	at least five AOD retrievals each day			
daily	0.51	0.59	0.60	0.60
monthly averages (at least 3 matched days for each site-month)	0.41	0.67	0.69	0.67
yearly averages (at least 10 matched days for each site)	0.19	0.69	0.71	0.67

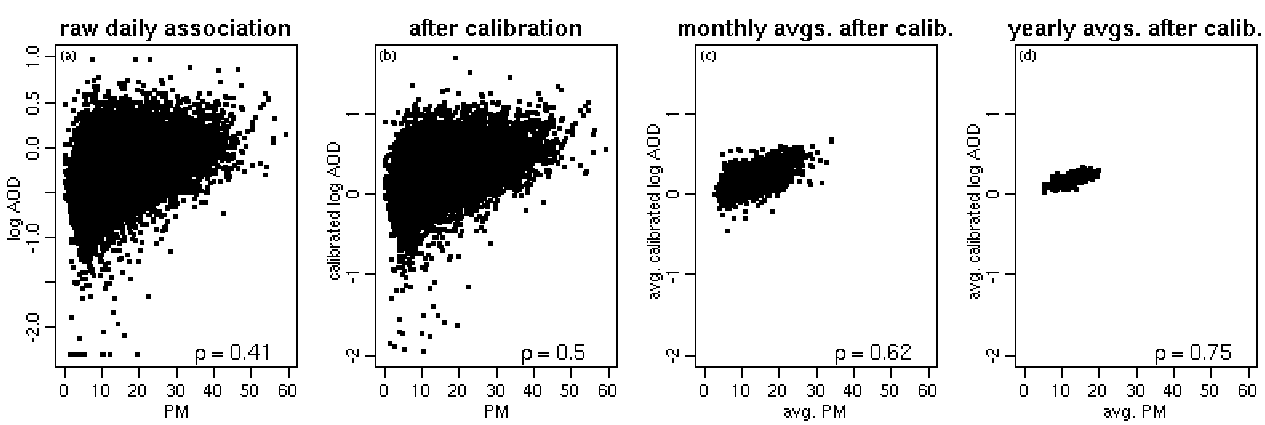

Figure 5: Scatterplots of various AOD proxies (any number of retrievals in a day) against $PM_{2.5}$, excluding winter: (a) raw daily AOD (b) calibrated daily AOD, (c) monthly average calibrated AOD for site-months with at least three matched pairs, and (d) yearly average calibrated AOD for sites with at least 10 matched pairs. In (a) and (b), one outlying point with $PM_{2.5} = 77.3$ is omitted.

down where long-term $PM_{2.5}$ is high and vice versa. What these confounders might be is unclear, but their effect as seen in Table 1 is marked. The calibration, which is primarily driven by the spatial term (see Section 3.2.2), is able to account for the confounding. However, we caution that the correlation of calibrated AOD and $PM_{2.5}$ at the yearly level appears to be primarily driven by large-scale spatial patterns in both variables. The implication is that calibrated AOD may not help to improve long-term predictions relative to a $PM_{2.5}$ prediction model without AOD that relies on large-scale spatial smoothing of the monitoring data plus information from GIS and meteorological covariates.

Finally, we considered whether the model comparison process based on cross-validation could itself result in overfitting and give us overly optimistic estimates of the association between calibrated AOD and $PM_{2.5}$. We calculated correlations between $PM_{2.5}$ values in the held-out tenth set and matched calibrated AOD. The correlations are not substantially lower (between 0 and 0.05 less) than the results for cross-validation on the nine sets as reported in Table 1, indicating that the correlations are not inflated by overfitting.

Smooth regression functions Figs. 6 and 7 show the fitted smooth regression functions of time, RH, PBL, and space for each of the three seasons. We interpret these functional relationships conditional on $PM_{2.5}$ being in the model. For a given concentration of $PM_{2.5}$, as expected, AOD increases with increasing PBL, since a higher PBL means the AOD retrieval is integrating over a longer column of air in which the concentration of $PM_{2.5}$ is likely reasonably constant. For a given concentration of $PM_{2.5}$, AOD increases with increasing RH because of the particle growth effect of humidity, which increases AOD relative to ground-level $PM_{2.5}$, as the latter is measured as dry mass. In general when RH is greater than 60-70%, we see a upward trend in all three seasons, suggesting that same particle dry mass has increasing light extinction capabilities with RH. This may be due to the fact that the growth effect of hygroscopic particles such as sulfate and certain organic carbon species become more substantial with increasing RH (*21*). Note that the wiggliness in the regression functions for PBL and RH likely reflects overfitting from not fully accounting for within-site correlation in (1). This could be done using random effects for each site; we do not pursue a more sophisticated approach because, as we describe in Section 3.2.2, RH and PBL are relatively unimportant compared to the spatial function in the model. The spatial patterns indicate that, holding $PM_{2.5}$ concentration constant, we see low values of AOD over the Ohio River valley and Appalachian Mountain region. In other words, in these regions, AOD is not as high as one would expect based on the concentrations

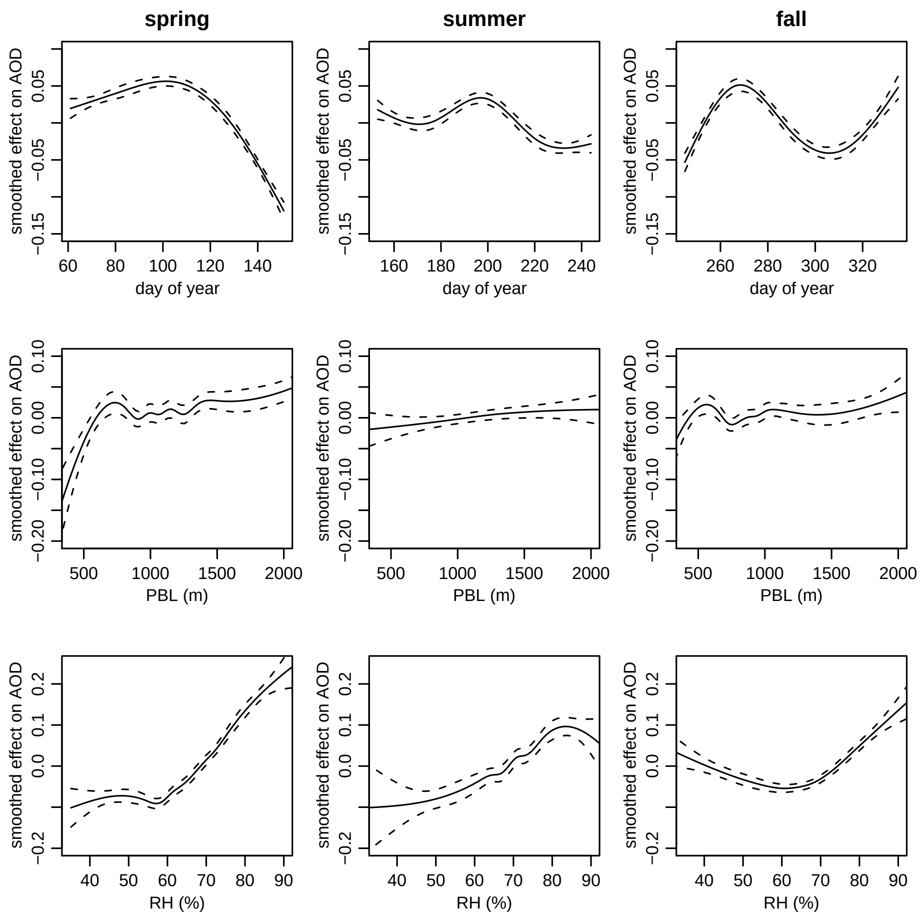

Figure 6: Fitted smooth regression relationships for time, $\hat{f}_t(t)$ (top row), PBL, $\hat{f}_{\mathrm{PBL}}(\mathrm{PBL})$(middle row) and RH, $\hat{f}_{\mathrm{RH}}(\mathrm{RH})$ (bottom row) by season with larger values indicating that AOD is high relative to $\mathrm{PM}_{2.5}$ for that value of the regression variable.

of $\mathrm{PM}_{2.5}$ in those areas. This may occur because large local emissions from power plants in the region increase the ratio of ground-level $\mathrm{PM}_{2.5}$ to AOD. Spatial patterns may also be caused by variability in aerosol type and variability in meteorology, to the extent that is not captured by the RH and PBL measures, as well as differences in the satellite viewing angle.

3.2.2 Alternative models

In the previous section, we used the simple arithmetic average, $\bar{a}_{it}$, with homoscedastic (i.e., constant variance) error in (1). As an alternative, we first consider a model that uses a more sophisticated time series-based estimator of daily AOD, $\hat{a}_{it}$ (8):

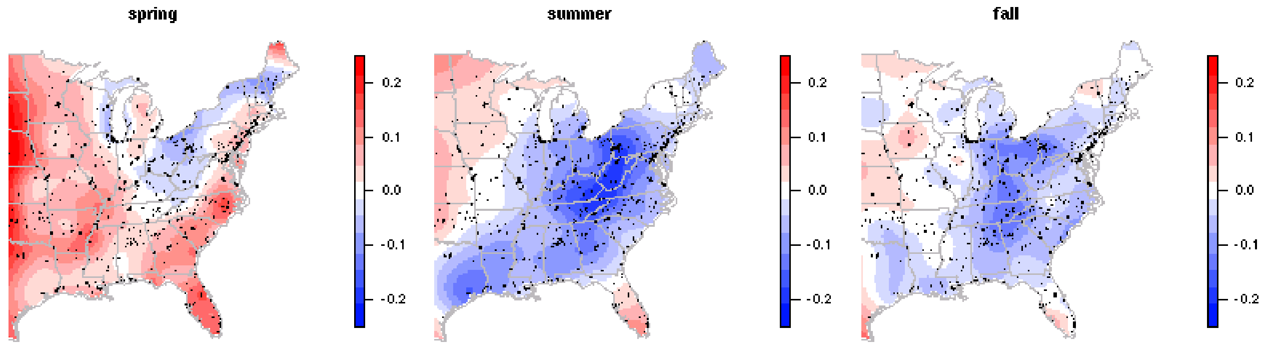

Figure 7: Fitted smooth spatial surfaces, $\hat{g}(s)$, by season. Blue indicates that AOD is low relative to $PM_{2.5}$ and red the converse. Points are the matched AOD-$PM_{2.5}$ sites.

$$\log \hat{a}_{it} \sim \mathcal{N}(\mu + g(s_i) + f_t(t) + f_{\text{PBL}}(\text{PBL}_{it}) + f_{\text{RH}}(\text{RH}_{it}) + \beta \text{PM}_{it}, \sigma^2 V(\hat{a}_{it}) + \tau^2). \tag{4}$$

This approach (whose derivation and accompanying uncertainty estimates are described in the supplementary material), accounts for the pattern of missing retrievals using weighting derived from the autocorrelation structure, downweighting retrievals that are close in time to other retrievals and upweighting retrievals that are isolated from other retrievals. The heteroscedastic variance accounts for the varying levels of certainty in the daily AOD estimates caused by having different numbers of AOD retrievals in a day (and by the time pattern of available retrievals). $V(\hat{a}_{it})$ is derived in (9). τ^2 accounts for the inherent noise in the relationship between AOD and $PM_{2.5}$ that would be present even without any missing retrievals. The term σ^2 is the proportionality constant that is missing from (9) and, with τ^2, is estimated in the model fitting. Table 1 (column (c)) includes a tabulation of the correlations from the time series approach for comparison with calibration based on the simple arithmetic average. The correlations improve only marginally, and since fitting (4) is much more computationally intensive, we proceed with the simpler homoscedastic model (1) that uses the average of the available AOD retrievals in the calibration and ignores the varying uncertainty.

Model (1) fits $f_t(t)$ as a smooth function of time, with about four effective degrees of freedom for each season. We also fit a model allowing a much less smooth function of time, which can account for short-term changes in the relationship between AOD and $PM_{2.5}$. This model overfits, with lower correlations between calibrated AOD and $PM_{2.5}$ (about 0.04 lower than those shown in Table 1). We also considered removing $f_t(t)$ from the model entirely. This change slightly reduced correlations compared to the model (1). While we continue to include time in the model, we note that accounting for temporally-varying bias seems to be of limited importance, probably because any factors that change the relationship over time do not affect the entire eastern U.S. all at once, while $f_t(t)$ can only represent changes over time affecting the entire spatial domain.

Next we considered different approaches to including the meteorological functions in the model. In the basic model, we used the average of RH and PBL over UTC times 12:00, 15:00, 18:00, and 21:00 to roughly match the time range of AOD retrievals. We also considered the use of RH and PBL as the average of only using UTC times 15:00 and 18:00 and as the value only at UTC time 18:00, to more closely match the period of maximum PBL during each day (PBL increases rapidly during late morning, so times of 15:00 and 18:00 are generally the highest values during a given 24-hour period). Both of these specifications had very little effect on the correlations, nor did using log PBL (following (*12*)) in place of PBL.

We also considered a simplified model with only a spatial bias function,

$$\log \bar{a}_{it} \sim \mathcal{N}(\mu + g(s_i) + \beta \mathrm{PM}_{it}, \tau^2). \tag{5}$$

which has the benefit of not requiring one to obtain meteorological information for the calibration. Table 1 demonstrates that the simple model performs well compared to the basic model (1). While Fig. 6 and model assessment results (not shown) indicate that time, RH and PBL are significant predictors of AOD, they do not explain enough variability in AOD such that the calibration model improves substantially by including these functions. We believe the much greater importance of the spatial function than the meteorological functions is related to the confounding effect we discuss in Section 3.2.1.

Spatial variation in the relationship between AOD and $\mathrm{PM}_{2.5}$ may be related to varying reflectivity, particularly between rural, vegetated areas and urban areas. As a proxy for reflectivity, we considered adding smooth regression functions of road density and population density but found they had little impact on the model fit, with the functions estimated to be essentially flat, indicating no relationship with AOD.

In the supplementary material, we consider whether the multiplicative bias, β, might vary spatially, finding some evidence for this, but no particular interpretable pattern.

3.3 Association between $\mathrm{PM}_{2.5}$ and AOD availability

In using AOD as a proxy for $\mathrm{PM}_{2.5}$, one danger is that the missingness of the AOD may itself be informative about $\mathrm{PM}_{2.5}$ and that using only available AOD retrievals may bias predictions of $\mathrm{PM}_{2.5}$. Days with few or

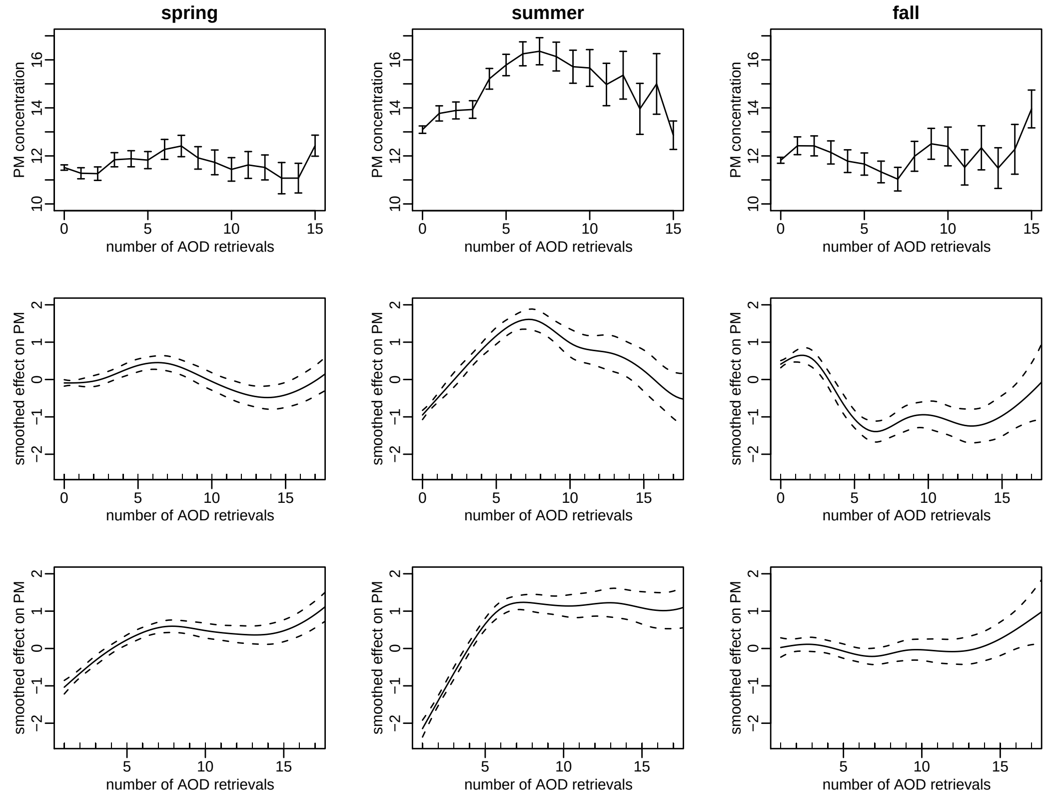

Figure 8: (Top row) Average $PM_{2.5}$ concentration (with 95% confidence interval) as a function of number of AOD retrievals by season. Data at the x-axis value of 15 are for 15 or more retrievals. (Middle row) Smoothed regression relationship between number of AOD retrievals and $PM_{2.5}$ by season, controlling for location, time, and meteorology (6). (Bottom row) Smoothed regression relationship between number of AOD retrievals and $PM_{2.5}$ by season when at least one retrieval is made in day, controlling for location, time, meteorology, and average AOD in a model of the form (6) but with a smooth term of average daily AOD also included.

no AOD retrievals may have systematically lower or higher average $PM_{2.5}$ than days with many retrievals because missingness is associated with meteorological conditions that are also associated with pollution levels. To investigate this possible association, we analyze the distribution of $PM_{2.5}$ as a function of the proportion of missing AOD retrievals.

First we consider the mean $PM_{2.5}$ by season as a function of the number of AOD retrievals. Fig. 8 (top row) indicates that in summer, there is a marked difference in $PM_{2.5}$ concentrations as a function of AOD retrievals, with the highest concentrations on days with approximately 5-10 retrievals and lower concentrations for days with few retrievals. Spring shows a somewhat similar, but less marked pattern, while fall shows little systematic pattern.

Since $PM_{2.5}$ varies in space and time, as does missingness, the association between missingness of AOD retrievals and $\text{PM}_{2.5}$ may occur merely because both missingness and $\text{PM}_{2.5}$ are separately associated with location and time. Therefore, we attempt to control for space and time, as well as meteorology as measured by PBL and RH, when assessing the relationship between $\text{PM}_{2.5}$ and missingness, by fitting the following generalized additive model separately for the spring, summer, and fall seasons.

$$\log \text{PM}_{it} \sim \mathcal{N}(\mu + g(s_i) + f_t(t) + f_{\text{PBL}}(\text{PBL}_{it}) + f_{\text{RH}}(\text{RH}_{it}) + f_n(n_{\text{AOD},it}), \tau^2) \tag{6}$$

where $n_{\text{AOD},i,t}$ is the number of AOD retrievals for location i and day t. In Fig. 8 (middle row) we see the fitted smooth regression function, $\hat{f}_n(n_{\text{AOD}})$, for each of the three seasons, indicating a nonlinear relationship between number of retrievals and $\text{PM}_{2.5}$, with $\text{PM}_{2.5}$ increasing with increasing number of retrievals, reaching a peak, and then declining as the number of retrievals increases. This suggests that after controlling for other factors affecting $\text{PM}_{2.5}$, there is still a relationship between missingness and $\text{PM}_{2.5}$.

For those days with at least one retrieval, we can ask if after controlling for measured AOD, there is any association between the number of retrievals and $\text{PM}_{2.5}$. Adding a smooth function of average AOD, $f_a(\bar{a}_{it})$ to (6) did not remove the association between missingness and $\text{PM}_{2.5}$ (Fig. 8, bottom row), although it did change the relationships somewhat, with spring and particularly summer showing increases in $\text{PM}_{2.5}$ with increasing number of retrievals and then levelling off with a larger number of retrievals. Fall shows little relationship of $\text{PM}_{2.5}$ to number of retrievals after accounting for the observed AOD.

In the analyses above, results broken out by subregions of the eastern US (northeast, eastern midwest, western midwest, southeast, and south-central, all within our defined region east of 100°W) suggest some heterogeneity in the relationship (not shown). This suggests the need to account for location in understanding the relationship between missingness and $\text{PM}_{2.5}$. The upper midwest particularly deviates from the patterns in Fig. 8.

These results suggest that predictive modeling of $\text{PM}_{2.5}$ based on GASP AOD should take the number of retrievals on a day into account as providing additional information about $\text{PM}_{2.5}$ concentrations. In particular, not accounting for missingness during the summer is likely to upwardly bias one's estimates of $\text{PM}_{2.5}$ as days with few or no AOD retrievals on average have low $\text{PM}_{2.5}$ concentrations. Of course on any individual day, clouds may prevent retrieval when $\text{PM}_{2.5}$ concentrations are high.

4 Discussion

We report the first comparison of GASP AOD with ground-level $PM_{2.5}$, building upon the expanding literature comparing AOD from MODIS and MISR with ground-level $PM_{2.5}$. We build calibration models that result in moderately strong correlations of calibrated AOD with $PM_{2.5}$ except during winter. Correlations increase with averaging over longer time periods when using the calibrated AOD. This stands in stark contrast to correlations between time-averaged raw AOD and $PM_{2.5}$, for which correlations decrease markedly with averaging, presumably because of confounding from variables that vary spatially and are correlated with long-term $PM_{2.5}$. Our results also suggest that there is useful information even from days with a single GASP AOD retrieval, both for estimating $PM_{2.5}$ for individual days and for providing additional information within a longer-term average. We point out that whether AOD retrievals are missing does not occur at random with respect to $PM_{2.5}$ concentrations. Our results are consistent with those of (*22*), who report that $PM_{2.5}$ concentrations are on average 15% higher when averaged over days with MODIS retrievals compared to averaging over all days.

Initial results from work in progress that directly compares MISR, MODIS and GASP AOD, all for 2004, suggest that correlations of MISR and MODIS AOD with $PM_{2.5}$ are somewhat higher at the daily resolution, with correlations in the range of 0.55 to 0.65 compared to 0.50 for GASP AOD when using days with any number of GASP AOD retrievals. Restricting the use of GASP AOD to days with more retrievals increases daily correlations to the level of MODIS and MISR, at the cost of loss of information. Given the limitations of the GOES instrument, the fact that the GASP AOD correlations are not too much lower than for MISR and MODIS indicates the promise of GASP AOD for use as a proxy for $PM_{2.5}$.

Critically, the comparison of daily correlations for those days with matched AOD retrievals and $PM_{2.5}$ measurements does not take into account the much greater data density of GASP AOD. The half-hourly temporal coverage provides much more opportunity for avoiding clouds at least once during the day and for averaging over multiple retrievals in a day. This can result in a more robust estimate that averages over noisiness in the retrieval and over temporal variability in pollution during the day. In addition, the geostationary orbit ensures that retrievals are attempted each day. To assess the potential importance of data density, we matched valid individual daily retrievals to a set of 632 AQS sites in the eastern United States, removing sites that were very close to one another (we allowed only one site per GOES pixel). We calculated the proportion of days in a month with a valid AOD retrieval for each of the three satellite instruments for each site. Using the sites as a set of locations roughly reflective of population in the eastern United States, we report that MISR (MODIS) provided a valid retrieval on average only 4% (14%) of the days in a given month at a given location. In contrast, GOES provided a valid retrieval on average 40% of the days (21% of days if only considering days with at least five retrievals). Note that in calculating valid retrievals for GOES, we assumed no valid retrievals in winter because of the lack of association between GASP AOD and $PM_{2.5}$ found in this work. Restricting to non-winter, the advantage of GOES is even more striking, with 53% of days with a valid retrieval (28% when requiring at least five retrievals in a day) compared to 16% for MODIS and 4% for MISR. Since our goal is to estimate monthly average $PM_{2.5}$ as an average across all days in the month the greater temporal coverage of GASP AOD should result in proxy values that are much more representative of $PM_{2.5}$ over all the days in the month. In ongoing work, we are assessing this quantitatively. Finally, in the supplementary material, we provide evidence that some of the criteria used to select valid GASP AOD retrievals might be relaxed to provide even more retrievals.

In contrast to the importance of its high temporal resolution, GASP AOD does not appear to provide real improvement in spatial resolution. While GASP AOD is available at higher nominal spatial resolution than MISR and MODIS, the lack of improved correlations when matched to monitors within the pixel suggests that the higher nominal resolution of GASP AOD does not provide a significant advantage, presumably because of instrument differences. Consistent with this, (*12*, *13*) found higher daily correlations than found here when averaging over multiple MODIS and MISR pixels. It is difficult to separate the effect of instrument differences from the different spatial resolutions of the AOD retrievals of the instruments to understand whether finer retrieval resolution would improve correlations with ground-level $PM_{2.5}$.

Of perhaps equal importance to its high temporal coverage, GASP AOD provides the possibility of a long-term record, allowing us to create a proxy for $PM_{2.5}$ starting in November 1994, when the GOES-8 satellite retrievals are first available. Dense $PM_{2.5}$ ground monitoring only began in 1999, so GASP AOD provides one of the few proxies for $PM_{2.5}$, apart from PM_{10} measurements and scattered observations and small datasets, for the period 1995-1998. For epidemiological work, the addition of data useful for exposure estimation for four years could greatly increase statistical power to detect health effects.

This work is part of a larger project in which we will use GASP, MODIS and MISR AOD integrated with ground-level $PM_{2.5}$ monitoring in a statistical model to estimate $PM_{2.5}$ at high spatial resolution across the eastern United States. Our results here indicate the potential of GASP AOD as a proxy for $PM_{2.5}$ and suggest that after calibration, we may be able to use GASP AOD as part of the model with a simple linear relationship to $PM_{2.5}$. Further conclusions as to the relative usefulness of the different AOD products as proxies for $PM_{2.5}$ will be informed by the statistical modeling. This will involve a base model built using $PM_{2.5}$ monitoring data and GIS and meteorological covariates and expanded models that also include AOD retrievals. We believe the ultimate test is whether the addition of AOD retrievals improves upon predictions that could be made without the remote sensing information.

Acknowledgments

This work was supported by grant 4746-RFA05-2/06-7 from the Health Effects Institute. The authors thank Doug Dockery for helping to initiate the collaboration between NOAA and Harvard, Steve Melly for GIS processing, and Jeff Yanosky for suggestions.

References

1. Dockery, D.; Pope, III, C.; Xu, X.; Spengler, J.; Ware, J.; Fay, M.; Ferris, B.; Speizer, F. An assocation between air pollution and mortality in six U.S. cities. *New England Journal of Medicine* **1993**, *329*, 1753–1759.

2. Abbey, D.; Ostro, B.; Peterson, F.; Burchette, R. Chronic respiratory symptoms associated with estimated long-term ambient concentrations of fine particulates less than 2.5 microns in aerodynamic diameter (PM2.5) and other air pollutants. *Journal of Exposure Analysis and Environmental Epidemiology* **1995**, *5*, 137–159.

3. Pope, C., III; Thun, M.; Namboodiri, M.; Dockery, D.; Evans, J.; Speizer, F.; C.W. Heath, J. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *American Journal of Respiratory and Critical Care Medicine* **1995**, *151*, 669–674.

4. Pope, C., III; Burnett, R.; Thun, M.; Calle, E.; Krewski, D.; Ito, K.; Thurston, G. Lung cancer, cardiopulmonary mortality and long-term exposure to fine particulate air pollution. *JAMA* **2002**, *287*, 1132–1141.

5. Pope, C., III; Burnett, R.; Thurston, G.; Thun, M.; Calle, E.; D. Krewski, e. a. Cardiovascular mortality and long- term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. *Circulation* **2004**, *109*, 71–77.

6. Yanosky, J.; Paciorek, C.; Schwartz, J.; Laden, F.; Puett, R.; Suh, H. Spatio-temporal modeling of chronic PM10 exposure for the Nurses' Health Study. *submitted to Atmospheric Environment* **2007**, *submitted*, .

7. Samet, J.; Dominici, F.; Curriero, F.; Coursac, I.; Zeger, S. Particulate air pollution and mortality: findings from 20 U.S. cities. *New England Journal of Medicine* **2000**, *343*, 1742–1757.

8. Dominici, F.; Peng, R.; Bell, M.; Pham, L.; McDermott, A.; Zeger, S.; Samet, J. Fine particles, air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of the American Medical Association* **2006**, *295*, 1127–1135.

9. Chu, D.; Kaufman, Y.; Zibordi, G.; Chern, J.; Mao, J.; Li, C.; Holben, B. Global monitoring of air pollution over land from the earth observing system-terra moderate resolution imaging spectroradiometer (MODIS). *Journal of Geophysical Research - Atmospheres* **2003**, *108*, D21,4661.

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

11. Engel-Cox, J.; Holloman, C.; Coutant, B.; Hoff, R. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment* **2004**, *38*, 2495–2509.

12. Liu, Y.; Sarnat, J.; Kilaru, V.; Jacob, D.; Koutrakis, P. Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing. *Environmental Science and Technology* **2005**, *39*, 3269–3278.

13. 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 Environment* **2007**, *107*, 33–44.

14. Pelletier, B.; Santer, R.; Vidot, J. Retrieving of particulate matter from optical measurements: a semiparametric approach. *Journal of Geophysical Research* **2007**, *112*, D06208.

15. Kahn, R.; Banerjee, P.; McDonald, D.; Diner, D. Sensitivity of multiangle imaging to aerosol optical depth and to pure-particle size distribution and composition over ocean. *Journal of Geophysical Research* **1998**, *103*, 32195–32213.

16. Prados, A.; Kondragunta, S.; Ciren, P.; Knapp, K. GOES aerosol/smoke product (GASP) over North America: comparisons to AERONET and MODIS observations. *Journal of Geophysical Research-Atmospheres* **2007**, *112*, D15201.

17. Knapp, K.; Haar, T. V.; Kaufman, Y. Aerosol optical depth retrieval from GOES-8: Uncertainty study and retrieval validation over South America. *Journal of Geophysical Research* **2002**, *107*, D4055.

18. Rogers, R.; Kondragunta, S.; McQueen, J.; Moshary, F.; Gross, B.; Hoff, R. Validation of ETA-CMAQ-modeled planetary boundary layer height with elastic lidar. *submitted* **2006**, submitted.

19. Chin, M.; Diehl, T.; Ginoux, P.; Malm, W. Intercontinental transport of pollution and dust aerosols: implications for regional air quality. *Atmospheric Chemistry and Physics* **2007**, *7*, 5501–5517.

20. Wood, S. *Generalized Additive Models: An Introduction with R*; Chapman & Hall: Boca Raton, 2006; p 391.

21. Chin, M.; Ginoux, P.; Kinne, S.; et al., Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements. *Journal of the Atmospheric Sciences* **2002**, *59*, 461–483.

22. Koelemeijer, R.; Homan, C.; Matthijsen, J. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmospheric Environment* **2006**, *40*, 5304–5315.

23. Schabenberger, O.; Gotway, C. *Statistical Methods for Spatial Data Analysis*; Chapman & Hall: Boca Raton, 2005; p 488.

Supplementary Information

A Time series estimates of daily AOD and associated uncertainty

One estimate of daily AOD is the simple arithmetic average of the available AOD retrievals. Our paper focuses on this estimate becasue of its simplicity and because the estimator described below does not substantially improve the calibration, as discussed in Section 3.2. However, in other settings, accounting for correlation in estimating long-term averages may be important. Here we outline the approach.

The disadvantage of using the simple arithmetic average is that it does not account for the temporal correlations between half-hourly values. Standard statistical theory indicates that a better estimator (one with less variability) can be obtained by accounting for the correlations and that an estimate of the uncertainty of the estimated daily average AOD should account for the temporal correlation as well.

We want to estimate the integrated AOD across the time period during which observations are available. Letting $a(h)$ represent AOD as a function of the time of day, we wish to estimate $a_d = \frac{1}{|t_2 - t_1|} \int_{t_1}^{t_2} a(h) dh$ for each day d. In spatial statistics the best linear unbiased predictor (BLUP) of the integrated value is the so-called block-kriging estimator, which relies on calculating covariances between intervals and single points in time (in this temporal setting). A numerical approximation is to predict $a(h)$ at a set of times, say a fine grid of times covering the interval $(t_1 = t_{min} - 15m, t_2 = t_{max} + 15m)$ where t_{min} and t_{max} are the first and last times that the solar zenith angle is less than 70 and where we extend the time window by half the time interval between observation times (15 min) so that all prediction times are within 15 minutes of a possible retrieval. We then approximate the integral as the average of the predictions at each time point on the fine grid based on a time series model

$$\hat{a}_d = \frac{1}{N} \sum_{i=1}^{N} \hat{a}(h_i) \tag{7}$$

where $\hat{a}(h_i)$ is the best linear unbiased predictor (BLUP) for AOD at time h_i. The BLUP must account for the correlation between AOD at different times; by doing so, the prediction $\hat{a}(h_i)$ is a weighted average of AOD values from nearby times. The overall estimator weights observations that are widely separated from other observations more than observations for which the most recent and nearest times in the future have available AOD values, as these provide somewhat redundant information. After exploratory analysis using time series of AOD for days with at least 10 observations, an AR(1) time series model appears appropriate for most days and locations. It appears that the autoregressive parameter in the AR(1) model varies slightly as a function of the number of AOD observations available but $\rho = 0.3$ seems to be a good compromise value for the correlation between observations one-half hour apart. This correlation is lower than one would expect for the true aerosol optical depth over time; we suspect the low autocorrelation is due noisiness in the satellite-retrieved AOD as a measurement of true AOD. The kriging model assumes the AOD observations over time at the prediction grid times (which include the observation times as well) follow a normal distribution, $a \sim \mathcal{N}(\mu 1, \sigma^2 \Sigma)$ where $\Sigma_{ij} = \rho^{2|h_i - h_j|}$. The prediction, $\hat{a} = (\hat{a}(h_1), \ldots, \hat{a}(h_N))$, takes the form of the simple kriging estimator (*23*),

$$\begin{aligned} \hat{a} &= \hat{\mu} 1_p + \Sigma_{21} \Sigma_{11}^{-1} (a_{\text{present}} - \hat{\mu} 1_n) \\ \hat{\mu} &= (1_n^T \Sigma_{11}^{-1} 1_n)^{-1} 1_n^T \Sigma_{11}^{-1} a_{\text{present}} \end{aligned} \tag{8}$$

where Σ_{11} is the correlation matrix (a submatrix of Σ) for the available data, a_{present}. Σ_{21} is the correlation between the predictions at the fine grid of times and the times of the available data and is calculated in the same manner as Σ_{11}. Following R. Smith (UNC Department of Statistics, unpublished), one can also derive the full prediction covariance matrix as

$$\begin{aligned} V(\hat{a}) \;\propto\; & \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} + (1_p^T \Sigma_{11}^{-1} 1_p)^{-1} \cdot \\ & (1_p 1_p^T - 1_p 1_n^T \Sigma_{11}^{-1} \Sigma_{12} - (1_p 1_n^T \Sigma_{11}^{-1} \Sigma_{12})^T + (\Sigma_{21} \Sigma_{11}^{-1} 1_n)(\Sigma_{21} \Sigma_{11}^{-1} 1_n)^T) \end{aligned} \tag{9}$$

where $\Sigma_{22} = \Sigma$ is the correlation between all the prediction times on the fine grid. The proportionality comes from leaving out a term, σ^2, common to all the predictions. Our estimate of $\hat{a}_d$ has variance proportional to $\frac{1}{N^2} 1_p^T V(\hat{a}) 1_p$, which we use as our estimate, $V(\hat{a}_d)$.

Because of the relatively low autocorrelation of $\rho = 0.3$ between half-hourly values, the resulting estimates of $\hat{a}_d$ do not vary substantially from $\bar{a}_d$, though the relative variances (ignoring σ^2) are somewhat different than $1/n$, the variance estimator for $\bar{a}_d$ (also ignoring σ^2).

B Spatially-varying multiplicative bias

We considered whether the multiplicative bias, β in (1), might vary spatially, fitting the model

$$\log \bar{a}_{it} \sim \mathcal{N}(\mu + g(s_i) + f_t(t) + f_{\text{PBL}}(\text{PBL}_{it}) + f_{\text{RH}}(\text{RH}_{it}) + (\beta + \beta(s))\text{PM}_{it}, \tau^2) \tag{10}$$

fitting an average effect, β, and also a spatially-varying bias, $\beta(s)$. This model can also be fit with the gam() function in R. The fitted model indicates that there is substantial spatially smooth variation in the multiplicative scaling, with the standard deviation of the fitted $\beta(s)$ across the sites equal to 0.0049, 0.0043 and 0.0068 for spring, summer and fall respectively, which is substantial variation relative to the the estimates, $\hat{\beta}$, of 0.016, 0.016, and 0.013 for the three seasons. Fig. 9 shows the estimates of $g(s)$ and $\beta + \beta(s)$, with

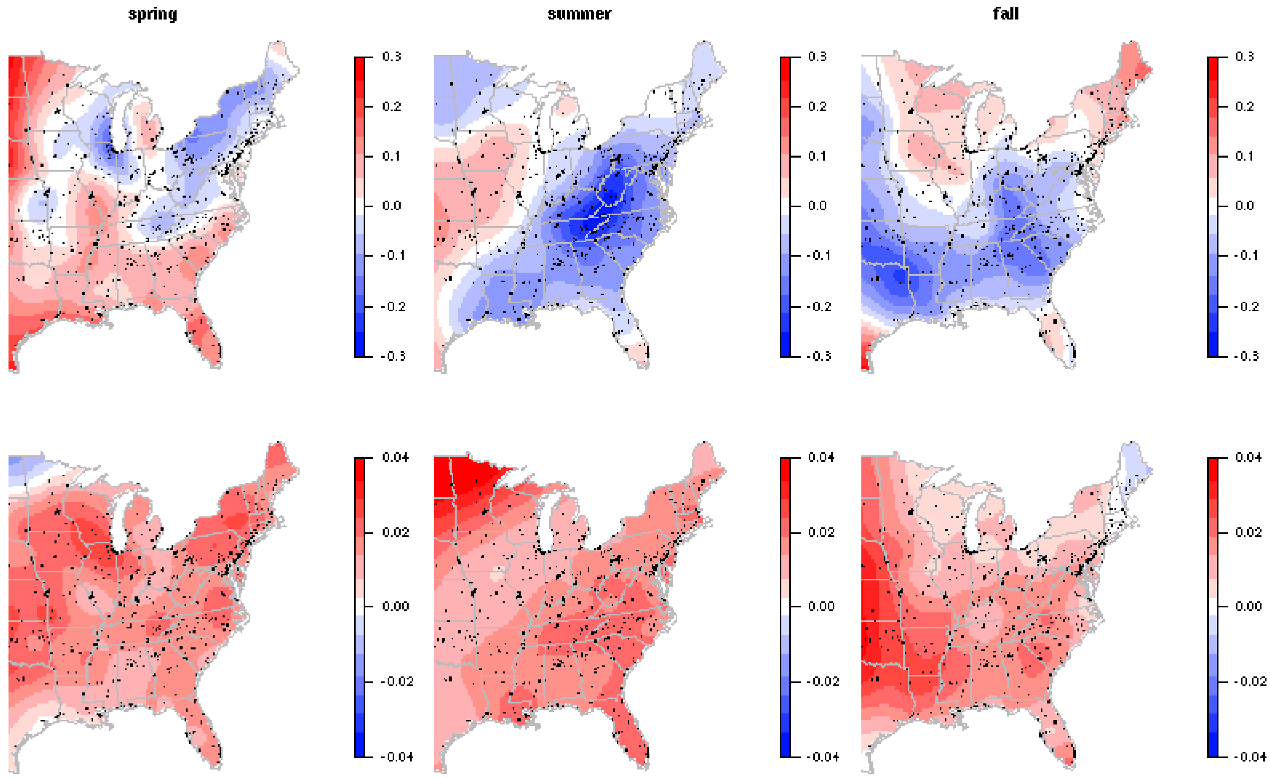

Figure 9: Fitted spatial functions for model (10) by season: additive functions, $\hat{g}(s)$ (top row) and multiplicative scaling functions, $\hat{\beta} + \hat{\beta}(s)$ (bottom row). Points are matched AOD-$PM_{2.5}$ sites.

evidence that the spatial patterns change somewhat between seasons. The overall patterns in the additive spatial function are similar to those estimated in the base model with the lower than expected AOD over the Appalachian Mountains/Ohio Valley, while the variability in the multiplicative scaling shows no particular interpretable pattern. Based on the multiplicative model, one could try to use the following calibration

$$a^*_{it} = \frac{1}{\hat{\beta} + \hat{\beta}(s)} \left(\log \bar{a}_{it} - \hat{\mu} - \hat{g}(s_i) - \hat{f}_t(t) - \hat{f}_{\text{PBL}}(\text{PBL}_{it}) - \hat{f}_{\text{RH}}(\text{RH}_{it}) \right). \tag{11}$$

However, when $\hat{\beta} + \hat{\beta}(s) \approx 0$, the model is indicating there is little relationship between AOD and $PM_{2.5}$ and there are some extreme calibrated values, a^*_{it}. Instead, in our use of calibrated AOD (2) in the larger project, we plan to allow for spatially-varying multiplicative bias directly in the statistical model rather than in the calibration step used to preprocess the AOD retrievals.

C Assessing the usefulness of AOD observations of uncertain quality

The processing of AOD retrievals produces a number of quality flags that may be used to screen out retrievals of poor quality, which might be biased or merely very noisy estimates of true AOD. These standard criteria used by NOAA to screen the retrievals are to require the following conditions for a valid retrieval: AOD value less than 10, AOD standard deviation less than 0.15, surface reflectivity greater than 0.01 and less than 0.15, channel 1 visible reflectivity greater than zero, aerosol signal greater than 0.01, and no clouds detected by the cloud screening in a 5 by 5 array of cells centered on the pixel of interest (Cloudsum=25). Given the availability of the gold standard $PM_{2.5}$ data, for which we would like GASP AOD to serve as a proxy, we can consider relaxing or making more stringent these standard quality criteria. The goal is to see if stronger associations with $PM_{2.5}$ can be obtained, or if equivalent associations can be obtained but with an increase in the number of usable retrievals. Note that we need to be cautious of finding stronger associations with stricter criteria merely because the stricter criteria result in removing AOD-$PM_{2.5}$ pairs that while less strongly associated are still associated with $PM_{2.5}$, which in a statistical prediction model would amount to throwing away proxy data with useful, albeit more variable, information. Since our focus is on potential relaxation of the criteria, we address this by comparing correlations calculated based only on matched pairs for days with at least one AOD retrieval under the stricter standard criteria.

We consider relaxing the following individual quality flag criteria one at a time: 1.) AOD standard deviation less than 0.30 rather than 0.15; 2.) Cloudsum>20 rather than Cloudsum=25; 3.) Cloudsum>15 rather than Cloudsum=25; 4.) solar zenith angle < 75 rather than zenith angle < 70; 5.) solar zenith angle < 80 ; 6.) solar zenith angle < 85; 7.) surface reflectivity <20 rather than <15; and 8.) surface reflectivity <25. Comparing only matched pairs for days with at least one AOD under the standard criteria, Table 2 shows correlations of AOD and $PM_{2.5}$ for the various criteria, excluding winter. The results suggest that relaxing the standard deviation criterion produces lower associations; this criterion serves to screen out retrievals when neighboring pixels have very different retrieved values, potentially because of cloud contamination. In contrast, relaxing the cloudsum criterion has limited effect when more than 20 of the pixels in the surrounding 5 by 5 array are cloud free, suggesting little information is added or lost from augmenting daily AOD averages based on these additional retrievals. Further relaxation of the cloudsum criterion appears to result in loss of information. Relaxing the surface reflectivity criterion decreases correlations. In contrast, relaxing the zenith angle criterion increases the associations between the AOD proxies and $PM_{2.5}$. Even relaxing so

Table 2: Correlations between AOD and $PM_{2.5}$ under different criteria for AOD validity for different temporal resolutions. All values are based on matched pairs for which there is at least one daily retrieval under the strictest (the standard) criteria. p-values from paired t-tests are indicated as (*) p<0.01; (**) p<0.001; (***) p<0.0001. Each test compares the squared model residuals from the regression of $PM_{2.5}$ on the AOD proxy based on the standard criteria (i.e., the top row results) to the squared model residuals from the regression of $PM_{2.5}$ on the AOD proxy based on one of the alternative criteria, to see if the mean squared residuals are substantially different under the alternative criteria.

	daily, raw AOD	daily calibrated AOD	monthly averages (at least 3 matched days for each site-month)	yearly averages (at least 10 matched days for each site)
Standard criteria	0.408	0.502	0.617	0.745
Relax std. dev. criterion	0.402*	0.486***	0.598***	0.743
Relax Cloudsum criteria (>20)	0.411*	0.502	0.617	0.746
Further relax cloudsum (>15)	0.410	0.498*	0.612	0.738
Relax zenith angle (<75)	0.423***	0.520***	0.629***	0.747
Further relax zenith angle (<80)	0.428***	0.530***	0.638***	0.751
Further relax zenith angle (<85)	0.427***	0.532***	0.637***	0.739
Relax reflectivity criterion (<20)	0.379**	0.494***	0.600***	0.722***
Relax reflectivity criterion (<25)	0.379**	0.492***	0.594***	0.716***

far as to include observations with zenith angle less than 85 degrees seems to increase associations in all but the yearly averaging, for which the association decreases but not significantly so. One note of caution is that (*16*) found higher mean square error in GASP AOD compared to AERONET AOD early and late in the day compared to the middle of the day (although correlations were no lower during these times), providing empirical evidence that GASP AOD may be less accurate as a measurement of AOD at high solar zenith angles.

Table 3 shows the increase in the number of retrievals and the number of days with several thresholds for the number of retrievals under the various criteria, indicating that the relaxed criteria admit a sizable increase in retrievals.

We can also consider correlations between AOD and $PM_{2.5}$ for new matched pairs that become available when relaxing the criteria, namely locations for which there was no AOD retrieval on the day under the stricter criteria. These new matched pairs are almost always based on a single AOD retrieval during the day, so a point of comparison is the correlation between the calibrated AOD under the standard criteria and $PM_{2.5}$ for days with only one matched pair, which is 0.38. The correlations for the new matched pairs are 0.37 and 0.35 when increasingly relaxing the cloudsum criterion; 0.30 when relaxing the standard deviation criterion;

Table 3: Percentage increase in number of retrievals under different criteria for AOD validity excluding winter, all compared to the standard criteria. Note that these only reflect retrievals that match $PM_{2.5}$ data and are meant only to give a rough estimate of the effect of the criteria on the number of retrievals.

	Number of half-hourly retrievals	Number of days with at least one retrieval	Number of days with at least three retrievals	Number of days with at least five retrievals
Relax std. dev. criterion	21	11	10	42
Relax Cloudsum criterion (>20)	9	11	7	33
Further relax Cloudsum (>15)	13	15	11	38
Relax zenith angle (<75)	14	11	9	38
Further relax zenith angle (<80)	24	15	15	49
Further relax zenith angle (<85)	34	23	20	58
Relax reflectivity criterion (<20)	24	14	15	50
Relax reflectivity criterion (<25)	28	15	17	54

0.33, 0.33, and 0.27 when increasingly relaxing the zenith angle criterion; and 0.26 in both cases of relaxing the reflectivity criterion. Given the calibration results in Table 2, it's somewhat surprising that relaxing the cloudsum criterion seems to outperform relaxing the zenith angle criterion when considering only new daily observations made available because of the relaxed criteria. In all cases, the positive correlations suggest that there is information about $PM_{2.5}$ available in the discarded observations that do not satisfy the standard criteria.

Next we consider making the zenith angle criteria more strict. Not surprisingly given that relaxing this criteria seems worthwhile, making it more strict decreased the correlations between the AOD proxies and $PM_{2.5}$ (generally by about 0.02). We also considered setting all negative AOD values to zero or excluding negative observations, the latter following (*16*). Setting the negative values to zero slightly decreased correlations while excluding such observations markedly decreased correlations (generally by about 0.04), so we suggest using the negative values as reported rather than truncating or excluding them when one's goal is use of AOD as a proxy for $PM_{2.5}$.