Walter A. Rosenblith New Investigator Award
RESEARCH REPORT

HEALTH EFFECTS INSTITUTE

Number 167
May 2012

Assessment and Statistical Modeling of the Relationship Between Remotely Sensed Aerosol Optical Depth and $PM_{2.5}$ in the Eastern United States

Christopher J. Paciorek and Yang Liu

Includes a Commentary by the Institute's Health Review Committee

Assessment and Statistical Modeling of the Relationship Between Remotely Sensed Aerosol Optical Depth and $PM_{2.5}$ in the Eastern United States

Christopher J. Paciorek and Yang Liu

with a Commentary by the HEI Health Review Committee

Research Report 167

Health Effects Institute

Boston, Massachusetts

Trusted Science • Cleaner Air • Better Health

Publishing history: The Web version of this document was posted at *www.healtheffects.org* in May 2012.

Citation for document:

Paciorek CJ, Liu Y. 2012. Assessment and Statistical Modeling of the Relationship Between Remotely Sensed Aerosol Optical Depth and $PM_{2.5}$ in the Eastern United States. Research Report 167. Health Effects Institute, Boston, MA.

 Cameographics, Belfast, Me., Compositor. Printed by Recycled Paper Printing, Boston, Mass. Library of Congress Number for the HEI Report Series: WA 754 R432.

Cover paper: made with at least 55% recycled content, of which at least 30% is post-consumer waste; free of acid and elemental chlorine. Text paper: made with 100% post-consumer waste recycled content; acid free; no chlorine used in processing. The book is printed with soy-based inks and is of permanent archival quality.

CONTENTS

ABOUT HEI

The Health Effects Institute is a nonprofit corporation chartered in 1980 as an independent research organization to provide high-quality, impartial, and relevant science on the effects of air pollution on health. To accomplish its mission, the institute

- Identifies the highest-priority areas for health effects research;
- Competitively funds and oversees research projects;
- Provides intensive independent review of HEI-supported studies and related research;
- Integrates HEI's research results with those of other institutions into broader evaluations; and
- Communicates the results of HEI's research and analyses to public and private decision makers.

HEI typically receives half of its core funds from the U.S. Environmental Protection Agency and half from the worldwide motor vehicle industry. Frequently, other public and private organizations in the United States and around the world also support major projects or research programs. HEI has funded more than 280 research projects in North America, Europe, Asia, and Latin America, the results of which have informed decisions regarding carbon monoxide, air toxics, nitrogen oxides, diesel exhaust, ozone, particulate matter, and other pollutants. These results have appeared in the peer-reviewed literature and in more than 200 comprehensive reports published by HEI.

HEI's independent Board of Directors consists of leaders in science and policy who are committed to fostering the public–private partnership that is central to the organization. The Health Research Committee solicits input from HEI sponsors and other stakeholders and works with scientific staff to develop a Five-Year Strategic Plan, select research projects for funding, and oversee their conduct. The Health Review Committee, which has no role in selecting or overseeing studies, works with staff to evaluate and interpret the results of funded studies and related research.

All project results and accompanying comments by the Health Review Committee are widely disseminated through HEI's Web site (*www.healtheffects.org*), printed reports, newsletters and other publications, annual conferences, and presentations to legislative bodies and public agencies.

ABOUT THIS REPORT

Research Report 167, *Assessment and Statistical Modeling of the Relationship Between Remotely Sensed Aerosol Optical Depth and $PM_{2.5}$ in the Eastern United States,* presents a research project funded by the Health Effects Institute and conducted by Dr. Christopher J. Paciorek of the Department of Biostatistics at the Harvard School of Public Health, Boston, Massachusetts, and Dr. Yang Liu of the Rollins School of Public Health of Emory University, Atlanta, Georgia. This research was funded under HEI's Walter A. Rosenblith New Investigator Award Program, which provides support to promising scientists in the early stages of their careers. The report contains three main sections.

The HEI Statement, prepared by staff at HEI, is a brief, nontechnical summary of the study and its findings; it also briefly describes the Health Review Committee's comments on the study.

The Investigators' Report, prepared by Paciorek and Liu, describes the scientific background, aims, methods, results, and conclusions of the study.

The Commentary is prepared by members of the Health Review Committee with the assistance of HEI staff; it places the study in a broader scientific context, points out its strengths and limitations, and discusses remaining uncertainties and implications of the study's findings for public health and future research.

This report has gone through HEI's rigorous review process. When an HEI-funded study is completed, the investigators submit a draft final report presenting the background and results of the study. This draft report is first examined by outside technical reviewers and a biostatistician. The report and the reviewers' comments are then evaluated by members of the Health Review Committee, an independent panel of distinguished scientists who have no involvement in selecting or overseeing HEI studies. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, as necessary, to revise their report. The Commentary reflects the information provided in the final version of the report.

HEI STATEMENT

Synopsis of Research Report 167

Assessment of the Relationship Between Satellite-Based Estimates and Measurements of $PM_{2.5}$ in the Eastern United States

INTRODUCTION

Over the past decade, satellite-based estimates of ground-level pollution have emerged as a potentially important source of information on human exposure to health-damaging pollutants such as nitrogen dioxide and fine particulate matter ($PM_{2.5}$; PM with an aerodynamic diameter of 2.5 µm or smaller). Estimates of the concentration of ground-level $PM_{2.5}$ can be made from satellite calculations of aerosol optical depth (AOD) — a measure of light extinction by aerosols in the total atmospheric column calculated from measurements of light scattering at various wavelengths. Health effects researchers have begun to use satellite-based estimates of $PM_{2.5}$ from AOD calculations in both epidemiologic research and risk assessment; however, the accuracy and precision of estimates provided by different satellite-based estimators and the circumstances in which such estimates might make the most important contributions remain uncertain.

Dr. Christopher J. Paciorek of the Harvard School of Public Health submitted an application under Request for Applications 05-2, "The Walter A. Rosenblith New Investigator Award," which was established to provide support for an outstanding investigator beginning his or her independent research career. In his proposed study, "Integrating Monitoring and Satellite Data to Estimate $PM_{2.5}$ Exposure and Its Chronic Health Effects in the Nurses' Health Study," Paciorek planned to reanalyze data on the chronic health effects of $PM_{2.5}$ in the Nurses' Health Study, a large epidemiologic cohort study, by integrating AOD measurements from satellite data with ground monitoring data to improve the exposure-assessment modeling. The HEI Health Research Committee urged Paciorek to focus on the estimates of exposure, rather than the epidemiologic analysis. Ultimately, HEI funded the current study, "Integrating Monitoring and Satellite Data to Retrospectively Estimate Monthly $PM_{2.5}$ Concentrations in the Eastern United States," which began in 2006.

The overall objective of Paciorek's study was to assess the ability of approaches that use satellite-based measurements of AOD from the National Aeronautics and Space Administration's (NASA's) multiangle imaging spectroradiometer (MISR) and moderate resolution imaging spectroradiometer (MODIS) satellites to fill spatial and temporal gaps in existing monitoring networks in the eastern United States. To accomplish this, the investigators developed statistical models for integrating monitoring, satellite, and geographic information system (GIS) data to estimate average monthly ambient $PM_{2.5}$ concentrations. They then applied those models across the eastern United States at various times during the period from 2000 to 2006 at a fine spatial resolution. Their goal was to better understand temporal and spatial variation in $PM_{2.5}$ and then determine how accurately it could be characterized based on satellite, monitoring, and GIS data. They developed and applied statistical methods to quantify (1) how uncertainties in exposure estimates based on ground-level monitoring data might be reduced by adding satellite-based monitoring data and (2) how systematic discrepancies between a proxy measure — such as AOD as a proxy for $PM_{2.5}$ — could be identified and accounted for. They also explored the potential for two additional sources of information, the Geostationary Operational Environmental Satellite (GOES) and the Community Multiscale Air Quality (CMAQ) atmospheric-chemistry model, to improve $PM_{2.5}$ exposure estimates for years in which satellite measurements from MISR and MODIS are unavailable.

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr. Christopher J. Paciorek at Harvard School of Public Health, Boston, Massachusetts, and Dr. Yang Liu at Rollins School of Public Health, Emory University, Atlanta, Georgia. Research Report 167 contains both the detailed Investigators' Report and a Commentary on the study prepared by the Institute's Health Review Committee.

APPROACH

The investigators assembled publicly available data for the eastern United States from satellites, ground-level air pollution monitors, meteorologic observations, and land-use databases; these included NASA's MODIS and MISR satellites for AOD measurements, the U.S. Environmental Protection Agency's (EPA's) Interagency Monitoring of Protected Visual Environments $PM_{2.5}$ monitoring network, GOES, the U.S. EPA's CMAQ model for emissions-based estimates of $PM_{2.5}$, the U.S. National Oceanic and Atmospheric Administration's North American Regional Reanalysis database, and the Multi-Resolution Land Characteristics Consortium.

They then analyzed the relationship between $PM_{2.5}$ monitoring data and AOD measurements in both space and time. First they estimated the correlation between $PM_{2.5}$ and AOD both before and after calibration with meteorologic factors and at different temporal and spatial scales in order to understand discrepancies between AOD and $PM_{2.5}$. Then they conducted analyses in which (1) AOD data were included as variables in statistical models to predict $PM_{2.5}$ concentrations, and (2) the discrepancies between proxies (such as AOD) and $PM_{2.5}$ were determined. They also explored the spatial relationship between AOD and $PM_{2.5}$ using new methods they had developed for this purpose.

RESULTS

The investigators report that satellite-based AOD estimates did not improve $PM_{2.5}$ predictions for the eastern United States as compared with predictions from other geospatial models. Although AOD data were temporally correlated with $PM_{2.5}$ measurements, correlations of long-term spatial averages were relatively weak unless they were adjusted statistically for the discrepancy between AOD and $PM_{2.5}$. Although statistical models that combined AOD, $PM_{2.5}$ observations, and land-use and meteorologic variables were highly predictive of $PM_{2.5}$, AOD itself contributed little to the predictive power of those models over and above the other variables.

Further, the investigators report that substituting $PM_{2.5}$ estimates from the U.S. EPA's CMAQ model for AOD data also did not improve the ability of the multivariable models to predict the measured $PM_{2.5}$ data, and that attempts to use data from the GOES satellite to correct for interference with AOD from surface reflectance (brightness), a factor that may affect the relationship of AOD with $PM_{2.5}$, were largely unsuccessful.

Using statistical models that accounted for potential discrepancies between AOD and $PM_{2.5}$ at both large and small spatial scales was an important determinant of predictive ability in this study. Models that did not account for discrepancies at small spatial scales had poor predictive ability for measured $PM_{2.5}$. The investigators attributed this lack of predictive ability to the fact that their analysis did not account for spatial differences in the vertical profile of the aerosol and to the effects of the variable degree of reflectance of the Earth's surface.

Paciorek and his colleague, Yang Liu, concluded that the inability of AOD data to improve the spatial prediction of monthly and yearly average $PM_{2.5}$ in the eastern United States is a result of the spatial discrepancy between AOD estimates and measured $PM_{2.5}$, particularly at smaller spatial scales. They stress the importance of explicitly accounting for such discrepancies in statistical models that use proxy estimates, such as AOD, in order to distinguish the $PM_{2.5}$ "signal" in the proxy measure (AOD) from the "noise" contributed by the discrepancy between the proxy and ground-level $PM_{2.5}$. They concluded that there is little evidence that current satellite-based estimates of AOD can improve the prediction of ground-level $PM_{2.5}$ at small-to-moderate scales in the eastern United States, and they argue that until more evidence regarding the reliability of satellite-derived AOD data is available, it is premature to use such data in epidemiologic studies as a proxy for $PM_{2.5}$. They note, however, that more promising results might be obtained in areas that have levels of ambient $PM_{2.5}$ higher than those in the eastern United States, such as some developing countries, because the AOD signal would be relatively stronger in such locales than the noise contributed by background surface reflectance.

INTERPRETATION AND CONCLUSIONS

On the basis of this report, several conclusions regarding the application of satellite-based estimates of $PM_{2.5}$ air pollution in health effects research seem warranted. First, the use of raw AOD estimates as a proxy for $PM_{2.5}$ measurements in health effects research should be avoided in general. Rather, approaches that combine information from multiple sources — remote sensing, model-based estimates,

and ground-level measurements — may offer the most promise. Recent studies that have taken the latter approach have reported that satellite-based estimates can explain a fair amount of between-city variability in $PM_{2.5}$, but that the estimates need to be combined with other data to explain within-city variation.

Applications to health effects research should include, to the extent possible, evaluations of the relationship between satellite-based estimates and monitoring data; researchers should try to quantify the contribution of satellite-based estimates to the total exposure measurement error in epidemiologic effect estimates.

Satellite-based estimates can potentially play an important role in the evaluation of exposure to, and health effects of, short-term episodes of high levels of air pollution from vegetation fires or dust events, especially in areas with limited or no monitoring, where satellite-based estimates could help define the spatial extent of the exposure.

In areas of the world where ground-based measurements of air pollution are not likely to be collected for the foreseeable future, satellite-based data may help address the needs of epidemiologic research and public health–based risk assessment. Satellite remote sensing may offer promise for providing information on exposure to $PM_{2.5}$ at regional-to-global scales, especially in places with the highest levels of pollution and the greatest estimated burden of disease attributable to it. However, there are limitations to and outstanding questions about the accuracy and precision with which ground-level aerosol mass concentrations can be inferred from satellite remote sensing, including the level of global variation in the relationship between AOD and $PM_{2.5}$ at specific satellite overpass times during cloud-free conditions. Addressing these issues will require a systematic effort that includes measuring ground-level $PM_{2.5}$ in selected global regions to identify the factors that most affect the relationship between satellite-based and ground-level estimates.

INVESTIGATORS' REPORT

Assessment and Statistical Modeling of the Relationship Between Remotely Sensed Aerosol Optical Depth and $PM_{2.5}$ in the Eastern United States

Christopher J. Paciorek and Yang Liu

Department of Biostatistics, Harvard School of Public Health (C.P.); Department of Environmental and Occupational Health, Rollins School of Public Health, Emory University (Y.L.)

ABSTRACT

Research in scientific, public health, and policy disciplines relating to the environment increasingly makes use of high-dimensional remote sensing and the output of numerical models in conjunction with traditional observations. Given the public health and resultant public policy implications of the potential health effects of particulate matter (PM*) air pollution, specifically fine PM with an aerodynamic diameter $\leq$ 2.5 µm ($PM_{2.5}$), there has been substantial recent interest in the use of remote-sensing information, in particular aerosol optical depth (AOD) retrieved from satellites, to help characterize variability in ground-level $PM_{2.5}$ concentrations in space and time. While the United States and some other developed countries have extensive PM monitoring networks, gaps in data across space and time necessarily occur; the hope is that remote sensing can help fill these gaps. In this report, we are particularly interested in using remote-sensing data to inform estimates of spatial patterns in ambient $PM_{2.5}$ concentrations at monthly and longer time scales for use in epidemiologic analyses. However, we also analyzed daily data to better disentangle spatial and temporal relationships. For AOD to be helpful, it needs to add information beyond that available from the monitoring network. For analyses of chronic health effects, it needs to add information about the concentrations of long-term average $PM_{2.5}$; therefore, filling the spatial gaps is key. Much recent evidence has shown that AOD is correlated with $PM_{2.5}$ in the eastern United States, but the use of AOD in exposure analysis for epidemiologic work has been rare, in part because discrepancies necessarily exist between satellite-retrieved estimates of AOD, which is an atmospheric-column average, and ground-level $PM_{2.5}$.

In this report, we summarize the results of a number of empirical analyses and of the development of statistical models for the use of proxy information, in particular satellite AOD, in predicting $PM_{2.5}$ concentrations in the eastern United States. We analyzed the spatiotemporal structure of the relationship between $PM_{2.5}$ and AOD, first using simple correlations both before and after calibration based on meteorology, as well as large-scale spatial and temporal calibration to account for discrepancies between AOD and $PM_{2.5}$. We then used both raw and calibrated AOD retrievals in statistical models to predict $PM_{2.5}$ concentrations, accounting for AOD in two ways: primarily as a separate data source contributing a second likelihood to a Bayesian statistical model, as well as a data source on which we could directly regress.

Previous consideration of satellite AOD has largely focused on the National Aeronautics and Space Administration (NASA) moderate resolution imaging spectroradiometer (MODIS) and multiangle imaging spectroradiometer (MISR) instruments. One contribution of our work is more extensive consideration of AOD derived from the Geostationary Operational Environmental Satellite East Aerosol/Smoke Product (GOES GASP) AOD and its relationship with $PM_{2.5}$. In addition to empirically assessing the spatiotemporal relationship between GASP AOD

This Investigators' Report is one part of Health Effects Institute Research Report 167, which also includes a Commentary by the Health Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. Christopher J. Paciorek, Department of Statistics, 367 Evans Hall, University of California, Berkeley, CA 94720; email: *paciorek@stat.berkeley.edu*.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83234701 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

* A list of abbreviations and other terms appears at the end of the Investigators' Report.

and $PM_{2.5}$, we considered new statistical techniques to screen anomalous GOES reflectance measurements and account for background surface reflectance.

In our statistical work, we developed a new model structure that allowed for more flexible modeling of the proxy discrepancy than previous statistical efforts have had, with a computationally efficient implementation. We also suggested a diagnostic for assessing the scales of the spatial relationship between the proxy and the spatial process of interest (e.g., $PM_{2.5}$).

In brief, we had little success in improving predictions in our eastern–United States domain for use in epidemiologic applications. We found positive correlations of AOD with $PM_{2.5}$ over time, but less correlation for long-term averages over space, unless we used calibration that adjusted for large-scale discrepancy between AOD and $PM_{2.5}$ (see sections 3, 4, and 5). Statistical models that combined AOD, $PM_{2.5}$ observations, and land-use and meteorologic variables were highly predictive of $PM_{2.5}$ observations held out of the modeling, but AOD added little information beyond that provided by the other sources (see sections 5 and 6). When we used $PM_{2.5}$ data estimates from the Community Multiscale Air Quality model (CMAQ) as the proxy instead of using AOD, we similarly found little improvement in predicting held-out observations of $PM_{2.5}$, but when we regressed on CMAQ $PM_{2.5}$ estimates, the predictions improved moderately in some cases. These results appeared to be caused in part by the fact that large-scale spatial patterns in $PM_{2.5}$ could be predicted well by smoothing the monitor values, while small-scale spatial patterns in AOD appeared to weakly reflect the variation in $PM_{2.5}$ inferred from the observations. Using a statistical model that allowed for potential proxy discrepancy at both large and small spatial scales was an important component of our modeling. In particular, when our models did not include a component to account for small-scale discrepancy, predictive performance decreased substantially. Even long-term averages of MISR AOD, considered the best, albeit most sparse, of the AOD products, were only weakly correlated with measured $PM_{2.5}$ (see section 4). This might have been partly related to the fact that our analysis did not account for spatial variation in the vertical profile of the aerosol. Furthermore, we found evidence that some of the correlation between raw AOD and $PM_{2.5}$ might have been a function of surface brightness related to land use, rather than having been driven by the detection of aerosol in the AOD retrieval algorithms (see sections 4 and 7). Difficulties in estimating the background surface reflectance in the retrieval algorithms likely explain this finding.

With regard to GOES, we found moderate correlations of GASP AOD and $PM_{2.5}$. The higher correlations of monthly and yearly averages after calibration reflected primarily the improved large-scale correlation, a necessary result of the calibration procedure (see section 3). While the results of this study's GOES reflectance screening and surface reflection correction appeared sensible, correlations of our proposed reflectance-based proxy with $PM_{2.5}$ were no better than GASP AOD correlations with $PM_{2.5}$ (see section 7).

We had difficulty improving spatial prediction of monthly and yearly average $PM_{2.5}$ using AOD in the eastern United States, which we attribute to the spatial discrepancy between AOD and measured $PM_{2.5}$, particularly at smaller scales. This points to the importance of paying attention to the discrepancy structure of proxy information, both from remote-sensing and deterministic models. In particular, important statistical challenges arise in accounting for the discrepancy, given the difficulty in the face of sparse observations of distinguishing the discrepancy from the component of the proxy that is informative about the process of interest. Associations between adverse health outcomes and large-scale variation in $PM_{2.5}$ (e.g., across regions) may be confounded by unmeasured spatial variation in factors such as diet. Therefore, one important goal was to use AOD to improve predictions of $PM_{2.5}$ for use in epidemiologic analyses at small-to-moderate spatial scales (within urban areas and within regions). In addition, large-scale $PM_{2.5}$ variation is well estimated from the monitoring data, at least in the United States. We found little evidence that current AOD products are helpful for improving prediction at small-to-moderate scales in the eastern United States and believe more evidence for the reliability of AOD as a proxy at such scales is needed before making use of AOD for $PM_{2.5}$ prediction in epidemiologic contexts. While our results relied in part on relatively complicated statistical models, which may be sensitive to modeling assumptions, our exploratory correlation analyses (see sections 3 and 5) and relatively simple regression-style modeling of MISR AOD (see section 4) were consistent with the more complicated modeling results.

When assessing the usefulness of AOD in the context of studying chronic health effects, we believe efforts need to focus on disentangling the temporal from the spatial correlations of AOD and $PM_{2.5}$ and on understanding the spatial scale of correlation and of the discrepancy structure. While our results are discouraging, it is important to note that we attempted to make use of smaller-scale spatial variation in AOD to distinguish spatial variations of relatively small magnitude in long-term concentrations of ambient $PM_{2.5}$. Our efforts pushed the limits of current technology in a spatial domain with relatively low $PM_{2.5}$ levels and

limited spatial variability. AOD may hold more promise in areas with higher aerosol levels, as the AOD signal would be stronger there relative to the background surface reflectance. Furthermore, for developing countries with high aerosol levels, it is difficult to build statistical models based on $PM_{2.5}$ measurements and land-use covariates, so AOD may add more incremental information in those contexts. More generally, researchers in remote sensing are involved in ongoing efforts to improve AOD products and develop new approaches to using AOD, such as calibration with model-estimated vertical profiles and the use of speciation information in MISR AOD; these efforts warrant continued investigation of the usefulness of remotely sensed AOD for public health research.

1. INTRODUCTION

Research in scientific, public health, and policy disciplines relating to the environment increasingly makes use of high-dimensional remote sensing and numerical-model output in conjunction with traditional observations. The remote-sensing and model output help to fill in the inevitable gaps in space and time between observations. In some cases, remote sensing or model output might be accurate enough to be considered error free and to be treated as equivalent to the observations, but in many settings, the output is better considered as a noisy proxy for the process of interest. In this case, the output might be considered to be a combination of signal for the process of interest and noise, which is sometimes called discrepancy. In addition to the obvious scientific difficulties posed by the discrepancy, it poses related statistical difficulties. Primary among these is how to separate the signal from the noise, given sparse (and also often noisy) observations and complicated spatiotemporal dependence structures in both the discrepancy and in the signal portion of the proxy. Even assessing the amount and type of information in the proxy is difficult in the spatiotemporal context.

SATELLITE AOD AS A PROXY FOR $PM_{2.5}$

In this report we explore these issues in the specific context of using remote sensing of AOD as a proxy for ground-level $PM_{2.5}$. AOD is a measure of the light-extinction capability of ambient particles in the entire atmospheric column. We report a number of analyses of the spatiotemporal relationships between AOD and $PM_{2.5}$ and of statistical modeling to combine observations, land-use and meteorologic information, and proxy values to predict $PM_{2.5}$ in space and time in the eastern United States. Estimating $PM_{2.5}$ in space and time is important for public health and public policy because of the evidence from observational studies that long-term chronic exposure to $PM_{2.5}$ is associated with adverse health outcomes, including cardiac events (Pope et al. 2002; Laden et al. 2006; Miller et al. 2007; Puett et al. 2009). Of course, these critical studies are just one piece of our understanding of the relationship between $PM_{2.5}$ and adverse health outcomes. Observational studies of acute health effects, natural experiments, and research on animal toxicology add to our understanding. While the magnitude of any causal relationship is likely small, the fact that so much of the U.S. and global populations are exposed to the levels of particulates implicated in these studies makes the health effects of $PM_{2.5}$ in aggregate of major concern and the subject of governmental regulation.

In the United States, $PM_{2.5}$ is monitored regularly via a network of approximately 1000 monitors operated by the states and other governmental bodies in collaboration with the U.S. Environmental Protection Agency (U.S. EPA). Even this extensive effort has major spatiotemporal gaps, with many monitors reporting every third day rather than every day, and inevitable spatial gaps, particularly in rural areas. Some components of $PM_{2.5}$, such as black carbon, vary at fine spatial scales, so even in well-monitored urban areas monitoring does not adequately capture within-city variability in $PM_{2.5}$. Satellite AOD, if well correlated with ground-level $PM_{2.5}$, carries promise for filling some of these gaps, particularly in filling the spatial gaps that are consequential for analyses of chronic health effects. Numerical-model output is another possibility, though we focus on it much less in this report. In addition, if methods could be developed to make use of satellite AOD (or model output), they would be of particular importance in the large parts of the world with little monitoring.

A recent review provides an overview of aerosol remote sensing (Hoff and Christopher 2009). Our report makes use of AOD products from three satellite instruments. The MODIS and MISR instruments are aboard the Terra satellite platform and began operation in March 2000. Terra's polar orbit gives full coverage of the globe at regular intervals, with retrievals in the eastern United States at a constant daily time point (10:30–10:45 AM local time). MODIS provides AOD retrievals at a nominal spatial resolution of 10 km, with retrievals in the northeast United States every 1 to 2 days depending on location, although MISR is considered to be more precise because of its multiangle viewing ability (Liu et al. 2007a). However, MISR has lower nominal spatial resolution (17.6 km) and provides retrievals only every 4 to 7 days, depending on location, because of its narrow viewing angle. GASP AOD is a product from GOES, so it is available every half hour

during daytime, and provides nominal 4-km resolution. However, the instrument has limitations that reduce the accuracy of the AOD retrievals, including coarse spectral resolution, fixed viewing geometry, and lack of a near-infrared channel for surface characterization. AOD retrievals from satellites in general suffer from errors in measurement of true AOD, with these errors primarily related to cloud and surface contamination.

Another concern, beyond measurement error in remotely sensed AOD as a retrieval of the true AOD, is that AOD is at best a proxy for ground-level $PM_{2.5}$ because of fundamental differences in the definitions of the two quantities. By definition, AOD estimates the light-extinction capability of ambient particles in the entire atmospheric column, while $PM_{2.5}$ measures the dry mass concentration of ground-level particles. These definitions give rise to several key differences between the quantities. First, AOD is an integrated-column quantity, while public health interest lies in $PM_{2.5}$ concentrations at the surface. Second, AOD values are affected not only by the abundance of particles, but also by atmospheric humidity, which changes the sizes of hygroscopic particles such as sulfate, nitrate, and sea salt. AOD has a very robust relationship with the mass concentration of the particles if the vertical distribution of particles, meteorology, and particle size and composition are known (Chin et al. 2004), but in reality this is almost impossible to achieve in a large spatial domain. Third, AOD values retrieved by polar-orbiting sensors such as MISR and MODIS are snapshots of atmospheric conditions, while $PM_{2.5}$ in our analysis is a 24-hour average. Finally, AOD retrievals are made at the scale of satellite pixels and are areal by nature, which makes it difficult to compare that data with point-level $PM_{2.5}$ measurements.

The hope that satellite AOD can inform $PM_{2.5}$ estimation and its availability from NASA's MODIS and MISR instruments have led to numerous analyses of the relationship between AOD and $PM_{2.5}$ and of methods to adjust for differences between the two. A number of studies in recent years have found correlations between AOD and $PM_{2.5}$ at various scales in time, space, and space–time (Wang and Christopher 2003; Engel-Cox et al. 2004; Liu et al. 2005; Koelemeijer et al. 2006; van Donkelaar et al. 2006; Liu et al. 2007a; Pelletier et al. 2007). Liu and colleagues (2004a) proposed to adjust AOD based on the ratio of AOD to $PM_{2.5}$ as estimated by an atmospheric-chemistry model, and Liu and colleagues (2005) built a regression model to adjust AOD based on spatial and temporal patterns, relative humidity (RH), the planetary boundary layer (PBL) height, and other factors affecting the relationship between AOD and $PM_{2.5}$. More recently, van Donkelaar and colleagues (2010) estimated long-term average $PM_{2.5}$ globally using MODIS and MISR AOD calibrated following the approach of Liu and colleagues (2004a) combined with screening out values from a given instrument in regions where that instrument's AOD is poorly correlated with ground measurements. Hoff and Christopher (2009) concluded that there were still significant obstacles to making use of satellite AOD for air quality purposes, in particular the existing uncertainty in the AOD retrieval itself, high spatial variability in the AOD–$PM_{2.5}$ relationship (likely related to spatiotemporal variability in PM properties), cloud cover, the inability to estimate AOD over bright surfaces, the range of spatial and temporal scales for different air quality needs, and the mismatch of spatial and temporal scales of AOD and what is needed for health analysis. Our analysis considers some of these obstacles, analyzing cloud cover and GASP AOD (see section 3), developing a new approach to screen out cloud-induced errors and adjust for surface reflectance in GOES retrievals (see section 7), quantifying spatial variability in the AOD–$PM_{2.5}$ relationship through our empirical calibration (see sections 3, 4, and 5), and focusing throughout on the importance of spatial and temporal scales and the mismatch of the scales AOD and $PM_{2.5}$.

The known scientific reasons for the discrepancy between AOD and $PM_{2.5}$ and the aforementioned results in the literature motivated us to consider a number of important issues in the $PM_{2.5}$–AOD context that also arise in other environmental contexts in which combining information and making use of proxies is desirable. First, spatial and temporal scales are critical. To use AOD as a proxy for pollution monitoring and epidemiologic purposes, we need to know how well AOD (or calibrated versions of AOD) reflects $PM_{2.5}$ at various spatial scales so that we know the circumstances in which we can rely on it. Similarly, we need to know how the association between AOD and $PM_{2.5}$ changes with temporal scales. Analyses that report correlations reflecting both spatial and temporal associations make it difficult to know what aspects of variability in $PM_{2.5}$ are captured by AOD. For chronic health effects, we need to know about spatial correlation without the influence of short-term temporal variability in both AOD and $PM_{2.5}$. Temporal correlations at individual monitoring stations may tell us little about the spatial correlations of long-term averages. Finally, we argue that the true long-term average of $PM_{2.5}$ is an integral over a given time period. The fact that an AOD retrieval is missing may itself be informative about the $PM_{2.5}$ level because of the effect of meteorology on both missingness and $PM_{2.5}$. Therefore, associations of long-term average AOD and $PM_{2.5}$ where the $PM_{2.5}$ averages include only values that match AOD

retrievals, rather than all $PM_{2.5}$ values representative of the true long-term mean, are likely overly optimistic estimates of the associations.

Given the emphasis in this report on spatial scale, we briefly discuss our terminology here. When we refer to small-scale variation, we mean patterns and processes that vary over kilometers and tens of kilometers, e.g., reflecting differences within urban areas and between urban areas and surrounding rural areas. When we refer to large-scale variation, we mean variation at the scale of hundreds and thousands of kilometers, reflecting differences between large portions of the domain of interest, e.g., differences between large areas of Pennsylvania, when analyzing our focal mid-Atlantic region of the United States, or between regions of the United States. In cases in which we wish to distinguish scales more carefully, moderate-scale variability refers to differences at the scale of a few hundred kilometers, e.g., between-urban-area variations within regions of the country. Finally fine-scale variation refers to variations at the scale of meters to a few kilometers, reflecting within-neighborhood variations. Given that the monitoring data readily allow us to estimate large-scale variations (Yanosky et al. 2008, 2009), our hope was that AOD would help to inform estimation of small-scale patterns (albeit not at the scale of individual kilometers that are below the nominal resolution of the retrievals) and of moderate-scale patterns up to a few hundred kilometers.

STATISTICAL MODELING

Our general statistical perspective is to consider that there is a true underlying spatial surface (process) of $PM_{2.5}$ for any given temporal aggregation (instantaneous, hourly, daily, monthly, yearly, multiyear) and that observations are unbiased estimates of $PM_{2.5}$ for the time period captured by the measurement at a location, but not necessarily at nearby locations. The true surface is likely highly spatially variable because of the fine-scale effects of roads, point sources, buildings, etc. This spatial heterogeneity may cause observations to be very different from those of a local spatial average, even over tens or hundreds of meters. Our ultimate objective in exposure estimation is to estimate the true underlying surface, understanding that we can only hope for reasonably accurate estimation at some degree of spatial aggregation. In this report, we attempt to estimate this surface statistically and assess whether AOD can improve our estimation. Next, we briefly discuss the advantages and drawbacks of a statistical approach to the problem.

The fine-scale spatial heterogeneity of $PM_{2.5}$ makes it difficult to validate AOD as a proxy because of the spatial scale mismatch between an AOD pixel and $PM_{2.5}$ observations at a specific location. An imperfect match might reflect the spatial misalignment or the discrepancy between AOD and true $PM_{2.5}$ averaged over the pixel. This difficulty helped lead to our use of statistical modeling to complement exploratory analyses. While there is no gold standard for estimating areal average $PM_{2.5}$, our statistical modeling attempted to improve upon the use of the observations by adjusting for the effects of local sources and of preferential placement of monitors in areas with high $PM_{2.5}$.

A proxy such as AOD might be used simply as an explanatory variable in a regression model, but a common approach in the recent statistical literature, building on the initial work by Fuentes and Raftery (2005), has been to treat the proxy as a second data source that reflects the true underlying surface (of $PM_{2.5}$ in our case), but with additive and multiplicative discrepancies. Much of our statistical work concentrated on how to represent the additive discrepancy and on the need to account for potential small-scale correlation in the discrepancy. Modeling such correlation can be difficult computationally, and there are serious inferential issues in statistically decomposing the proxy into signal plus noise. Although such issues are complicated, we feel that considering them lies at the heart of a serious attempt to statistically combine information sources when the proxy is likely to have a complex discrepancy structure. Furthermore, for a proxy like AOD that can have missing values, some sort of modeling is required regardless of whether AOD is considered as an outcome or as an explanatory variable. At the same time, complicated statistical models contain their own dangers. It can be difficult to understand how they translate data into inference and prediction, and results may not be robust to model misspecification in the face of model assumptions that are difficult to check and the complexity of real-world data. To the extent that these concerns are salient, simple use of a proxy as a regressor, while potentially less informative and difficult when the proxy has many missing values, may be safer than more complicated modeling approaches such as those presented here and in the statistical literature.

In our modeling, we built on work by Yanosky and colleagues (2008 and 2009) and Paciorek and colleagues (2009), and to improve the prediction of $PM_{2.5}$ concentrations, we made extensive use of covariates representing land use, topography, emissions sources, and meteorology, as well as of spatial smoothing, to capture moderate- and large-scale variation. These efforts relied on the use of publicly available data in a geographic information system (GIS), albeit at some computational expense. Our perspective is that given the discrepancy concerns related to AOD, we need to see that it improves prediction over and above what is possible in its absence when we make use of these other sources of information. Statistical modeling satisfied

this goal, while more straightforward correlation calculations did not. Of course, unlike in the United States, such rich information is not available in most countries, but in countries without sufficient available $PM_{2.5}$ data, it is also difficult to assess and account for the discrepancy between AOD and $PM_{2.5}$.

OVERVIEW OF THE REPORT

Section 2 gives the specific aims of our work based on the initial proposal, with revisions to the aims developed in light of initial results after two years. At that point we decided to focus more on statistical-modeling techniques to combine proxies and observations and to develop methods to calibrate raw reflectance output from GOES as a new proxy for $PM_{2.5}$. In section 3 we report on the associations between GASP AOD and $PM_{2.5}$ at various spatial and temporal scales, and we report on a calibration approach (Paciorek et al. 2008) that builds on Liu and colleagues (2005) to account for the discrepancy between GASP AOD and $PM_{2.5}$. Section 4 considers the association between long-term average MISR AOD and $PM_{2.5}$, in hopes that long-term averaging of what is considered to be the most accurate satellite AOD retrieval would provide a good proxy for chronic $PM_{2.5}$ exposure. Section 5 presents our basic statistical model and the results of combining MODIS and GASP AOD and observations in conjunction with land-use and meteorologic measurements (Paciorek and Liu 2009). Section 6 describes how we extended the methodology of section 5. We address a key shortcoming in the model's representation of the proxy discrepancy term, propose a model-based diagnostic of proxy discrepancy at different spatial scales, and assess model performance in simulations. This section also describes how we applied this methodology using MODIS AOD as the proxy, as well as using $PM_{2.5}$ concentrations estimated by the CMAQ atmospheric-chemistry model as the proxy. Section 7 shows how we developed and assessed a new proxy for $PM_{2.5}$ based on GOES reflectance measurements, given weak-to-moderate correlations of GASP AOD with $PM_{2.5}$ and the lack of improvement in $PM_{2.5}$ predictions when GASP AOD was used (section 5). In section 8, we present general conclusions from our analyses, discussing the reasons for the discrepancy between AOD and $PM_{2.5}$ and new approaches under investigation for improving AOD as a proxy for $PM_{2.5}$. Our statistical conclusions focus on the importance of spatial and temporal scales, concerns about measurement error in epidemiologic analysis, and the need for robust statistical-modeling approaches that account for complicated proxy discrepancy when information sources are combined.

2. SPECIFIC AIMS

INITIAL SPECIFIC AIMS

Advances in spatial modeling and GIS technology, combined with the availability and demonstrated utility of satellite proxy data for $PM_{2.5}$ estimation, present the opportunity for integrated estimation of $PM_{2.5}$ for use in health effects analyses. Bayesian statistical techniques provide a natural framework for the integration.

In our revised proposal, we proposed using a combination of satellite and U.S. EPA monitoring data to estimate $PM_{2.5}$ concentrations, with the following specific aims:

1. Development of Bayesian statistical models to integrate monitoring, satellite, and GIS data to estimate monthly ambient $PM_{2.5}$ concentrations at high spatial resolution. These techniques were to account for important challenges such as satellite cloud cover, irregular temporal data sampling by satellites and ground monitors, and bias in satellite AOD as a proxy for $PM_{2.5}$.
2. Estimation of monthly $PM_{2.5}$ across the eastern United States for the period 2000 through 2006 at a fine spatial resolution (10 km $\times$ 10 km or finer). These estimates were to combine available $PM_{2.5}$ monitoring data and satellite measurements of AOD from the MODIS and MISR instruments. The estimated concentrations were to be made available publicly through HEI for potential use in a variety of analyses of exposures and health effects.
3. Understanding the temporal and spatial heterogeneities in $PM_{2.5}$ and our ability to characterize them based on satellite, monitoring, and GIS data. In particular, our aim was to quantify temporal and spatial variability in $PM_{2.5}$ and to explore the possibility of using spatial patterns at one point in time as a proxy for spatial patterns at other points in time.
4. Using a measurement-error framework, to quantify the reduction in exposure uncertainty and the potential reduction in health effects uncertainty provided by the use of both satellite and monitoring data, and comparing that exposure uncertainty with the uncertainty associated with using only monitoring data.
5. Assessing the promise of two additional sources of information to improve exposure estimates for 1995 through 1999, by carrying out a pilot study for the year 2001. MODIS and MISR data are not available for the years before 2000, and monitoring data are limited, so our aim was to assess the potential of GOES data to estimate $PM_{2.5}$. We also aimed to assess the

utility of adjusting the satellite estimates of $PM_{2.5}$ using vertical-profile information from the CMAQ atmospheric-chemistry model. If successful, these pilot investigations could form the basis for future proposals, leveraging the work done for this proposal.

REVISED SPECIFIC AIMS FOR THE THIRD GRANT YEAR

1. Development of a statistical framework for accounting for systematic proxy discrepancy. We wanted to develop a general statistical framework to be described in this report, with the AOD–$PM_{2.5}$ situation as a key example.
2. Analysis of GOES raw reflectance data and the development of methods to screen and calibrate reflectance as a proxy for $PM_{2.5}$. We aimed to directly use GOES channel 1 reflectance data, adjusting for background surface reflectance using statistical methods.
3. Assessment of the degree to which CMAQ vertical-profile information can be used to improve calibration of AOD to $PM_{2.5}$.

PROGRESS ON SPECIFIC AIMS

This report summarizes the results of our research. Here, we briefly summarize our work in the context of progress on the specific aims. With regard to initial aim 1 and revised aim 1, we developed Bayesian methods to integrate monitoring, satellite, and GIS and meteorologic information, in particular proposing a new approach to accounting for proxy discrepancy (sections 5 and 6). Given the lack of improvement in predicting $PM_{2.5}$ concentrations when AOD information was incorporated, we did not attempt to achieve initial aims 2 and 4. With regard to initial aim 3, through exploratory data analysis and the statistical-integration models, we considered the relationship between AOD and $PM_{2.5}$ in space and time, at various scales. However, we did not directly consider the use of spatial patterns at one point in time as a proxy for spatial patterns at other points in time. For initial aim 5, we carried out extensive assessment of GOES-based GASP AOD and included GASP AOD in our statistical-integration models. With regard to revised aim 2, we developed a new screening procedure for GOES reflectance measurements and a new method for estimating background surface reflectance. We also considered the use of corrected reflectance as a proxy for $PM_{2.5}$. Finally, with regard to revised aim 3 (also part of initial aim 5) — because of time constraints and feedback from HEI's Research Committee — we did not explore the use of CMAQ vertical-profile information to calibrate AOD, but we made extensive use of CMAQ-estimated $PM_{2.5}$ in our statistical discrepancy modeling work, assessing the direct use of CMAQ $PM_{2.5}$ as a proxy for $PM_{2.5}$ in the eastern United States during 2001.

3. SPATIOTEMPORAL ASSOCIATIONS BETWEEN GOES AOD RETRIEVALS AND GROUND-LEVEL $PM_{2.5}$

INTRODUCTION

Epidemiologic studies have provided evidence that chronic exposure to PM is related to increased mortality, as well as outcomes such as ischemic heart disease, dysrhythmias, heart failure, cardiac arrest, and lung cancer (Dockery et al. 1993; Pope et al. 2002, 2004; Miller et al. 2007). Studies of the chronic health effects of $PM_{2.5}$ have relied on the spatial heterogeneity of $PM_{2.5}$ concentrations to estimate the effects. A combination of spatial modeling and land-use regression can improve estimation of concentrations by the use of covariate information to help estimate concentrations at locations far from monitors and at scales smaller than can be achieved through smoothing of the monitoring data (Yanosky et al. 2008). However, efforts suffer from the sparse spatial representation in the monitoring network. Evidence of the health effects of acute exposure to $PM_{2.5}$ (e.g., Samet et al. 2000; Dominici et al. 2006) has relied on temporal heterogeneity in $PM_{2.5}$, but the every-third- (or every-sixth-) day schedule at many monitors reduces statistical power in such time series studies.

Remote sensing holds promise for adding information for exposure estimation, particularly spatial information in areas — primarily suburban and rural — far from monitors and temporal information on days without monitoring. Satellite-derived AOD has been correlated with ground level $PM_{2.5}$ (Wang and Christopher 2003; Engel-Cox et al. 2004; Liu et al. 2005; Liu et al. 2007a; Pelletier et al. 2007), specifically with particles with diameters ranging from 0.05 to 2 μm (Kahn et al. 1998), which is roughly the definition of $PM_{2.5}$ (particles with aerodynamic diameter $\leq$ 2.5 μm), the size fraction on which current EPA regulatory efforts focus. These correlations occur despite the mismatch in vertical detail between total-column aerosol, as estimated by AOD, and ground-level $PM_{2.5}$, the level of interest for health effects studies. One approach to help account for the mismatch has been calibration via a regression model based on season, spatial location, and meteorologic information (Liu et al. 2005).

Efforts to use AOD as a proxy for $PM_{2.5}$ have concentrated on the MISR and MODIS instruments. These are on polar-orbiting satellites and sample individual locations in

the eastern United States via a single snapshot either every 4 to 7 days (MISR) or every 1 to 2 days (MODIS). This infrequent sampling combined with a high proportion of missing retrievals because of cloud cover has resulted in sparse coverage in space and time, which hinders the use of the MODIS and MISR instruments in health effects studies. Geostationary satellites provide much more complete data; the GASP AOD provides observations every 30 minutes on a nominal 4-km grid, which provides finer nominal resolution than the MODIS or MISR grids. However, the GASP AOD retrievals are less precise than those from the polar-orbiting instruments because of the coarse spectral resolution and fixed viewing geometry of the sensor (Prados et al. 2007). Since GASP AOD retrievals use only the visible channel, 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. Despite these limitations, GASP AOD retrievals have been reasonably well correlated with Aerosol Robotic Network (AERONET) ground measurements of total-column aerosol (Knapp et al. 2002; Prados et al. 2007), although the retrievals of MODIS and especially MISR are more precise (Knapp et al. 2002). To date, no studies have been done on the relationship between GASP AOD and ground-level $PM_{2.5}$.

In this part of the study we assessed the potential of GASP AOD to act as a proxy for daily, monthly, and yearly ground-level $PM_{2.5}$, focusing on these time scales because of their relevance for epidemiologic studies. First, we assessed the basic strength of the association between GASP AOD and $PM_{2.5}$ in space and time. We built flexible regression-style models to calibrate daily AOD to $PM_{2.5}$ based on meteorologic, spatial, and temporal effects. We compared the association of calibrated AOD with daily, monthly, and yearly average $PM_{2.5}$. Our goal was to understand the association of GASP AOD with $PM_{2.5}$ and to show how to calibrate GASP AOD to increase its utility, not to physically interpret our statistical modeling of AOD. Finally, we assessed whether the absence of an AOD retrieval was associated with the $PM_{2.5}$ concentration, to determine if bias was induced by ignoring the missingness and simply using the available retrievals.

Note that much of this section of our study was published earlier (Paciorek et al. 2008).

DATA AND METHODS

Data

We made use of GASP AOD from GOES-12 (East) imager data for the year 2004, provided by the U.S. National Oceanic and Atmospheric Administration (NOAA). Prados and colleagues (2007) described the GOES-12 imager data and GASP AOD algorithm in detail. In brief, AOD was calculated from a single visible channel (520–720 nm) based on assumptions about surface reflectivity and atmospheric and aerosol properties, while the cloud mask was determined from infrared channels 2 (3.9 μm) and 4 (10.7 μm) and the visible channel. The pixel centroids of the AOD retrievals are nominally on a 4-km grid, but the distance between centroids is not generally 4 km. Retrievals were attempted every half hour, but cloud cover and high surface reflectivity led to many missing observations.

GASP AOD (henceforth referred to as AOD in this section) retrievals were available during daylight, from 10:45 through 23:45 Coordinated Universal Time; i.e., Greenwich Mean Time (UTC). In our core analysis, we followed NOAA's criteria for screening valid AOD observations, which are described in Appendix B (available on the HEI Web site). That appendix provides sensitivity analyses suggesting we could relax some of these criteria. Negative retrievals occurred due to errors in the estimation of surface reflectivity when AOD was low. Unlike Prados and colleagues (2007), we made use of negative retrievals in the hope that they would indicate low AOD. Appendix B shows that including these retrievals provided useful information.

To assess the relationship between $PM_{2.5}$ and AOD, we matched monitoring data from the U.S. EPA Air Quality System (AQS) to the nearest GOES pixel, omitting a small number of monitors for which the nearest pixel centroid was closer to another monitor. Since we used AOD as the dependent variable in our regression modeling, this step avoided duplicate AOD values. We then selected days for which the U.S. EPA monitor reported a $PM_{2.5}$ concentration. Our interest was in fine-resolution estimation of $PM_{2.5}$, so unlike other analyses we used individual pixels instead of aggregating AOD across adjoining pixels. We calculated a daily estimate of AOD as the simple average of the available retrievals, but considered another approach, described in Appendix B. We had 99,159 matched daily observations, of which 46,684 had at least one valid AOD retrieval during the day of the observation. Table 1 shows summary statistics for the matched AOD retrievals and $PM_{2.5}$ observations for days with at least 1 valid AOD retrieval.

We used $PM_{2.5}$ data with U.S. EPA parameter 88101. This excluded most $PM_{2.5}$ data from Interagency Monitoring of Protected Visual Environments (IMPROVE) sites, which are generally in very rural areas. This approach avoided issues of comparability of the AQS and IMPROVE observations and focused our calibration efforts on data from populated areas. We included all observations regardless of any quality flags in the data record, at the suggestion of U.S. EPA personnel (Michael Papp, 2006, personal communication) who indicated that all data reported

Table 1. Summary Statistics for GASP AOD Retrievals (Unitless) and $PM_{2.5}$ Observations (μg/m^3) for 2004 by Day and Location for Days with at Least One Valid AOD Retrieval[a]

Geographic Region / Time Period	Sample Size	Median	(5%, 95%)
GASP AOD			
Contiguous U.S.			
All seasons	56,385	0.13	(−0.06, 0.50)
Winter	7,709	0.15	(0.01, 0.44)
Spring	15,719	0.17	(−0.03, 0.52)
Summer	19,099	0.13	(−0.07, 0.55)
Fall	13,858	0.07	(−0.08, 0.40)
Eastern U.S.[b]			
All seasons	46,684	0.13	(−0.06, 0.51)
Winter	6,541	0.15	(0.01, 0.43)
Spring	13,361	0.18	(−0.03, 0.52)
Summer	15,454	0.13	(−0.08, 0.57)
Fall	11,328	0.08	(−0.08, 0.41)
$PM_{2.5}$ Observations			
Contiguous U.S.			
All seasons	56,385	10.6	(4.0, 27.4)
Winter	7,709	10.7	(4.2, 25.1)
Spring	15,719	10.4	(4.2, 22.5)
Summer	19,099	11.6	(4.2, 30.1)
Fall	13,858	9.7	(3.6, 27.7)
Eastern U.S.[b]			
All seasons	46,684	11.1	(4.3, 27.3)
Winter	6,541	10.4	(4.5, 21.5)
Spring	13,361	10.7	(4.4, 22.4)
Summer	15,454	12.9	(4.7, 31.0)
Fall	11,328	9.8	(3.8, 27.2)

[a] Excludes 11 matched pairs for which the $PM_{2.5}$ observations have concentrations greater than 100 μg/m^3.

[b] Locations east of 100° W longitude.

to the AQS should be valid data. For simplicity, we used only data with the parameter occurrence code equal to 1, thereby including only the primary monitor at a site. The AQS data are the primary data used for estimating exposure in epidemiologic studies, so we considered them as the gold standard in this study while acknowledging that the ground measurements were not error free (instrument error variance was approximately 1.5 based on colocated monitors, relative to a trimmed variance of the measurements of 60). We also made use of local land-use and monitoring objective information that the U.S. EPA provides for many of the monitors.

For meteorologic information, we concentrated on the PBL (i.e., mixing height) and RH as the key variables that affect the relationship between $PM_{2.5}$ and AOD (Liu et al. 2005). The PBL was 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. A 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 RH, resulting in greatly increased light-extinction efficiency. Since $PM_{2.5}$ is measured as dry particle mass (measured at a controlled RH of approximately 40%), we expected higher RH to be associated with a larger ratio of AOD to $PM_{2.5}$. We used the North American Regional Reanalysis (NARR) meteorologic fields; the NARR assimilates available data

with a state-of-the-art meteorologic model to estimate meteorologic parameters every 3 hours on a 32-km grid covering North America. Rogers and colleagues (2006) reported that the NARR PBL is highly correlated with light detection and ranging (LIDAR) measurements, although in urban areas the correlation decreases. We used data from 12:00, 15:00, 18:00, and 21:00 UTC to match the time range of the AOD retrievals, and we took the inverse-squared distance-weighted averages of the values from the four grid points closest to each EPA monitor as simple estimates of the already-smooth model-output fields at the monitor sites. See Appendix A for more details on the data sources and manipulations.

In principle, since AOD is at the half-hourly resolution, we might have calibrated it to hourly $PM_{2.5}$ measurements from the AQS. However, there are far fewer hourly monitors than daily monitors, limiting our ability to calibrate AOD to hourly $PM_{2.5}$, and there is no Federal Reference Method (FRM) for hourly $PM_{2.5}$. The relationship of hourly $PM_{2.5}$ (averaged for the day) to daily FRM measurements can vary by location and season, as was evident in the AQS data, which suggested the need to calibrate hourly $PM_{2.5}$ to daily $PM_{2.5}$. The need to calibrate occurs in part because of the loss of semivolatile compounds in $PM_{2.5}$ that is observed to occur in continuous $PM_{2.5}$ measurements. For simplicity and because our interest was in relationships between $PM_{2.5}$ and AOD at time scales longer than hourly (note that the U.S. EPA air quality standard is a 24-hour average), we restricted our analysis to daily associations based on data from daily monitors using the FRM and daily average AOD.

Associations of AOD and $PM_{2.5}$ were weak in the western United States (see Spatiotemporal Associations Between Daily AOD and $PM_{2.5}$, under Analyses), so most of our results are reported for the area east of 100° W.

Statistical Calibration Model

We built a statistical regression model to understand the factors that modify the relationship between AOD and $PM_{2.5}$. Using our model, we calculated a calibrated AOD variable that was more strongly associated with $PM_{2.5}$. Ultimately, as part of our larger research effort, the calibrated AOD would be used in a statistical prediction model for $PM_{2.5}$.

Liu and colleagues (2005) treated $PM_{2.5}$ as the dependent (response) variable in their regression models, using the logarithmic transformation of both AOD and $PM_{2.5}$ to create an additive model on the log scale. Here, we considered log AOD as the dependent variable and regressed on $PM_{2.5}$ and other factors, treating AOD as observed data. We took this approach in part 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 issues are difficult to account for if AOD is considered to be the independent variable. Using the less-error-prone variable as the independent variable also helped avoid bias (toward 0) in the estimated regression coefficient. Furthermore, our goal was to understand AOD as a proxy for $PM_{2.5}$. Using $PM_{2.5}$ as the dependent variable would have placed the focus on including other variables that help to explain $PM_{2.5}$, which is the focus of a large body of environmental health research, rather than on the assessment of the association of $PM_{2.5}$ and AOD. In our models, the $PM_{2.5}$ values observed at the monitors stood in for true $PM_{2.5}$, ignoring any monitor instrument error.

We considered a variety of models, building on Liu and colleagues (2005) but using smooth regression functions (Wood 2006) in place of linear functions and indicator variables for region and season. After a model-comparison process, described in Appendix B, we arrived at the final model, in which daily average AOD for location i and day of year t, $\bar{a}_{it}$, is distributed:

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

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), the PBL, and the RH, respectively. β is a multiplicative bias coefficient that scales from units of $PM_{2.5}$ to unitless AOD, PM_{it} is the matched $PM_{2.5}$ measurement, and τ^2 is the error variance. We used the simple average of the available AOD retrievals for each day, $\bar{a}_{it}$, but, as described in Appendix B (available on the HEI Web site), we also considered a more sophisticated approach. Since there were negative retrievals, we added 0.6 to each daily mean and then log transformed the resulting values (the minimum value was −0.5, so adding 0.6 avoided a long left tail after transformation). Using $\bar{a}_{it}$ in place of $\log \bar{a}_{it}$ gave similar results, but the residuals were slightly more skewed. Fitting separate models of the form of equation 1 for each season, we estimated β to be 0.0018 (95% confidence interval [CI] 0.0010–0.0027) in winter, an order of magnitude smaller than 0.0164 (CI, 0.0157–0.0170) in spring, 0.0164 (CI, 0.0158–0.0169) in summer, and 0.0129 (CI, 0.0123–0.0134) in fall. Because of the near-0 coefficient for winter, we chose to fit models only for spring, summer, and fall. We fit separate versions of the model (equation 1) for

each season to facilitate computations with such a large dataset and to allow the relationships to vary by season.

The model (equation 1) can be fit in the statistical software R, using the gam() function, designed for fitting generalized additive models (Wood 2006). The software uses penalized splines to parameterize the smooth functions of time, space, and covariates, allowing for nonlinear relationships. Multiple penalty terms were estimated from the data using extensions of generalized cross-validation (GCV) (Wood 2006), thereby ensuring that the functions were sufficiently smooth to avoid overfitting and provide generalizability. However, based on the model comparisons described in Appendix B, we restricted the flexibility of the function of time, which gave fitted functions, $\hat{f}_t(t)$, with 3 to 4 degrees of freedom for each season.

Having fit the model and estimated the smooth functions, we created a calibrated AOD variable, a^*_{it}, by subtracting the values of all the fitted functions from the observed value, $\log \bar{a}_{it}$, except for $PM_{2.5}$:

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

Our hope was that by adjusting for factors that modified the relationship of AOD and $PM_{2.5}$, the calibrated AOD (which we note was on a different scale than the raw AOD) would be more strongly associated with $PM_{2.5}$ than would raw AOD and would have a reasonably linear relationship with $PM_{2.5}$. Linearity was preserved when we averaged to longer time scales, as in

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

This may produce more robust proxy estimates of $PM_{2.5}$ that average over short-term fluctuations. We could then use time-averaged a^*_{it} as one of a collection of covariates to develop a statistical prediction model with $PM_{2.5}$ as the dependent variable (e.g., Yanosky et al. 2008), where the scaling represented by β_0 and β_1 could be estimated within the prediction model.

To address potential overfitting in our regression model, 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. 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 gave us cross-validated values of calibrated AOD for the nine sets that we could correlate with the held-out $PM_{2.5}$ observations.

ANALYSES

Retrieval Availability

To compare the density of GASP AOD data with that of MODIS and MISR AOD, we examined the retrievals for each of the three instruments for all AQS sites in the eastern United States and for each month calculated the proportion of days with at least one valid retrieval for each instrument; we used the sites as a rough reflection of the population distribution. We matched each of the AQS sites to the nearest GOES pixel and, on a day-to-day basis, to pixels within a nominal pixel radius of 12.4 km for MISR and 7.1 km for MODIS. On average, MISR provided a valid retrieval only 5% of the days in a given month at a given location; MODIS, 11%. In contrast, GOES provided a valid retrieval on average 39% of the days (21% of the days if only days with at least 5 retrievals were considered), including winter data but assuming that none of the winter retrievals were valid because of the lack of association between GASP AOD and $PM_{2.5}$ during that season (see Calibration Model Results, below). If only nonwinter data were considered, 52% of days had a valid GOES retrieval (28% when requiring at least 5 retrievals in a day) compared to 12% for MODIS and 5% for MISR.

Spatiotemporal Associations Between Daily AOD and $PM_{2.5}$

Figure 1 plots raw correlations calculated over time for each pixel–monitor match. The correlations were much higher in the eastern than in the western United States, as has been found for both MODIS (Engel-Cox et al. 2004) and MISR (Liu et al. 2005) retrievals, and as was expected based on higher surface reflectivity in the (less-vegetated) western United States. Also, with the exception of California, a higher proportion of the aerosol in the West is in the free troposphere, with less local anthropogenic pollution than in the East (Chin et al. 2007). Correlations varied by season, with lower correlations (and few retrievals) during winter and with summer, spring, and fall correlations similar to the all-season results (see Appendix B, available on the HEI Web site). There were no clear and substantial relationships between the correlations and local information about each AQS site, such as land-use type, population density, monitoring objective, or local emissions. The results were robust with respect to various thresholds for 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 were lower when fewer retrievals were required for a day.

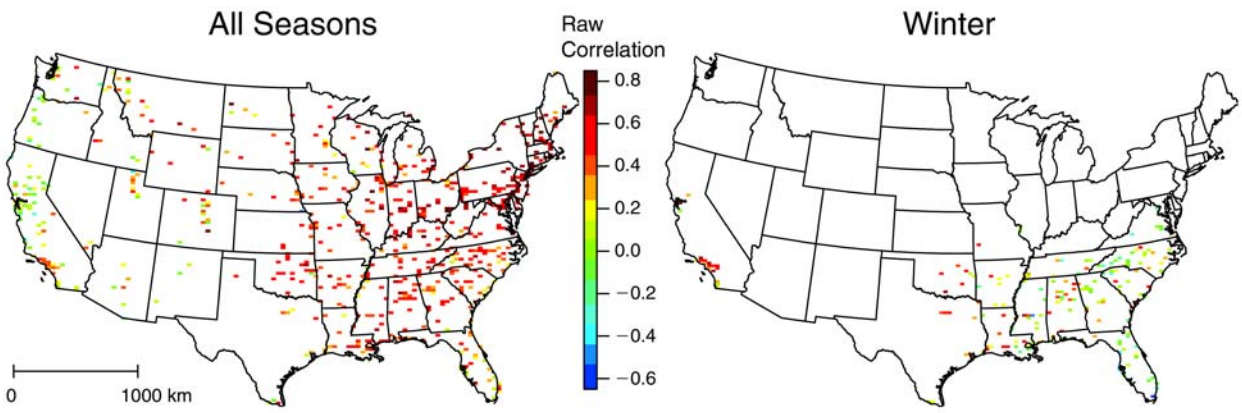

Figure 1. Temporal correlations at individual sites of daily average $PM_{2.5}$ concentrations with the average of half-hourly GASP AOD retrievals for all seasons (left map) and winter only (right map) in 2004. Plots are based on site–days with at least 3 AOD retrievals, and only locations with at least 10 days of matched pairs are shown.

Because of the low correlations in the western United States, we restricted our analyses to locations east of 100° W, which includes most of the counties violating the U.S. EPA air quality standard for $PM_{2.5}$ with the exception of those in California.

Correlations over space (Figure 2) were less strong than those over time, suggesting that AOD can better distinguish high from low $PM_{2.5}$ over time at fixed locations than over space at fixed times. This may have been related to spatially varying factors such as average reflectivity, aerosol type, or local meteorology that obscure, and potentially confound, the relationship between AOD and $PM_{2.5}$. The results were again robust with respect to various thresholds. Our calibration work (see the following subsection) suggests that the relationship between AOD and $PM_{2.5}$ varied by location, which helps to explain the low cross-sectional correlations seen here; after calibration based on location, the associations improved. The correlations again tended to be lower in winter; however, this cannot be explained by the minimal differences in the variability of $PM_{2.5}$ and AOD between winter and the other seasons. Another possibility is that changes in reflectance with vegetation loss in winter may not be accurately captured by the retrieval algorithm.

Calibration Model Results

Here we focus on key results with respect to the importance of calibration and the relationships between AOD (either raw or calibrated) and $PM_{2.5}$ at different temporal resolutions. We calculated correlations at the daily scale as well as after averaging across available matched pairs within a month and within a year for each pixel–monitor match. As correlations were calculated only for days for which both the AOD proxy and $PM_{2.5}$ data were available, we overstate the predictive ability of AOD for true monthly and yearly average $PM_{2.5}$ concentrations, since there were many days with no AOD retrievals. Our correlation results measured the ability of the AOD proxies to mirror heterogeneity in $PM_{2.5}$ over space and time.

On the daily scale, model-calibrated AOD (equation 2) was more strongly correlated with $PM_{2.5}$ than was the raw daily log average, $\log \bar{a}_{it}$ (Table 2). More importantly, without calibration, we could not average over time and achieve more robust relationships; we discuss this surprising result below. Requiring a minimum number of

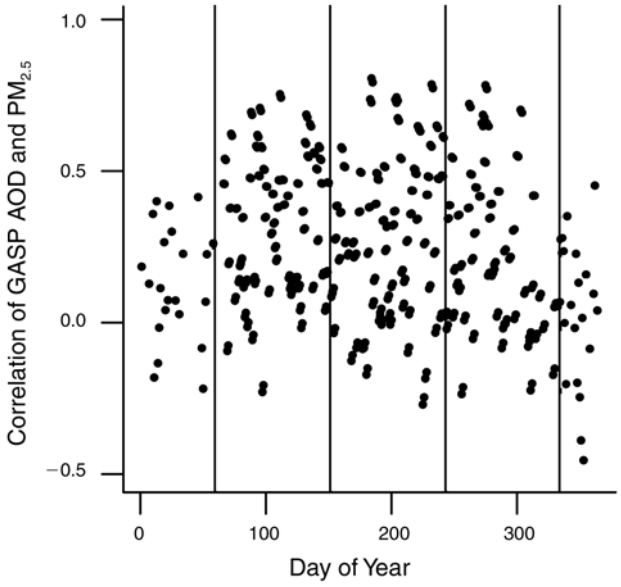

Figure 2. Spatial correlations of GASP AOD with $PM_{2.5}$ concentrations by day of year in the eastern United States in 2004. Each correlation requires at least 3 retrievals per site–day. Only days with at least 30 such sites are included. Vertical lines divide the seasons.

AOD retrievals in a day improved associations over the shorter time periods. However, over the yearly period, reducing the number of days with matched pairs — as compared with using all days with at least 1 retrieval — resulted in year-long averages that were less robust and a decrease in correlations. Therefore, we suggest that analyses that average to monthly or yearly resolution should include all available AOD retrievals. Comparisons using the held-out tenth set indicated that the model-selection process described in Appendix B resulted in little overfitting.

The reduction in correlations between raw AOD and $PM_{2.5}$ when data were averaged over time mirrors the fact that correlations over time, holding space fixed, tended to be stronger than correlations over space, holding time fixed (see the previous subsection, Spatiotemporal Associations Between Daily AOD and $PM_{2.5}$ Concentrations). The within-site relationships between AOD and $PM_{2.5}$ were positive, but across sites and most noticeably at the yearly resolution, AOD was only weakly associated with $PM_{2.5}$. The most likely explanation is that there were spatially varying confounders that obscured the long-term average relationship between $PM_{2.5}$ and AOD, driving long-term average AOD down where long-term $PM_{2.5}$ was high and vice versa. This might have been related to local effects at the locations of monitors, some of which are purposely sited near local sources, which could increase the $PM_{2.5}$ point measurement relative to the AOD value at the grid cell. While the source of this confounding is unclear, its effect was marked. The calibration, which was primarily driven by the spatial term (see Appendix B), was 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 when added to a $PM_{2.5}$ prediction model that relies on large-scale spatial smoothing of the monitoring data plus information from GIS and meteorologic covariates.

In Appendix B, we compare the results with alternative models considered in our model-selection process, demonstrating that the model has good relative performance.

Figure 3 shows the fitted smooth regression functions of time, the PBL, the RH, and space for each of the three seasons. We interpret these functional relationships conditional on $PM_{2.5}$ being in the model. As expected, for a given concentration of $PM_{2.5}$, AOD increased with increasing PBL, since a higher PBL meant the AOD retrieval was integrating over a longer column of air with reasonably homogeneous $PM_{2.5}$ concentrations. AOD increased with increasing RH (when the RH was greater than 60%–70%). The particle-growth effect of humidity increases the light-extinction capability of the particles, thereby increasing AOD relative to ground-level $PM_{2.5}$, as the latter is measured as dry mass. This nonlinearity might have occurred because the growth effect of hygroscopic particles becomes more substantial with increasing RH (Chin et al. 2002). The wiggliness in the regression functions for the PBL and RH likely reflects a moderate amount of overfitting from not fully accounting for within-site correlation (equation 1). The spatial patterns indicate that, holding $PM_{2.5}$ constant, the values for AOD were low for the Ohio River valley and the Appalachian Mountains region. The lower-than-expected values for AOD might have occurred because large local emissions from power plants in the region increased the ratio of ground-level $PM_{2.5}$ to AOD. The spatial patterns might also have been caused by additional meteorologic factors, variability in aerosol type, and differences in satellite viewing angle.

Table 2. Correlations of Both Raw and Calibrated GASP AOD with $PM_{2.5}$ at Different Temporal Resolutions, Excluding Winter, for 2004 in the Eastern U.S.

Temporal Resolution of Correlations	Raw AOD[a] $\log \bar{a}_{it}$	Calibrated AOD[a] $\left(a^*_{it}\right)$
Any Number of AOD Retrievals in a Day		
Daily	0.41	0.50
Monthly averages (at least 3 matched days for each site–month)	0.34	0.62
Yearly averages (at least 10 matched days for each site)	0.17	0.75
At Least 5 AOD Retrievals Each Day		
Daily	0.51	0.59
Monthly averages (at least 3 matched days for each site–month)	0.41	0.67
Yearly averages (at least 10 matched days for each site)	0.19	0.69

[a] The two AOD proxies are (1) raw AOD, calculated using the log average daily AOD, and (2) calibrated AOD.

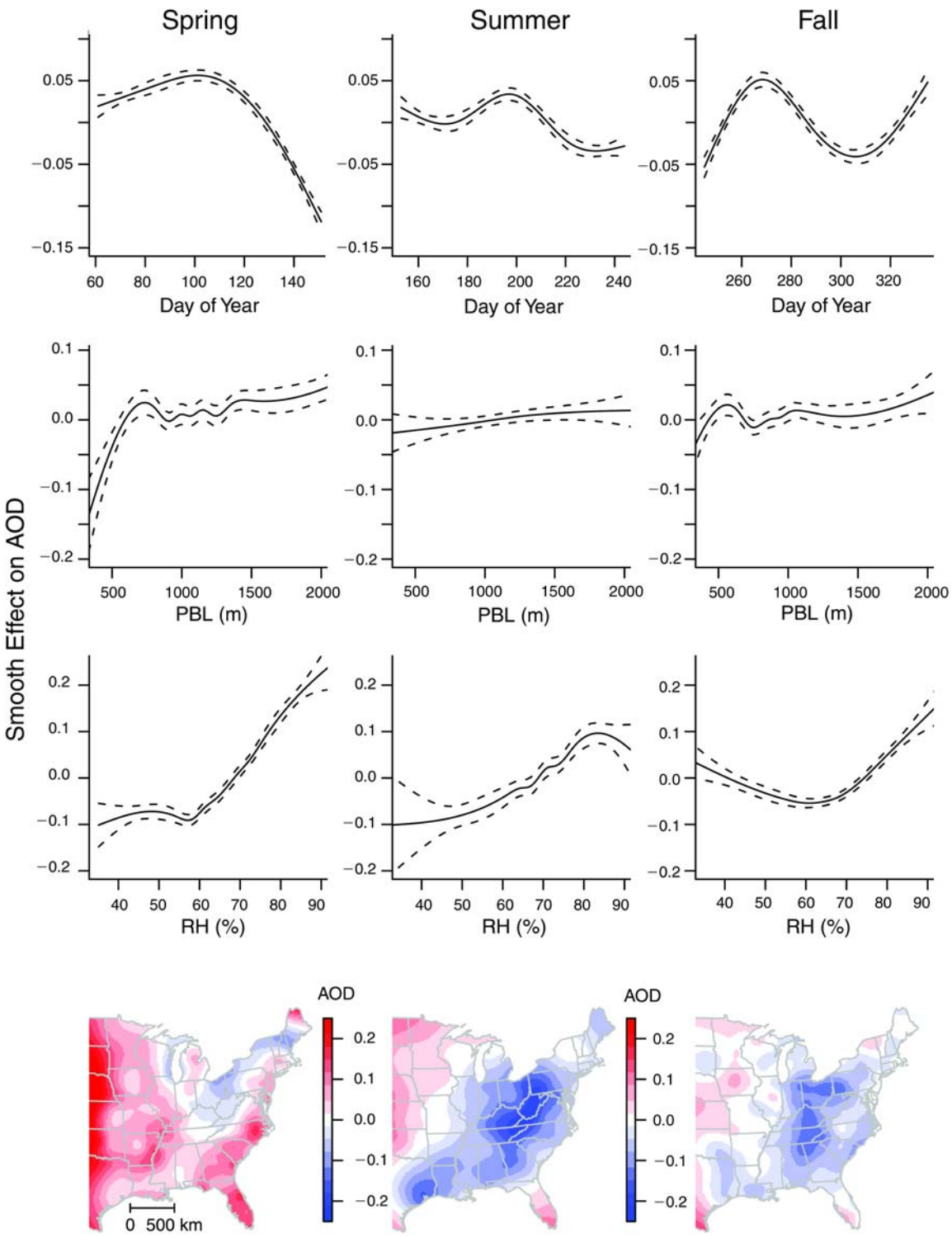

Figure 3. Fitted smooth regression relationships for the effects on AOD of time, $\hat{f}_t(t)$ (top row); the PBL, $\hat{f}_{PBL}(PBL)$ (second row); the RH, $\hat{f}_{RH}(RH)$ (third row); and space, $\hat{g}(s)$ (bottom row), by season for the eastern United States in 2004 based on the statistical model (equation 1). Larger values indicate that AOD was high relative to $PM_{2.5}$ for that value of the regression variable. In the spatial plots, red indicates that AOD was high relative to $PM_{2.5}$, and blue the converse. On the figures in the first three rows, dashed lines indicate pointwise 95% CIs.

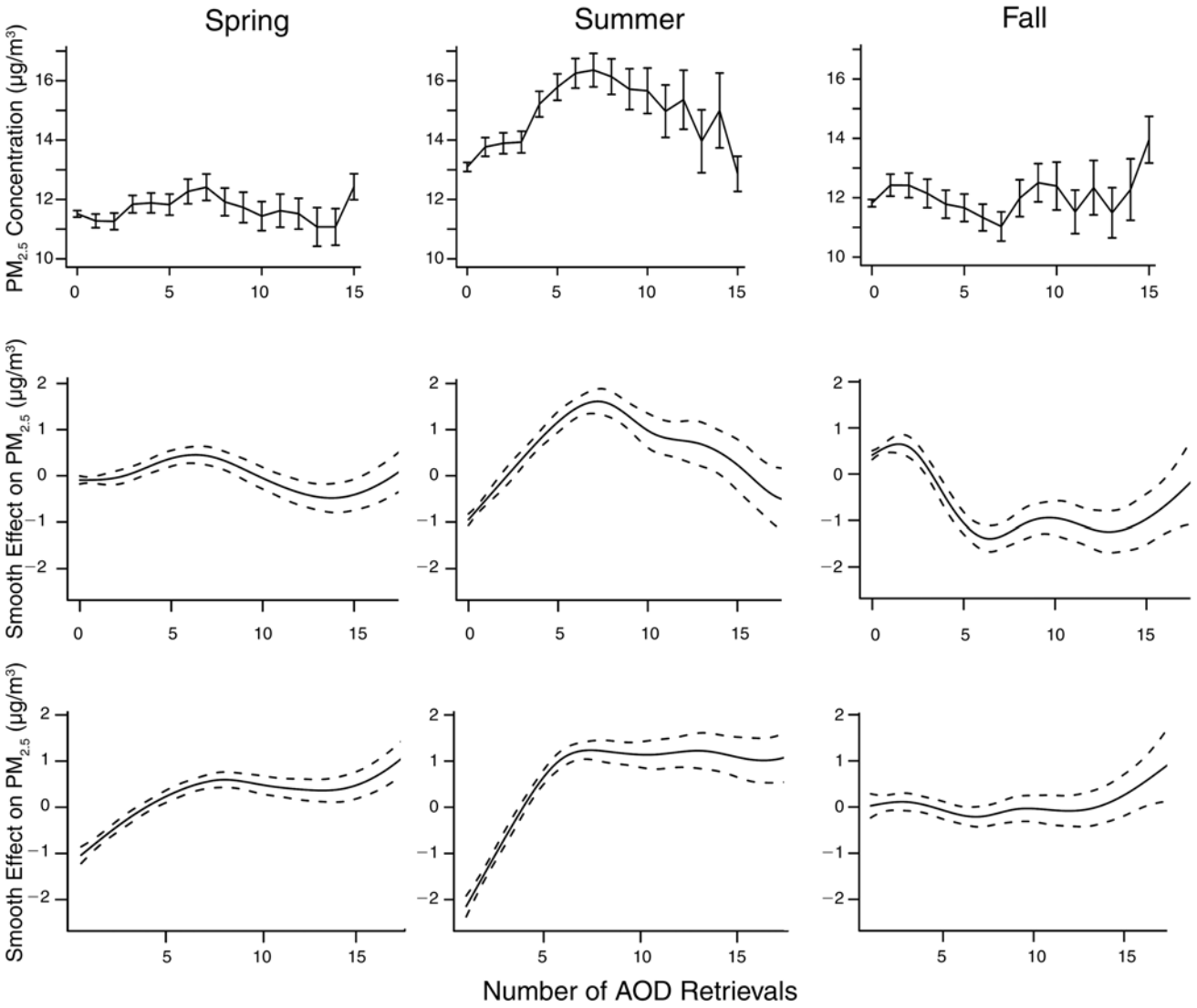

Figure 4. Relationship of the number of AOD retrievals in a day with 24-hour average $PM_{2.5}$ concentrations in the eastern United States for 2004. The top row shows the average $PM_{2.5}$ concentration in µg/m³ (with 95% CI) as a function of the number of GASP AOD retrievals by season. Data at the x-axis value of 15 are for 15 or more retrievals. The middle row shows the smooth regression relationship between the number of AOD retrievals and $PM_{2.5}$ concentration by season, controlling for location, time, and meteorology (equation 4). The bottom row shows the smooth regression relationship between the number of AOD retrievals and $PM_{2.5}$ by season for those days when at least 1 retrieval was made, controlling for location, time, meteorology, and average AOD (in a model of the form of equation 4), but with a smooth term of average daily AOD also included.

Note that the fact that the RH and the PBL varied in both space and time helped avoid concurvity and allowed us to separate the effects of the RH and PBL from the effects of space and time.

Association Between $PM_{2.5}$ and AOD Availability

Missingness of AOD retrievals might be associated with meteorologic conditions that are also associated with pollution levels. Missing retrievals are primarily caused by cloud cover, which itself is a function of meteorologic conditions that may be correlated with $PM_{2.5}$ levels. If days with few or no AOD retrievals have systematically lower or higher average $PM_{2.5}$ than days with many retrievals, then missingness of AOD is itself informative about $PM_{2.5}$. If so, then using only available AOD retrievals could bias predictions of $PM_{2.5}$. We investigated this by analyzing the distribution of $PM_{2.5}$ as a function of the proportion of missing AOD retrievals.

The top row of Figure 4 shows the mean $PM_{2.5}$ by season as a function of the number of AOD retrievals. In summer, there was a marked difference in $PM_{2.5}$ concentrations as a function of AOD retrievals, with the highest concentrations

on days with approximately 5 to 10 retrievals and lower concentrations for days with few retrievals. Spring showed a somewhat similar, but less marked pattern, while fall showed little systematic pattern.

Since $PM_{2.5}$ varies in space and time, as does missingness, the association between missingness and $PM_{2.5}$ might have occurred merely because both missingness and $PM_{2.5}$ are separately associated with location and time. We attempted to control for space, time, and meteorology by fitting the following generalized additive model separately for the spring, summer, and fall seasons:

$$\log \mathrm{PM}_{it} \sim N(\mu + g(s_i) + f_t(t) + f_{\mathrm{PBL}}(\mathrm{PBL}_{it}) + f_{\mathrm{RH}}(\mathrm{RH}_{it}) + f_n(n_{\mathrm{AOD},it}), \tau^2), \quad (4)$$

where PM_{it} is the $PM_{2.5}$ concentration, $n_{\mathrm{AOD},it}$ is the number of AOD retrievals for location i and day t, and the remaining notation follows that of equation 1. In the middle row of Figure 4 we see the fitted smooth regression function, $\hat{f}_n(n_{\mathrm{AOD}})$, for each of the three seasons, indicating a nonlinear relationship between number of retrievals and $PM_{2.5}$ concentrations, with $PM_{2.5}$ 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 $PM_{2.5}$, there was still a relationship between missingness and $PM_{2.5}$.

For those days with at least 1 retrieval, we could also control for the AOD magnitude. Adding a smooth function of average AOD, $f_a(\bar{a}_{it})$, to model 4 did not remove the association between missingness and $PM_{2.5}$ (Figure 4, bottom row), although it did change the relationships somewhat, with spring and particularly summer showing increases in $PM_{2.5}$ with increasing number of retrievals, and then $PM_{2.5}$ leveling off with a larger number of retrievals. Fall showed little relationship between $PM_{2.5}$ and number of retrievals after accounting for the observed AOD.

These results suggest that predictive modeling of $PM_{2.5}$ based on AOD should take the number of retrievals in a day into account, as it provides additional information about $PM_{2.5}$ concentrations. In particular, not accounting for missingness during the summer is likely to upwardly bias one's estimates of $PM_{2.5}$ because days with few or no AOD retrievals on average have low $PM_{2.5}$ concentrations.

DISCUSSION

We report the first comparison of GASP AOD with ground-level $PM_{2.5}$, building upon the expanding literature comparing MODIS and MISR AOD with $PM_{2.5}$. We built calibration models that resulted in moderately strong correlations of calibrated AOD with $PM_{2.5}$ in the eastern United States during the spring, summer, and fall. Correlations increased with averaging over longer time periods when we used the calibrated AOD. We also point out that the occurrence of missing AOD retrievals was not random with respect to $PM_{2.5}$ concentrations, as has also been reported for MODIS AOD (Koelemeijer et al. 2006).

The higher nominal spatial resolution of GASP AOD compared to MODIS and MISR AOD did not appear to provide real improvement in spatial resolution, presumably because of instrument differences. The daily correlations of GASP AOD with $PM_{2.5}$ found in this study were lower than those found by Liu and colleagues (2005 and 2007a), who averaged over multiple MODIS and MISR pixels. However, given the limitations of the GOES instrument, the fact that the GASP AOD correlations were not too much lower than those for MISR and MODIS AOD indicates the promise of GASP AOD as a proxy for $PM_{2.5}$.

Critically important is that the comparisons of daily matched $PM_{2.5}$ and AOD data do 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 a day and for more robust estimates of daily pollution by averaging over noisiness in the retrievals and over temporal variability in pollution during the day. To estimate monthly average $PM_{2.5}$ concentrations, the greater data density should result in proxy values that are more representative of $PM_{2.5}$ over all the days in the month than would data from MODIS and MISR, which we assess quantitatively in later sections of this report. In Appendix B (available on the HEI Web site), we provide evidence that some of the criteria we used to select valid GASP AOD retrievals might be relaxed to provide even more retrievals.

Of perhaps equal importance as its dense temporal coverage, GASP AOD provides the possibility of a long-term record starting in November 1994, when the GOES-8 satellite retrievals were first available. Dense $PM_{2.5}$ ground monitoring only began in 1999, so GASP AOD provides one of the few proxies for $PM_{2.5}$ for the period of 1995 through 1998. For epidemiologic work, four years of improvements in exposure estimation could increase statistical power to detect health effects.

Note that in the period since the work with GASP AOD described here was completed and published (Paciorek et al. 2008), the GOES team has developed several improvements to the GASP AOD retrievals, including correction of an error in the azimuth angle definition, an improved method for estimating surface reflectance, improved calculation of the AOD standard deviation, and inclusion of scattering angle in the screening criteria. Because of time limitations and the availability of only 3 months of the new GASP AOD retrievals, we did not rerun our analyses based on the newer data.

4. CALIBRATION AND ASSESSMENT OF MISR AOD AS A PROXY FOR LONG-TERM AVERAGE $PM_{2.5}$

INTRODUCTION

Given the poor performance of MODIS and GASP AOD in helping to predict monthly and yearly average $PM_{2.5}$ concentrations at smaller spatial scales (see sections 5 and 6) and our inability to consider MISR AOD in the statistical modeling when analyzing only a single year, because of the limited number of retrievals, we consider here the relationship between long-term average $PM_{2.5}$ and MISR AOD over the eastern United States. Another reason for doing this is that MISR AOD is considered to be of higher quality than MODIS or GASP AOD because the MISR is a multispectral instrument with narrow bandwidth, and it makes use of multiple viewing angles.

In previous work, Dr. Liu and colleagues derived a regression model for predicting $PM_{2.5}$ concentrations from MISR AOD that included a number of regression terms relating to meteorologic variables known to affect the relationship of $PM_{2.5}$ and AOD (the PBL and RH) as well as spatial and temporal variability (Liu et al. 2005). Liu and colleagues (2004a) derived an AOD proxy based on MISR AOD, corrected for vertical-profile information (at coarse spatial resolution) based on an atmospheric-chemistry model, and van Donkelaar and colleagues (2010) extended the approach to derive a multiyear average AOD proxy based on both MISR and MODIS global AOD retrievals.

In this study, we considered several regression approaches for calibrating $PM_{2.5}$ and AOD and then assessed the correlation between the resulting AOD proxies and $PM_{2.5}$ concentrations over the eastern United States for the period 2000 through 2006.

DATA AND METHODS

We considered the period 2000 through 2006 and the area of the eastern United States east of 100° W longitude; we omitted the western United States because of the poor correlations of AOD and $PM_{2.5}$ in this region found in other analyses (Engel-Cox et al. 2004; Liu et al. 2005; section 3 of this report, but note van Donkelaar et al. 2010). For $PM_{2.5}$ observations, as in other analyses, we used data from the AQS and IMPROVE networks, considering all monitors with a relatively complete record during the seven years (at least 300 daily values and no gaps longer than 60 days), which gave us 503 monitors in the eastern United States. We related all data sources to a common 4-km grid, which is described in Appendix A and used in the remainder of this report.

We downloaded meteorologic data products from the NARR for the seven-year period, extracting the 15:00 and 18:00 UTC observations of the PBL and the RH. These times matched most closely the time when the satellite that hosted the MISR instrument passed over the eastern United States. Since the RH throughout the atmospheric column may be a more useful quantity for adjusting the raw AOD retrievals, we calculated the RH based on the NARR specific humidity (SH) values at pressure levels from 1000 hectopascals (hPa) to 700 hPa (in 25-hPa increments) and used the simple average of the RH at the NARR pressure levels, since pressure levels decline approximately linearly with altitude. Seven hundred hPa corresponds approximately to 3000 m, so this average approximated the average RH over the lowest 3000 m of the atmosphere, roughly corresponding to the PBL. Given the relatively low elevations in the eastern United States, we also ignored the variability in pressure at the surface, which meant that in regions with higher elevation, we made use of estimated SH values for below the earth's surface. Note that surface RH taken directly from the NARR and average RH calculated from SH, as above, showed only a moderate correlation of 0.42. The conversion from SH to RH was done using the NASA standard atmospheric structure, as coded in the GEOS-Chem atmospheric-chemistry model, based on the NARR vertical profiles of temperature and pressure:

$$\begin{aligned} ESAT &= 10^{23.5518-\frac{2937.4}{T}} \times T^{-4.9283} \\ SH_{mb} &= 1.6072 \times SH \times P \\ \mathrm{RH} &= 100\frac{SH_{mb}}{ESAT}, \end{aligned} \tag{5}$$

where P is pressure in millibars (mb), SH is specific humidity as a percentage, SH_{mb} is specific humidity in mb, T is temperature in Kelvin, and $ESAT$ is the saturation water vapor pressure in mb.

To assign the PBL and RH values to each cell of the 4-km grid for either 15:00 or 18:00 UTC, we calculated inverse distance-weighted (IDW) averages of the NARR values (which were provided on a 32-km grid) at the four NARR pixel centroids nearest to each 4-km-grid cell. In our final models, we used the PBL and RH values at 18:00 UTC because the correlations with AOD were somewhat higher for that hour than for 15:00 UTC (0.40 vs. 0.38 for SH-derived column-average RH, 0.13 vs. 0.08 for ground-level RH, and 0.24 vs. 0.04 for the PBL).

MISR instruments are aboard the Terra satellite platform, whose polar orbit has been providing full coverage of the globe at regular intervals, starting in March 2000, with retrievals in the eastern United States at a constant daily time point (10:30–10:45 AM local time). MISR AOD

retrievals are at a nominal spatial resolution of 17.6 km with retrievals in the eastern United States every 2 to 9 days depending on location (Liu et al. 2005). AOD cannot be retrieved below clouds, so cloud-filtering algorithms use the infrared portion of the spectrum to detect and omit obscured observations (Engel-Cox et al. 2004). Errors and uncertainties in the filtering can lead to erroneous AOD retrievals, and high surface reflectivity can also prevent retrievals. We screened the MISR AOD (version 22) retrievals and included only retrievals with AOD less than 1.5, those classified as successful retrievals under clear conditions, and those with heterogeneous surface conditions. We assigned the MISR AOD values to the cells of the 4-km grid based on the MISR pixel nearest to a given cell and avoided assigning values to grid cells outside a given satellite orbit swath. Finally, we associated each $PM_{2.5}$ observation with the MISR values assigned to the cell in the 4-km grid in which the monitor was located. Table 3 presents summary statistics for daily and long-term average MISR AOD retrievals and $PM_{2.5}$ observations.

Koelemeijer and colleagues (2006) considered the calibration

$$\mathrm{AOD}^{*} = \frac{\mathrm{AOD}}{\mathrm{PBL}\, f\,(\mathrm{RH})}\ , \qquad (6)$$

which on the log scale is equivalent to log AOD* = log AOD − log PBL − log f(RH), where the function $f(\bullet)$ is derived from measurements involving the increase of the aerosol-extinction cross section with the RH. Liu and colleagues (2005) developed a prediction model for $PM_{2.5}$ on the east coast of the United States, based on MISR AOD and meteorologic variables that likely modified the relationship, regressing log $PM_{2.5}$ on log AOD, log PBL, the RH, and categorical variables for the distance from the coast (more or less than 100 km), site location (rural, urban, or suburban), season, and region (New England, mid-Atlantic, or south Atlantic).

Based on this previous work, we considered the following calibrations of individual MISR AOD retrievals with 24-hour average $PM_{2.5}$ observations matched as described above, where models 1, 2, and 3 are variants on the Liu and colleagues (2005) approach and models 4 and 5 on the relationship in Koelemeijer and colleagues (2006). Note further that log transformations of $PM_{2.5}$ and AOD gave us models that more closely satisfied standard regression assumptions.

1. Model 1: We used log $PM_{2.5}$ as the outcome and regressed on log AOD, the RH, dummy variables for the four seasons, and a smooth regression term for log PBL. The R^2 for this model was 0.30, and the regression coefficient for log AOD was 0.47.
2. Model 2: We fit a model similar to model 1, but we added a smooth term for calendar day and a smooth spatial surface, to attempt to adjust for temporal and spatial variation in the relationship between $PM_{2.5}$ and AOD. The R^2 for this model was 0.45, and the regression coefficient for log AOD was 0.45.
3. Model 3: We fit a model as in model 2, but we excluded log AOD to assess how much of the predictive power of this approach came merely from associations of $PM_{2.5}$ with meteorology, time, and space. The R^2 for this model was 0.26.
4. Model 4: We modeled log AOD as the outcome, regressing on log $PM_{2.5}$, the SH-derived RH, dummy variables for the four seasons, and a smooth regression term for log PBL. The R^2 for this model was 0.54, and the regression coefficient for log $PM_{2.5}$ was 0.48.
5. Model 5: We fit a model similar to model 4, but we added a smooth term for calendar day and a smooth spatial surface, to attempt to adjust for temporal and spatial variation in the relationship between $PM_{2.5}$ and AOD. The R^2 for this model was 0.59, and the regression coefficient for log $PM_{2.5}$ was 0.51.

In contrast to the relationship in Koelemeijer and colleagues (2006) (equation 6), we considered a smooth term of log PBL and a linear term of RH based on graphical assessment. For models 4 and 5, we used the SH-derived column-average RH because this value was much better correlated with AOD than surface RH (0.40 vs. 0.13). However, for models 1, 2, and 3, we used surface RH because $PM_{2.5}$ was the outcome variable.

Given that the outcome for models 1, 2, and 3 was $PM_{2.5}$ concentration, our proxies for predicting $PM_{2.5}$ from these models were simply predictions from the fitted model at the locations of our 503 monitors for all days over the

Table 3. Summary Statistics for MISR AOD Retrievals (Unitless) and $PM_{2.5}$ Observations (μg/m^3) for 2000–2006 in the Eastern U.S.[a] for Both Daily Averages (Matched by Day and Location) and Long-Term Averages (Matched by Location)

Data Source	Sample Size	Median	(5%, 95%)
Daily Values			
MISR AOD	35,338	0.10	(0.03, 0.37)
$PM_{2.5}$	35,338	11.0	(4.0, 27.3)
Long-Term Averages			
MISR AOD	73,238	0.13	(0.08, 0.22)
$PM_{2.5}$	503	13.5	(9.1, 16.7)

[a] Locations east of 100° W longitude.

seven years, based on meteorology at the monitor sites on those days. For models 4 and 5, our goal, as in section 3, was to fit a model with an approximately linear relationship between log AOD and log $PM_{2.5}$ and subtract the influence of the other variables from log AOD to create a proxy variable that we hoped would be well correlated with and linearly related to log $PM_{2.5}$. The fitted model 4 was

$$\begin{aligned}\log \mathrm{AOD}_{it} &= \hat{\beta}_0 + \hat{\beta}_{\mathrm{PM}} \log \mathrm{PM}_{it} \\ &\quad + \hat{\gamma}_{\mathrm{season}(t)} + \hat{\beta}_{\mathrm{RH}} \mathrm{RH}_{it}^{\mathrm{column}} \\ &\quad + \hat{f}_{\mathrm{PBL}}(\log \mathrm{PBL}_{it}) + e_{it},\end{aligned} \tag{7}$$

where AOD_{it}, PM_{it}, $\mathrm{RH}_{it}^{\mathrm{column}}$, and PBL_{it} are the AOD, PM, RH, and PBL values for location i and day t, and e_{it} is the error term. $\hat{\gamma}_{\mathrm{season}(t)}$ represents the season dummy terms, and $\hat{f}_{\mathrm{PBL}}$ is the fitted smooth regression term for log PBL.

Based on this fitted model, our calibrated proxy was

$$\begin{aligned}a_{it}^{*} &= \log \mathrm{AOD}_{it} \\ &\quad - \left(\hat{\gamma}_{\mathrm{season}(t)} + \hat{\beta}_{\mathrm{RH}} \mathrm{RH}_{it}^{\mathrm{column}} + \hat{f}_{\mathrm{PBL}}(\log \mathrm{PBL}_{it})\right).\end{aligned} \tag{8}$$

The calibrated proxy based on model 5 also subtracted additional spatial and temporal components, $\hat{g}(s_i) + \hat{f}_t(t)$.

The daily proxy values for each calibration procedure were exponentiated to return to the original scale and then averaged for each location over the seven years.

Exploratory analysis suggested that long-term average MISR AOD correlated with land use, so we calculated the proportions of land-cover classes in the cells of the 4-km grids based on the National Land Cover Database (see Appendix A for more details). We then assessed the relationships among MISR AOD, $PM_{2.5}$, and land cover using regression models run on the $PM_{2.5}$, AOD, and land-cover values assigned to the cells of the 4-km grids, omitting cells with fewer than 50 AOD retrievals. We focused on forested areas, defined by summing the proportion of land in each cell classified as deciduous, evergreen, and mixed forest and woody wetland.

RESULTS

The correlation of the 503 long-term averages with average raw MISR AOD, for AOD averages based on at least 20 retrievals, was 0.19 (95% CI, 0.11–0.28); it was 0.26 (CI, 0.18–0.35) when at least 50 retrievals were required. We next assessed the degree to which matching in time mattered, by considering only monitors for which there were at least 20 AOD retrievals matched by day with $PM_{2.5}$ observations: The correlation was 0.32 (CI, 0.23–0.41), compared with 0.25 (CI, 0.15–0.34) using the same set of locations but without matching by day. This suggests that the lack of temporal matching over time did decrease the utility of AOD information somewhat, since our goal was to predict true long-term average $PM_{2.5}$ concentrations. Recall that we had screened our monitors so that the averages for the data from these monitors would be good estimates of true long-term average $PM_{2.5}$. We caution that the reported 95% CIs are appropriate with regard to each individual correlation coefficient (albeit not accounting for spatial correlation), but they are not particularly helpful for comparing correlation coefficients because the two statistics are not independent.

On average, 103 MISR AOD retrievals were associated with each of the 503 monitors, and the number of retrievals ranged from 8 to 189. When only the 458 monitors for which there were at least 50 AOD retrievals were included, the correlation of the model 1 proxy with $PM_{2.5}$ concentrations was 0.06 (95% CI, −0.03 to 0.15), whereas for the model 2 and 3 proxies, it was 0.73 (CI, 0.68–0.77). This suggests that simple calibration with meteorology was not successful. In contrast, including the spatial term greatly improved the correlation, but when AOD was left out of the model, the correlation was the same. This indicates that AOD did not add spatial information to the proxy and that the correlation was driven by simply estimating the smooth spatial variation in $PM_{2.5}$. This suggests that the drop in R^2 seen in the calibration models when AOD was removed was driven by the temporal, rather than spatial, association of AOD and $PM_{2.5}$. This comparison underscores the importance of understanding the marginal contribution of AOD to the prediction of $PM_{2.5}$ in such calibration approaches.

When we considered the models that used AOD as the outcome, we found that the correlation of the model 4 proxy with $PM_{2.5}$ was 0.23 (CI, 0.14–0.32), whereas the proxy for model 5, which included the spatial term in the calibration, had a correlation with $PM_{2.5}$ of 0.57 (CI, 0.51–0.63). Given the results for models 2 and 3, this suggested that in model 4 AOD in and of itself was of limited help in predicting $PM_{2.5}$ and that the improvement in the correlation came from the spatial information in the calibration. We would have hoped that the spatial surface merely corrected for the large-scale discrepancy between AOD and $PM_{2.5}$ and allowed smaller-scale information in AOD to be used to predict $PM_{2.5}$. In assessing the models further, we found that the spatial correlation of the spatial surface estimate fit in model 5, $\exp(\hat{g}(s_i))$, with $PM_{2.5}$ was −0.43 (CI, −0.49 to −0.35), which indicated that the spatial correction was itself correlated with $PM_{2.5}$, with larger values of $PM_{2.5}$ scaled down on average. This made it hard to assess the degree to which the information about $PM_{2.5}$ was

coming from AOD rather than from $PM_{2.5}$ being correlated spatially at large scales with the discrepancy between AOD and $PM_{2.5}$. Although we did not assess these MISR AOD proxies in the context of our statistical prediction modeling, in our work with MODIS and GASP AOD calibrated in this fashion (sections 5 and 6), the AOD proxies were found not to improve prediction of $PM_{2.5}$, despite correlations between proxy values and $PM_{2.5}$ of the magnitude seen here for the model 5 MISR AOD proxy.

Figure 5 shows maps of long-term average $PM_{2.5}$ and raw AOD. Given the low correlation between $PM_{2.5}$ and raw AOD, the lack of spatial alignment is not surprising. Figure 6 shows the full surface of raw AOD and indicates that the areas of low AOD aligned closely with forested areas, whereas higher AOD was associated with agricultural and developed areas. The correlation between the proportion of forest in a cell and long-term average AOD was −0.476 (CI, −0.479 to −0.473). Note that the areas that were omitted because they had fewer than 50 AOD retrievals (in the Appalachians and across areas near the Canadian border and Great Lakes) also tended to have low average AOD and to be forested. We investigated whether the association of land use and $PM_{2.5}$ drove the AOD–$PM_{2.5}$ relationship or whether some of the association of AOD and $PM_{2.5}$ was driven by surface characteristics that correlated with $PM_{2.5}$, rather than by the direct relationship of aerosol and $PM_{2.5}$. Based on the data from the cells containing the 503 monitors, we found a correlation of −0.23 (CI, −0.32 to −0.15) between the forested proportion of the cell and $PM_{2.5}$, a correlation of 0.26 (CI, 0.18–0.35) for AOD and $PM_{2.5}$ (as above), and a correlation of −0.45 (CI, −0.52 to −0.37) between the proportion forested and AOD. This suggests AOD was more strongly related to land cover (forest vs. nonforest) than to $PM_{2.5}$ and that the AOD–land cover relationship may have partially driven the AOD–$PM_{2.5}$ relationship, rather than the reverse. Regressions of log AOD on $PM_{2.5}$ and the proportion forested, both individually and jointly, supported this hypothesis. The R^2 was 0.23 when we regressed log AOD on the forest variable alone and 0.09 with a regression coefficient of 0.020 (CI, 0.014–0.026) when we regressed on $PM_{2.5}$ alone. However, when we regressed on both variables, the R^2 was 0.27 with a smaller regression coefficient for $PM_{2.5}$ of 0.013 (CI, 0.008–0.019), suggesting that a portion of the AOD–$PM_{2.5}$ relationship appeared to be driven by the effect of land cover on AOD. Fortuitously, both AOD and $PM_{2.5}$ were negatively correlated with forest cover: If the correlations had been in opposite directions, the association of surface reflectance and AOD could have weakened the relationship of AOD and $PM_{2.5}$. To the extent that AOD is merely standing in for land cover, we can directly account for land cover in statistical prediction models of $PM_{2.5}$. Of course, spatial misalignment between $PM_{2.5}$ point observations and the large MISR pixels could have driven down the AOD–$PM_{2.5}$ relationship relative to the AOD–land use relationship, but we note that Figure 5, left

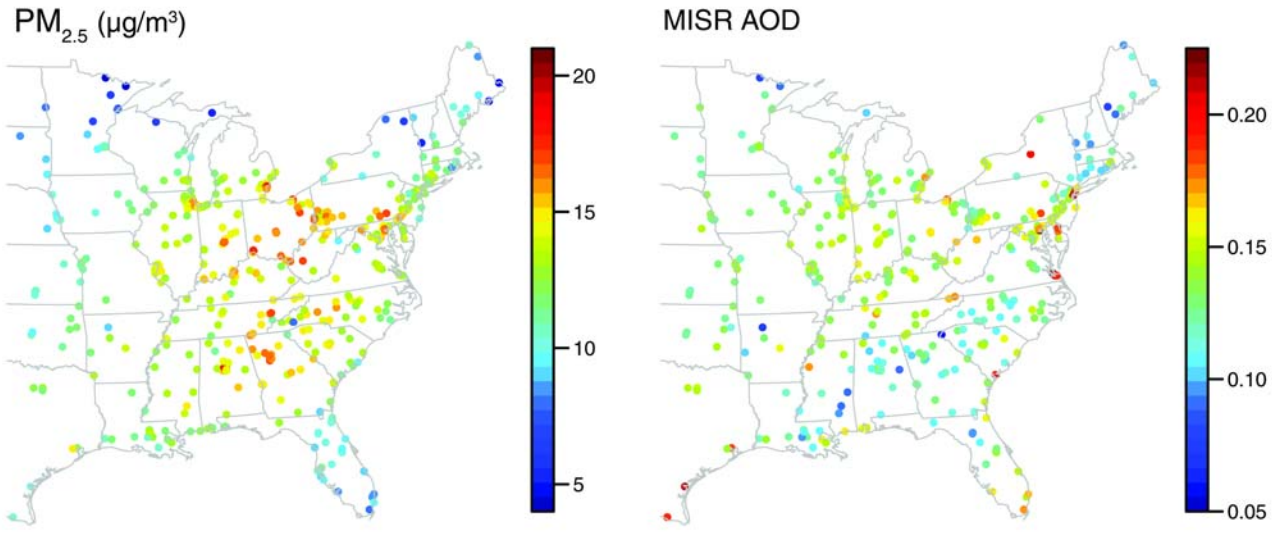

Figure 5. Long-term (2000–2006) average $PM_{2.5}$ at AQS and IMPROVE monitoring stations (µg/m³) (left map) and long-term average MISR AOD at the same stations (right map). AOD values based on fewer than 50 retrievals over the seven-year period are omitted. A small number of MISR AOD values are truncated to 0.225 to accentuate the heterogeneity of values in the right map.

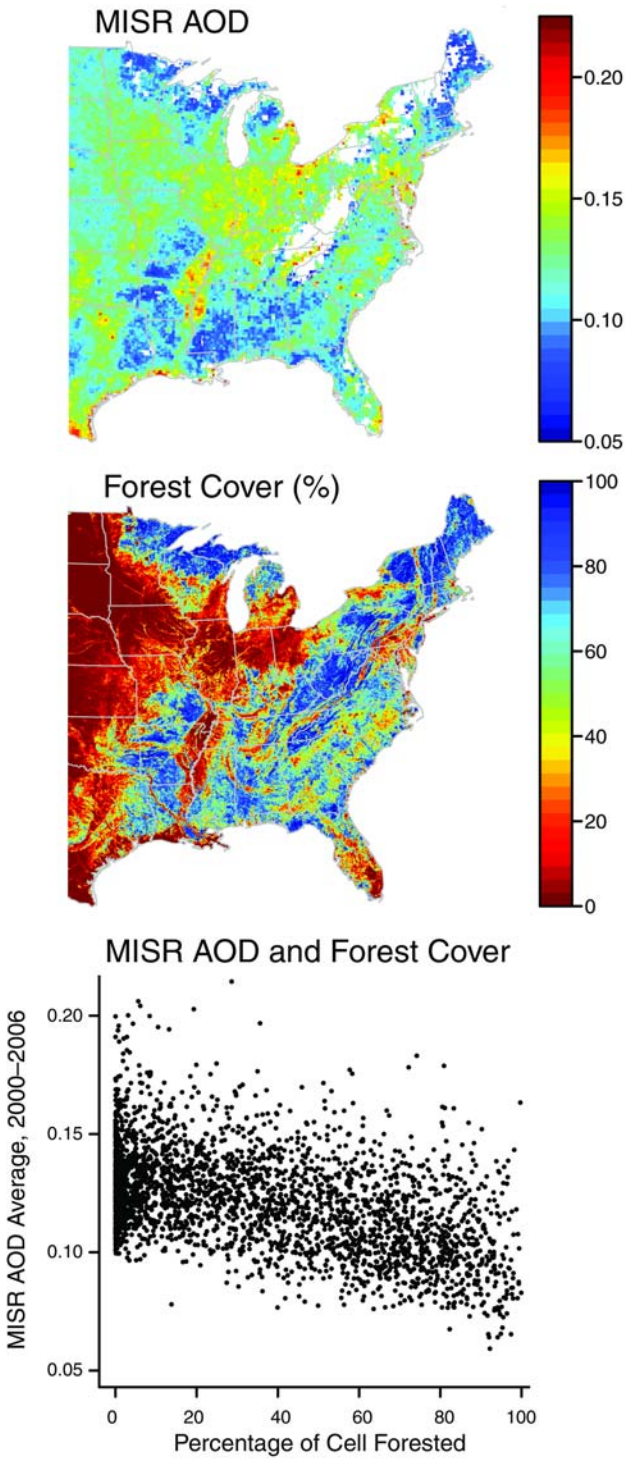

Figure 6. The relationship of forest cover and MISR AOD. The top map shows long-term (2000–2006) average MISR AOD at all locations. AOD values based on fewer than 50 retrievals over the seven-year period are omitted, and a small number of AOD values are truncated to the range 0.05 to 0.225. The middle map shows the percentage of forest cover in 4-km-grid cells. The graph shows AOD as a function of the percentage of forest cover for each cell of a random sample of 5000 grid cells. A small number of points with AOD values outside the range of 0.05 to 0.225 are not shown.

map, suggests that much of the long-term average $PM_{2.5}$ variability was at moderate-to-large scales. Further assessment of this issue would involve accounting for the effects of the vertical profile, the RH, and the PBL on the AOD–$PM_{2.5}$ relationship.

In terms of the geographic effects of this phenomenon, note that the band across the southern United States from Mississippi through Georgia to North Carolina where AOD (Figure 5, right map) was low relative to $PM_{2.5}$ (Figure 5, left map) corresponds to relatively forested areas (Figure 6, middle map). Several areas of high AOD shown in Figure 6, top map, that might not have corresponded to high $PM_{2.5}$ (Figure 5, left map) are Iowa, southern Minnesota, and along the Mississippi River, where Arkansas borders with Tennessee and Mississippi.

DISCUSSION

Our results suggest that even after calibration with meteorologic variables known to affect the AOD–$PM_{2.5}$ relationship (Liu et al. 2005), long-term average proxy values derived from MISR AOD were weakly related to $PM_{2.5}$. Spatial calibration, which played an empirical role in our analysis similar to the role of the atmospheric-chemistry model in Liu and colleagues (2004a) and van Donkelaar and colleagues (2010), was critical to improving correlations between AOD proxies and ground-level $PM_{2.5}$, as was seen also for GASP AOD (section 3). The atmospheric-chemistry model used in those studies had a resolution of 2×2.5 degrees, and in our work any calibration was necessarily limited by the resolution of the monitoring data. Thus, such approaches are of limited help in accounting for smaller-scale discrepancy, which our results suggest is important in limiting the correlation of AOD and $PM_{2.5}$. Of course, discrepancy at the finest scales, that of the pixel size and below, is a result of the scale mismatch, with areal AOD values necessarily unable to capture local variability in point-level $PM_{2.5}$ observations. But the results here suggest the presence of substantial discrepancy at small scales that are larger than this finest of scales. We note that our interest in small-scale information, driven by the desire to identify health effects within and between cities, rather than the health effects associated with regional variation in $PM_{2.5}$ (Pope et al. 2002; Zeger et al. 2007), pushed the limits of the available AOD products. It remains unclear what factors were most important in causing the apparent discrepancy at these smaller scales, after we corrected for larger-scale discrepancy. As we discuss in sections 7 and 8, one open question is whether smaller-scale information on vertical-profile variation would have helped to explain and correct for some of the discrepancy.

In our results, we noted that the spatial variation in long-term average MISR AOD closely followed patterns of surface reflectivity, suggesting that surface reflectance contamination was an important contributor to the discrepancy at smaller scales. An associated question was whether some of the association of AOD and $PM_{2.5}$ was driven by the simple correlation of AOD retrievals with land use and not with AOD acting as a direct assessment of $PM_{2.5}$ concentrations at the surface. Similar spatial patterns were seen in the AOD averages and the final proxy of van Donkelaar and colleagues (2010). Satellite AOD products are known to have difficulty distinguishing aerosol signals from land signals when aerosol levels are low, so surface contamination is likely an important factor in the difficulty in making use of small-scale variation in AOD as a proxy for $PM_{2.5}$. In the United States, in particular, the magnitude of satellite AOD error is large relative to the signal (Liu et al. 2004b).

van Donkelaar and colleagues (2010) reported more encouraging results for larger spatial domains. For the United States, they reported larger correlations than seen in this study, with the exception of the correlation of $PM_{2.5}$ with our model 5 proxy. The most important difference between their results and ours that likely explains much of the difference in correlations is that they calculated correlations over the entire United States, which has a broader range of $PM_{2.5}$ values than the eastern United States and a simple contrast of generally high values in the East and low values in the West. This likely led to increases in the proportion of variation they explained. In addition, their consideration of long-term averages was based on daily values of $PM_{2.5}$ and AOD proxies coincident in time, which increased correspondence relative to comparing to true long-term average $PM_{2.5}$. Interestingly, if we manipulate the calibration $PM = \eta AOD$ in van Donkelaar and colleagues (2010), we see that $\exp(-\hat{g}(s))$ in our model 5 is essentially the scaling term, η (but without an overall scaling constant that puts the quantities in units of $PM_{2.5}$). The negative correlation of the spatial calibration term for model 5 with PM (see Results, above, in this section) is consistent with the apparent positive correlation between η of their equation 1 and $PM_{2.5}$ that is evident in Figures 2 and 3 of their report.

5. LIMITATIONS OF REMOTELY SENSED AEROSOL AS A SPATIAL PROXY FOR $PM_{2.5}$

INTRODUCTION

Epidemiologic studies provide evidence that chronic exposure to PM is related to increased mortality and morbidity (Dockery et al. 1993; Pope et al. 2002; Miller et al. 2007). Studies of the chronic health effects of $PM_{2.5}$ rely on spatial heterogeneity of $PM_{2.5}$ concentrations to identify the effects. Spatial statistical modeling combined with land-use regression can improve estimation of concentrations at smaller scales and in areas far from monitors by using land use and meteorological information (Yanosky et al. 2008; Paciorek et al. 2009), but efforts still suffer from the spatial sparsity of the monitoring network.

Remote sensing holds promise for adding spatial information for estimating exposures, particularly in areas far from monitors, which tend to be suburban and rural (e.g., Figure 7, described below), although reviewers of this report have commented that such areas might have more fine-scale variability than urban areas because of local sources and topography. Satellite-derived AOD has been correlated with ground level $PM_{2.5}$ (Wang and Christopher 2003; Engel-Cox et al. 2004; Liu et al. 2005; Koelemeijer et al. 2006; Liu et al. 2007a; Pelletier et al. 2007; section 3 of this report). These correlations occur despite the vertical

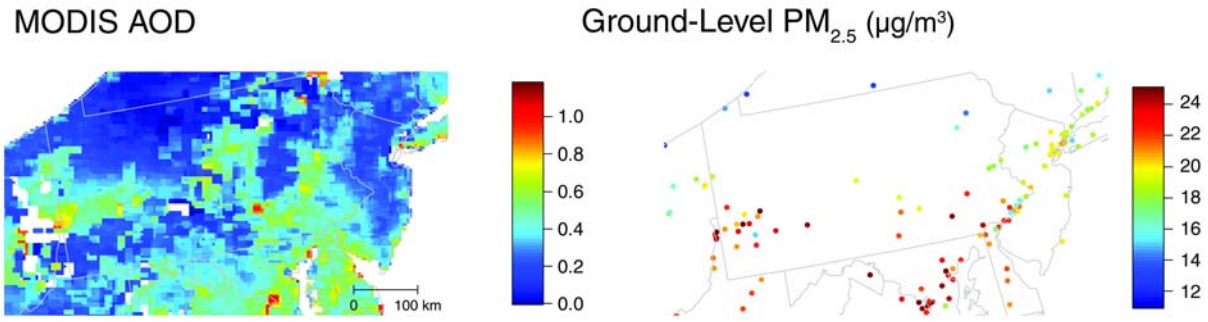

Figure 7. Monthly average MODIS AOD (left map) and ground-level $PM_{2.5}$ concentrations (µg/m³) (right map). The values are from July 2004 for the U.S. mid-Atlantic region defined by this study.

mismatch of total-column aerosol, as quantified by AOD, and ground-level $PM_{2.5}$ concentrations, the level of interest for health studies, and despite the temporal mismatch of 24-hour averages of $PM_{2.5}$ and daytime (often single-snapshot) AOD. These results and success in using AOD to document pollution episodes (Wang and Christopher 2003; Al-Saadi et al. 2005) have led to excitement about using AOD as a proxy, standing in for $PM_{2.5}$, or for using it in combination with ground measurements to better predict $PM_{2.5}$. Our work focused on improving empirical prediction, rather than physical explanation, of the spatial patterns of $PM_{2.5}$.

Most studies of the AOD–$PM_{2.5}$ association have focused on temporal (longitudinal) correlations or have not distinguished spatial (cross-sectional) from temporal correlations. However, estimating spatial heterogeneity is critical for estimating chronic exposure to $PM_{2.5}$. Correlations of long-term averages using matched daily (or hourly) values (e.g., van Donkelaar et al. 2006) do not take into account the large number of missing retrievals, resulting from orbit patterns, cloud cover, and surface reflectivity, that can seriously compromise the association between available AOD and long-term average $PM_{2.5}$ concentrations. Finally, but critically, simple correlations do not tell us if AOD improves predictions within a statistical model that already uses information on meteorology, land use, and regional variation. We are not aware of any such analysis of the use of AOD for exposure estimation.

Here we report on both raw empirical results and statistical modeling of the relationship between AOD and $PM_{2.5}$ and on the ability of AOD retrievals to improve predictions of ground-level $PM_{2.5}$ in the eastern United States, focusing on the mid-Atlantic region. We take a public health perspective, in which good estimates of $PM_{2.5}$ concentrations over an entire specified spatial region and time period are needed as an input in epidemiologic analysis. We first showed positive, but moderate and variable, correlations at various temporal scales. The correlations did not improve when we looked at longer-term averages over all the days in a period of time. We introduced a statistical model that treated AOD as proxy data for $PM_{2.5}$, estimating a $PM_{2.5}$ prediction surface that reflected both the $PM_{2.5}$ and AOD data. This model was highly sensitive to assumptions about the structure of the discrepancy between AOD and $PM_{2.5}$. The results suggested that there were systematic, spatially correlated differences between AOD and $PM_{2.5}$ and that AOD should be disregarded in predicting $PM_{2.5}$. We confirmed this result using a simpler model, with $PM_{2.5}$ data as the gold standard, by regressing $PM_{2.5}$ on AOD and numerous other predictors, which showed that the use of AOD in an already successful prediction model offered no gain in predictive power.

METHODS

Data

All analyses were for the year 2004. Associations of AOD and $PM_{2.5}$ were weak in the western United States (Engel-Cox et al. 2004; Liu et al. 2005; section 3 of this report), so we focused on the eastern United States. Our daily exploratory analyses used data east of 100° W longitude. To limit computations with large remote-sensing datasets, our longer-term analyses, including the statistical modeling, focused on a mid-Atlantic region encompassing Pennsylvania and New Jersey (see Figure 7), which contains the major metropolitan areas of New York, Philadelphia, Washington DC, Baltimore, and Pittsburgh as well as large rural areas in the north. The heterogeneity in population density and the presence of large point-source emissions from power plants and industrial plants in the southwestern portion of the mid-Atlantic region provided a test region with substantial variability in pollution.

We used AOD retrievals from three satellite instruments: MODIS, MISR, and GASP. The MODIS and MISR instruments are aboard the Terra satellite platform, whose polar orbit has been providing full coverage of the globe at regular intervals, beginning in March 2000, with retrievals in the eastern United States at a constant daily time point (10:30–10:45 AM local time). Both MISR (primarily version 15, at 558 nm) and MODIS (Collection 5, at 550 nm) provided retrievals of AOD, a dimensionless measure of light extinction over the entire vertical column of air through the atmosphere (also known as aerosol optical thickness), which were available through NASA. AOD generally ranged from 0 to 5, with values greater than 1 associated with heavy haze. MISR AOD retrievals were at a nominal spatial resolution of 17.6 km with retrievals in the northeast United States every 4 to 7 days depending on location (Liu et al. 2005). MODIS provided AOD retrievals at a nominal resolution of 10 km, with each location covered every 1 to 2 days (Wang and Christopher 2003, Engel-Cox et al. 2004). AOD cannot be retrieved below clouds, so cloud-filtering algorithms use the infrared portion of the spectrum to detect and omit obscured observations (Engel-Cox et al. 2004). Errors and uncertainties in the filtering could have led to erroneous AOD retrievals, and high surface reflectivity could also have prevented retrievals.

GASP AOD (interpolated at 550 nm) was calculated from GOES-12 (East) imager data; the NOAA provided their most recent (as of 2007) version. GASP AOD was at a nominal spatial resolution of 4 km, but retrievals were less precise than MODIS or MISR AOD retrievals because of the coarse spectral resolution and fixed viewing geometry of the sensor (Prados et al. 2007). Retrievals were attempted every

half-hour during daylight, 10:45 to 23:45 UTC, but again, cloud cover and high surface reflectivity could have led to many missing observations. We used daily average GASP AOD, regardless of the number of retrievals.

We used 24-hour averages of gravimetric FRM measurements from the U.S. EPA AQS with parameter 88101, omitting a small number of IMPROVE monitors, which are generally placed where few people live. While hourly data were better matched in time to the MODIS and MISR snapshots, the number of hourly monitors was limited and there is no FRM for hourly measurements. Also, our interest was in the relationship of $PM_{2.5}$ and AOD at monthly and yearly periods.

In our statistical modeling, we used GIS-based and meteorologic covariates to help explain $PM_{2.5}$ variation, following Yanosky and colleagues (2008). Covariates that could have helped predict $PM_{2.5}$ at fine spatial scale included distance to major roads in three road classes (Class 1, primary road with limited access; Class 2, primary road without limited access; and Class 3, secondary and connecting roads, state and county highways). We also used the point locations of year 2002 primary $PM_{2.5}$ emissions from the U.S. EPA's 2002 National Emissions Inventory (NEI). Other covariates were calculated using a GIS at the resolution of the 4-km grid used in our statistical modeling. These included road density for the three road classes, population density, and elevation at the cell centroid. As a measure of non-point-source emissions in each cell, we assigned the density (total emissions divided by county area) of the area-level primary $PM_{2.5}$ emissions, from the 2002 NEI, in the county of the cell centroid. Meteorologic variables were based on the NARR (Mesinger et al. 2006) fields, available at 32-km resolution every 3 hours. For each 3-hour value and each grid cell, we computed the IDW average of the NARR values from the four NARR points nearest to the cell centroid. Values were then averaged for the month. Our second statistical model used wind speed and temperature, but we also considered the RH, PBL height, mean sea-level pressure, and precipitation.

We also made use of a calibrated AOD variable (see section 3), which accounted for systematic effects of the PBL, the RH, the season, and time-invariant regional variation that modified and obscured the relationship between daily $PM_{2.5}$ concentrations and daily AOD. The calibration was achieved by regressing daily 2004 AOD values, from across the eastern United States, on daily $PM_{2.5}$ and the variables just mentioned, matched in space and time. By including time-invariant regional variation, we caused the long-term average AOD and long-term average $PM_{2.5}$ to more closely match at large spatial scales, necessarily increasing correlations of $PM_{2.5}$ and AOD. We hoped that by including this spatial term we could adjust for large-scale differences between AOD and $PM_{2.5}$, which would allow us to exploit common patterns of AOD and $PM_{2.5}$ at smaller spatial scales, to the extent that they existed.

Table 4 shows summary statistics for the three types of AOD retrievals and for the $PM_{2.5}$ observations. Appendix A provides more details on the data.

Exploratory Analyses

Our goal in the exploratory analyses was to understand the association between AOD and $PM_{2.5}$ at different temporal aggregations, to assess the potential of AOD to help predict chronic $PM_{2.5}$ exposure. We started by considering daily associations, when AOD and $PM_{2.5}$ were matched such that both types of data were available for a given day and location, mirroring analyses in the published literature. We matched available $PM_{2.5}$ 24-hour averages with AOD retrievals from the pixel nearest the associated $PM_{2.5}$ monitor for each of the three satellite instruments, omitting a small number of monitors for which the nearest pixel centroid was closer to another monitor. Our interest was in fine-resolution estimation of $PM_{2.5}$. Therefore, unlike for other analyses, we used individual pixels instead of aggregating AOD across adjoining pixels.

When considering the prediction of long-term average $PM_{2.5}$ for epidemiologic analyses of chronic exposure, missing AOD retrievals cause researchers to rely on a subset of days with AOD retrievals (determined by weather conditions that also affect $PM_{2.5}$ levels, so AOD patterns represent only cloud-free conditions) to estimate monthly or yearly pollution. On average, MODIS, MISR, and GOES retrievals were available over the land area of the entire mid-Atlantic region 16%, 4%, and 38%, respectively, of the days in 2004. Also, the MODIS and MISR AOD snapshots taken every day at the same time might not have matched average daily ground-level $PM_{2.5}$ pollution. To assess the long-term spatial relationship of AOD and $PM_{2.5}$, we considered associations of yearly $PM_{2.5}$ and AOD, relating average AOD from available retrievals to average $PM_{2.5}$ based on all available $PM_{2.5}$ monitoring data, not just $PM_{2.5}$ data matched by day to AOD retrievals. These associations eliminated temporal correlations within a site that could obscure the spatial association. However, simple yearly averaging did not account for the differential frequency of successful AOD retrievals over the seasons in the year (which resulted in the overweighting of summer AOD values) nor did it allow us to consider monthly associations, so we also report results at the monthly level.

For the monthly analyses, we focused on MODIS and GOES retrievals because of the extremely low availability

Table 4. Summary Statistics for MISR, MODIS, and GASP AOD Retrievals (Unitless) and $PM_{2.5}$ Observations ($\mu g/m^3$) for 2004 for Both Daily Averages (Matched by Day and Location) in the Eastern U.S. and Monthly Averages of $PM_{2.5}$ Observations in the Mid-Atlantic Region

Data Source	Sample Size	Median	(5%, 95%)
Daily Values, Eastern U.S.[a]			
MISR AOD	4,381	0.09	(0.03, 0.38)
$PM_{2.5}$ matched to MISR	4,381	10.5	(3.7, 26.2)
MODIS AOD	11,291	0.07	(−0.04, 0.46)
$PM_{2.5}$ matched to MODIS	11,291	10.2	(3.8, 26.6)
GASP AOD	46,684	0.13	(−0.06, 0.51)
$PM_{2.5}$ matched to GASP	46,684	11.1	(4.3, 27.3)
Monthly Averages, Mid-Atlantic[b]			
$PM_{2.5}$, all seasons	1,842	13.0	(8.2, 21.0)
$PM_{2.5}$, winter	460	11.2	(7.4, 16.8)
$PM_{2.5}$, spring	463	13.3	(8.8, 19.0)
$PM_{2.5}$, summer	459	17.0	(11.9, 24.5)
$PM_{2.5}$, fall	460	11.3	(7.8, 18.3)

[a] Locations east of 100° W longitude.

[b] Monthly average AOD values are not included because these values were processed before use in the statistical models presented in section 5.

of MISR retrievals, which would have required us to interpolate AOD values over very large areas with no retrievals. Even after we averaged to the month, some locations were found to have no retrievals, so we used a statistical smoothing model to estimate an AOD surface for each month. This model used a computationally efficient Markov random field (MRF) representation of a thin plate spline (TPS) (Rue and Held 2005; Yue and Speckman 2010; see also section 6 and Appendix C, available on the HEI Web site, for more detail) that readily fit the AOD retrieval data for each month (there were as many as 15,000 observations in a month). We used a heteroscedastic residual variance that accounted for the differing number of retrievals in different locations. While the model has the flexibility to perform either substantial or little smoothing, in practice because the AOD values did not show a lot of white-noise-like behavior, the smoothed fields looked similar to the raw fields, but with imputed values where no data were available.

Statistical Modeling

Our exploratory analyses did not account for complications such as differing numbers of $PM_{2.5}$ observations and AOD retrievals by location and heterogeneity in $PM_{2.5}$ at scales smaller than the AOD pixels. Most importantly, because the correlations of AOD and $PM_{2.5}$ might have reflected variability in $PM_{2.5}$ that could have been predicted by other sources of information, such as land use, meteorology, or estimation of large-scale regional variation through spatial smoothing of monitored values, they could have overstated the usefulness of AOD as a predictor in light of other readily available information. To address these issues, we turned to formal statistical modeling, analyzing the mid-Atlantic region. Both the model described above and this model were specified in a Bayesian context and were fit by standard Markov chain Monte Carlo (MCMC) methods. We did not use MISR data because of the limited number of retrievals.

Using AOD as Proxy Data Recent statistical efforts have focused on combining multiple sources of information by treating the sources as reflecting a true, unknown spatial process (Fuentes and Raftery 2005; Gelfand and Sahu 2009; McMillan et al. 2010). Accordingly, we fit statistical models for individual months. $PM_{2.5}$ and AOD observations were considered to be separate data sources that reflected the unknown $PM_{2.5}$ surface for a given month. The first stage of the model contained two likelihood terms representing the probabilistic relationships of the $PM_{2.5}$ and AOD data to the underlying processes and covariates.

For $PM_{2.5}$, for an individual month, we specified the likelihood for the $PM_{2.5}$ concentration at location i, PM_i, as

$$PM_i = y_i \sim N(P_{s(i)} + \sum_k f_k(z_{k,i}), \sigma^2_{y,i}), \quad (9)$$

where the core of the model was the unknown true pollution surface that we wanted to estimate, represented on a 4-km grid as P_s, where s indexed grid cells. We represented the monthly averages of available 24-hour concentration measurements in terms of the gridded pollution surface, locating individual observations, y_i, within the appropriate grid cell, $s(i)$. $f_k(z_{k,i})$ were smooth regression functions that reflected the effects of local covariates, z_k, that affected $PM_{2.5}$ at scales below 4 km, a decomposition similar to that of Beelen and colleagues (2007). In particular, we used the distance to the nearest Class 1 and Class 2 roads, forcing the effect to be 0 beyond 500 m (Zhou and Levy 2007). By modeling the effect of nearby roads (and point emissions; see below), we attempted to account for differences between AOD and $PM_{2.5}$ caused by within-pixel heterogeneity captured by $PM_{2.5}$ monitors but smoothed over in the AOD pixel values. We thus assessed the ability of AOD to capture spatial heterogeneity of $PM_{2.5}$ at small scales (tens of kilometers) but not at fine scales (meters to kilometers). The error variance, $\sigma^2_{y,i}$, reflected various components of uncorrelated error and accounted for the varying number of daily observations by location. The likelihood term for AOD is presented later, in equation 11.

The unknown pollution process on the grid was represented as

$$P_s = \mu + \sum_k h_k(w_{k,s}) + g_s, \quad (10)$$

where μ was an overall mean and $h_k(w_{k,s})$ were smooth regression functions of grid-scale covariates: the density of Class 1, Class 2, and Class 3 roads, population density, elevation, and non-point-source area emissions. g_s was a smooth spatial term representing residual spatial structure unaccounted for by covariates, in particular regional variation. Since we fit the model individually for each month, we omitted meteorologic covariates, which tended to be spatially smooth and whose influence would have been difficult to separate from g_s, causing their influence to be reflected in the estimate of g_s. Also included in the model was a term that accounted for the effect of point emissions within 100 km of a monitor. This term represented the effect of multiple point sources at a given receptor location (i.e., a monitor or prediction point) as the emission-strength-weighted sum of a smooth distance effect evaluated for each individual source–receptor pair. The distance function was a universal function representing the effect of a single source of unit strength on a receptor as a smooth function of the distance between source and receptor. It was estimated from the data based on a new statistical approach that leveraged the additive structure of mixed model representations of splines, as described in Appendix D (see the HEI Web site). This term was used both as a covariate affecting the individual $PM_{2.5}$ observations based on the point location of each monitor (see equation 9) and as a covariate affecting the gridded process, P_s (based on averaging over a subgrid of 16 points within each cell).

We specified the AOD retrieval in an individual month, AOD_s, as reflecting the unknown $PM_{2.5}$ process,

$$AOD_s = a_s \sim N(\beta_0 + \beta_1 P_s + f_{cloud}(z_{cloud,s}), \sigma^2_{a,s}), \quad (11)$$

up to additive, β_0, and multiplicative, β_1, bias, with a smooth regression function of cloud cover, $f_{cloud}(z_{cloud,s})$, where $z_{cloud,s}$ was the monthly average proportion of cloud-free retrievals in the cell, based on the GOES algorithm for retrieving clouds. This was included to help account for bias from retrievals that were systematically missing because of clouds (Koelemeijer et al. 2006; section 3 of this report). $\sigma^2_{a,s}$ reflected various components of uncorrelated error and accounted for the varying number of daily retrievals by location. A complicating factor was that for different satellite orbits on different days, the MODIS pixels shifted spatially. Therefore, we considered the overlap of all the pixels in an orbit with the 4-km grid, assigning to each grid cell, s, the value of the MODIS pixel in which the cell centroid fell. Taking the retrievals assigned to each cell, we then calculated a monthly average for each cell, resulting in a_s. The pixel locations of GOES retrievals were constant over time, so we calculated a monthly average and then assigned each grid cell the weighted average AOD of the GOES pixels that the cell overlapped, weighted by the area of overlap. While simplistic, we believe these approaches caused minimal distortion in the AOD values used in the modeling because of the reasonably smooth local variation in daily AOD values from pixel to pixel. Note that we assumed in this model that any difference between AOD and $PM_{2.5}$ was spatially uncorrelated noise, which resulted in an estimate of P_s that reflected the spatial structure in both $PM_{2.5}$ and AOD observations.

Note, however, that maps of monthly average AOD show strong spatial structure (e.g., Figure 7, left map) with limited spatially uncorrelated noise (i.e., white noise) apparent. This spatial structure might have been caused in part by systematic, spatially correlated differences between

AOD and $PM_{2.5}$, rather than having been a reflection of spatial structure in ground-level $PM_{2.5}$. Factors likely to have contributed to such differences, which would have operated even if AOD had been measured perfectly, included the spatial structure of pollution aloft above the PBL and daily spatial patterns of missing retrievals because of clouds, which would have resulted in an aggregate effect at the monthly level. Of course, AOD is not measured perfectly (Knapp et al. 2002; Remer et al. 2005), as reflected in moderate correlations between monthly average MODIS and GOES AOD and caused in part by spatial variability in surface reflectivity and $PM_{2.5}$ composition. The summed effect of all these differences, which we refer to as systematic discrepancy, could be substantial and it, rather than pixel-scale white noise, could be the dominant factor explaining the low correlations of AOD with $PM_{2.5}$ seen in our exploratory analyses. Models that treat AOD as a proxy for $PM_{2.5}$ without accounting for potential systematic discrepancies might predict spatial patterns of $PM_{2.5}$ that do not match reality. We assessed the sensitivity of our results to assumptions about systematic discrepancies by including an additive spatial bias term, φ_s, that was represented at the grid scale and replaced the bias constant, β_0, in equation 11. Models with such a term allow for the possibility that AOD retrievals might be providing information about spatial processes specific to the retrievals that does not reflect spatial patterns in ground-level $PM_{2.5}$ concentrations. We estimated φ_s using a penalized TPS approach that penalized complex spatial surfaces, thereby favoring simple surfaces if the data could be sufficiently explained by a smooth surface (Ruppert et al. 2003). Such an approach was also used for the other smooth terms in the model, fit naturally within the Bayesian context with the level of smoothing determined by the data. For computational reasons and because the key results are best visualized in model fits of individual months, we fit the model separately for each of the 12 months.

The advantage of this modeling approach was that it naturally treated AOD retrievals as data and allowed for missing retrievals. By considering different assumptions about spatial bias, we could assess the concordance of spatial patterns of AOD and $PM_{2.5}$ and investigate the assumption that the spatial pattern in AOD represented a signal that was informative about $PM_{2.5}$.

Using AOD as a Predictor of $PM_{2.5}$ We also considered a model in which AOD was used as a predictor on the right-hand side of a regression-style model and the $PM_{2.5}$ data were treated as the gold standard. This had the benefit of directly calibrating $PM_{2.5}$ to AOD and, if there were little empirical association, discounting AOD as a predictor of $PM_{2.5}$.

In this model, $PM_{2.5}$ observations were modeled as in equation 9,

$$\mathrm{PM}_{it} = y_{it} \sim N(P_{s(i),t} + \sum_k f_k(z_{k,i}), \sigma^2_{it}), \tag{12}$$

whereas the unknown smooth pollution process, $P_{s,t}$, was similar to equation 10, but included AOD, $A_{s,t}$, as a predictor:

$$P_{s,t} = \mu_t + \beta_{1,t} A_{s,t} + \sum_k h_k(w_{k,s,t}) + g_{s,t}. \tag{13}$$

This model was fit simultaneously to all 12 months, which were indexed by the t subscript. For simplicity, we assumed that $g_{s,t}$, the residual spatial structure, was not correlated over time, which eased computations. Previous work has suggested that month-to-month correlation is limited and that including correlation would do little to improve predictions (Paciorek et al. 2009), so this assumption should not have affected our ability to assess whether AOD could improve $PM_{2.5}$ predictions. We allowed $\beta_{1,t}$ to vary in an unstructured way with time, in case the relationship of AOD and $PM_{2.5}$ varied by season (section 3). We limited the covariates, w_k (some of which did not vary with time), to population density, elevation, area emissions, point emissions, density of Class 3 roads, wind speed, and temperature. We also included monthly average cloudiness as a covariate, to help account for bias resulting from missing AOD retrievals.

One downside to our approach was that it required AOD values at all locations. We used the MRF representation of a TPS, described in Exploratory Analyses in section 5, to smooth the observed AOD retrievals and make predictions, $A_{s,t}$, at unobserved locations. Preprocessing of pixel-level AOD values to align with the 4-km grid was done as described previously.

RESULTS

Exploratory Analyses

Correlations between daily $PM_{2.5}$ and AOD (matched by day and location) that reflected both temporal and spatial associations were higher than correlations for individual days, which reflected only spatial associations (Table 5). The spatiotemporal associations roughly matched those that have been used in the literature as evidence of the potential of AOD as a proxy for $PM_{2.5}$ (e.g., Engel-Cox et al. 2004; Liu et al. 2005; section 3 of this report). Figure 8 illustrates the daily data used to calculate purely spatial associations, showing the spatial mismatch between MODIS AOD patterns and $PM_{2.5}$ levels seen at monitors. In this example, the mismatch did not appear to be driven by local variability in $PM_{2.5}$, as the levels were fairly consistent

Table 5. Correlations for 2004 of Daily AOD with Matched 24-hr $PM_{2.5}$ for the Eastern U.S. and of Yearly Averages of AOD with $PM_{2.5}$, Matched in Space But Not by Day, for the Mid-Atlantic Region

	Raw AOD			Calibrated AOD[a]		
	MODIS	MISR	GOES	MODIS	MISR	GOES
Daily Values, Eastern U.S.						
Temporal plus spatial variation: Overall correlation of daily values across all sites and days	0.60	0.50	0.38	0.64	0.57	0.40
Spatial variation only: Average of daily spatial correlations[b]	0.35	0.30	0.23	0.45	0.32	0.29
Yearly Averages, Mid-Atlantic Region[c]						
Spatial variation only: Correlation of yearly averages	0.09	0.25	−0.07	0.49	0.22	0.53

[a] Calibrated AOD has been adjusted to account for the effects of PBL, RH, season, and regional variation on the relationship between daily AOD and $PM_{2.5}$.

[b] Only days with at least 20 matched sites are included.

[c] Yearly averages reflect all available AOD retrievals and all available 24-hr average $PM_{2.5}$ concentrations. Yearly results include only sites with at least 100 daily $PM_{2.5}$ observations and exclude one site with high $PM_{2.5}$ levels outside Pittsburgh that is just downwind of a major industrial facility.

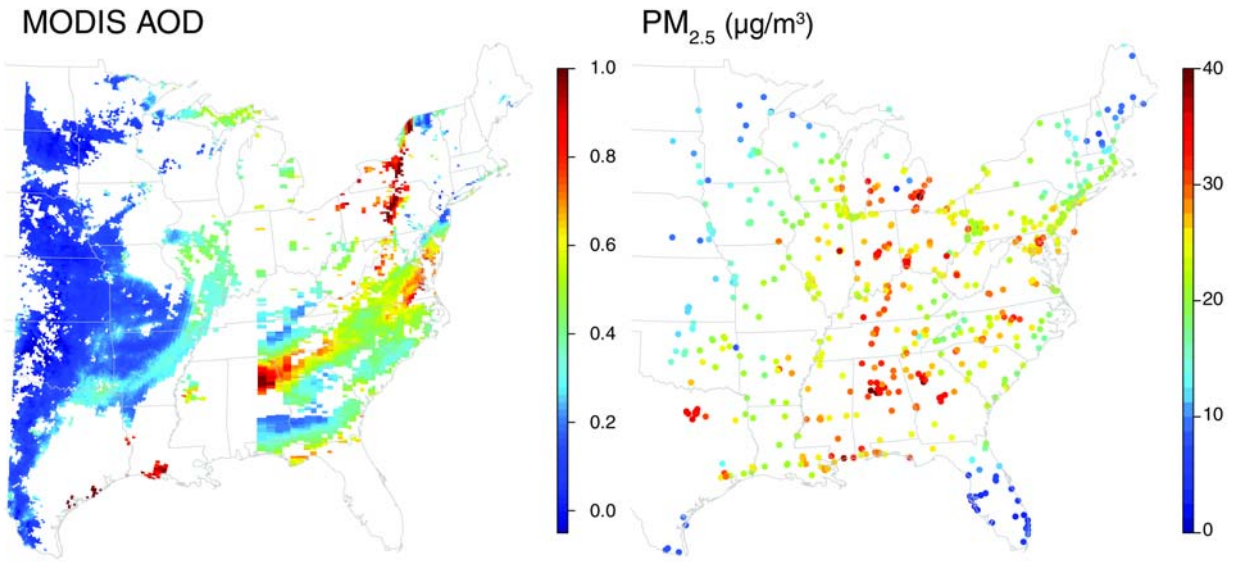

Figure 8. MODIS AOD (from two overpasses; left map) and 24-hour average $PM_{2.5}$ observations (µg/m³) from a single day, July 20, 2004 (right map). The maps illustrate spatial mismatch on a day when $PM_{2.5}$ observations were spatially uniform within metropolitan areas and over broader regions.

within individual metropolitan areas as well as over larger regions. Using AOD directly in our analyses did not account for meteorologic factors and systematic temporal and spatial variability that modified the relationship between AOD and $PM_{2.5}$, so we also considered the calibrated version of AOD, which somewhat improved the correlations (Table 5).

Table 5 shows near-0 correlations of yearly average $PM_{2.5}$ based on all available 24-hour values (everyday or every-third-day sampling) with AOD from available retrievals. Note that for monitors reporting only every 3 days, missing $PM_{2.5}$ values contributed to noise in the associations seen here. After calibration, AOD was moderately correlated with $PM_{2.5}$. The calibration included an overall spatial term that adjusted for any large-scale regional mismatch between AOD and $PM_{2.5}$ that was consistent over the year. This term was responsible for much of the increase in correlation after calibration, as it necessarily caused the large-scale patterns of long-term average AOD and $PM_{2.5}$ to more closely match. We hoped that correcting for such large-scale mismatch would allow us to explore whether there was independent information in AOD for predicting smaller-scale patterns of $PM_{2.5}$, a question that we answered in the statistical modeling in this section.

Correlations were higher in the warmer months, but were moderate at best (Figure 9). Results were similar when locations were restricted to those with at least 3 days in a given month with AOD retrievals. The poor correlations resulted in part from the limited availability of retrievals, particularly during winter, seen in gray in Figure 9.

Based on taking the 12 smoothed monthly values for AOD and averaging to get a yearly value at each location, Table 6 shows near-0 correlations of raw AOD and moderate correlations of calibrated AOD with yearly average $PM_{2.5}$. By averaging using equal weights for each month, we attempted to account for the varying availability of retrievals in different seasons. During the warmer part of the year (avoiding months with few retrievals), correlations of calibrated AOD with $PM_{2.5}$ increased, but those of raw AOD did not. If we considered only monitors with at least 300 observations (shown in the last row of Table 6) during the year (i.e., monitors reporting daily, with little missing data), correlations for the calibrated AOD were similar, whereas for raw AOD they were higher but still moderate in magnitude. The increased correlations might have been related to the fact that monitors that sample daily are more likely to be in locations with high $PM_{2.5}$ concentrations; that is, these monitors were more likely to be categorized by EPA as monitors sited to monitor areas of high $PM_{2.5}$ concentrations (18% of daily-sampling monitors but only 5% of non-daily-sampling monitors).

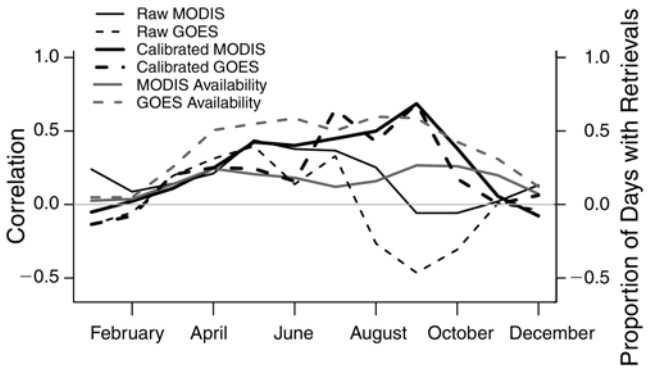

Figure 9. Correlations (across space) of monthly average smoothed AOD with spatially matched $PM_{2.5}$. The results (from 2004) are for locations in the mid-Atlantic region with at least five daily $PM_{2.5}$ values during the month (thin lines represent raw AOD; thick lines, calibrated AOD). Gray lines show the proportion of days in a month with a successful retrieval, averaged over land areas in the region. For consistency, the results exclude the site outside Pittsburgh excluded from Table 5, but this exclusion has little effect here.

Table 6. Correlations (Across Space) of 2004 Yearly and Warm-Season (April–October) Average AOD and Spatially Matched $PM_{2.5}$ (Sites with at Least 100 Daily $PM_{2.5}$ Observations) for the Mid-Atlantic Region[a]

	Raw AOD		Calibrated AOD	
	MODIS	GOES	MODIS	GOES
Correlation of yearly values	0.19	−0.06	0.53	0.44
Correlation for April–October	0.05	−0.19	0.65	0.70
Correlation of yearly values, daily monitors only	0.43	0.27	0.45	0.40

[a] Yearly averages are computed from monthly data. For AOD, these are produced by spatial smoothing of the available monthly data, thus filling in missing values at the monthly level. Results exclude the site outside Pittsburgh excluded in Table 5.

Using AOD as Proxy Data: Sensitivity to Systematic Discrepancies

Figure 10 illustrates using MODIS AOD data from July 2004 to make model-based predictions of $PM_{2.5}$ and estimates of φ_s (based on equations 9, 10, and 11), allowing different amounts of complexity in φ_s, the additive spatial bias term. When the model omitted the spatial bias term (row 1), which represents AOD as reflecting $PM_{2.5}$ with only simple additive and multiplicative bias, predictions of $PM_{2.5}$ strongly tracked the spatial patterns of AOD (i.e.,

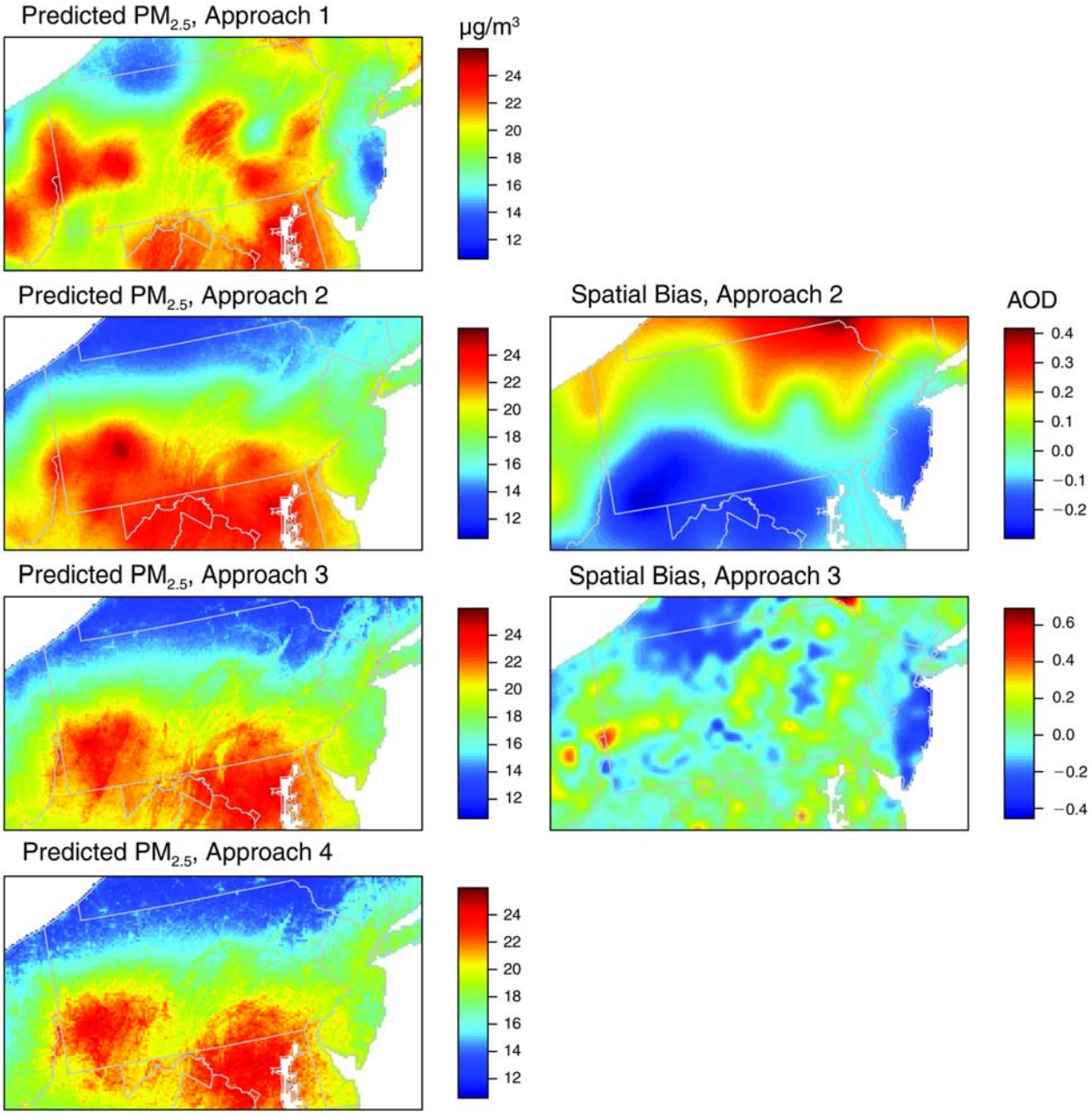

Figure 10. Sensitivity of predicted $PM_{2.5}$ ($\mu g/m^3$) to the characterization of spatial discrepancy. The left column shows $PM_{2.5}$ predictions for models in which AOD and $PM_{2.5}$ observations are treated as data reflecting a common unknown $PM_{2.5}$ process, using calibrated MODIS AOD for July 2004. Approach 1 excludes the spatial discrepancy term, φ_s, thereby treating AOD as a simple proxy for $PM_{2.5}$, with simple additive and multiplicative bias. In approach 2, φ_s is constrained to be a somewhat smooth process, with a maximum of 55 degrees of freedom (a penalized spline with 55 knots). In approach 3, φ_s is relatively unconstrained, with a maximum of 755 degrees of freedom. Approach 4 does not use AOD. The right column shows the estimated φ_s surfaces for approaches specified in the left column. Approach 1 does not include spatial bias, φ_s (hence, no top-right subplot), and approach 4 does not use AOD, so the spatial bias term, φ_s, is not involved in the model (hence, no bottom-right subplot).

Figure 7, left map). As spatial bias was introduced (row 2) and more flexibility in the spatial bias term was allowed (row 3), predictions increasingly tracked the $PM_{2.5}$ observations (i.e., Figure 7, right map) and also tracked results from a model fit without AOD (row 4). The fit of the penalized spline model did not stabilize on a smooth bias surface. When the bias term was forced to be smooth, the model was unable to adequately represent the AOD data based on the $PM_{2.5}$ surface, the smooth bias, and the white-noise error. This suggests that $PM_{2.5}$ and AOD observations shared few common spatial patterns and that true $PM_{2.5}$ was best modeled solely based on ground-level $PM_{2.5}$ data, with AOD variability modeled separately. This is seen in row 3, where the model essentially disregarded AOD in predicting $PM_{2.5}$ and attributed most of the variability in AOD to φ_s. Using GOES AOD or raw AOD with data for the other 11 months led to similar results. In summary, systematic discrepancies were considerable and critical to include, and predictions were very sensitive to assumptions about the discrepancy term. If the spatial bias were estimated to be a relatively smooth process, able to be resolved by having $PM_{2.5}$ and AOD data in the same region, the modeling approach would provide a means to improve $PM_{2.5}$ prediction: combining the data sources while accounting for the bias. However, these results suggest the bias process was not smooth and could not be adequately estimated without more dense $PM_{2.5}$ data, which were not available and would have largely obviated the need to use AOD as a proxy.

Using AOD as a Predictor: Effects on Predictive Ability

Predictive ability at both the monthly and yearly resolutions did not improve when we added either calibrated MODIS or GOES AOD to the models (equations 12 and 13) already containing the other covariates (Table 7). When we excluded the other covariates (except the GOES cloud term for consistency in comparing the AOD and no-AOD models) and accounted for spatial variability solely based on spatial smoothing of the observations within the model framework, the addition of AOD still resulted in essentially no improvement in predictions (Table 7). The results were similar when data from locations that were most likely affected by very local sources were omitted (Table 7, population exposure monitors columns). Sensitivity analyses (results not shown) indicated that AOD had a similarly limited effect on predictive power when raw AOD was used, when observations were restricted to monitors in areas with sparse monitoring, or when observations

Table 7. Cross-validation R^2 (Mean Squared Prediction Error) for Predictions of Yearly and Monthly Average $PM_{2.5}$ from Statistical Prediction Models With and Without Calibrated AOD and Other Predictors of $PM_{2.5}$ Concentrations (After Other Predictors) for the Mid-Atlantic Region for 2004[a]

	Yearly Averages[b]		Monthly Averages	
Model	All Monitors (n = 151)	Population Exposure[c] Monitors (n = 130)	All Monitors (n = 1793)	Population Exposure[c] Monitors (n = 1542)
Models Including Land-Use, Emissions, and Meteorologic Predictors				
No AOD	0.580 (1.04)	0.570 (0.93)	0.827 (2.71)	0.839 (2.48)
With calibrated MODIS AOD	0.573 (1.06)	0.564 (0.94)	0.825 (2.73)	0.839 (2.50)
With calibrated GOES AOD	0.572 (1.06)	0.563 (0.95)	0.825 (2.73)	0.838 (2.50)
Models Without Land-Use, Emissions, and Meteorologic Predictors[d]				
No AOD	0.463 (1.33)	0.456 (1.18)	0.794 (3.22)	0.810 (2.94)
With calibrated MODIS AOD	0.467 (1.32)	0.459 (1.17)	0.794 (3.22)	0.810 (2.94)
With calibrated GOES AOD	0.467 (1.33)	0.458 (1.17)	0.794 (3.22)	0.810 (2.94)

[a] For a given location, only months for which the location has at least four $PM_{2.5}$ daily values are included. Results exclude one site with high $PM_{2.5}$ values outside Pittsburgh that is just downwind of a major industrial facility.

[b] Yearly average results include only locations with at least 6 months of available $PM_{2.5}$ data.

[c] The "population exposure" designation assigned to monitors by the U.S. EPA indicates that such monitors are not likely to be affected by large local sources.

[d] These models do include the GOES cloudiness term, for consistency in comparisons of the AOD and no-AOD models.

were restricted to monitors performing daily sampling (which avoided the extra noise caused by missing monitor values). Note that the higher predictability of monthly compared with yearly $PM_{2.5}$ levels shown in Table 7 occurred because of the importance of temporal variation, which was easy to estimate based on the monitoring. The results of including AOD were consistent with the estimates of $\beta_{1,t}$, which were small in magnitude, with wide uncertainty intervals that included 0. Correlations of predictions with and without AOD were greater than 0.999, indicating that including AOD in the exposure modeling when doing an epidemiologic analysis would have negligible impact.

DISCUSSION

We urge caution in assuming that currently available remotely sensed AOD can help improve exposure estimation for $PM_{2.5}$. We urge particular caution in using AOD to estimate spatial heterogeneity of $PM_{2.5}$ where there is little ground-level $PM_{2.5}$ data for ground truthing, because of the lack of a strong spatial correlation between available AOD retrievals and long-term average $PM_{2.5}$. In a setting in which reasonably dense $PM_{2.5}$ data were available, our statistical-modeling results indicated little or no improvement in the prediction of long-term average $PM_{2.5}$ when AOD was added. To the extent that raw correlations of AOD and $PM_{2.5}$ reflected the ability of AOD to capture some of the pattern in $PM_{2.5}$, our results suggest that patterns could be better estimated by simple spatial smoothing of the available $PM_{2.5}$ data and regression on other predictors, rendering the AOD information extraneous. Koelemeijer and colleagues (2006) found much stronger correlations of yearly average MODIS AOD and PM $\leq$ 10 μm in aerodynamic diameter (PM_{10}) in Europe; this may have been related to their focus on rural background sites, their larger spatial domain, and the greater variability in their PM_{10} concentrations.

Remote sensing is of particular interest in developing countries with little monitoring (e.g., Kumar et al. 2008), but our results suggest that spatial patterns seen in AOD may poorly reflect spatial patterns in ground-level $PM_{2.5}$. Without evidence of strong correlations over space, as opposed to purely temporal correlations, the use of AOD to determine the spatial heterogeneity of $PM_{2.5}$ may be misleading. Given that our results reflected our focus on a region of moderate size, it is possible that AOD could be more helpful for larger regions, although daily spatial correlations over the eastern United States were relatively weak (Table 5) and previous work showed moderate long-term correlations at best over the United States (Rush et al. 2004). As we discuss in detail in sections 7 and 8, an important limitation of our approach was that we did not make use of vertical-profile information, found elsewhere to be of use for larger-scale calibration of AOD to $PM_{2.5}$ (at 2×2.5 degrees in Liu et al. 2004a and van Donkelaar et al. 2010). However, our calibrated AOD made use of a large-scale spatial adjustment to better align AOD and $PM_{2.5}$, so it seems unlikely that such large-scale vertical-profile information would have changed our primary results. Nonetheless, for capturing larger-scale patterns in larger spatial domains, AOD that is calibrated based on vertical-profile information may be more successful (e.g., Liu et al. 2004a; van Donkelaar et al. 2010). Also, AOD might be helpful for estimating temporal heterogeneity, but missing AOD retrievals are a major problem.

One might ask whether AOD is useful under specific conditions or in specific locations, such as for pollution episodes (e.g., Wu et al. 2006). It is not clear how important such episodes are for predicting long-term average $PM_{2.5}$ or how to include such information only under the circumstances in which it is predictive of $PM_{2.5}$. If AOD is useful in some but not all circumstances, the practical challenge in making use of it is the need of epidemiologists for exposure estimates without gaps in space or time, often over large domains and long periods of time.

Systematic discrepancies such as what we saw in the satellite AOD proxy for $PM_{2.5}$ can easily be misleading because the spatial structure seen in the proxy leads one to think that the patterns reflect real patterns in the process of interest. In this study, the evidence suggested that much of this structure did not represent the true structure of $PM_{2.5}$. Such systematic discrepancies have arisen in other contexts (Campbell 1996; Robinson 2004). It seems likely that the output of deterministic models used to estimate atmospheric processes, including pollution, such as the output of the widely used CMAQ model, contain systematic errors that result in correlated errors in model output, either because of errors in inputs or because of aspects of the system under study that are not captured by the model.

Note that since the work with GASP AOD described here was completed and published (Paciorek and Liu 2009), the GOES team has developed several improvements to the GASP AOD retrieval, including correction of an error in the azimuth angle definition, an improved surface-reflectance-estimation method, and improved calculation of the AOD standard deviation, as well as the inclusion of scattering angle in the screening criteria. Because of time limitations and the availability of only 3 months of the new GASP AOD retrievals, we did not rerun the analyses for this study using the newer data.

6. FLEXIBLE SPATIAL LATENT VARIABLE MODELING FOR COMBINING INFORMATION SOURCES WHILE ACCOUNTING FOR SYSTEMATIC ERRORS IN PROXIES

INTRODUCTION

There has been substantial interest recently in combining observations at spatial point locations with proxy information from remote sensing and the output of numerical models, to improve the prediction of spatial and spatiotemporal surfaces, particularly in the area of air quality. Building on the work of Fuentes and Raftery (2005), statisticians have proposed a number of modeling approaches, often termed data fusion, using Bayesian hierarchical spatial models to combine information sources; the goals have included air-quality management, forecasting pollution, and exposure prediction for health effects analysis. Critically, model output and remote-sensing retrievals often produce surfaces that are highly spatially correlated, but some of this correlation may represent spatially correlated error (also termed discrepancy or offset, particularly if the proxy is not designed to estimate the focal process of interest but some other related quantity) with regard to the quantity of interest. For example, a numerical model may overpredict a pollutant over a wide area because of shortcomings in information on emissions sources, while cloud or surface contamination may result in spatially correlated errors in satellite retrievals.

In the statistical formulations of the problem, the possibility of proxy discrepancy has generally been acknowledged, but previous modeling efforts have often placed strong constraints on the structure of the discrepancy for reasons of identifiability or computational feasibility. Fuentes and Raftery (2005) proposed the following general model with their proxy (numerical-model output) treated as data via a second likelihood,

$$\begin{aligned} Y_i &= L(s_i) + \varepsilon_i \\ A_j &= \varphi(s_j) + \beta_1(s_j)L(s_j) + e_j. \end{aligned} \tag{14}$$

The model relates both the gold-standard observations, Y_i, and the proxy values, A_j (A for auxiliary), to the latent, true process of interest (the focal process), $L(s)$. Here, we suppress any change-of-support manipulations in defining $L(s_j)$ when s_j is an area. $\varphi(s)$ and $\beta_1(s)$ are additive and multiplicative bias terms, and ε_i and e_j are error terms. In general, it is difficult to identify both $\varphi(s)$ and $\beta_1(s)$, and Fuentes and Raftery (2005) used a scalar β_1 and took $\varphi(s)$ as a simple polynomial in the spatial coordinates. Critically, since the additive bias has a very low dimensional representation, this approach assumes that all of the small- and moderate-scale spatial structure in the proxy is signal with respect to $L(s)$. In section 5 we used a structure similar to this one, with remote-sensing retrievals playing the role of the proxy, but chose a reduced-rank spline basis for additional flexibility in modeling the discrepancy. We found that model fitting was sensitive to the number of basis functions, with increasingly better fits as the number of basis functions increased, such that computational complexity prevented fitting a model with enough basis functions to model $\varphi(s)$ adequately. Other recent work has used such moderately flexible specifications for quantities analogous to $\varphi(s)$: Fuentes and colleagues (2008) used a small number of basis functions, whereas McMillan and colleagues (2010) used b-splines in two dimensions.

The dangers in limiting the flexibility of the discrepancy representation are that systematic discrepancy will bias the prediction of the spatial process of interest in subdomains and that correlated uncertainty will not be properly acknowledged. In short, spatially correlated discrepancy in the proxy may look like signal because of the dependence structure, but will cause spurious features in the prediction surfaces. Gold-standard data can help to assess the potential for discrepancy at scales resolved by the data. For large scales (relative to the data density), one can hope to estimate and adjust for the discrepancy. At smaller scales, one can at best hope to discount, but not adjust for, proxy information if the data indicate that the proxy is poorly related to the focal process. The key to this effort lies in using a sufficiently flexible model specification for the spatial discrepancy term.

We propose to use a computationally efficient MRF specification that is sufficiently flexible to model discrepancies at a variety of spatial scales. This MRF specification (Rue and Held 2005; Yue and Speckman 2010) approximates a TPS while retaining the sparse precision matrix structure of more widely used MRFs, such as conditional autoregressive (CAR) models based on neighborhood adjacencies; sparse structure allows for efficient computation. For proxy variables that are often very large in dimension (numerical-model output and remote-sensing observations on large grids), modeling $\varphi(s)$ efficiently is critical. The ability of the proposed MRF specification to capture variation at a range of spatial scales stands in contrast to (1) reduced-rank basis function approaches (Kammann and Wand 2003; Ruppert et al. 2003; Banerjee et al. 2008) that can capture large-scale structure but require many knots with associated computational slowdowns for small-scale structure, and (2) traditional CAR models that, while computationally tractable in capturing small-scale local variability, show small-scale variation when attempting to

represent large-scale processes for which small-scale variability is absent (see MRF Specification later in this section). Given the ability of the proposed model structure to capture or omit variability at various scales, we propose an approach to assessing the scales of discrepancy as a standard part of any effort involving data fusion.

The potential for improving prediction using data fusion, particularly for filling in spatial regions with little or no data using the rich spatial structure of the proxy, is appealing and lies behind the recent growth in its interest to scientists, statisticians, and government agencies such as the U.S. EPA. While correlation between the proxy and the gold standard is usually cited as a reason why data fusion is potentially useful, such correlation does not demonstrate that including the proxy will improve prediction relative to a model that spatially smooths the gold-standard data and that may leverage additional covariates to improve predictive performance. For example, if the proxy only correlates with the gold standard at large scales, one may well be able to simply smooth the gold-standard data to capture the large-scale variation, with little contribution from the proxy. A complicated proxy error structure, as opposed to unrealistic white noise, may be difficult to represent and to account for and may result in correlated errors in predictions. The additional covariates may be able to account for small-scale variation and render the proxy less useful even if it does contain real small-scale information.

Thus, from an applied perspective, our primary concern is that overly constrained discrepancy terms implicitly assume that the proxy is useful and data fusion successful; publication pressures can also contribute to a desire to demonstrate that data fusion improves prediction. In analyses with constrained representation of the discrepancy, the change in the predictions resulting from inclusion of the proxy may primarily reflect discrepancy rather than increased information. In contrast, in section 5 we found no improvement in $PM_{2.5}$ prediction using several satellite aerosol products, nor did McMillan and colleagues (2010) in the case of output from the U.S. EPA's CMAQ model. Sahu and colleagues (2009) found a statistically significant regression coefficient for the relationship between CMAQ output and their process of interest, but the magnitude of the coefficient was small, and there appears to have been little evidence that including the proxy improved prediction. Berrocal and colleagues (2009) did find improvement in the prediction of ozone, relative to ordinary kriging without covariates, when they included the CMAQ proxy as a covariate.

Our primary methodological contribution is to present an approach that allows one to focus on these questions: Is the proxy useful and data fusion successful? At what scales is it useful and at what scales does it give erroneous predictions? Can we use the real information in the proxy and discount the influence of the discrepancy component? Our particular applied goal was to assess and, if they were helpful, use aerosol remote-sensing retrievals and the output of an atmospheric-chemistry model to better predict spatial patterns in monthly average $PM_{2.5}$ concentrations across the eastern United States, improving upon current spatiotemporal modeling efforts that combine smoothing with land-use and meteorologic covariates (Yanosky et al. 2009; Paciorek et al. 2009; Szpiro et al. 2010). Better estimation of ambient concentrations is critical for understanding chronic health effects (including those for susceptible subpopulations), air pollution regulation, and risk assessment. We illustrate our methodology in the following related analyses: (1) spatial analysis of monthly $PM_{2.5}$ in 2004 in the mid-Atlantic United States based on remote sensing of AOD, (2) spatiotemporal analysis of monthly $PM_{2.5}$ in 2001 in the mid-Atlantic United States based on output from the CMAQ model, and (3) spatial analysis of monthly $PM_{2.5}$ in 2001 in the eastern United States based on CMAQ output. We assessed improvements in predictive performance relative to models that do not make use of the proxy and analyzed the scales of the discrepancy between proxies and the gold standard.

MODEL AND METHODS

In this subsection we outline our basic modeling approach; technical details of the model specification and MCMC fitting methods are provided in Appendix C, available on the HEI Web site.

Spatial Latent Variable Model

We propose the basic spatial model,

$$\begin{aligned} Y_i &= \beta_y(x_{y,i}) + P_{Y,i}^T L + \varepsilon_i \\ A_j &= \varphi(s_j) + \beta_1 P_{A,j}^T L + e_j, \end{aligned} \tag{15}$$

where the notation follows equation 14, with L the vectorized representation of a spatial latent variable represented on a fine grid. $\beta_y(x_{y,i})$ is a regression function (easily generalizable to a sum of functions) that represents sub-grid-scale variability, and $P_{Y,i}$ and $P_{A,j}$ are rows of mapping matrices that pick off elements of L to map to the observations and proxy values. P_A may also weight the focal process values to account for spatial misalignment of the proxy and base grid (which is particularly relevant for irregular remote-sensing grids). By relating point-level measurements to a grid-based latent process, our model

has some of the flavor (and the computational advantages) of the measurement-error model of Sahu and colleagues (2009), though $\beta_y(\bullet)$ allows for modeling of subgrid heterogeneity.

We then represent

$$L(x,s) = \sum_p \beta_{L,p}(x_{L,p}) + g(s) \quad (16)$$

as the sum of multiple regression terms, $\beta_{L,p}(x_{L,p})$, where $x_{L,p}$ is the pth covariate, and the remaining spatial variation, $g(s)$, where s is a spatial location. $\beta_y(\bullet)$ and $\beta_{L,p}(\bullet)$ could be simple linear terms or spline-based regression terms, to capture nonlinear relationships with covariates. In such a case, we use the mixed-model formulation of a penalized TPS (cubic radial basis functions in one dimension), placed in a Bayesian framework, where the unknown variance component of the exchangeable basis coefficients controls the amount of smoothing (Ruppert et al. 2003; Crainiceanu et al. 2005). When the covariates are able to represent most of the small-scale variation in the focal process, $g(\bullet)$ need only explain large-scale variation, so one approach is to represent $g(\bullet)$ using a penalized TPS (Kammann and Wand 2003; Ruppert et al. 2003), as we did for the analyses for the mid-Atlantic region (see these sections below: Spatial Analysis of MODIS AOD in the Mid-Atlantic Region and Spatiotemporal Analysis of CMAQ Output in the Mid-Atlantic Region). Alternatively, when the knot-based representation of $g(\bullet)$ requires so many knots that computations bog down, either because of small-scale process variation or a large domain, $g(\bullet)$ can be represented as an MRF, as described next for $\varphi(\bullet)$ and used in the analysis of CMAQ output in the eastern United States (see Spatial Analysis of CMAQ Output in the Eastern United States, below).

For $\varphi(\bullet)$, we use the MRF approximate representation of a TPS, $\varphi \sim N_{m-3}(0,(\kappa Q)^{-})$, described in detail in the next subsection. Q is the MRF weight matrix, with rank $m-3$ — hence the use of the generalized inverse. κ is a precision parameter. This representation is the key to our approach and our conceptualization of the critical bias issue in data fusion; it allows the discrepancy to represent either smooth large-scale variation or wiggly small-scale variation depending on the data, with the MRF approximation providing for computational feasibility in dealing with high-dimensional proxy variables. We worked with up to 17,500 pixels (the value of m) in our analyses. Note that integrating over φ produces a spatially correlated proxy-error process, so the distinction between representing spatial variation in the mean (as in splines) or in the covariance (as in kriging) is artificial.

MRF Specification

MRF models, such as standard CAR models, often use simple binary weights in which direct neighbors are given a weight of 1 and all other locations are given 0 weight (e.g., standard CAR models). However, such models have realizations with unappealing properties. Rue and Held (2005) characterized the intrinsic Gaussian CAR model in one dimension as a discretely observed Wiener process (i.e., one-dimensional Brownian motion), and Besag and Mondal (2005) showed that the intrinsic Gaussian CAR model on a fine two-dimensional grid approaches two-dimensional Brownian motion (the de Wijs process) asymptotically as the grid resolution increases. Brownian motion has continuous but not differentiable sample paths, so it is not surprising that the process realizations of standard CAR models are locally heterogeneous, as seen next, regardless of the value of the variance component for the process.

A more flexible alternative that has received little attention but is potentially widely applicable is an MRF whose weight structure is motivated by the smoothness penalty that induces the TPS (what we refer to as the TPS-MRF) (Rue and Held 2005; Yue and Speckman 2010). For a regular grid, Figure 11 compares the neighborhood and weight structure of the standard CAR model with binary weights to that of the TPS-MRF model. Note that the presence of both positive and negative values in the precision matrix bears similarity to the oscillations and negative values in the equivalent kernels for spline smoothing and Gaussian-process smoothing (Silverman 1984; Sollich and Williams 2005). In the section Derivation of the MRF Weight Matrix for the Thin Plate Spline Approximation in Appendix C (see the HEI Web site), we show the full set of values that define the weight matrix, Q, for the TPS-MRF, including the boundary effects. Figure 12 shows the fitted posterior mean surface under the standard CAR and TPS-MRF models for a simulated dataset (unitless) whose true surface is very smooth. Note the local heterogeneity induced by using the inappropriate CAR model. To demonstrate the need for smooth-process specifications for proxy error, consider the smooth variation in the CMAQ output in Figures 16 and 17. (For more details on Figures 16 and 17, see Spatiotemporal Analysis of CMAQ Output in the Mid-Atlantic Region and Spatial Analysis of CMAQ Output in the Eastern United States, respectively, in section 6 under Simulations.)

Spatiotemporal Latent Variable Model

For a spatiotemporal extension to the model, we allow $g_t, t = 1,\ldots,T$ to vary smoothly in time with a common mean and autoregressive temporal dependence, while we treat

A

	−1	
−1	4	−1
	−1	

B

		1		
	2	−8	2	
1	−8	20	−8	1
	2	−8	2	
		1		

Figure 11. Neighborhood structure and precision-matrix elements for a single row of the precision matrix. The elements are represented spatially relative to the focal grid cell (a nonboundary cell). Matrix A is for a standard MRF model with binary weights, with the weight only on the nearest neighbors in the cardinal directions. Matrix B is for an MRF model with weights that are based on the TPS approximation.

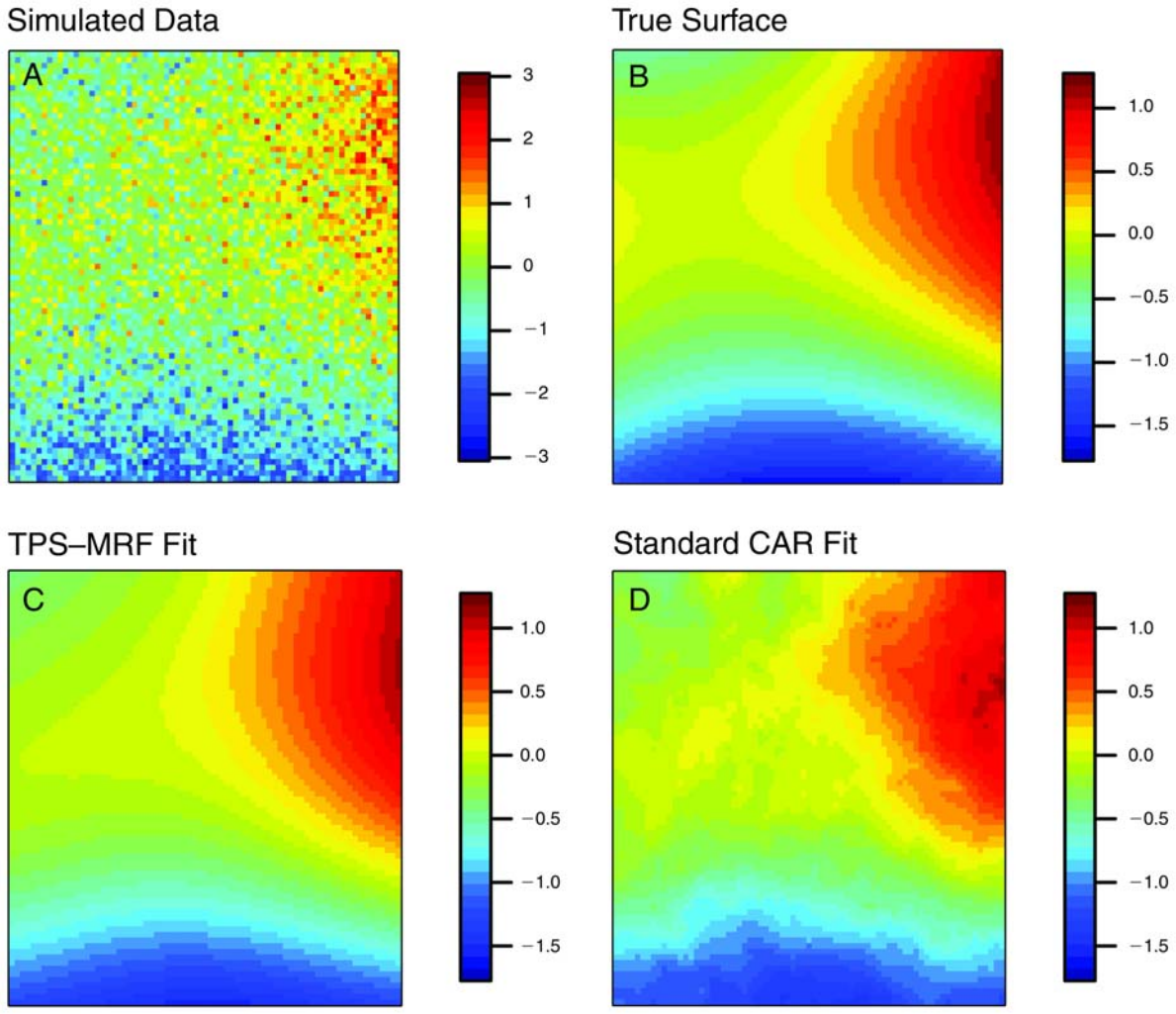

Figure 12. Assessment via simulation of the TPS-MRF and standard CAR models. Image **A** shows the simulated data, which is based on white noise added to a smooth true surface, image **B**. Image **C** shows the posterior mean under the MRF model that approximates a TPS, and image **D** shows the posterior mean under a standard CAR neighborhood structure.

φ_t as exchangeable with a common mean, thus capturing overall spatial variation. The inclusion of a common mean in both cases allows for appropriate estimation of long-term average uncertainty (Stein and Fang 1997).

We start with the basis matrix representation of $g(\bullet)$, $g = Z_g b_g$, $b_g \sim N(0,\sigma^2_{b_g} I)$ (Ruppert et al. 2003; Crainiceanu et al. 2005), where Z_g is the basis matrix, b_g the coefficients, and $\sigma^2_{b_g}$ the variance of the coefficients. We then use an autoregressive model for the coefficient for each knot, $b_{g,k} \sim N(u_k 1,(1-q)\sigma^2_{b_g} R(\rho))$, which after marginalizing over the mean parameters (one for each knot, exchangeable across knots), $u_k \sim N(0,q\sigma^2_{b_g} I)$, gives us the representation $b_{g,k} \sim N_T(0,\sigma^2_{b_g}(qJ + (1-q)R(\rho)))$, where J is a matrix of 1's, $R(\rho)$ is an autoregressive [AR](1) autocorrelation matrix parameterized by decay parameter ρ, and $q \in (0,1)$ weights the exchangeable and autocorrelated contributions to the variance. If we consider all the coefficients across times and knots jointly, we have $g = \{g_1,\ldots,g_T\} = Z^*_g b^*_g$ where Z^*_g is blocked.

For the temporal structure of the discrepancy term, a simple approach is to treat the discrepancy at the time points, φ_t, as conditionally independent given a common mean, φ_{mean}. This gives the model,

$$\begin{aligned} \varphi_t &\sim N_{m-3}(\varphi_{\text{mean}},(\kappa_1 Q)^-) \\ \varphi_{\text{mean}} &\sim N_{m-3}(0,(\kappa_2 Q)^-), \end{aligned} \tag{17}$$

with two precision parameters, κ_1 and κ_2. We then marginalize over the common mean to derive an improper joint prior, $\varphi = \{\varphi_1,\ldots,\varphi_T\} \sim N_{(m-3)T}(0,(D(\kappa_1,\kappa_2)\otimes Q)^-)$, for the discrepancy terms for all the months together, where D^{-1} is a T by T matrix with the diagonal elements $\frac{1}{\kappa_1}+\frac{1}{\kappa_2}$ and the off-diagonal elements $\frac{1}{\kappa_2}$. As an alternative to the specification above, which was used in our analyses, one could specify an autoregressive structure of order 1 [AR(1)] in addition to a common mean, which gives an improper joint prior with similar structure but with D^{-1} having the diagonal elements $\frac{1}{(1-\rho^2_\varphi)\kappa_1}+\frac{1}{\kappa_2}$ and the off-diagonals $D^{-1}_{ij} = \frac{\rho^{|i-j|}_\varphi}{(1-\rho^2_\varphi)\kappa_1}+\frac{1}{\kappa_2}$, where ρ_φ is the correlation parameter.

We used the exchangeable model to analyze the CMAQ model output for the 12 months of 2001 in the mid-Atlantic region, for which there appeared to be little autocorrelated temporal structure to the discrepancy. Given the lack of a relationship between the remote-sensing retrievals and the gold-standard observations seen in the individual monthly analyses (see Spatial Analysis of MODIS AOD in the Mid-Atlantic Region under Examples, below), we did not carry out a spatiotemporal analysis for the remote-sensing data. Note that if we had, computational complexity would have slowed fitting considerably.

Scenarios for the Spatial Scales of Discrepancy

Discrepancy Scenarios One important goal of including the spatial discrepancy term is to understand the scales at which the proxy and the process of interest are well correlated and at which they are poorly correlated. We posit a range of potential relationships. These represent extreme scenarios, so in any practical problem, the reality is likely a combination of scenarios.

- Scenario A (white-noise discrepancy): The spatial structure in the proxy mirrors that in the focal process, but there is fine-scale discrepancy at the scale of pixels that can be treated as white noise. Under this scenario, there is no need for $\varphi(\bullet)$ given the white-noise error structure, $\{e_j\}$. Smoothing over the proxy gives information about the process of interest.
- Scenario B (small-scale discrepancy): The proxy accurately reflects the focal process at large scales, but there is smaller-scale correlated discrepancy. Under this scenario, models without a sufficiently flexible discrepancy term may treat the discrepancy as signal since it is not white noise. If the discrepancy is large enough in magnitude, one would expect the estimate of β_1 to be attenuated, hindering the use of the large-scale information in the proxy. In contrast, with a flexible representation of $\varphi(\bullet)$, the model can treat the discrepancy as having short-range spatial dependence, thereby ignoring this component of variation in the proxy. Note that without dense data, the model cannot correct for the small-scale discrepancy. Also, unless there are large spatial gaps in the observations, it's not clear if using the proxy to help estimate large-scale variation improves upon simply smoothing the gold-standard data.
- Scenario C (large-scale discrepancy): The proxy accurately reflects small-scale variation in the focal process, but there is large-scale discrepancy. In this case, one can correct for this large-scale mismatch by estimating $\varphi(\bullet)$ (needing only moderate amounts of gold-standard data), and one can predict small-scale variation in the focal process from small-scale variation in the proxy. This scenario is accommodated by other

approaches in the data-fusion literature that constrain the discrepancy to vary only at larger spatial scales.

- Scenario D (uninformative proxy): The proxy and focal process are at best weakly related at all scales, and β_1 is near 0. In this case, a model without a flexible discrepancy term may have trouble representing the proxy reasonably. Without a flexible discrepancy term in the case of limited gold-standard data, the focal process prediction may be driven largely by the proxy, and β_1 may be estimated to be far from 0. This allows the variation in the proxy to be explained by a spatial process rather than by white noise (with higher accompanying posterior density, which is analogous to a penalized likelihood setting), even at the expense of increasing the error variance for the observations.

Note that in our analyses we saw evidence that scenario D was in operation for the remote-sensing proxy, whereas for the numerical-model proxy for the eastern United States, scenario B held to some degree, albeit with some large-scale discrepancy as well. Scenario C presumably requires a highly accurate proxy that is somehow only biased because of large-scale discrepancy. While researchers generally hope that a proxy will have high resolution, it's not clear how realistic that is.

Note that when $\varphi(\bullet)$ is estimated as a large-scale process, it is analogous to including spatial variation in the mean, which could be thought of as correcting for spatial bias. When $\varphi(\bullet)$ is estimated as a small-scale process, it is analogous to accounting for spatial variation through the variance term, with a short spatial range. Given the equivalence between a stochastic process in the mean and integrating over that process to move the variation into the covariance, we believe the distinction between representation in the mean and variance was artificial, and therefore we focus on understanding the scale of the discrepancy. In all cases within a given subdomain, one can think of the discrepancy as causing local bias in the predicted surface, if it is treated as signal.

Spatial Discrepancy Diagnostic To assess the spatial scales of the discrepancy term, we proposed to use a spatial variogram-based diagnostic introduced by Jun and Stein (2004) for assessing numerical-model performance relative to observations. Briefly, a variogram quantifies the strength of association of spatial data as a function of distance, calculating and plotting the average squared difference between observations as a function of the distance between them. Jun and Stein (2004) calculated variograms for the model output, observations, and model error (defined as model output minus observations) and proposed the ratio of the model-error variogram to the sums of the variograms for model output and observations as a diagnostic of the spatial variation in the observations captured by the model output as a function of distance. If the model output and observations were independent, then the variogram of the error would be the sum of the variograms for output and observations, so the ratio would be 1 on average. Our analog for assessing the discrepancy based on the output of our statistical model (equation 15) is, as a function of distance, d,

$$R(d) = \frac{\text{Variog}(\varphi; d)}{\text{Variog}(\varphi + \beta_1 L; d) + \text{Variog}(\beta_1 L; d)}, \qquad (18)$$

which we obtained by plugging the analogous quantities from our model into the diagnostic proposed by Jun and Stein (2004), using $\beta_1 L$ as our best estimate of the focal process, scaled to the units of the proxy. The interpretation is the proportion of the variation in the proxy that is accounted for by the discrepancy term, as a function of spatial scale. The diagnostic has the following appealing extremal properties: When $\beta_1 = 0$ or if φ offsets all the variation in the focal process at a given scale, $R(d) = 1$, indicating that all of the variability in the proxy at the scale of distance d is explained by discrepancy. When $\varphi = 0$, or has no variation at a given scale, then $R(d) = 0$, indicating that all of the variability in the proxy at the scale is explained by the process of interest. Ideally, this quantity would be near 0 for all scales, but if there were more discrepancy at large than at small scales, we would expect the ratio to increase with distance, and if there were more discrepancy at small than at large scales, to decrease with distance.

Marginalization and MCMC Sampling

The models described in the subsections Spatial Latent Variable Model and Spatiotemporal Latent Variable Model, under Model and Methods in this section, contain two spatial processes that can trade off in explaining variation in the proxy. In addition, in these models there is cross-level dependence between the spatial-process values and the hyperparameters controlling spatial-process dependence and variation. This raised concerns about MCMC convergence and mixing if the processes were proposed separately or a process and its hyperparameters were proposed separately. Poor mixing in such situations is common, and marginalization over the process or joint sampling of a process and its hyperparameters (Rue and Held 2005; Paciorek 2007) is often used, when possible, to improve mixing. Here, the high dimensionality of the processes complicated matters further. We specifically proposed models that allow marginalization over the process values, with efficient sampling of the process possible off-line because of the process representation.

In particular, for the spatial model (see Spatial Model Structure in Appendix C, available on the HEI Web site), we marginalized first over the MRF for φ and then over the basis coefficients for the regression terms so that we could use sparse matrix manipulations when sampling the remaining parameters with the Metropolis algorithm. When the focal process was represented as a reduced-rank TPS, its process values were part of this second marginalization, and when the focal process was also represented as an MRF (Subnational Model Structure in Appendix C), we marginalized over it in conjunction with φ. In the spatiotemporal model, our structure for φ_t allowed us to integrate over the discrepancy processes and then marginalize over the basis coefficients so that we could again make use of sparse matrix computations in sampling the remaining parameters, as described in Spatiotemporal Model Structure in Appendix C. Note that even with these efforts, sampling was slow; for the space-only models for the mid-Atlantic, run times were on the order of 6 hours, while for the space-only eastern United States models and the spatiotemporal mid-Atlantic models, run times were on the order of 3 to 5 days. Much of this time was required to achieve good mixing of the variance components for the MRF and for the spline basis coefficients without requiring the use of highly informative priors.

SIMULATIONS

Methods

To assess the characteristics of the model structure and the decomposition of variability between the discrepancy term and the focal process, we fit a simplified version of the spatial model to simulated data under a variety of scenarios:

1. Large- and small-scale discrepancies present with $\beta_1 = 1$
2. Large- and small-scale discrepancies with $\beta_1 = 0$
3. No spatial discrepancy, $\beta_1 = 1$
4. Large-scale discrepancy only, $\beta_1 = 1$
5. Small-scale discrepancy only, $\beta_1 = 1$
6. Sparse observations ($n = 40$), with some small-scale discrepancy and a minor amount of large-scale discrepancy, $\beta_1 = 1$

The data were simulated based on the spatial context of our mid-Atlantic analyses, with a similar number of observations, $n = 171$ (except for scenario 6). We performed the simulations with linear effects of a small number of land-use covariates, with the covariates taking the same values as those in our mid-Atlantic domain, but with some of the covariates transformed differently in the data-generation process than in the fitting process. We also simulated residual spatial surfaces, $g(\bullet)$, and, at two different scales, discrepancy, $\varphi(\bullet)$, all based on Gaussian-process realizations (with range parameters chosen to achieve the desired scale of variability). Note that the surface-generation model differed from the spatial-process representation in our model.

The model structure was a simplification of the structure shown in Appendix C (on the HEI Web site). In particular, we included only the linear effects of covariates. The proxy was fully observed over the entire 175×100 4-km grid with no misalignment, but we calculated the likelihood only for land-based pixels. The error structure for both the observations and proxy was a simple homoscedastic normal-error structure, except that multiple observations within a grid cell were modeled as in Appendix C.

Results

In general, the model was able to predict spatial variation in the true process reasonably well, but it did not exploit the information in the proxy effectively. Predictive performance generally decreased when the proxy was included, compared to fixing $\beta_1 = 0$, with the R^2 decreasing from 0.81 when the proxy was ignored to between 0.70 and 0.72 when it was included in scenarios 1, 4, and 5. In scenario 2, the model correctly estimated that $\beta_1 \approx 0$ and did not lose any predictive power. In scenario 3, the R^2 also decreased, to 0.70, with the model attributing some of the signal to the discrepancy term.

This suggests that the added information in the proxy was offset by the model's inability to sufficiently screen out discrepancy in the proxy. Except in scenario 2, including the proxy as a predictor, rather than a second likelihood, did improve prediction (to R^2 values of 0.84, 0.82, 0.90, 0.87, and 0.83 in the first five scenarios, respectively). In general, assuming $\varphi = 0$ or that φ could be represented as very smoothly varying in space led to much worse predictive performance, except in scenario 3, for which the value of R^2 was 0.91 with φ excluded and 0.95 with φ represented smoothly. The one case in which the full model outperformed a model that ignored the proxy was scenario 6, for which the R^2 was 0.34 in the full model and less than 0 in the model without the proxy (i.e., the sum of squared errors was larger than the empirical variation in the true focal process values). In scenario 6, using the proxy as a regressor rather than using the two-likelihood approach again resulted in better performance, with $R^2 = 0.49$. In all scenarios, the estimated β_1 in the full model and the estimated regression coefficient for the proxy when it was used as a regression term were attenuated relative to the true value of 1. Finally, the estimated discrepancy surfaces reasonably matched the true discrepancy (not shown). We conclude that not including the discrepancy term is dangerous when discrepancy is present, but that the proposed

model was not able to exploit the signal present in the proxy to improve prediction when the gold-standard data were relatively dense.

Figure 13 shows the discrepancy scale diagnostic (see Spatial Discrepancy Diagnostic under Scenarios for the Spatial Scales of Discrepancy in this section) for the full model under each scenario. In general, there appeared to be some, but imperfect, information in the diagnostic about the scale of the discrepancy, with the expected relationship with scale given the data generation for a given scenario. That is, when there is more discrepancy at large than at small scales, the ratio should increase with distance, and when there is more discrepancy at small than at large scales, it should decrease. However, even when there was no true discrepancy at a particular scale, the diagnostic estimated discrepancy, which was caused by the attenuated estimates of β_1. This suggests that one should treat the diagnostic as indicative, rather than conclusive, about the scales of discrepancy. Also, the diagnostic results suggest caution in interpreting the diagnostic at the shortest distances, given the observed drop in the diagnostic value even for the scenarios with small-scale discrepancy (scenarios 1, 5, and 6; Figure 13).

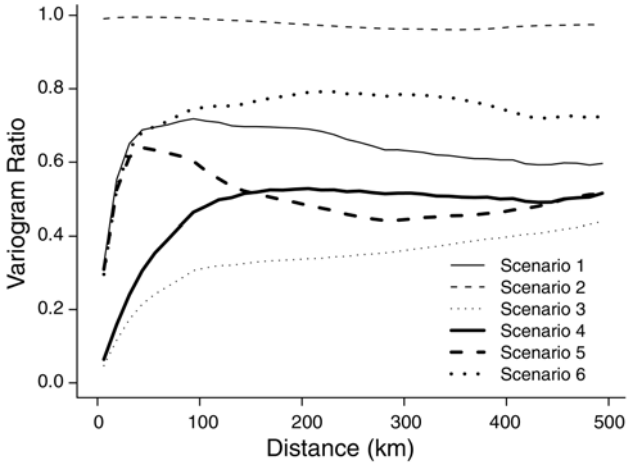

Figure 13. Discrepancy scale diagnostic estimated using the full statistical prediction model fit to data simulated under six scenarios. These are large- plus small-scale discrepancy (1); uninformative proxy (2); no spatial discrepancy (3); large-scale discrepancy only (4); small-scale discrepancy only (5); and sparse data with some small-scale and minor large-scale discrepancy (6).

EXAMPLES

Given our interest in using data fusion to try to improve exposure estimates and thus gain a better understanding of the health effects of chronic exposure to $PM_{2.5}$, and given the general scientific and public health interest in $PM_{2.5}$, we modeled fine PM in the eastern United States and assessed the use of remotely sensed AOD retrievals and CMAQ output as proxies for monthly $PM_{2.5}$. In our first analysis (see Spatial Analysis of MODIS AOD in the Mid-Atlantic Region, below), we fit separate spatial models for the mid-Atlantic region for each month of 2004 with MODIS AOD as our proxy. In contrast to previous work in this domain (see section 5) — which used a knot-based TPS for $\varphi(\bullet)$ and found that the discrepancy was estimated to be increasingly wiggly as more knots were added, until fitting became impossible — here we used the TPS-MRF for $\varphi(\bullet)$. We then considered CMAQ output in the mid-Atlantic as a proxy and jointly fit all 12 months of 2001 (full CMAQ output was only available to us for that year) with our spatiotemporal model, as well as with individual monthly spatial models, to understand the effects of including temporal correlation (see Spatiotemporal Analysis of CMAQ Output in the Mid-Atlantic Region, below). Finally, in our third analysis, we fit separate monthly spatial models for 2001 for the entire eastern United States, to assess the CMAQ proxy at larger scales and better understand the scale decomposition and discrepancy term (see Spatial Analysis of CMAQ Output in the Eastern United States, below). We restricted our analysis to output for every third day and to only those ground monitors that reported observations every day or every third day. This removed temporal misalignment caused by missing $PM_{2.5}$ observations from monitors that reported less frequently than every third day.

For simplicity and because we used monthly averages, we did not transform the outcome, in contrast to other work on $PM_{2.5}$ (Smith et al. 2003;Yanosky et al. 2009), but log or square-root transformations would have been good alternatives. For regression terms that represent sources, it makes more sense to employ additivity on the original scale, whereas for modifying variables, such as meteorologic conditions, log transformation to scale multiplicatively makes more sense. Achieving additivity and multiplicativity in a single model is not easily accomplished. Note that the long right tail of CMAQ output might have been accommodated in the mean for the proxy through φ because the output was spatially correlated, including the values in the tail. As in Paciorek and colleagues (2009), residuals from the various models indicated long-tailed behavior, reasonably characterized by *t* distributions with approximately 5 degrees of freedom, albeit with right skew. Predictive performance suggested that the influence of outliers was not extreme, but the use of a *t* distribution for the observation errors would be worth exploring. However, this would have been somewhat difficult to do in the context of the additive error structure we derived based on components of variability in the $PM_{2.5}$ measurements (see Appendix C available on the HEI Web site). Finally, our exploration suggested that the instrument error variance increased with the magnitude of

the measurement, which is something that would be worth considering more carefully for daily modeling than for the modeling of monthly averages done here.

In the next subsections, we outline our basic model and highlight the results; the model structure and computational details are provided in detail in Appendix C. We report ninefold cross-validation prediction results, retaining a tenth set to assess the degree of overfitting that may have been a result of our model selection process (Draper and Krnjacic 2006). We also show plots that illustrate the estimated spatial surfaces of $PM_{2.5}$ and of the discrepancy and the proposed discrepancy scale diagnostic.

In Table 8, we show correlations of raw observations and the proxy values, based on monthly and yearly aggregations. MODIS AOD was moderately correlated with the observations, whereas CMAQ output in the mid-Atlantic region was poorly correlated with the observations and CMAQ output in the eastern United States was moderately correlated with them. Much of the yearly average correlation for MODIS AOD was induced by the calibration, and our model results (shown in the following subsection) suggest that the correlations did not translate into improved prediction. See section 4 for more discussion of this issue.

Data

We used 24-hour average gravimetric FRM measurements from the U.S. EPA AQS with parameters 88101 and 88502, including IMPROVE monitors. For likelihood calculations, monitor locations were associated with the cell in the 4-km grid in which they fell.

We used MODIS AOD retrievals from 2004, matching the analysis in section 5. MODIS is on the Terra satellite platform, whose polar orbit has been providing full coverage of the globe at regular intervals, beginning in March 2000, with retrievals in the eastern United States at a constant daily time point (10:30–10:45 AM local time). MODIS provided AOD retrievals at a nominal resolution of 10 km with each location covered every 1 to 2 days (Wang and Christopher 2003; Engel-Cox et al. 2004). AOD cannot be retrieved below clouds, so cloud-filtering algorithms use the infrared portion of the spectrum to detect and omit obscured observations (Engel-Cox et al. 2004). Errors and uncertainties in the filtering could have led to erroneous AOD retrievals, and high surface reflectivity could also have prevented retrievals.

We obtained CMAQ output for 2001 from a 36-km model run over the entire United States, provided by Atmospheric and Environmental Research (AER) and funded by the Electric Power Research Institute (EPRI). Hourly CMAQ output was available for a regular 36-km grid (a Lambert conformal conic projection) with 14 vertical levels. CMAQ relies on meteorologic and emissions-inventory inputs, and was designed for short-term air quality applications, so its long-term average output has received limited evaluation.

We used GIS-based and meteorologic covariates to help explain $PM_{2.5}$ variation, following Yanosky and colleagues (2008). Covariates that may have helped predict $PM_{2.5}$ at fine spatial scale included distance to major roads in two road classes (Class 1 and Class 2), which were calculated for each monitor location and included as smooth spline terms $\beta_{y,p}(x_{y,p})$, $p = 1,2$ (see equation 15). Other covariates were calculated using a GIS at the resolution of the 4-km grid and modeled using the smooth spline terms $\beta_{L,p}(x_{L,p})$, $p = 1,\ldots6$ (see equation 16). These included road density for three road classes (Class 1, Class 2, and Class 3), population density, and elevation at the cell centroid. As a measure of the non-point-source emissions in each cell, we assigned to the cell the density (total emissions divided by county area) of the area-level primary $PM_{2.5}$ emissions from the U.S. EPA's 2002 NEI in the county of the cell centroid.

We also used the point locations of 2002 primary $PM_{2.5}$ emissions from the 2002 NEI. The distance to point-source emissions (from sources that emitted more than 5 tons per year) was handled using a term that accounted for the

Table 8. Simple Correlations of Observations and Proxy Values (All Available Values, Not Matched by Day) Using Proxy Data for the Pixels in Which $PM_{2.5}$ Observations Fell

Time Scale	Mid-Atlantic, 2004, MODIS AOD	Mid-Atlantic, 2001, CMAQ	Eastern U.S.[a], 2001, CMAQ
Monthly[b]	0.58 (0.60 uncalibrated)	0.26	0.52
Yearly[c]	0.60 (0.26 uncalibrated)	0.095	0.56

[a] Locations east of 100° W longitude.

[b] Only monthly averages for location–months with at least five daily observations are included.

[c] Yearly results include only locations having at least 9 months with at least five daily observations. Yearly averages are calculated as the mean of the 12 monthly averages.

effect of point-source emissions within 100 km. This term represented the effect of multiple point sources at a given receptor location (i.e., a monitor or prediction point) as the emission-strength-weighted sum of a smooth distance effect evaluated for each individual source–receptor pair. The distance function is a universal function representing the effect of a single source of unit strength on a receptor as a smooth function of the distance between the source and receptor. It was estimated from the data, based on a new statistical approach that leverages the additive structure of mixed-model representations of splines, as described in Appendix D (available on the HEI Web site). This term was used to describe within-grid-cell variability of monitor observations and, using an approximation to the integral of the effect over the entire cell, grid-cell average values.

Meteorologic variables were based on the NARR fields (Mesinger et al. 2006), which were available at 32-km resolution every 3 hours. For each 3-hour value and each grid cell, we computed an IDW average of the NARR values from the four NARR points nearest the cell centroid. We then averaged the values for the month. Our second statistical model used wind speed and temperature, but we also considered the RH, PBL height, mean sea-level pressure, and precipitation.

Table 9 shows summary statistics for the monthly average $PM_{2.5}$ observations and CMAQ $PM_{2.5}$ predictions used for the three analyses described next. More details on the data are available in Appendix A.

Table 9. Summary Statistics for Monthly Averages of $PM_{2.5}$ Observations (μg/m^3) and CMAQ-based $PM_{2.5}$ Predictions[a,b]

Data Source	Time Period	Sample Size	Median	(5%, 95%)
$PM_{2.5}$, mid-Atlantic, 2004	All year	1,955	12.9	(7.9, 20.9)
	Winter	485	11.0	(6.9, 16.7)
	Spring	488	13.0	(8.5, 19.0)
	Summer	489	16.9	(11.6, 24.3)
	Fall	493	11.2	(7.4, 18.2)
$PM_{2.5}$, mid-Atlantic, 2001	All year	2,061	14.0	(9.1, 23.3)
	Winter	512	14.2	(8.3, 23.7)
	Spring	510	12.9	(8.8, 17.8)
	Summer	517	19.3	(10.9, 25.8)
	Fall	522	12.9	(9.2, 17.4)
CMAQ, mid-Atlantic, 2001	All year	2,232	16.0	(10.3, 30.0)
	Winter	558	16.9	(9.5, 36.1)
	Spring	558	14.0	(9.7, 26.8)
	Summer	558	17.1	(11.5, 30.0)
	Fall	558	15.9	(11.0, 29.0)
$PM_{2.5}$, eastern U.S.[c], 2001	All year	8,309	12.7	(7.1, 21.0)
	Winter	1,961	12.2	(6.6, 21.8)
	Spring	2,079	12.5	(7.4, 17.8)
	Summer	2,096	15.4	(9.1, 23.3)
	Fall	2,173	11.6	(6.6, 17.0)
CMAQ, eastern U.S.[c], 2001	All year	38,448	13.2	(5.6, 24.4)
	Winter	9,612	12.8	(6.3, 26.5)
	Spring	9,612	13.0	(5.5, 23.8)
	Summer	9,612	13.7	(4.9, 23.3)
	Fall	9,612	13.5	(5.9, 24.0)

[a] Excluding CMAQ grid cells with more than 40% overlap with the Atlantic Ocean or Lake Erie.

[b] AOD values are not included because these values were processed before use in the statistical models presented in section 6.

[c] Locations east of 100° W longitude.

Spatial Analysis of MODIS AOD in the Mid-Atlantic Region

For this analysis, we used the 4-km base grid (175×100 4-km cells for the mid-Atlantic region) as the resolution of $L(\bullet)$, $g(\bullet)$, and $\varphi(\bullet)$. AOD was misaligned with respect to this grid, and a further complicating factor related to the MODIS instrument was that for different satellite orbits on different days, the pixels shifted spatially. Therefore, we considered the overlap of all the pixels in an orbit over the 4-km grid and assigned to each grid cell, s, the value of the MODIS pixel in which the cell centroid fell. Using the retrievals assigned to each cell, we computed a monthly average for each cell. More sophisticated approaches are possible (Mugglin et al. 2000), but for our purposes, this ad hoc realignment retained the essential character of the AOD retrievals and reduced computations. Missing AOD retrievals because of cloud cover were naturally handled by treating AOD as data, using equation 15. We attempted to account in part for missingness due to cloud cover (section 3) by including in the additive mean for the proxy a regression term, $\beta_a(x_a)$, that is a smooth function of average cloud cover over the month for each location, from the GOES cloud screening retrieval, thereby making a missing-at-random assumption. $g(\bullet)$ was modeled using the TPS mixed-model representation (Kammann and Wand 2003; Ruppert et al. 2003). We did not use meteorologic covariates (obtained from the NARR data product), as these proved to have high concurvity with the estimated residual spatial variation, g. This concurvity was driven by the fact that the covariates varied primarily at large scales, which was presumably caused in part by the coarse 32-km resolution of the NARR grid.

We considered both raw MODIS AOD and a "calibrated" MODIS AOD that adjusted off-line for the effects of boundary layer height and the RH, as well as large-scale temporal (seasonal) and spatial adjustments, based on a comparison of daily $PM_{2.5}$ and AOD values from colocated monitors and AOD pixels (sections 3, 4, and 5). Given the poor results with MODIS AOD in this part of the analysis and the poor results with GASP AOD in other parts of the analysis (see section 5), we did not include GASP AOD in our assessment here.

Once the model was fit, the estimated regression coefficients for the latent $PM_{2.5}$ as an explanatory variable for the proxy were essentially 0 (Figure 14A), indicating that the estimation found no relationship between the proxy and the estimated latent process, thus discounting the proxy in predicting $PM_{2.5}$. The spatial discrepancy plot (Figure 14E) indicates that at all scales the proxy was explained by the discrepancy term rather than the latent process (see the next subsection for more discussion of Figure 14). These results are consistent with those found in section 5. Figure 15 shows predictions from the model when AOD was excluded and when included; the predictions were very similar and consistent with the estimated values of β_1 being near 0. Given this, it is not surprising that the cross-validative predictive assessment indicated that including the proxy did not improve predictive performance (Table 10). Note that yearly prediction appeared slightly worse when the proxy was excluded, but this result was likely within the uncertainty of the predictive assessment, which was difficult to quantify given the correlation structure of the data. The results were qualitatively similar when we considered only held-out monitors in more rural areas, suggesting that AOD retrievals were not adding information in areas far from monitors (Table 11).

Spatiotemporal Analysis of CMAQ Output in the Mid-Atlantic Region

For this analysis, we used the spatial domain described in the previous subsection, but we used the 12 months of 2001. We used the CMAQ 36-km-square grid over the domain as the grid resolution of $\varphi(\bullet)$ (19×11 cells), relating L to the proxy on that grid with a mapping matrix P_A that calculated a weighted average of the latent process values for the 4-km cells falling in each 36-km cell. We used the spatiotemporal formulation from the subsection Spatiotemporal Latent Variable Model, under Model and Methods in this section, to build dependence among the months for the $g_t(\bullet)$ surfaces and to build exchangeable dependence among the months for the $\varphi_t(\bullet)$ surfaces. We specified a priori independent regression parameters for each month, represented by $\beta_{1,t}$. Given the sensitivity of the model to the temporal dependence structure reported at the end of this subsection, we also fit simple monthly spatial models in the style shown in the subsection Spatial Latent Variable Model in this section, but otherwise retained the features discussed here.

For the spatiotemporal analysis, in addition to the temporally invariant covariates considered in the subsection Spatial Analysis of MODIS AOD in the Mid-Atlantic Region in this section, we considered the NARR-based meteorologic covariates. Since the observations did not occur on all the days in each month, we calculated average values for the meteorologic variables (using IDW interpolation based on the four NARR pixels nearest each observation) solely for the days in the month with observations, averaging the 3-hour NARR values, for use in the $PM_{2.5}$ model likelihood. For prediction and the CMAQ output likelihood, we used IDW interpolation to the grid-cell centroids, averaging over all the days in the month. Including meteorology did not improve prediction, perhaps because it varied only at the

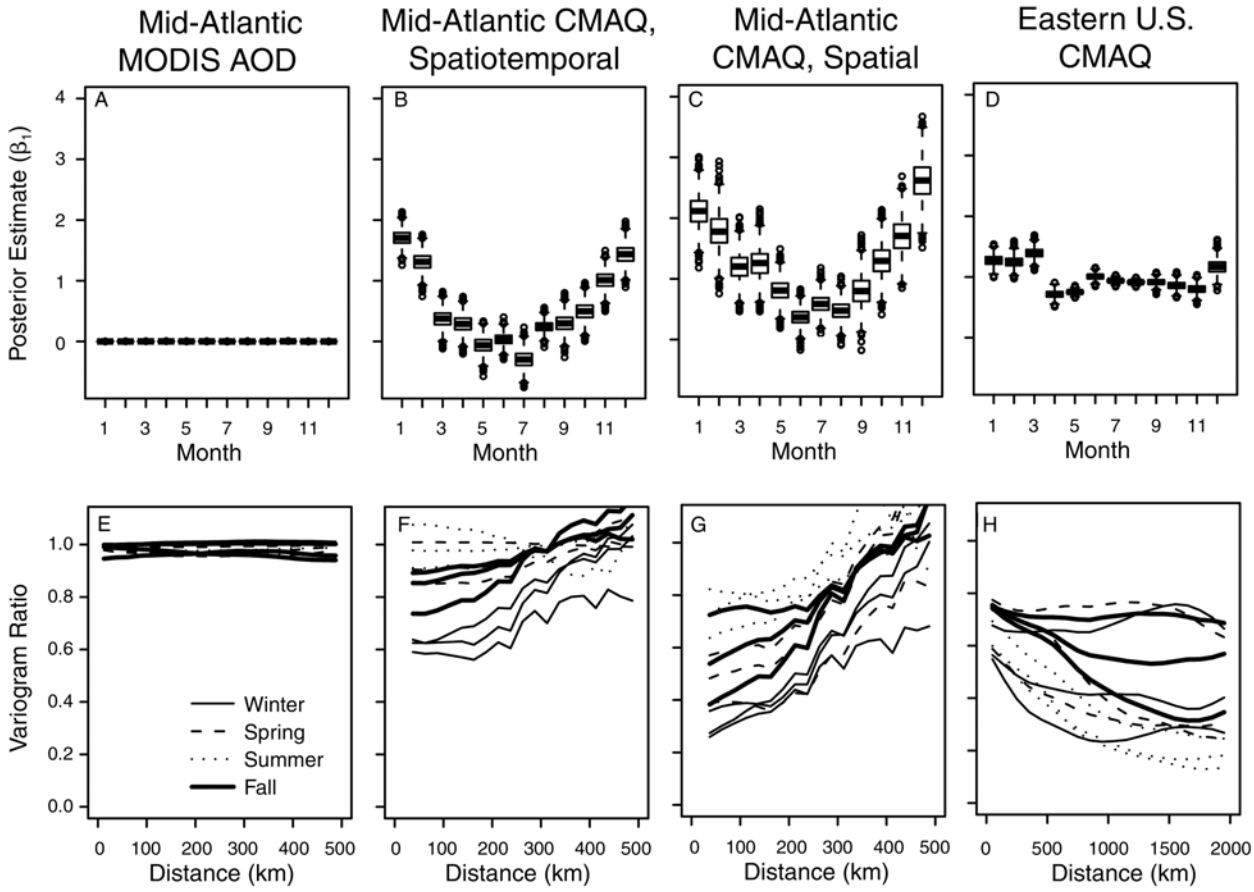

Figure 14. Boxplots of posterior estimates of the multiplicative bias, β_1 (top row) by month, and discrepancy scale diagnostic plots (bottom row), based on the statistical prediction model (equation 15). Plots are for MODIS AOD as the proxy in 2004 for the mid-Atlantic using the spatial model (far-left panels); CMAQ output as the proxy in 2001 for the mid-Atlantic using the spatiotemporal model (middle-left panels); CMAQ output as the proxy in 2001 for the mid-Atlantic using the spatial model (middle-right panels); and CMAQ output as the proxy in 2001 for the eastern United States using the spatial model (far-right panels).

Table 10. Cross-Validation R^2 (Root Mean Squared Prediction Error) for Monthly and Yearly Average $PM_{2.5}$ from Statistical Prediction Models, With and Without Proxy Information

Time Scale / Proxy Inclusion	Mid-Atlantic, 2004, MODIS AOD	Mid-Atlantic, 2001, CMAQ, Spatiotemporal Model	Mid-Atlantic, 2001, CMAQ, Monthly Spatial Models	Eastern U.S., 2001, CMAQ
Monthly prediction[a]				
With proxy	0.806 (1.80)	0.640 (2.60)	0.755 (2.14)	0.827 (1.71)
Without proxy	0.808 (1.79)	0.686 (2.42)	0.777 (2.04)	0.826 (1.72)
Yearly prediction[b]				
With proxy	0.670 (1.00)[c]	< 0[d] (1.97)[c]	0.503 (1.32)[c]	0.800 (1.21)
Without proxy	0.650 (1.03)[c]	0.169 (1.70)[c]	0.584 (1.20)[c]	0.835 (1.09)

[a] Including monthly averages based on at least five daily observations.

[b] Including yearly averages (averages of monthly values) based on at least 9 months with at least five daily observations.

[c] Excludes one site outside Pittsburgh just downwind of a major industrial facility.

[d] The squared correlation of held-out data and predictions is 0.473, but a graph of the observations vs. predictions is not centered on the one-to-one line, so the error sum of squares exceeds the total sum of squares.

Table 11. Cross-Validation R^2 for Monitors Generally Isolated from Other Monitors, for Monthly and Yearly Average $PM_{2.5}$ from Statistical Prediction Models, With and Without Proxy Information[a]

Time Scale / Proxy Inclusion	Mid-Atlantic, 2004, MODIS AOD	Mid-Atlantic, 2001, CMAQ, Spatiotemporal Model	Mid-Atlantic, 2001, CMAQ, Monthly Spatial Models	Eastern U.S., 2001, CMAQ
Monthly prediction[b]				
With proxy	0.830 (1.73)	0.626 (2.74)	0.721 (2.36)	0.739 (2.12)
Without proxy	0.830 (1.72)	0.588 (2.87)	0.761 (2.19)	0.781 (1.94)
Yearly prediction[c]				
With proxy	0.669 (1.17)	< 0[d] (1.99)	0.254 (1.67)	0.710 (1.58)
Without proxy	0.624 (1.25)	< 0[d] (1.94)	0.395 (1.51)	0.826 (1.22)

[a] Those in 4-km-grid cells with fewer than 187.5 people per square km, equaling 3000 people per cell.

[b] Including monthly averages based on at least five daily observations.

[c] Including yearly averages based on at least 9 months with at least five daily observations.

[d] A graph of the observations vs. predictions is not centered on the one-to-one line, so the error sum of squares exceed the total sum of squares.

large scales (given the monthly averaging and the resolution of the NARR grid) at which $g_t(\bullet)$ could stand in for it.

For both the spatiotemporal formulation and the individual spatial models fit separately to each month, the estimated values of β_1 were larger in winter than in summer (Figure 14B and C). The discrepancy scale diagnostic plot suggests that more of the variation in the proxy at small scales was related to the latent process than it was at large scales (Figure 14F and G). This likely relates to CMAQ having correctly placed hot spots around the urban areas (Figure 16). However, the model-estimated discrepancy surfaces indicated the need to substantially correct the CMAQ output at larger scales and also to discount the CMAQ-estimated hot spot in southeastern Pennsylvania and moderate the hot spot over New York City, as these were not supported in the observations (Figure 16). Figure 16 also illustrates that CMAQ-estimated $PM_{2.5}$ showed similar spatial patterns from month to month (represented in the exchangeable structure of the monthly discrepancy processes). This consistency over time contrasted with the spatiotemporal heterogeneity seen in raw observations, which suggests that the CMAQ model was not adequately capturing changing spatial patterns over time. For both models, cross-validative predictive performance when the proxy was excluded was as good as or better than it was when the proxy was included (Tables 10 and 11).

For the mid-Atlantic spatial domain, our spatiotemporal model appeared to predict less well than simple spatial models, particularly for the yearly averages, even when the proxy was excluded (Table 10). Apparently, by forcing the same covariate relationships for all months and by smoothing and shrinking the spatial residual terms, $g_t(\bullet)$, we greatly reduced predictive performance. More investigation is warranted to better understand why smoothing over time did not help with predictive performance, but presumably it was related to model misspecifications in the more complicated spatiotemporal model that did not match the structure of the data.

Spatial Analysis of CMAQ Output in the Eastern United States

For this analysis, we used the 4-km grid over the eastern United States (roughly the area east of 100° W longitude), giving a grid of 669×677 cells for representation of the covariates available at the scale of the 4-km grid. We used the 36-km CMAQ grid of 73×77 cells for $\varphi(\bullet)$. For this larger domain with more complexity in the pollution surface, we represented $g(\bullet)$ on this same 36-km grid using the TPS-MRF representation because the mixed-model TPS would require a computationally burdensome increase in the number of basis coefficients. To calculate the $PM_{2.5}$ model likelihood, we used the covariate values for the cell in the 4-km grid that each observation fell in and used the value of g from the cell in the 36-km grid in which the observation fell. For the CMAQ model likelihood, the CMAQ value for a given 36-km-grid cell was related to the covariate effects from the 4-km-grid cells by weighted averaging with weights based on the overlap between a CMAQ-grid cell and the cells of the 4-km grid.

We found that β_1 was estimated to be large (between 3 and 7), φ was strongly negatively correlated with L, and φ had very large negative values to offset the large positive

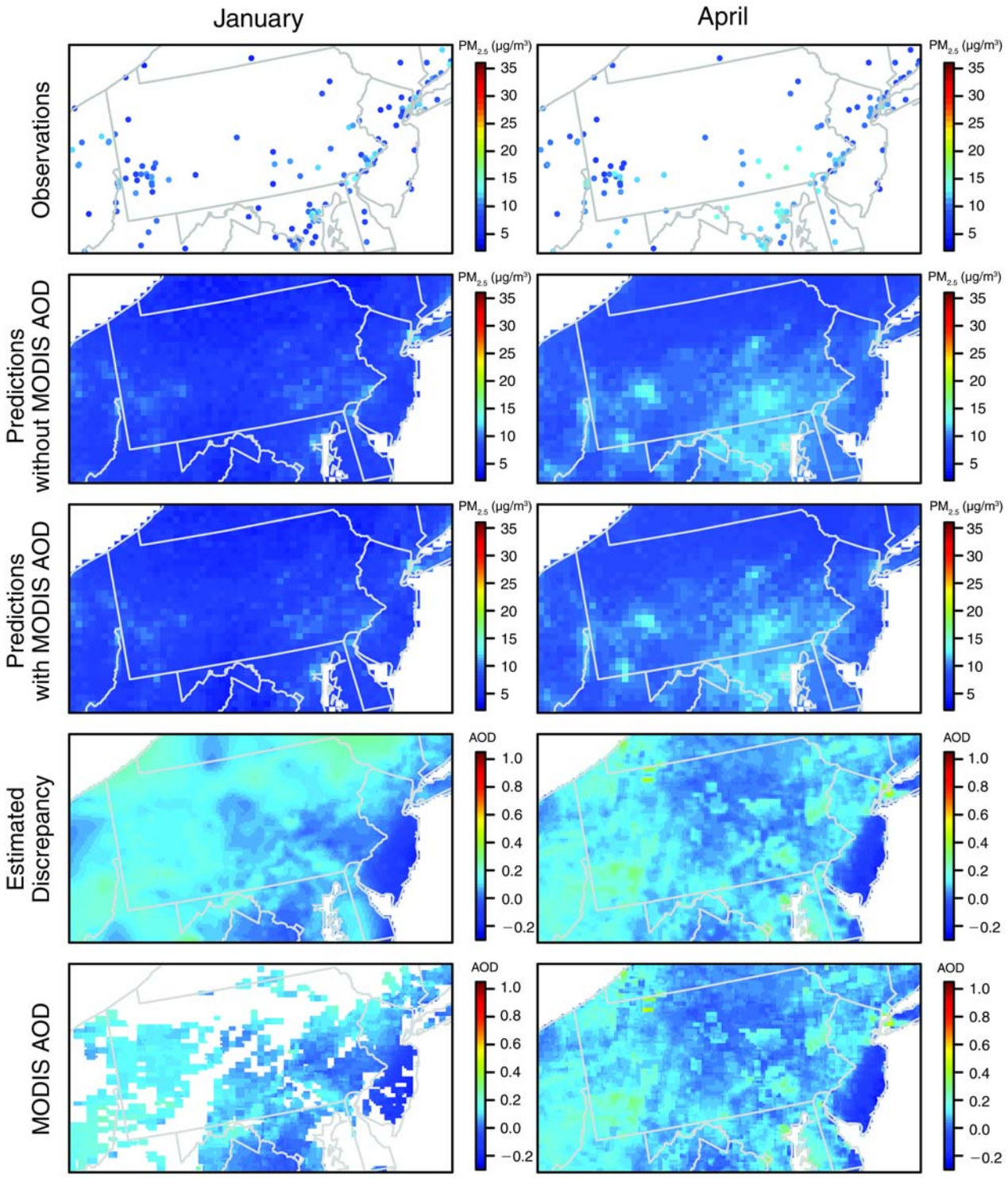

Figure 15. Spatial modeling of $PM_{2.5}$ using MODIS AOD in the mid-Atlantic region in 2004. The figure shows $PM_{2.5}$ observations (first row); model $PM_{2.5}$ predictions, excluding AOD (second row); model predictions including AOD (third row); estimated discrepancy in the model with AOD (fourth row); and AOD values (fifth row) for 4 months in 2004: January, April, July, and October. (*Figure 15 continues next page.*)

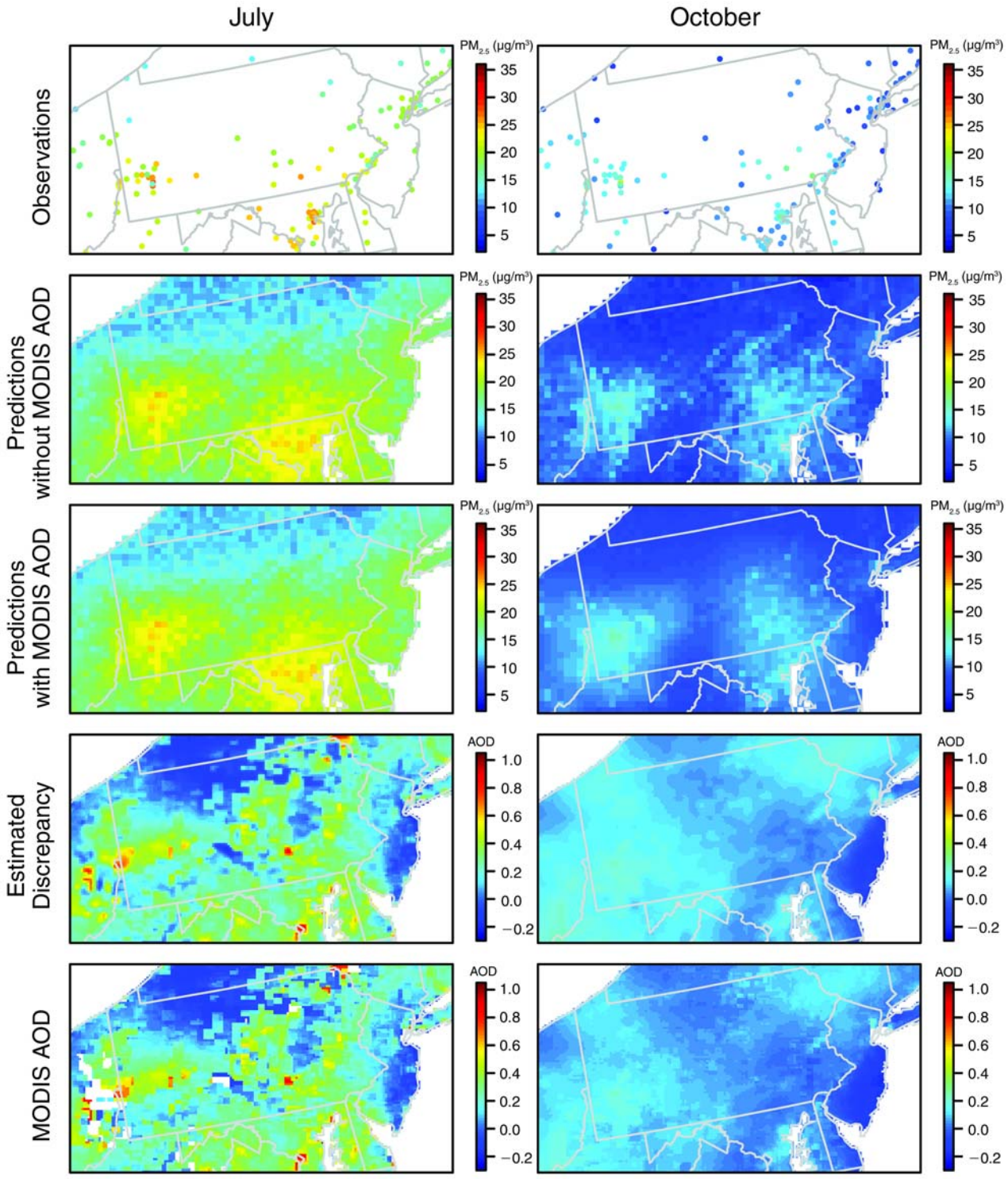

Figure 15. (*Continued*).

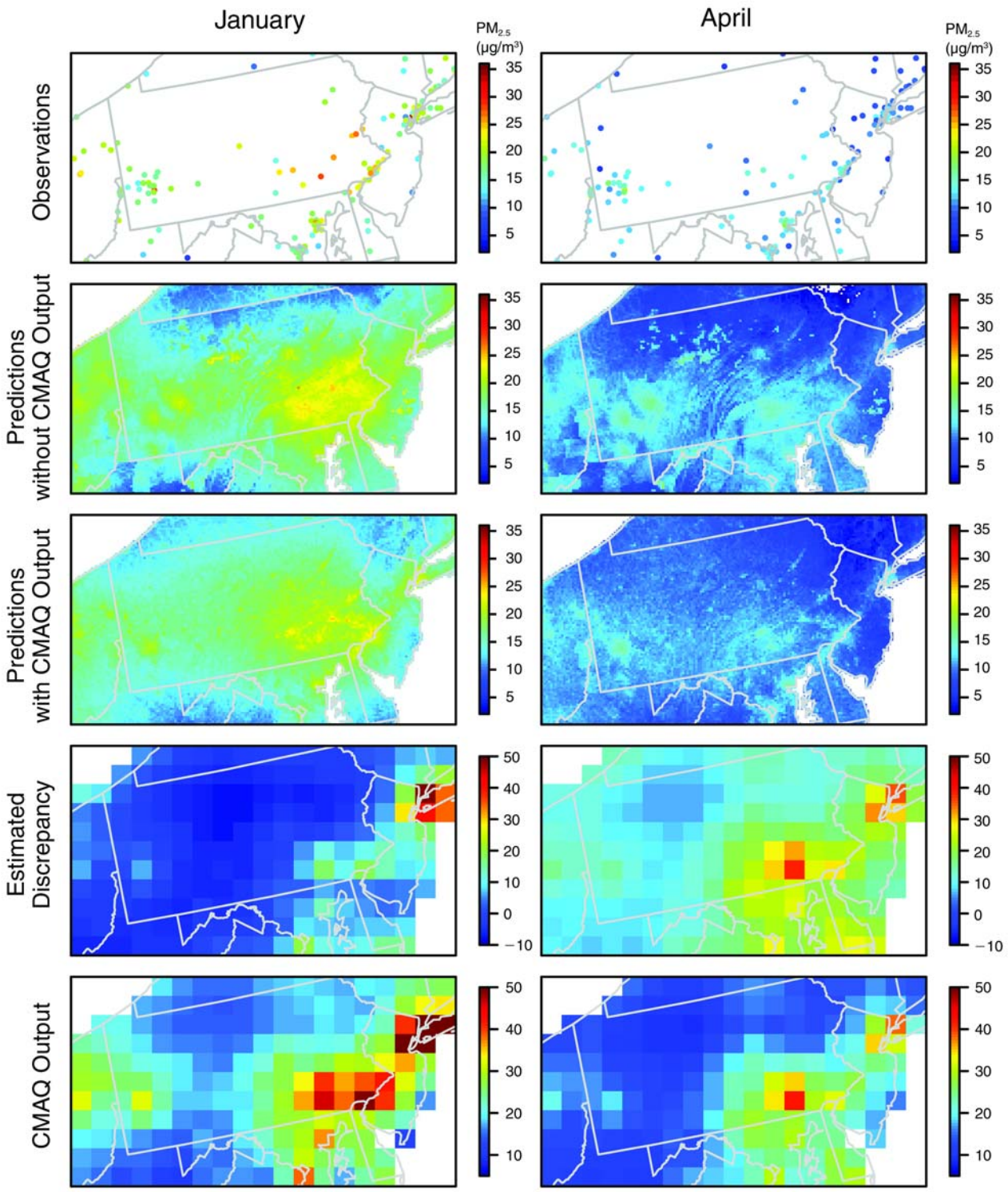

Figure 16. Spatiotemporal modeling of $PM_{2.5}$ using CMAQ output in the mid-Atlantic region in 2001. The figure shows $PM_{2.5}$ observations (first row); spatiotemporal model predictions of $PM_{2.5}$, excluding CMAQ values (second row); model predictions including CMAQ values (third row); estimated discrepancy in the model with CMAQ values (fourth row); and CMAQ values (fifth row) for 4 months in 2001: January, April, July, and October. Note that in January, some CMAQ values larger than 50 µg/m³ (up to 75 µg/m³) in the New York City area are truncated to 50 µg/m³. (*Figure 16 continues next page.*)

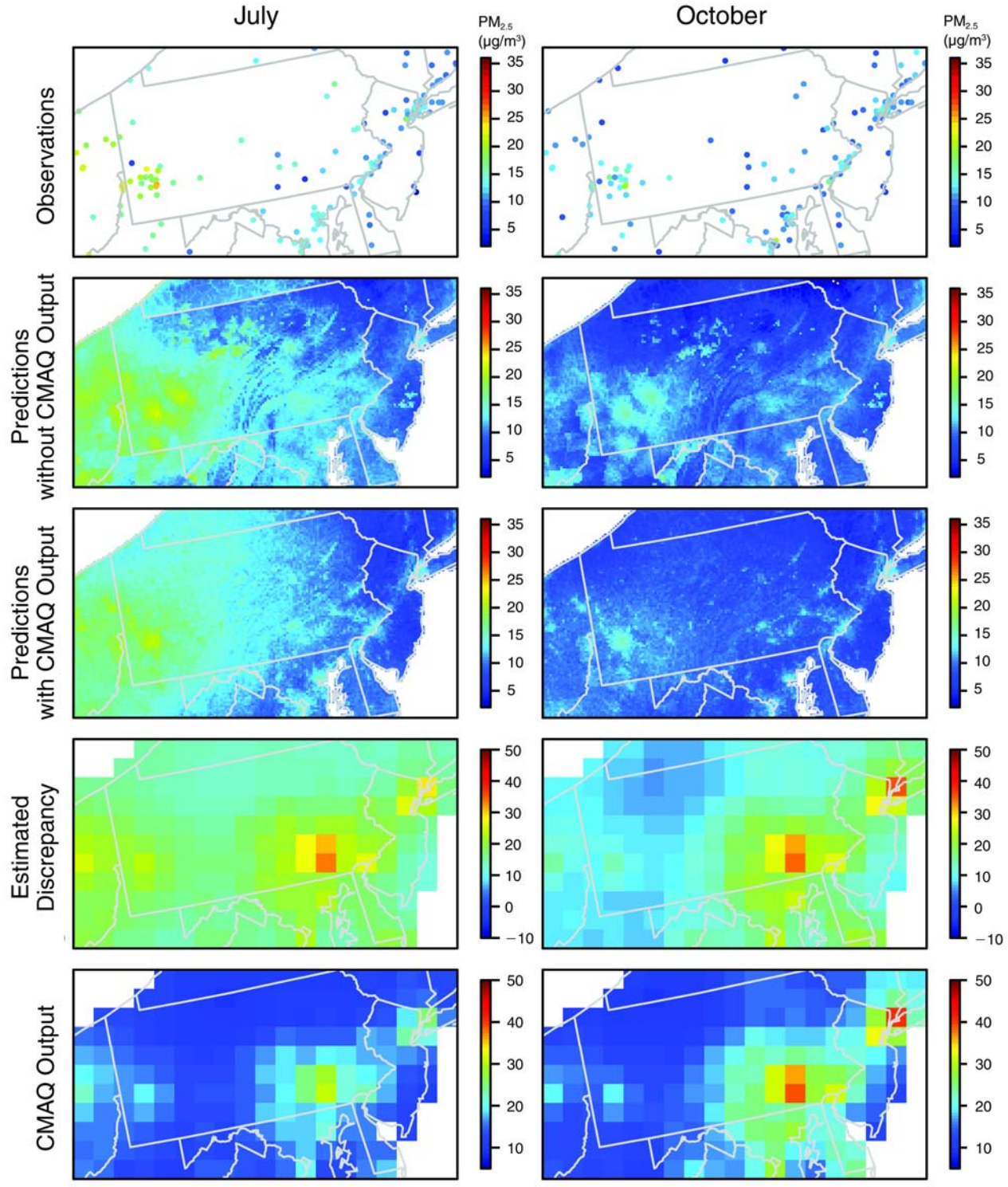

Figure 16. (*Continued*).

values of $\beta_1 L$. This appeared to be driven by the CMAQ model's overprediction in some urban areas, with CMAQ estimating the urban-to-rural gradient as being much stronger than was apparent in the observations. φ then adjusted for the effects of large values of β_1 elsewhere in the spatial domain. To address this in an ad hoc manner and identify φ as the orthogonal variation in the proxy not accounted for in the latent process, we carried out an ad hoc orthogonalization of φ and L within each step of the MCMC method. While a formal orthogonality constraint on φ and L was technically appealing, the simple ad hoc approach was effective in practice, and predictive results were similar with and without the orthogonalization. Other approaches to the problem might involve letting β_1 vary with covariates such as population density or truncating the very high CMAQ predictions.

Using this orthogonalized specification, the estimated values of β_1 were generally near 1, as one would hope (Figure 14D). In this larger domain, the variability in the CMAQ output at larger scales was more associated with the latent process than was the variability at smaller scales, suggesting that CMAQ was better able to resolve regional variability than more local variability (Figure 14H). This is consistent with the discrepancy surfaces shown in Figure 17, where one can see that the model corrected for hotspots at the scale of states or portions of states that showed no evidence of being real hotspots based on the observations, such as over Iowa and southern Minnesota and eastern North Carolina. Once again, the use of CMAQ output as a proxy did not improve predictive ability (Tables 10 and 11). Predictive performance was better than in the mid-Atlantic domain because of the greater range of natural variability in $PM_{2.5}$ over this much larger area. When we considered the use of meteorologic variables, the relationships were often the opposite of those expected based on scientific grounds, while prediction was little affected, once again presumably because of identifiability issues with respect to the residual spatial term, *g*.

As a final assessment, we included CMAQ output as a simple regression term and found the cross-validation R^2 (root mean squared prediction error [RMSPE]) to be 0.849 (1.60) for the monthly prediction and 0.849 (1.05) for the yearly prediction; both values were slightly better than for modeling without the proxy (Table 10) but suggested that the CMAQ output provided limited additional information given the other terms in the model. The slight improvement was matched when restricting predictive assessment to more rural sites, with a cross-validation R^2 (RMSPE) of 0.798 (1.87) for the monthly prediction and 0.834 (1.19) for the yearly prediction (compare these results to those in Table 11). The posterior mean regression coefficients for CMAQ-estimated $PM_{2.5}$ were between 0.48 and 0.89 for the 12 months, with the average of those 12 posterior means being 0.67.

We also considered whether CMAQ might be more helpful in a setting with sparse observations (such as was seen in our simulations) by artificially using only 100 monitors for our training dataset. However, we found that prediction for the remaining monitor locations was very poor because, with sparse $PM_{2.5}$ observations, we were not able to adjust for the CMAQ discrepancy as well as we could with more dense data. As above, including CMAQ output as a regressor in this case did help to improve prediction slightly as compared to a model that only made use of the observations.

DISCUSSION

Our model is a spatial latent variable model, in which the two spatial latent variables represent a decomposition of the proxy into signal, for the process of interest, and noise. The model calibrates the proxy to the observations implicitly; to do this well, a sufficient number of gold-standard observations is required. Identifiability in the model is an obvious concern given the attempt to decompose the proxy into signal and noise, with the unknown focal process treated as a regressor in the model, such as is done in a measurement error model. We fit a penalized model that identifies the discrepancy and the focal process based on a tradeoff between goodness of fit for the observations and for the proxy values and penalization of the latent spatial processes. This tradeoff is likely quite sensitive to model specification and to the relative richness of observational and proxy data and could be prone to fitting effects not fully understood in a complicated hierarchical setting. As a related comment, it would be worthwhile to fit a more robust likelihood specification to ensure that outliers in either the observations or the proxy (of most concern in our examples with regard to large values of the CMAQ proxy) do not overly influence the results.

Despite the concerns about identifiability and sensitivity to model structure, our view is that this decomposition is the fundamental task when using proxies. One doesn't know the quality of the proxy and must assess its quality, ideally as a function of scale. A priori constraints on the flexibility of the discrepancy process incorporated into current data fusion approaches make strong assumptions about proxy quality, and we believe that such assumptions contribute to overoptimism about the usefulness of proxy information and about the degree of certainty in resulting predictions based on the data fusion approach. Our model attempts to assess proxy quality and to make use of available information in the proxy without prejudging

whether the spatial correlation in the proxy reflects spatial signal in the process of interest.

We did our best to account for errors and local variability in the observations, as well as for spatial and temporal misalignment between observations and the proxy, to avoid concluding that a proxy was not helpful because of noise or fine-scale variability in the gold standard, but we could not completely rule out this possibility. However, we note that comparisons of daily maps of proxy values and observations indicated a great deal of mismatch between the two, particularly on days when the observations showed large-scale variation but little local variation. On such days, the patterns in true $PM_{2.5}$ were well identified from the observations, yet the proxies often did a poor job of capturing that variation. For AOD, this might have been related in part to the reliance on a snapshot at a single time to try to characterize a full-day average. It is possible that the proxy variables might have added useful information in the absence of our GIS-based covariates, but it would be difficult to account for local variation in the observations that was not necessarily reflected in a gridded proxy without the use of covariates in the model. Part of the purpose of our covariates was to account for the local variation that contributed to the discrepancy between the proxy and observations. Local variation that is not accounted for in the model structure is likely to mask any real relationship between the proxy and the true process of interest at the grid scale and may lead to a discounting of real information in the proxy.

In our spatiotemporal modeling, we used a simple spatiotemporal structure with the discrepancy process exchangeable across months but with no month-to-month temporal correlation, but we also suggest an AR(1) structure with equal computational feasibility as a potential alternative in future work. Berrocal and colleagues (2009) and Sahu and colleagues (2009) specified independence for their analogous latent processes, while McMillan and colleagues (2010) used a spatiotemporal b-spline basis. As seen in our results, modeling the dependence between months did influence our estimation, with the result being poorer predictive performance, suggesting that a simpler spatial model that is more straightforward to fit, interpret, and understand might be a better choice, though it wouldn't capture all the temporal dependence in the data.

This study's approach was to represent proxy information as data through an additional likelihood, extending the research of Fuentes and Raftery (2005). Berrocal and colleagues (2009) proposed instead to regress on the proxy. In our simulations, we found that, relative to the two-likelihood model, regressing on the proxy in some cases improved predictive ability and never decreased it; and it had the added advantage of greatly reduced computational complexity. So if one is purely interested in prediction and not in understanding the structure of the proxy, a simple regression approach may be best. An important drawback is that for proxies such as remote-sensing retrievals affected by cloud cover, missing observations are a problem. This then requires an imputation of some sort, which brings one back to modeling the proxy.

The second potential drawback is more subtle and involves difficulties in decomposing the proxy into scales when the relationship with the process of interest varies by scale. If the proxy captures small-scale patterns well but misses large-scale patterns, then the additive spatial-process term used in Berrocal and colleagues (2009) could adjust for this discrepancy, leveraging the proxy to improve prediction of small-scale variation. A potential concern arises when the proxy captures large-scale patterns well, but the small-scale patterns in the proxy are predominantly noise with respect to the process of interest, a likely scenario in real-world applications. In such a situation, the model would likely estimate a large regression coefficient for the proxy because of the association of the proxy and the observations at a large scale. The resulting prediction would be reasonable at the large scale but would also include all the small-scale spatial discrepancy in the proxy, which would be interpreted by the analyst as signal, with the proxy assumed to be of high quality at small scales. The spatial discrepancy term in the model would not be able to correct for this small-scale discrepancy because the sparsity of the observations wouldn't allow for estimation of and adjustment for this discrepancy. In such a situation, we'd like to decompose the scales of variability in the proxy, but the regression approach conditions on all the scales of the proxy. Alternatively, the small-scale discrepancy in the proxy might act as measurement error, causing attenuation in the regression coefficient, potentially to near 0, limiting any information gain from the correspondence between the proxy and the process of interest that occurs at larger spatial scales.

From an interpretation standpoint, the two-likelihood approach cleanly distinguishes between discrepancy and signal, at least in principle. A simpler alternative would be to explicitly decompose the proxy into two or more scales and include each component as a separate regression term, allowing the model to inform which scales are correlated with the observations. How to do this decomposition is an open question, but a simple approach would be to successively smooth the proxy with spatial smooth terms with fixed degrees of freedom.

We note that CMAQ output at nominal resolutions (12 km and 4 km) that are higher than those for the output

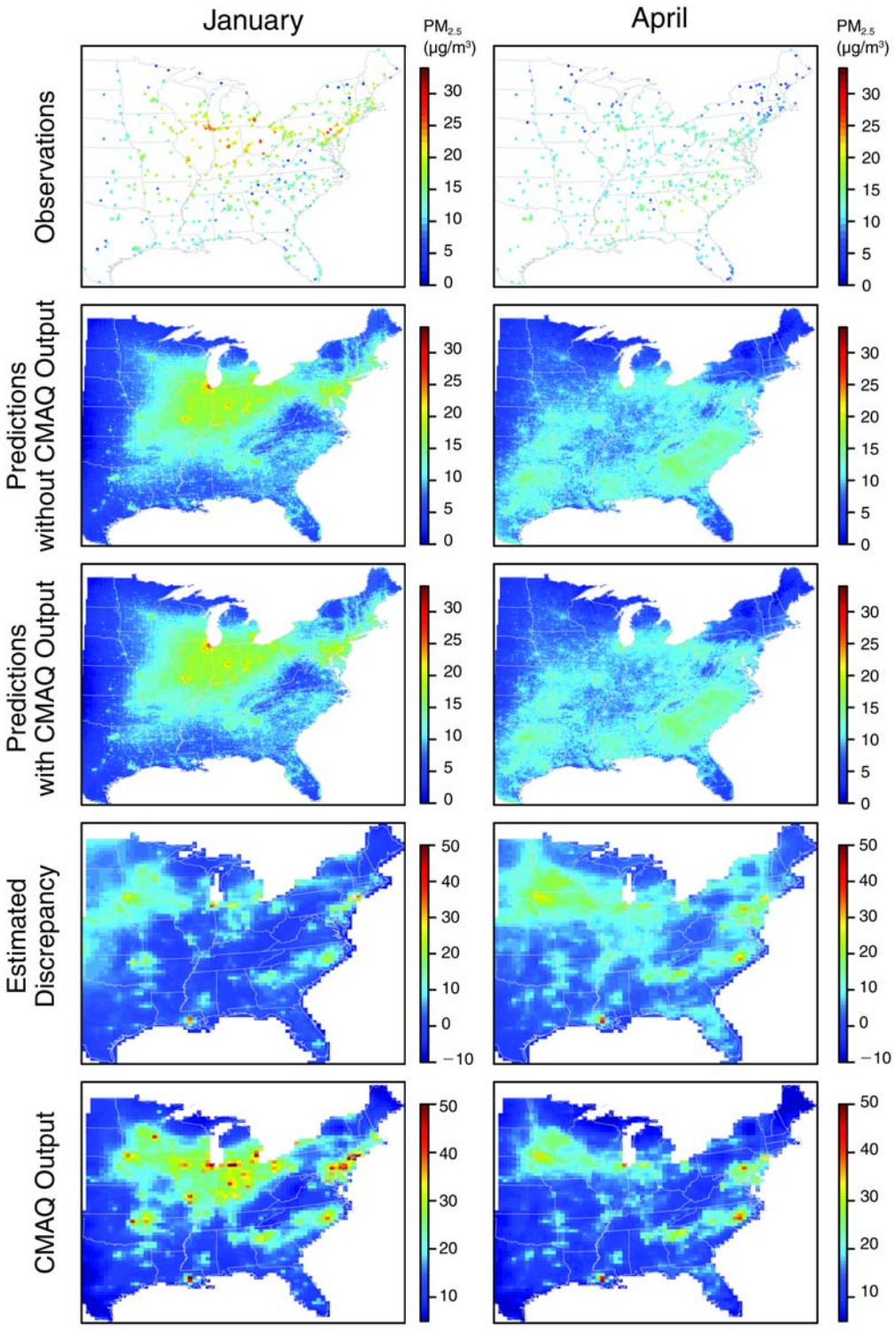

Figure 17. Spatiotemporal modeling of $PM_{2.5}$ using CMAQ output in the eastern United States in 2001. The figure shows $PM_{2.5}$ observations (first row); model $PM_{2.5}$ predictions, excluding CMAQ (second row); model predictions including CMAQ (third row); estimated CMAQ discrepancy (fourth row); and CMAQ values (fifth row) for 4 months in 2001: January, April, July, and October. Note that in January, some CMAQ values larger than 50 µg/m^3 (up to 75 µg/m^3) are truncated to 50 µg/m^3. (*Figure 17 continues next page.*)

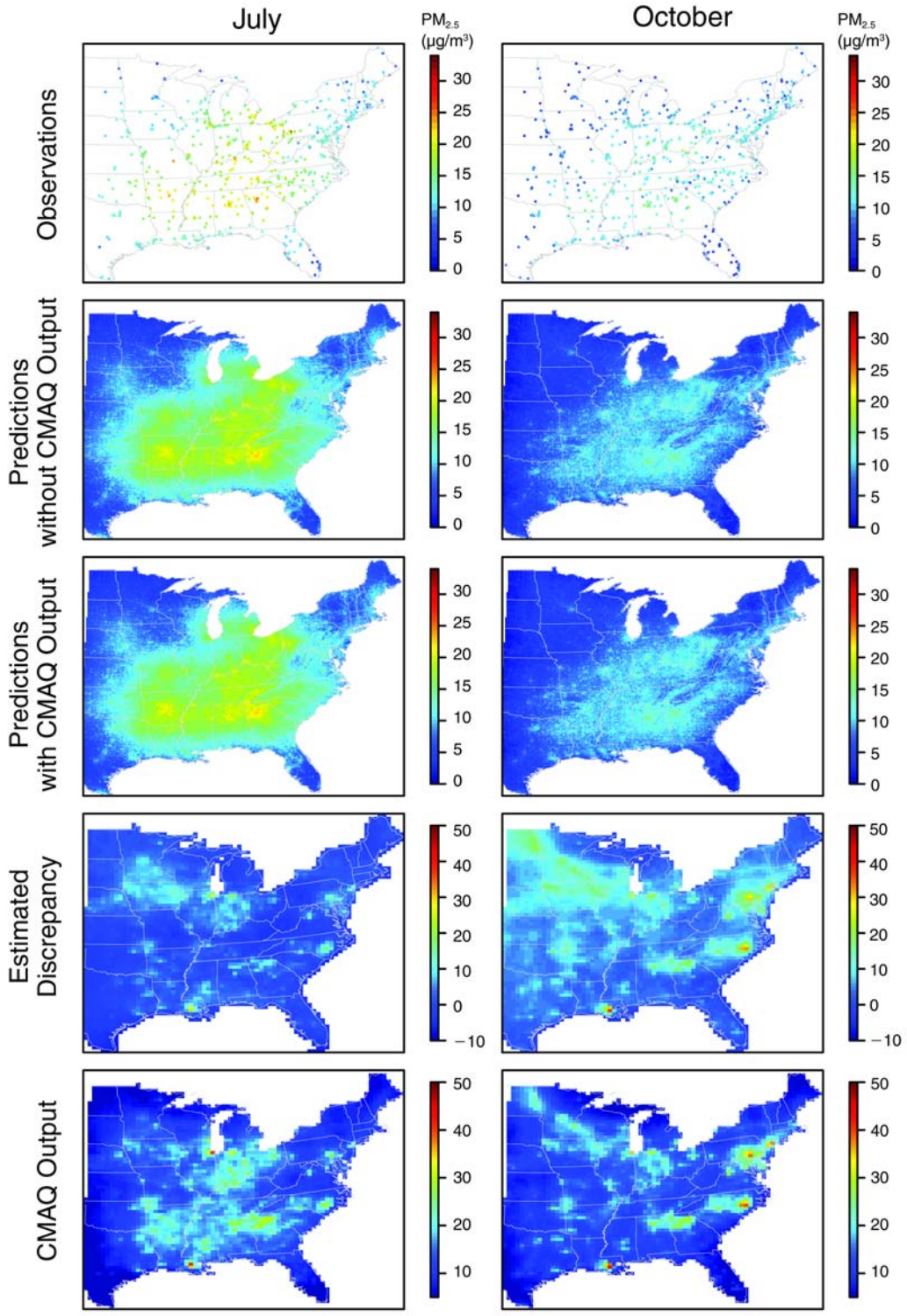

Figure 17. (*Continued*).

we used is available for some areas and time periods, but given the apparent small-scale discrepancy seen in this study, we caution against assuming that model output for smaller scales can actually resolve small-scale variation in $PM_{2.5}$. Similar concerns apply to remote-sensing products that provide nominally higher resolution: As seen in section 5, the GASP AOD product, available at a nominal 4-km resolution, did not help to improve $PM_{2.5}$ predictions.

7. USING GOES REFLECTANCE MEASUREMENTS AS A PROXY FOR GROUND-LEVEL $PM_{2.5}$

INTRODUCTION

In section 3 we saw that GASP AOD was well correlated with $PM_{2.5}$ only after spatial calibration, and the resulting correlation appears to have been driven largely by a statistical artifact: The spatial calibration surface was itself associated with $PM_{2.5}$. However, there was no evidence that the calibrated AOD provided information about $PM_{2.5}$ at smaller scales. When GASP AOD was included in statistical modeling of $PM_{2.5}$ (section 5), we saw no improvement in predictive ability. These discouraging results may have been caused in part by limitations of the GOES imager and the GASP AOD retrieval algorithm, including the coarse spectral resolution, fixed viewing geometry, and the lack of a near-infrared channel to characterize the surface accurately. In this section, we investigate two avenues for reducing noise to improve the performance of GOES-derived proxies in predicting ground-level $PM_{2.5}$.

First, previous work (Prados et al. 2007) and our own exploratory work have indicated that errors in cloud screening have led to noisy retrievals, particularly early and late in the day. The current GASP AOD algorithm attempts to remove anomalous retrievals in individual pixels when the standard deviation of GASP AOD in surrounding pixels is large. An alternative approach, which may be better able to screen out anomalies and avoid screening out true spatial heterogeneity, is to screen based on anomalies in time at individual pixels. Aerosol concentrations generally change smoothly over time, particularly over the area of a pixel, so abrupt changes (particularly increases, because cloud contamination is associated with high reflectance readings) are more likely to indicate contaminated retrievals. In this section, we investigate screening procedures based on the change in channel 1 reflectance over time, from one half hour to the next.

The current GASP AOD algorithm accounts for variability in surface reflectance — which varies widely based on land use and vegetation — by estimating the background based on the second-lowest channel 1 reflectance measurement over the past 28 days at each pixel and at each half hour of the day. This approach makes use of all channel 1 reflectance values, regardless of cloudiness or the aerosol levels over the 28-day period. To develop a background estimate that characterized reflectance measurements under low-pollution conditions, we linked the channel 1 reflectance values with ground-level $PM_{2.5}$ monitoring data and built statistical models to estimate the background.

Given our interest in deriving a proxy for ground-level $PM_{2.5}$, our approach was to directly use channel 1 reflectance — screened and corrected for background as described above — as a proxy, avoiding the extra complexity and assumptions (in particular about aerosol size and chemical composition) involved in the GASP AOD retrieval algorithm. The hope was to avoid the complications of assumptions that might actually decrease associations with ground-level $PM_{2.5}$.

METHODS

Data and Overview of the GASP AOD Retrieval Algorithm

We made use of measurements and derived retrievals from GOES-12 (East) imager data, provided by the NOAA, focusing on the times of 12:15 through 21:15 UTC as these are largely during daylight hours throughout the year. Prados and colleagues (2007) described the GOES-12 imager data and GASP AOD algorithm in detail. GOES data were at a nominal spatial resolution of 4-km and were obtained every half hour during daylight, from 10:45 through 23:45 UTC, as GOES is a geostationary satellite. NOAA's data product provides several quantities used in deriving GASP AOD. Channel 1 reflectance is the normalized total reflectance at a wavelength of 660 nm at the top of the atmosphere. The GASP AOD algorithm estimates the "mosaic" as the second smallest channel 1 reflectance value for a given pixel and half hour of the day over the previous 28 days, multiplied by the cosine of the zenith angle. This mosaic (and its normalized version, the clear-sky composite, which takes the mosaic and divides it by the cosine of the zenith angle) are the algorithm's estimate of reflectance at the top of the atmosphere under low aerosol conditions. The mosaic is used to estimate surface reflectance under the assumption that it represents surface effects

combined with background aerosol (assumed AOD of 0.02), gas absorption, and Rayleigh scattering effects. AOD is then calculated by subtracting an estimate of the surface reflectance (as well as gas absorption and Rayleigh scattering) from channel 1 reflectance and using an algorithm that makes assumptions about particle size, aerosol composition, and particle single scattering albedo.

NOAA's original screening criteria for GASP AOD excluded retrievals with any of the following characteristics: AOD greater than 10, surface reflectance greater than 0.15 or less than 0.01, signal (a variable describing retrieval sensitivity) less than 0.01, standard deviation of AOD (based on the focal pixel and its 24 surrounding pixels) greater than 0.15, clouds detected in the pixel or any of the surrounding 24 pixels, channel 1 reflectance less than 0, or solar zenith angle greater than 70° (which is related to reduced accuracy when the sun is low in the sky). The cloud mask is determined from infrared channels 2 (3.9 μm) and 4 (10.7 μm) and the visible channel.

Since our initial work with GASP AOD, described in sections 3 and 5, the GOES team, in 2009, finalized several improvements to the GASP AOD retrieval algorithm, including the correction of an error in the azimuth angle definition, an improved method for estimating surface reflectance, and an improved calculation of the standard deviation of AOD as well as inclusion of scattering angle in the screening criteria. The retrieval changes did not affect the channel 1 reflectance values, so they did not have an impact on our method development. However, the improved retrievals were available only for summer (June–August) 2004. Therefore, to compare the results of our new proxy relative to the improved GASP AOD, we used that time period for the assessment of our new method. We used 24-hour average $PM_{2.5}$ data from the U.S. EPA AQS, including IMPROVE data (parameter codes 88101 and 88502).

Our analysis covered the mid-Atlantic region and a large region in the southeastern United States centered on Atlanta and covering Georgia and much of Alabama, South Carolina, eastern Tennessee, and western North Carolina (shown in Figure 18). The surface reflectance modeling was limited to the mid-Atlantic.

More details on the data sources and manipulations are available in Appendix A.

Temporal Screening for Anomalous Reflectance Measurements

The GASP AOD retrieval algorithm has a critical screening criterion that screens out retrievals for which there is large local variability in AOD based on a 5×5 set of pixels surrounding and including the focal pixel. In part, this criterion seeks to avoid reporting retrievals that may be affected by undetected clouds, as cloud screening is more difficult with GOES than with MISR or MODIS because GOES is less sensitive in detecting water vapor. Exploratory graphics (not shown) suggested that it is this screening criterion that often catches anomalously high values of AOD that would otherwise be reported as valid retrievals.

Our approach was to consider the time series of channel 1 reflectance values for each pixel for a given day and screen out temporal, rather than spatial, anomalies. Figure 19 shows examples of time series plots for individual pixels. Examination of a large number of such plots for summer 2004 in the mid-Atlantic and southeast United States regions indicated several features of channel 1 reflectance. First, reflectance appeared to be anomalously high more often in the first and last time periods than in the other time periods. Next, not surprisingly, high values tended to be associated with clouds detected by the GOES cloud screen. Finally, in some pixels on some days, there were consistently low values of channel 1 reflectance over successive half hours, whereas, in others, channel 1 reflectance oscillated wildly. Unless a mass of pollution arrived at a location all at once and then departed, oscillations and sharp gradients likely represent errors of some sort, probably due to erroneous cloud screening (and cloud shadows) rather than real changes in AOD. In particular, higher values following sharp increases were likely caused by undetected cloud arrival, while lower values following sharp decreases may have been caused by the departure of clouds. Exploratory spatial plotting of the surface of retrievals in a given half hour (e.g., Figure 18) often showed areas with very high channel 1 reflectance and resulting high AOD retrievals (not shown). Large areas with anomalous values occurred more often early in the day, whereas later in the day anomalous values were often scattered in a more heterogeneous fashion amidst lower values, also noticeable in Figure 18. Similar spatial patterns were found in AOD as well (not shown). Notice that the anomalous areas were likely to have been caused by clouds: These areas shifted over time, often had distinct borders, and were in areas where cloud screening had screened out many other pixels. NOAA's AOD standard deviation criterion screens out many of the anomalously high values that remain after cloud screening, but additional high values often remain at the edges of the screened area and as speckled regions with scattered high values among both lower values and screened-out pixels.

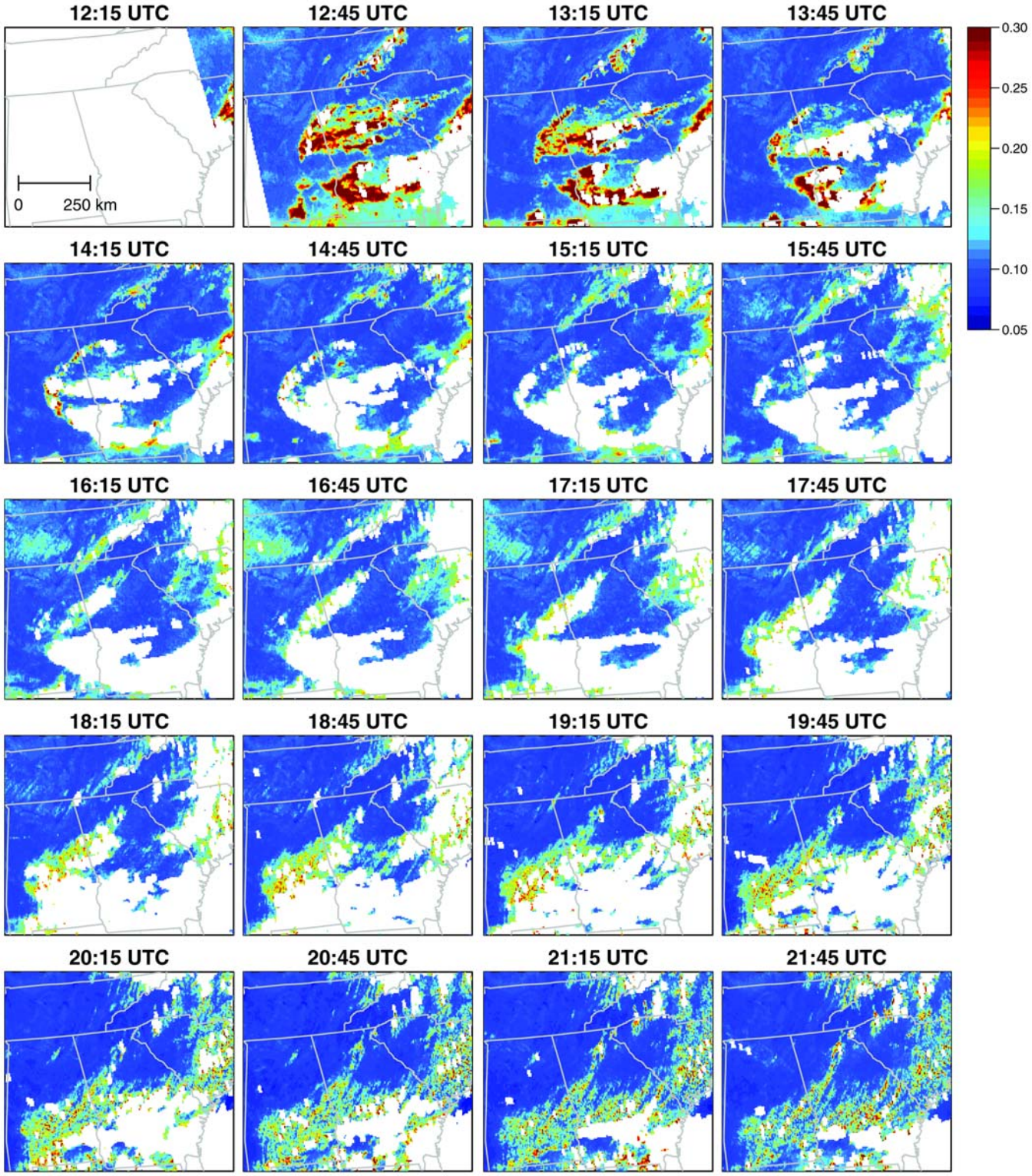

Figure 18. GOES channel 1 reflectance for half-hourly observations on August 7, 2004, in the southeastern United States. White areas are those excluded based on the GOES cloud screen (any clouds in the 5 × 5 grid of pixels surrounding a given pixel) and a solar zenith angle greater than 70°. Values larger than 0.30 (up to 0.40) are truncated to 0.30 to better visualize the differentiation of lower values.

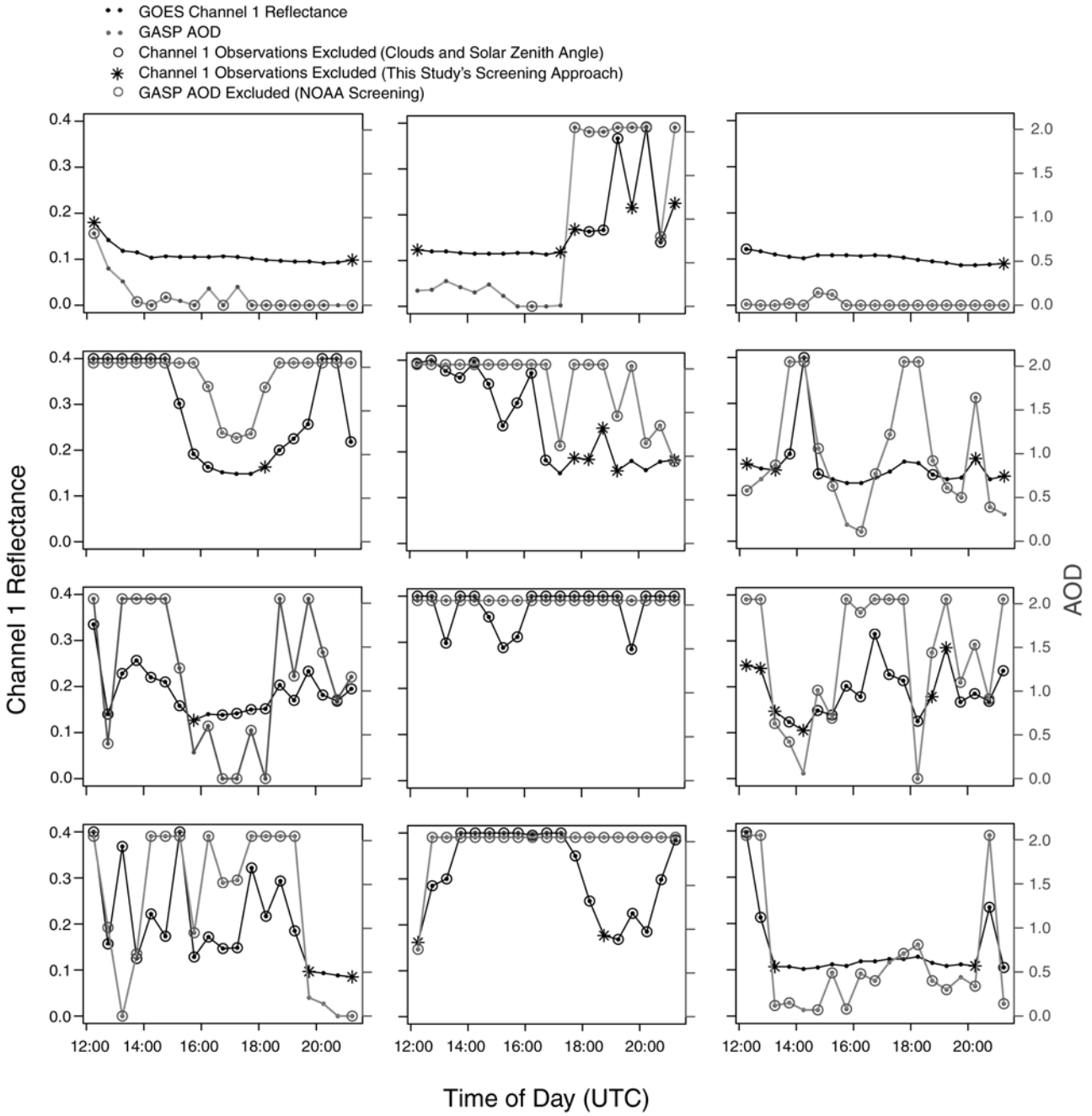

Figure 19. Alternate approaches to screening GOES channel 1 reflectance and GASP AOD illustrated in a random selection of 12 pixel–day combinations (June–August 2004) in the mid-Atlantic region. Channel 1 observations are first screened out based on the presence of clouds (any clouds in the 5×5 grid of pixels surrounding a given pixel) and solar zenith angle (greater than 70°) only. Additional observations are excluded based on our temporal screening approach. GASP AOD retrievals are screened using NOAA's original AOD screening approach. Negative AOD retrievals are truncated to 0.

Given this exploratory analysis and these resulting hypotheses, we considered the following screening criteria:

1. Avoid the earliest and latest half hours.
2. Avoid channel 1 values greater than a cutoff value, as most of these are likely related to cloud interference.
3. Avoid channel 1 values for times when a cloud was detected in the previous or the next half hour in the same pixel.
4. Avoid values for which the magnitude of the second difference (the difference of differences of adjacent measurements in time) exceeds a threshold.
5. Avoid values for which the change since the previous half-hourly value or with respect to the next half-hourly value exceeds a threshold.

As our first criterion, we avoided the earliest and latest half hours (12:15 and 21:15 UTC), as we could not calculate first and second differences or cloudiness in adjacent time periods for these boundary time points (criteria 3–5). Also, reflectance tended to be slightly higher on average during these periods than in the next and previous half hours, respectively, suggesting more erroneous retrievals. Figure 20, which illustrates the second criterion, shows the proportion of cloud detections as a function of binned channel 1 reflectance values. Note that estimated cloudiness increased with increasing channel 1 reflectance, suggesting that for large values of channel 1 reflectance, cloud contamination was highly likely. This suggests that even when clouds were not detected using the GOES cloud screen, there might have been contamination. We chose to screen out channel 1 reflectance values greater than 0.2, but in further studies other values would be worth considering. The third criterion was a reflection of the possibility

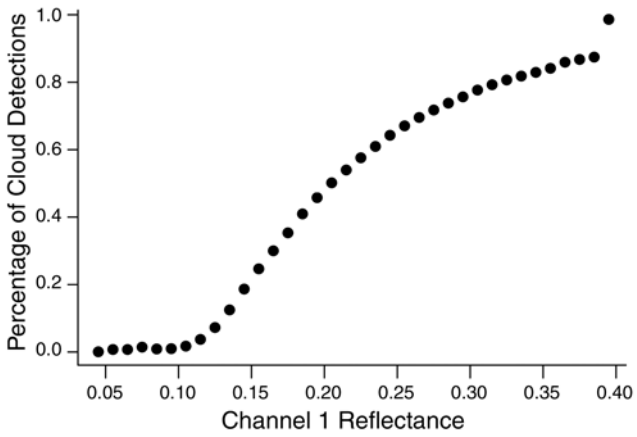

Figure 20. Proportion of time that clouds were detected in a pixel as a function of GOES channel 1 reflectance values for the pixel, for the mid-Atlantic region (2003–2006). The reflectance values were binned for plotting purposes.

that if a cloud was observed in a pixel in one half hour, the chance of undetected cloud contamination in neighboring time points was likely to have been higher.

As noted previously, Figure 19 shows example time series of channel 1 reflectance. Note the sharp spikes and large gradients present for many successive triplets or pairs of half-hourly values. We looked at the variability from half hour to half hour for pixel–days for which channel 1 reflectance stayed fairly constant for multiple successive time periods, and, based on that information, we chose cutoffs for the first and second differences in channel 1 reflectance (estimates of the first and second derivatives) that seemed likely to exclude anomalous values for channel 1 reflectance. For the second difference, we chose to exclude channel 1 reflectance values when the absolute value of the second difference was larger than 0.07. We calculated the first difference for each point relative to the previous half hour and the subsequent half hour. When an observation was larger than either the previous or next half hour by 0.03 or more, we excluded it, and when an observation was smaller than either the previous or next half hour by 0.05 or more, we excluded it. The reasoning behind these choices was that cloud contamination would likely have led to larger values, so we more strictly screened out values that were higher than their temporal neighbors. However, even if the trend were downwards, large changes suggest that cloud contamination or other issues might be affecting multiple retrievals over time, so we erred on the side of excluding values. We started the screening process by including all channel 1 reflectance measurements so that we could calculate our five screening criteria. Then, we applied both our new screening criteria and the following two of the original NOAA screening criteria: Retrievals were excluded when clouds were detected in the pixel or in the surrounding 24 pixels or when the solar zenith angle was greater than 70°, thereby avoiding retrievals that occurred when the sun was low in the sky (Prados et al. 2007). We also excluded retrievals with a scattering angle of less than 50° or greater than 170°, per NOAA's revised screening criteria, again seeking to avoid retrievals that occurred when the sun was low in the sky.

Note that our screening criteria were not based on any comparison with $PM_{2.5}$ values but solely on consideration of the temporal patterns in channel 1 reflectance and the association of those patterns with the GOES cloud screening criterion. We could, therefore, directly evaluate screened channel 1 values relative to $PM_{2.5}$ observations. We calculated the correlation of the $PM_{2.5}$ observations with both GASP AOD and channel 1 reflectance using the daily averages of available GOES retrievals in the mid-Atlantic and southeast regions. For both AOD and channel 1 reflectance, we considered screening out values based on

the original NOAA screening criteria (see Data and Overview of the GASP AOD Retrieval Algorithm, under Methods in section 7) and on our new criteria. Note that, when considering channel 1 reflectance, this involved screening the reflectance values based on NOAA criteria involving the derived AOD retrieval values. GOES pixels and $PM_{2.5}$ monitor locations were matched in space by finding the GOES pixel nearest to each monitor. Channel 1 reflectance and GASP AOD retrieval values for that pixel were averaged over the day for values not excluded by the screening criteria being used and were then matched by day to the 24-hour average $PM_{2.5}$ value for the monitor. We didn't consider longer-term averages at this stage. We left this to our modeling of surface reflectance, described next, which attempted to account for variability in surface reflectance.

Surface Reflectance Modeling

The current GASP AOD algorithm uses the second smallest value of the channel 1 reflectance for a given pixel and half hour of the day over the previous 28 days to estimate the mosaic, a proxy for reflectance unaffected by aerosol in the atmosphere. Our approach was to use the $PM_{2.5}$ monitoring data to determine when $PM_{2.5}$ was likely to be low for a given pixel and to build a model for the background surface reflectance based on channel 1 reflectance at those times. We built our model for the mid-Atlantic region for the time period of April 24, 2003, through December 31, 2006; the beginning date was when GOES-12 retrievals first became available.

To estimate $PM_{2.5}$ at all locations as best as possible, we first considered only every third day, when most of the monitors were operating, since many monitors operated only every third day. For each of these days, we fit a simple TPS with the amount of smoothing determined via GCV, as implemented in the gam() function in R. For each point in the 4-km grid, we predicted $PM_{2.5}$ and also calculated standard errors of prediction. Note that for the standard errors we included only uncertainty in the fitted mean, omitting the residual variance, in an attempt to estimate uncertainty in the surface and not in monitor-to-monitor variability. We did not use covariates such as land-use information in our model, as our main goal was to find areas and days for which all the $PM_{2.5}$ observations were low and to estimate this background $PM_{2.5}$ from simple distance-weighted averaging of the monitor values without attempting to estimate small-scale variability in $PM_{2.5}$. The assumption was that on days in which a whole area had low $PM_{2.5}$ concentrations, small-scale variability was limited.

Next, we explored models explaining channel 1 reflectance as a function of diurnal and seasonal variability, since reflectance is known to change as surface characteristics change as a result of surface greenness and as the angle of the sun varies with respect to the surface. We included only channel 1 reflectance values at locations and on days with predicted $PM_{2.5}$ of less than 10 $\mu g/m^3$ and a 95% prediction interval for $PM_{2.5}$ values whose upper bounds were less than 12.5 $\mu g/m^3$, to avoid fitting the model when we were less sure that $PM_{2.5}$ was low. We excluded all channel 1 values that did not pass our new screening approach as described in the preceding subsection.

We focused on two core models for channel reflectance,

$$\begin{aligned}\text{Model A: channel } 1_{i,d,h} &= \beta_0^{(1)} + \beta_{\text{PM}}^{(1)}\text{PM}_{i,d,h} \\ &\quad + f_{\text{season}(d)}(h) + f_{\text{doy}}(\text{doy}(d)) + \varepsilon_{i,d,h}\end{aligned} \tag{19}$$

and

$$\begin{aligned}\text{Model B: channel } 1_{i,d,h} &= \beta_0^{(2)} + \beta_{\text{PM}}^{(2)}\text{PM}_{i,d,h} \\ &\quad + f_{\text{joint}}(\text{doy}(d),h) + \varepsilon_{i,d,h},\end{aligned} \tag{20}$$

where β_0 is the intercept, $\varepsilon_{i,d,h}$ is the error term, d indexes the days from April 24, 2003, through December 31, 2006; doy(d) maps these days to the 365 days of the year; i indexes pixels; and h indexes half hours of observation. Model A has four smooth functions of time of day, one for each season, $f_{\text{season}(d)}(h)$, to capture changing diurnal patterns over the year, while $f_{\text{doy}}(\text{doy}(d))$ captures smooth variation over the course of the seasonal cycle. In contrast, model B allows the diurnal cycle to change smoothly over the days of the year, rather than discretely by season, using a tensor product of spline terms of day of year and half hour of day, $f_{\text{joint}}(\text{doy}(d),h)$. Both models were fit using the gam() function in R, based on the default penalized TPS specification with smoothing chosen by GCV. To reduce computations, we fit the model using a random subsample of 500,000 observations. The estimated smooth terms in the models are shown in Figure 21. Having fit the model, we calculated the fitted values of channel 1 reflectance in the full dataset, and we subtracted the fitted values from the actual reflectance values. Using the resulting residuals, we calculated the average residual for each pixel to estimate the spatial backgrounds, $\hat{g}_i^{(\text{A})}$ and $\hat{g}_i^{(\text{B})}$, for models A and B, respectively, having adjusted for temporal variability through the model fitting. We also considered stratifying the estimates of the spatial background by season. Some variation by season was seen, but the basic patterns were consistent across the seasons. A more sophisticated approach would have been to fit a spatiotemporal smooth to the residuals as a function of location and time of year,

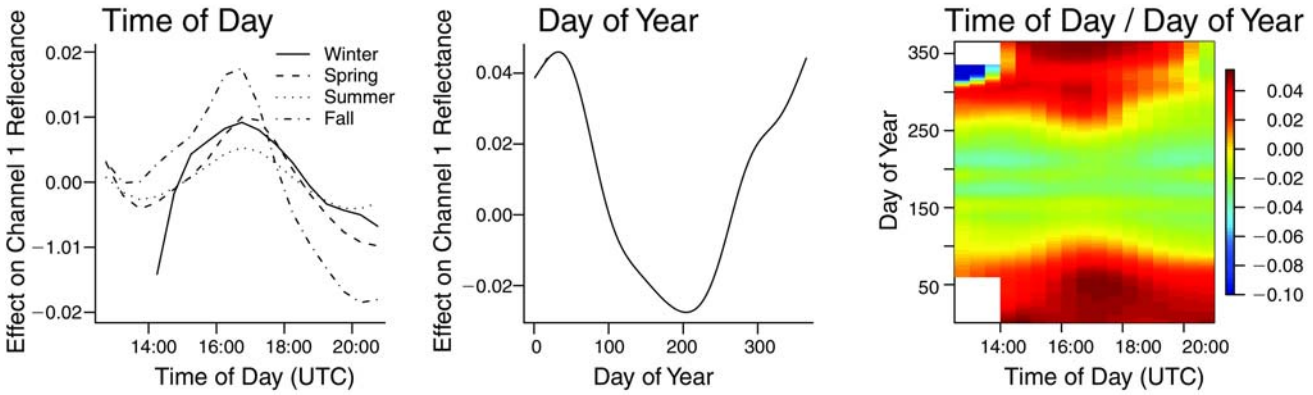

Figure 21. Effect on channel 1 reflectance of time of day and day of year for the mid-Atlantic region (2003–2006). Shown are the estimated smooth effect on GOES channel 1 reflectance of time of day by season in model A (equation 19; left panel); the estimated smooth effect of day of year in model A (middle panel); and the estimated bivariate smooth effect of time of day and day of year in model B (equation 20; right panel). In the left and middle panels, uncertainty bands are so close to the estimated smooths that they cannot be distinguished from them, except for some values, which would appear at the very beginning and end of the time-of-day plots in the left panel but are not shown.

to estimate the seasonally changing background smoothly rather than as step changes at March 1, June 1, September 1, and December 1.

To assess whether our models captured the main structure in the variability of channel 1 reflectance under low $PM_{2.5}$ conditions, we examined the residuals described above, assessing with further model fits and exploratory plots whether there was remaining temporal variability that varied systematically with longitude or the urban–rural gradient. We did not find any effects whose magnitude approached that of the terms in the models above, so we were satisfied that the models captured the important temporal structure in the data without ignoring important interactions between the temporal structure and other characteristics of the pixels.

We then assessed the results of the new surface reflectance derivation using data from the mid-Atlantic region for summer 2004. To assess model A as a means of adjusting channel 1 reflectance for background surface reflectance, we took the following step: For each GOES pixel matching a $PM_{2.5}$ monitor location (based on finding the GOES pixel centroid nearest each monitor location), we calculated a proxy for $PM_{2.5}$ defined as

$$\text{channel1}_{i,d,h} - \hat{g}_i^{(\text{A})} - \hat{f}_{\text{season}(d)}(h) - \hat{f}_{\text{doy}}(\text{doy}(d)) \quad (21)$$

for half-hourly channel 1 reflectance values not excluded based on our new screening approach. We then averaged values for each pixel over the half-hour segments of the day to derive a daily proxy to compare with 24-hour average $PM_{2.5}$ values. For model B, the proxy was defined similarly but with $\hat{g}_i^{(\text{B})}$ and the tensor product of day of year and half hour used in place of the additive half-hour and day-of-year terms used in model A. Note that this assessment included all days rather than a subset of every third day.

RESULTS

Temporal Screening

The example time series plots in Figure 19 illustrate channel 1 reflectance values screened out by our new criteria as points covered by black asterisks. Note that unrealistically high reflectance values, as well as values that were associated with large spikes and large gradients, were screened out. Table 12 shows spatiotemporal correlations of channel 1 reflectance and AOD with ground-level $PM_{2.5}$ based on the original and our new screening criteria. Note that in addition to applying the NOAA criteria to screen AOD retrievals and our criteria to screen channel 1 reflectance values, we considered applying our criteria to screen the AOD retrievals and applying the NOAA criteria to screen channel 1 reflectance. For the mid-Atlantic region, our simple screening of channel 1 reflectance produced correlations nearly as large as the correlation between GASP AOD and $PM_{2.5}$. For the southeastern region, while our screening greatly improved the correlations relative to the very low correlation of the NOAA-screened channel 1 reflectance, the resulting spatiotemporal correlation was lower than seen for AOD. For summer average spatial correlations (calculated based on daily values matched in time) for both regions (Table 12), our screening of channel 1 reflectance produced correlations as strong as those seen with GASP AOD. Note that much of the improvement from the screening appears to have come simply from removing channel 1 reflectance values exceeding 0.2.

Table 12. Spatiotemporal and Summer Average Spatial Correlations (95% CI) of $PM_{2.5}$ and GOES Proxies[a] Matched by Day and Location for the Mid-Atlantic and Southeastern Regions, Summer 2004

	Spatiotemporal Correlations		Spatiotemporal Correlations Four or More Proxy Values per Day[b]		Summer Average Spatial Correlations[c]	
	Mid-Atlantic	Southeast	Mid-Atlantic	Southeast	Mid-Atlantic[d]	Southeast
AOD, NOAA screening	0.51 (0.48–0.54)	0.54 (0.50–0.57)	0.53 (0.49–0.57)	0.57 (0.52–0.61)	0.51 (0.36–0.63)	0.26 (0.06–0.44)
AOD, our screening	0.43 (0.40–0.46)	0.37 (0.33–0.40)	0.52 (0.48–0.55)	0.49 (0.45–0.52)	0.03 (−0.14–0.20)	0.21 (0.03–0.38)
Channel 1, NOAA screening	0.30 (0.27–0.33)	0.06 (0.02–0.10)	0.40 (0.37–0.43)	0.22 (0.18–0.26)	0.40 (0.25–0.53)	0.27 (0.09–0.43)
Channel 1, our screening	0.46 (0.43–0.49)	0.35 (0.31–0.38)	0.53 (0.50–0.56)	0.47 (0.43–0.51)	0.51 (0.37–0.63)	0.28 (0.10–0.44)

[a] Both AOD and direct use of channel 1 reflectance.

[b] These columns show correlations only for days and locations for which at least four screened AOD or channel 1 values are available.

[c] Only locations with at least 9 days of coincident proxy and $PM_{2.5}$ data are included.

[d] Excludes one site outside Pittsburgh that is just downwind of a major industrial facility.

Note also that while we report 95% CIs for the correlations, it's not clear that these are helpful in comparing correlations between different proxies because the data used for CIs for a given proxy–$PM_{2.5}$ correlation are not independent of the data used for another proxy–$PM_{2.5}$ correlation.

Surface Reflectance Modeling

We next assessed the effects of our surface reflectance modeling in the mid-Atlantic region. Figure 22 shows the surface reflectance we estimated based on $\hat{g}^{A}(\bullet)$, the temporally averaged residuals from our channel 1 reflectance model A (equation 19), in comparison with a visual satellite image downloaded from Google maps. Note that the estimated surface reflectance closely follows the patterns in the surface image, with darker forested areas and lighter developed and cropland areas. These patterns also match the patterns seen on days with low $PM_{2.5}$ over the entire region (not shown), suggesting that our approach was doing a reasonable job in estimating the background surface reflectance.

$\hat{g}^{B}(\bullet)$ (not shown) was indistinguishable from $\hat{g}^{A}(\bullet)$, whereas estimates of the background stratified by season were similar to the full-year estimate but with some heterogeneity and with the areas in the snow shadow of Lake Erie (upper-left area of the Figure 22 plots) not estimable during winter because of a lack of reflectance values not excluded by screening (not shown). Figure 23 shows the GOES-reported mosaics at midday for a sample of several days, which indicate highly spatially and temporally variable estimates that seem unrealistic as proxies for the background surface reflectance, based on comparison with land-cover and land-use features.

Table 13 reports the correlations between $PM_{2.5}$ and our new channel 1 proxies (making use of both model A and model B) for the mid-Atlantic region. The correlations are directly comparable to those in Table 12 because the spatial and temporal domains were the same. We do not report results from the southeast because of the size of the computations involved in carrying out the surface reflectance estimation. We found that despite the plausibility of our estimated surface reflectance and of our screening criteria, the correlation of our surface-corrected channel 1 proxy with $PM_{2.5}$ was no better than the correlations of $PM_{2.5}$ with raw channel 1 reflectance or with GASP AOD, as shown in Table 12. For the summer average spatial correlations, correlations were lower than they were without the surface correction. Given that we restricted our models to days with low $PM_{2.5}$, a partial explanation for the decrease in correlation is that we removed the contribution of the background surface reflectance, which itself correlates with $PM_{2.5}$. The reason is that locations that are brighter on average even on days with low $PM_{2.5}$ tended to be more polluted in general, as seen in the last column of

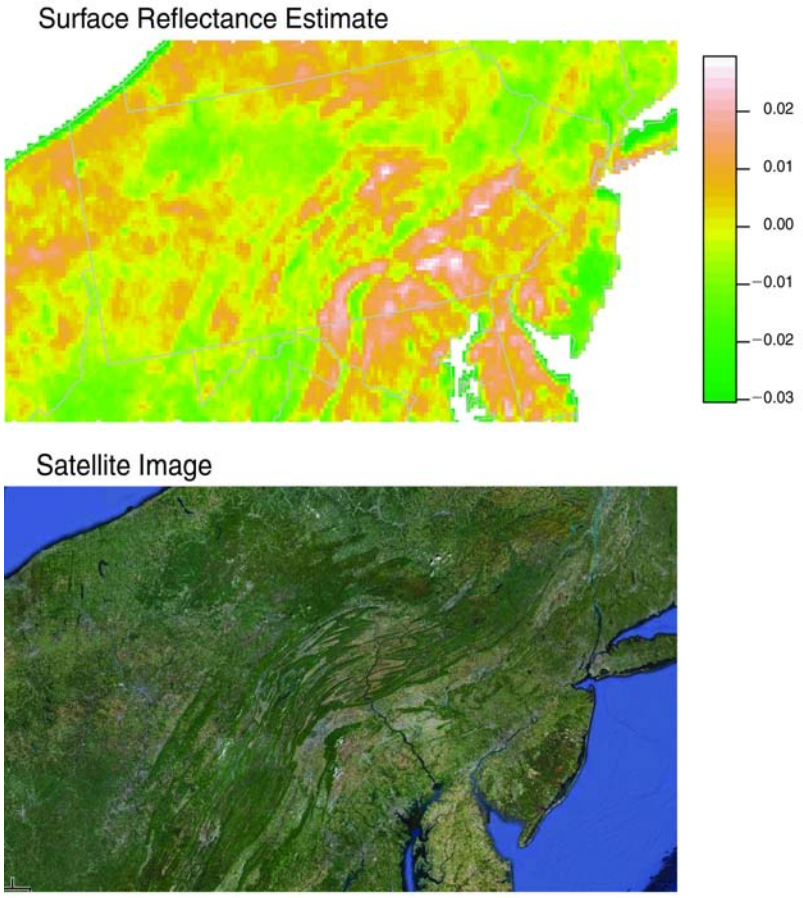

Figure 22. Estimated surface reflectance for GOES, based on model A (equation 19; left map), and a visual image from Google Maps of roughly the same area, but with a different map projection (right map). For the model A surface reflectance, some values along the coast and Lake Erie are truncated from below −0.03 to −0.03. (Aerial image © 2010 TerraMetrics, Inc.; *www.terrametrics.com*; used with permission.)

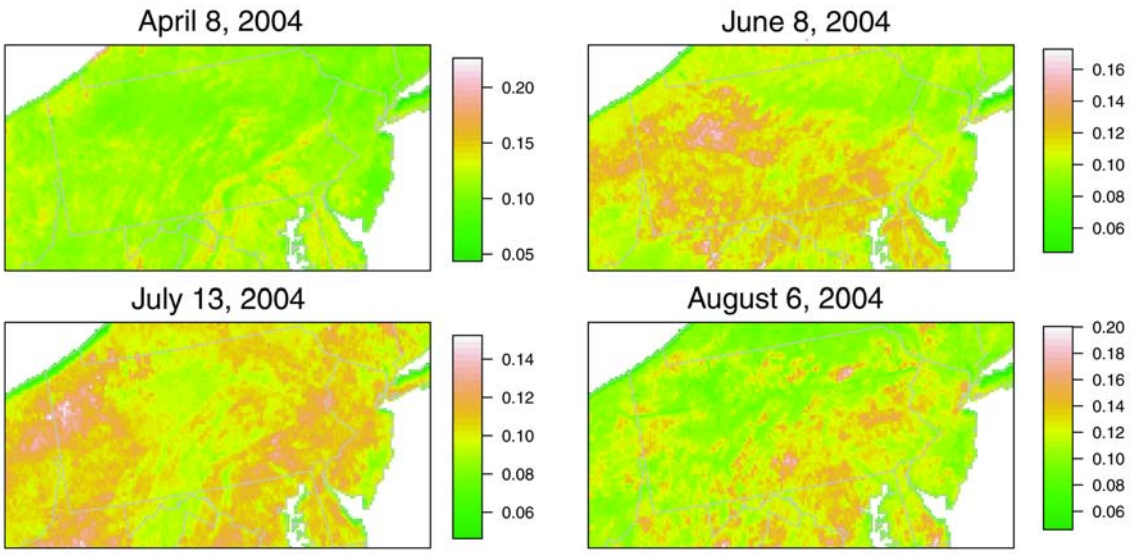

Figure 23. GOES-reported mosaic for 4 arbitrary days in 2004 at 17:45 UTC over the mid-Atlantic region. Note that the scale varies for the different subplots to emphasize spatial contrasts.

Table 13. This makes sense since the brighter areas are urban and agricultural areas, and the darker areas are forested. This suggests that some of the correlation of $PM_{2.5}$ with raw channel 1 reflectance (and possibly with GASP AOD since it is affected by any uncorrected contamination from the background surface reflectance) was driven simply by the fact that GOES measurements of reflectance correlated with land use, which correlated with $PM_{2.5}$, rather than by direct information on aerosols in the reflectance values. Figure 24 compares summer average $PM_{2.5}$ concentrations at the monitoring stations to summer averages of our new channel 1 proxy for pixels overlapping the monitoring sites. The proxy appeared to capture some of the $PM_{2.5}$ variation in the western portion of the region, finding the contrast between the Ohio Valley area and northwestern Pennsylvania, but it failed to capture the high $PM_{2.5}$ concentrations in the southeastern portion of the region.

Finally, note that the results based on the season-specific surface reflectance estimate for the summer (not shown) were very similar, suggesting that our time-invariant estimate of the background surface reflectance did not affect our results.

Table 13. Spatiotemporal and Summer Average Spatial Correlations of $PM_{2.5}$ and Surface-Corrected GOES Channel 1 Proxies[a] (95% CI) Matched by Day and Location for the Mid-Atlantic Region, Summer 2004

	Spatiotemporal Correlations	Spatiotemporal Correlations, Four or More Proxy Values per Day[b]	Summer Average Spatial Correlations[c,d]	Spatial Correlation of Estimated Surface Reflectance, $\hat{g}$, with $PM_{2.5}$[d,e]
Channel 1, Model A	0.46 (0.43–0.49)	0.55 (0.52–0.58)	0.29 (0.12–0.44)	0.29 (0.12–0.44)
Channel 1, Model B	0.43 (0.40–0.46)	0.49 (0.45–0.52)	0.27 (0.10–0.42)	0.30 (0.14–0.45)

[a] Two different models for the surface characterization are used.

[b] These columns show correlations only for days and locations for which at least four screened AOD or channel 1 values are available.

[c] Only locations with at least 9 days of coincident proxy and $PM_{2.5}$ data are included.

[d] Excludes one site outside Pittsburgh that is just downwind of an industrial facility.

[e] This calculation uses the same $PM_{2.5}$ locations and daily averages used in the summer average spatial correlations column (third data column), for consistency.

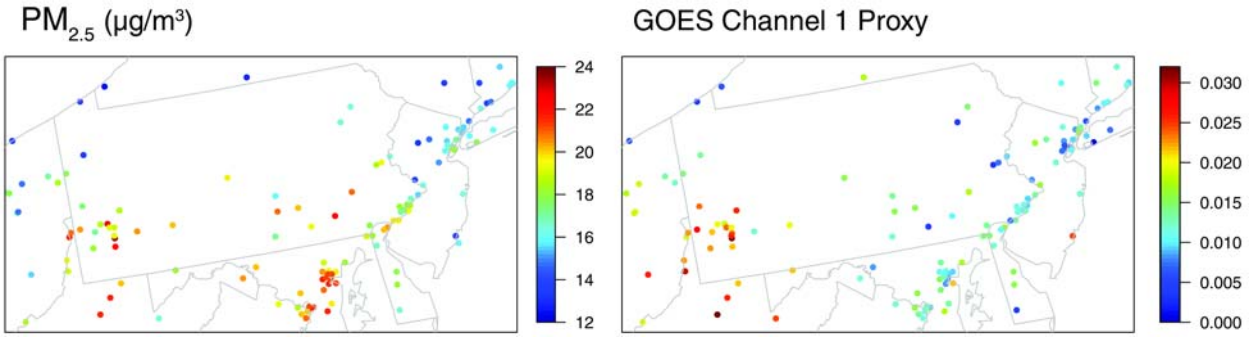

Figure 24. Summer 2004 average $PM_{2.5}$ concentrations at monitoring stations (left map) and summer 2004 averages of our new GOES channel 1 proxy (equation 21; right map) matched in space and by day to the monitoring data. Only locations with at least nine daily matches are included. For the left map, the data from one site near Pittsburgh, with a $PM_{2.5}$ concentration of 30.5 µg/m^3, are truncated to 24, to emphasize the contrasts.

DISCUSSION

Using a statistical model to account for temporal variability, we introduced methods to screen channel 1 reflectance based on temporal variability at a pixel on a given day and to estimate the background surface reflectance based on the average channel 1 reflectance at each location for times with low $PM_{2.5}$ levels. The temporal screening increased correlations between channel 1 reflectance and $PM_{2.5}$, in some cases to the level of the correlation between GASP AOD and $PM_{2.5}$. However, accounting for the background surface reflectance to adjust channel 1 reflectance resulted in decreased average spatial correlations for the summer, because the background surface reflectance itself correlated with temporally averaged $PM_{2.5}$. This suggested that the GOES reflectance measurements contained two components that were each modestly informative about $PM_{2.5}$: overall surface brightness and information on aerosol variability over space and time. It appears that some of the association seen between the satellite reflectance data and $PM_{2.5}$ came from spatial variability in the satellite measurements that was not driven by actual aerosol variability. It seems likely, given the concerns raised here about the estimation of the background surface reflectance in the GASP AOD retrieval algorithm, that some of the AOD–$PM_{2.5}$ association was also of this nature.

Note that the results reported here for the 3-month summer averages only assessed correlations of the various proxies and $PM_{2.5}$ when both were observed on the same day. Assessments based on the long-term average of the available proxies and the average of $PM_{2.5}$ over all days would likely have shown lower correlations because of missing proxy values, as is described in sections 3 and 4.

NOAA's GASP AOD algorithm was developed to compute AOD retrievals in a real-time fashion, relying only on measurements of a single half-hour snapshot. Our approach relied on having information on neighboring half-hourly retrievals during a specified day. Our approach, therefore, involved postprocessing but could be the basis of an alternative off-line retrieval. Our assessment was based simply on screening and adjusting channel 1 reflectance directly, rather than making use of AOD. While we feel that we have developed a useful surface reflectance adjustment, given the issues discussed, it is not clear that making use of this adjustment in the AOD retrieval algorithm would improve associations of GASP AOD with $PM_{2.5}$, despite the plausibility of our estimated surface reflectance surfaces. We do feel that removing the influence of the background would help lead to a better understanding of the degree to which $PM_{2.5}$–AOD correlations are driven by information in AOD on true aerosol variability as opposed to surface characteristics.

One factor that might have prevented larger correlations was that we did not account for the vertical profile of aerosols, which can introduce discrepancy between reflectance and $PM_{2.5}$ when the atmosphere is not well mixed and the vertical profile varies spatially. Given the scope of the project, we did not have the ability to run an atmospheric-chemistry model for the period of interest, and model output for long runs was not generally available from other research groups. Efforts to use model output that provides vertical-profile information at large spatial scales (e.g., 2×2.5 degrees in Liu et al. 2004a and van Donkelaar et al. 2010) have found success. However — particularly for the smaller of the two regions examined here, the mid-Atlantic — accounting for vertical-profile variation at such a large scale seems unlikely to improve the correlations substantially, in light of the small-scale discrepancy seen between our new reflectance-based proxy and $PM_{2.5}$. It remains an open question whether information on smaller-scale variation in the vertical profile would help to resolve some of the discrepancy seen at smaller scales. Vertical-profile information might be less important for longer-term averages than for short time scales affected by individual events that transport large amounts of $PM_{2.5}$. However, there may be important local spatial differences in the profile that persist over time, perhaps because of differences between rural and urban areas, coastal effects, topographic effects, or local emissions sources. The challenge is to obtain sufficiently accurate information on the vertical-profile variation at smaller scales for the full spatiotemporal domain being studied in a given epidemiologic analysis.

8. CONCLUSIONS

In this report we summarize the results of empirical analyses and of statistical-model development for the use of proxy information, in particular satellite AOD, in predicting $PM_{2.5}$ in the eastern United States. In brief, we had little success in improving predictions for use in epidemiologic applications. We found positive correlations of AOD with $PM_{2.5}$ data matched in time, but lower spatial correlation for long-term averages, unless we used calibration that adjusted for large-scale discrepancy between AOD and $PM_{2.5}$ (sections 3, 4, and 5). Statistical models that combined AOD, $PM_{2.5}$ observations, land-use and meteorologic variables, and a spatial smoothing component (analogous to kriging) were highly predictive of $PM_{2.5}$, but AOD added little information beyond that provided by the other sources (sections 5 and 6). In part this appears to have been caused by the fact that large-scale spatial patterns in $PM_{2.5}$ could be predicted well by

smoothing the monitor values, while small-scale spatial patterns in AOD appeared to weakly reflect variation in $PM_{2.5}$. Even long-term averages of MISR AOD, considered the most accurate, albeit the most sparse, of the AOD products over land, were only weakly correlated with and did not improve prediction of $PM_{2.5}$ (section 4). Thus, in locations where $PM_{2.5}$ observations are sufficiently dense to support statistical modeling with land-use covariates and smoothing or kriging, our results suggest that AOD would not substantially improve exposure predictions. However, the information contained in a proxy such as AOD might be sufficient to allow beneficial regression on the proxy when $PM_{2.5}$ observation density decreases.

Previous consideration of satellite AOD has largely focused on MODIS and MISR AOD; one contribution of our work is a more extensive consideration of GOES-derived GASP AOD and its relationship with $PM_{2.5}$. An advantage of the geostationary GOES data is the uniform, high temporal frequency of sampling. Given the moderate correlations of GASP AOD and $PM_{2.5}$, and the fact that higher correlations of long-term averages after spatial calibration primarily reflected the necessarily improved large-scale correlation (section 3), we considered new statistical techniques to screen anomalous GOES reflectance measurements and account for background surface reflectance (section 7). While the results appeared sensible, correlations of adjusted reflectance with $PM_{2.5}$ were no better than GASP AOD correlations with $PM_{2.5}$.

To what extent are our results generic to all three instruments, given our extensive work with GOES? MODIS, and particularly MISR with its multiangle capability, provide higher quality AOD retrievals than does GOES. However, our exploratory correlation analysis in section 5 indicated modest spatial correlation between AOD and $PM_{2.5}$ for all three instruments, even after large-scale spatial calibration that corrected for large-scale discrepancy and necessarily increased correlations. Our statistical modeling found that neither MODIS nor GOES AOD improved $PM_{2.5}$ predictions, raising concerns about the small-scale correspondence of AOD and $PM_{2.5}$ (section 5), but we note that this result was only for the mid-Atlantic region and for 2004. Our analysis of the relationship of long-term average MISR AOD and $PM_{2.5}$ over the eastern United States helped to broaden our results. Here, the limited number of retrievals from MISR was less of an issue, but we still found discrepancy to be a critical concern and did not find that MISR AOD improved prediction of long-term average $PM_{2.5}$ (section 4). Ideally, we would have had time to carry out a similar analysis using MODIS AOD, although the similar results for the exploratory analysis of MODIS and MISR in section 5 suggest that results for MODIS might not have been appreciably better than what we found for MISR. Thus, our results suggest that the issue of discrepancy is a generic one, particularly in contexts such as the eastern United States, where $PM_{2.5}$ levels are generally low and therefore show smaller spatial contrasts than elsewhere in the world.

STRATEGIES FOR IMPROVING THE USEFULNESS OF AOD AS A PROXY FOR $PM_{2.5}$

In our statistical analyses and modeling, we attempted to address the inherent differences between AOD and $PM_{2.5}$ (described in detail in section 1) to the best of our ability, but shortcomings in our approach no doubt contributed to the results summarized above. First, we included the PBL as an indicator of vertical mixing of particles to roughly account for the variability of the vertical profiles of particles in space and time, as this affects the near-surface portion of AOD that more directly relates to $PM_{2.5}$ concentrations. Second, we included information on the lower-atmosphere RH to roughly account for the influence of particle growth on AOD values. Our correlation analyses did not account for the spatial mismatch between areal AOD and point-level $PM_{2.5}$, but our statistical modeling attempted to account for the sub-pixel-scale effects of roads and emissions point sources on $PM_{2.5}$ monitor values. Finally — with regard to the temporal mismatch between AOD snapshots and 24-hour $PM_{2.5}$ measurements — because our focus was on long-term averages related to chronic health effects, we extensively examined monthly, yearly, and longer-term relationships of AOD and $PM_{2.5}$. We hoped that with sufficient AOD retrievals the effects of temporal mismatch would be minimized.

Should analyses of the AOD–$PM_{2.5}$ relationship match values in time or use all available values? Our analysis of MISR AOD suggested higher correlation when AOD and $PM_{2.5}$ values were matched by day rather than when all available values were used (section 4). For $PM_{2.5}$, long-term averages are likely an unbiased and highly accurate estimate of true $PM_{2.5}$ at the monitoring site, while AOD is subject to potentially informative missingness because of cloud cover, as seen in our analysis for GASP AOD (section 3) and in Koelemeijer and colleagues (2006) for MODIS AOD. Whether an AOD retrieval is missing is likely related to $PM_{2.5}$ concentration because $PM_{2.5}$ is strongly related to meteorologic conditions. Also, long-term average $PM_{2.5}$ at the MODIS and MISR overpass time (10:30 AM in the eastern United States) might differ from long-term average $PM_{2.5}$ over the full 24 hours of the day. However, many previous analyses have assessed the relationship of $PM_{2.5}$ and AOD based only on data matched by day or hour. To avoid overoptimism about AOD as a proxy for long-term

average $PM_{2.5}$, we suggest carrying out an unmatched analysis (which would include all available data) in addition to any matched analysis. Alternatively, one can supplement a matched analysis with an assessment of the extent to which average $PM_{2.5}$ — on days (or hours) with AOD retrievals — is representative of true long-term average $PM_{2.5}$.

Our statistical modeling using the proxy as data strongly discounted the proxy (sections 5 and 6). The results suggest that unless the raw correlations of proxy and observations are strong, including the proxy is unlikely to help prediction. An obvious concern is that limitations in the $PM_{2.5}$ observations, rather than just in AOD, played a role. $PM_{2.5}$ measurements are not error-free and even when they are error-free, bias could be induced because of preferential sampling (Diggle et al. 2010), such as when $PM_{2.5}$ monitors are overrepresented in highly polluted areas. This raises concerns about the degree to which observations can be treated as a true gold standard. Despite these concerns, we argue that the error in the measurements is likely to have a simpler structure than that in the proxy, with simple instrument error combined with very-fine-scale spatial heterogeneity that is not representative of the $PM_{2.5}$ process at a broader scale. In contrast, when the $PM_{2.5}$ observations showed robust (i.e., locally homogeneous) patterns at moderate scales on individual days, the proxy variables often failed to capture much of the pattern, suggesting that complicated discrepancy structure is a concern for the proxies. We attempted to account for $PM_{2.5}$ observational errors and preferential sampling through our modeling strategy. Similarly, we attempted to account for fine-scale variation in the point measurements that is necessarily smoothed over in the pixel-based AOD by using information on the distance from monitors to roads and point emissions sources. In the absence of a true gold standard at the pixel scale for the AOD retrievals, this modeling represented our best effort at constructing a gold standard for $PM_{2.5}$ exposure, albeit one that was sensitive to the choice of model structure.

The magnitude of the error in satellite AOD products is known to be large relative to average aerosol levels present in the United States, based on comparison with AERONET (Liu et al. 2004b). This highlights the difficulty in trying to use AOD to distinguish relatively small differences in long-term averages of ground-level $PM_{2.5}$ in different spatial locations. Distinguishing aerosols from the surface signal is a major challenge for satellite AOD retrievals and has particular implications for characterizing spatial variability in $PM_{2.5}$, with surface contamination likely to be having effects at multiple scales. For both MISR AOD (section 4) and GASP AOD (section 7), we saw evidence that some of the correlation between raw AOD and $PM_{2.5}$ might have been a function of uncorrected surface brightness related to land use, rather than having been driven by the detection of aerosol in the AOD retrieval algorithms. We suggest that future remote-sensing analyses consider the possibility that land use and its influence on surface reflectance that is not fully corrected for in the retrieval algorithm — rather than the direct relationship of $PM_{2.5}$ and AOD — might partially explain spatial correlations of $PM_{2.5}$ and AOD. In section 7 we proposed a statistical method to estimate long-term average surface reflectance — based on days with low $PM_{2.5}$ concentrations — that might be used in AOD retrieval algorithms.

In developing countries with higher aerosol levels and more spatial variability, such as China (Jiang et al. 2007), satellite AOD may be better able to capture spatial patterns in $PM_{2.5}$ even at the small scales, because the relative magnitude of the errors caused by uncorrected spatial variability in surface reflectance would be smaller relative to the higher aerosol levels. In addition, because it is more difficult to build models in these areas based on $PM_{2.5}$ measurements and land-use covariates, AOD may add more incremental information there than in the models we fit in our data-rich setting. However, given the potential for discrepancy, based on first principles (see section 1) and seen in our analyses, we encourage validation efforts to support the use of AOD in data-poor settings, perhaps with carefully targeted monitoring. For some air quality purposes, capturing large pollution gradients at large scales is likely to be sufficient, while for epidemiologic analysis, as discussed further in the subsection Using AOD for Public Health Research, we believe that the important questions are the magnitude of pollution gradients at smaller scales and the ability of AOD to resolve these.

In contrast to our empirical statistical calibrations — including use of the PBL and the RH — another way to handle some of the inherent differences between AOD and $PM_{2.5}$ is to use vertical profiles of atmospheric aerosol simulated in a model to scale column AOD to its surface portion. The improvements seen when such model profiles were used (Liu et al. 2004a; Liu et al. 2007a; Liu et al. 2007b; van Donkelaar et al. 2010) are not surprising because the vertical profile of particles is determined by the complex interaction of regional and local-scale meteorology, diurnal variation in emissions, pollution transport, and seasonal change in land-surface characteristics. The totality of these effects was not easily captured by the statistical calibration of AOD based on meteorologic variables that was done in this report, building on Liu and colleagues (2005). However, with the relatively dense $PM_{2.5}$ observations available to us, we were able to achieve smooth spatial calibration of AOD to $PM_{2.5}$ to adjust for

larger-scale spatial discrepancy and assess the importance of using large-scale calibration in place of model-simulated profiles. In general, such calibration led to improved correlations (sections 3, 4, and 5). However, while we hoped that correcting for large-scale discrepancy between AOD and $PM_{2.5}$ would improve overall spatial correlations by unmasking small-scale association between AOD and $PM_{2.5}$, we saw little evidence of this. Rather, it appeared that improvements in spatial correlation were primarily related to the correlation between $PM_{2.5}$ and the spatial adjustment surface. In effect, we empirically calibrated AOD so that AOD and $PM_{2.5}$ were more correlated at the large scale. It is not surprising then that spatial correlations increased at this scale, yet the improvement in correlation seemed related primarily to this scale.

Unfortunately, multiyear full-chemistry model simulations — even at relatively coarse spatial resolution (e.g., 2×2.5 degrees) — as well as data processing and storage were very expensive at the time of our analysis and beyond the scope of this project. As a result, we did not test whether using vertical profiles of aerosol that are simulated in models makes AOD a better proxy for $PM_{2.5}$. Further — at least for the GEOS-Chem model used in Liu and colleagues (2004a) and in van Donkelaar and colleagues (2010) — the model resolution of 2×2.5 degrees limits the degree to which this approach can help in improving small-scale correlations of AOD and $PM_{2.5}$. It remains an open question whether information on smaller-scale variation in the vertical profile would help to resolve some of the discrepancy between AOD and $PM_{2.5}$ seen at smaller scales, or whether other causes of discrepancy, such as background contamination, are more important. A major challenge in assessing this question lies in obtaining sufficiently accurate information on the vertical-profile variation at smaller scales for large domains. Such differences might be studied with Cloud Aerosol LIDAR and Infrared Pathfinder Satellite Observations (CALIPSO) satellite data, although the poor spatial coverage of the satellite would limit the use of its data in actually mapping the variation. Because global three-dimensional atmospheric-chemistry and transport models (CTMs) such as GEOS-Chem can now be run at 0.5 to 1-degree resolution in the continental United States, and regional CTMs such as the CMAQ model estimate vertical profiles of aerosol at 12- to 36-km resolution, these models are worthwhile areas for research, although obtaining good emissions inventories may pose a difficulty. Two reviewers of a draft of this report commented that using CMAQ for the profiles is probably not the best approach, given that CMAQ was developed initially for short-term air quality applications and has limitations in terms of emissions inputs and parameterizations. These may limit its ability to characterize both $PM_{2.5}$ concentrations (seen empirically in section 6) and vertical profiles of aerosol.

Making use of particle-speciation information is a recent direction showing promise. Because of its multiangle design, the MISR instrument is able to distinguish spherical from nonspherical particles and bright from light-absorbing particles, and it can distinguish particle size to some degree. This capability allows a rough separation between secondary particles (spherical particles, such as sulfate, nitrate, and combustion-related organic carbon) and natural-source particles (mineral dust). It also allows a separation between light-absorbing particles (e.g., black carbon and certain organic carbonaceous species) and non-light-absorbing particles (e.g., sulfate, nitrate, and sea salt). Liu and colleagues (2007c) found that MISR AOD can be decomposed into fractional AODs from each of the MISR-defined aerosol components just mentioned. They also found that regression models using fractional AODs as predictors of $PM_{2.5}$ performed significantly better than those using total AOD as the main predictor (Liu et al. 2007b; Liu et al. 2009). Unfortunately, speciation information is only available from MISR, and MISR retrievals are substantially sparser than MODIS or GOES retrievals. Another direction of recent research has been the work of Drury and colleagues (2008), who developed a MODIS retrieval algorithm that makes use of the rich particle-speciation estimates in a CTM. These retrievals provided a considerable improvement over operational MODIS AOD in comparisons with time-averaged AERONET AOD observations in the western and central United States.

Finally, MISR and MODIS retrievals could be performed at higher nominal spatial resolution with existing data but with modified algorithms. It is not certain whether such retrievals would provide higher real resolution that would improve predictions of $PM_{2.5}$ or whether the variation in the retrieved AOD at this scale would be dominated by discrepancy.

STATISTICAL CONSIDERATIONS

As the work of assessing remote sensing and model-based proxies continues, it is worthwhile to keep in mind a danger with proxies that relates to the modeling of discrepancy. Proxies tend to be spatially correlated. A standard conclusion based on scientific experience with traditional observational data, which tend to have uncorrelated errors, is that spatial correlation is suggestive of signal. Thus, a proxy that has little signal for the process of interest may still on its face look informative, because discrepancy is confused with signal. Furthermore, in assessing a proxy, as we did for AOD (section 3), one

cannot assume that temporal correlation provided evidence for spatial correlation, particularly for long-term averages (section 3). We found little evidence that either the AOD or CMAQ proxies helped improve prediction when the proxy was modeled as data, despite the correlations between proxy values and $PM_{2.5}$ observations. We believe that future efforts that consider remote-sensing output as proxies for an exposure of interest need to focus on the question of improved prediction, rather than simply showing correlation. Much of the reason for the lack of improved prediction appears to relate to the limited information in the proxies about $PM_{2.5}$ at smaller scales, namely in the 10- to 200-km range.

We note that our models relied on other predictor variables such as land use and meteorology, so we were already leveraging additional information above and beyond simple smoothing of the observation values, such as in kriging approaches. Such predictors are of course spatially correlated, so errors from such modeling techniques are also spatially correlated, but the regression conditioning ensures that the relationship of the predictors with $PM_{2.5}$ is calibrated with the observations. Furthermore, we have substantive reason to believe that such variables are proxies for $PM_{2.5}$ at fine scales because of their relationships to emissions and meteorologic effects. We accounted for larger-scale errors caused by large-scale discrepancies between these variables and $PM_{2.5}$ in the residual spatial process (sections 5 and 6). Given this success in regressing on other predictors, using a proxy variable as a regressor appears safer from the perspective of prediction (and helped improve predictions somewhat in our use of CMAQ-based $PM_{2.5}$; section 6). However, its usefulness is derailed when many missing values occur, and it doesn't help in understanding the scales of discrepancy in a proxy. A regression approach that resolves scales may be a fruitful direction for research.

In our statistical modeling (sections 5 and 6), we proposed statistical models to combine proxies and observations, with the critical component of the model being the characterization of the proxy discrepancy. Compared to previous statistical work, we focused on representing the discrepancy in a form that allowed either large-scale or small-scale discrepancy to be accounted for, rather than assuming that all small-scale spatial patterns in the proxy represented signal for the process of interest. The resulting decomposition of proxy into discrepancy and signal was necessarily a difficult task with concerns about statistical identifiability. Accounting for discrepancy (also called structural model error) has been receiving increasing attention in the context of deterministic models (Kennedy and O'Hagan 2001; Smith 2002; Goldstein and Rougier 2009). The error structures in such models are similarly complicated and are generally not well identified from available data because of scale issues and data sparsity. Another difficulty is in the use of such models for making predictions into the future, when no data are available to help characterize systematic model discrepancy; consider global climate model runs. Traditional statistical models are based on independent, identically distributed data, and traditional spatial statistics methods on Gaussian processes. In environmental problems, which increasingly combine deterministic models, remote sensing, and traditional observations, we need statistical approaches to bring together these sources of information and deal with complicated error structures in the proxies and models. Such contexts raise serious challenges for the specification of statistical models, with model misspecification difficult to detect and correct. In effect, our statistical models, particularly in hierarchical Bayesian settings, are beginning to behave like complicated deterministic models in which the behavior of (and biases in) the model system are not well understood by examining only the relationships of the components used to build the models.

As discussed, our emphasis on statistical modeling served several goals: (1) to account for local variations in $PM_{2.5}$ that could have obscured the relationship between pixel-scale AOD and true $PM_{2.5}$ at that scale, (2) to assess improvements in predictions from adding AOD retrievals to state-of-the-art statistical models that use spatial smoothing (which is analogous to kriging) and land-use regression components fit to the observational data, and (3) to assess and account for spatial discrepancy as a function of spatial scale. Given this emphasis and the aforementioned challenges of complicated statistical models, we note that the statistical-modeling results were complemented by and consistent with simpler approaches that did not rely heavily on the form of the statistical model: correlation analyses (sections 3, 4, and 5), more simple regression modeling of MISR AOD (section 4), and the use of AOD as a covariate in statistical models (section 5).

USING AOD FOR PUBLIC HEALTH RESEARCH

Our goal was to use AOD retrievals to improve predictions of $PM_{2.5}$ concentrations for use in epidemiologic analyses. In particular, we hoped to fill in spatial gaps in the monitoring network at small-to-moderate scales. In the United States, estimation of the large-scale variation in $PM_{2.5}$ is straightforward using the observations from the monitoring network, so there is little need for making use of AOD at that scale. However, because $PM_{2.5}$ health effects are likely small in magnitude, the presence of confounding factors is a critical issue in using observational

data to investigate causality. This is even more problematic when exposure is poorly ascertained, as the combined effects of confounding and exposure error are poorly understood and likely complex. Large-scale variation in $PM_{2.5}$ is likely confounded with other variables that vary at large scales, such as demographic patterns in factors such as diet, lifestyle, and socioeconomic status (Burnett et al. 2001; Pope et al. 2002; Cakmak et al. 2003; Zeger et al. 2007). This further supports our interest in better estimation of smaller-scale variation, namely the scale of variation within and between metropolitan areas. These concerns about confounding hold true not only in the United States but in other countries and regions of the world: Caution is warranted if health effects estimates are identified from pollution contrasts at the scale of countries or regions within larger countries. At the small and moderate scales, we found little evidence that AOD was strongly related to $PM_{2.5}$ (sections 3, 4, and 7) and no evidence that it was helpful for prediction (sections 5 and 6) in the eastern United States, an area with relatively low levels of, and small spatial contrasts in, $PM_{2.5}$. Therefore, we believe that a better understanding of the reliability of AOD as a proxy at such scales, and more generally of the physical contexts in which AOD can be useful, is needed before relying on AOD as a proxy for $PM_{2.5}$ in epidemiologic contexts. Analyses and models that treat AOD as a proxy for $PM_{2.5}$ should account for the possibility of systematic discrepancy.

Measurement error is well understood in the statistical literature in cases with a simple error structure (Carroll et al. 2006). Exposure estimation for epidemiologic analysis raises difficult issues because the exposure estimates from statistical prediction models and from highly structured (in space and time) proxies such as AOD and CTM output have a complicated error structure that is difficult to characterize (Zeger et al. 2000; Gryparis et al. 2009). We do not sufficiently understand how this kind of error affects estimates of chronic health effects. In particular, we do not know how systematic differences between spatiotemporal patterns of a proxy and those inferred from ground monitors would affect epidemiologic results, especially bias in health effects estimation. Furthermore, there is no way to assess the impact of measurement error in light of the health effect estimates obtained using a given exposure-estimation approach, and the complicated error structure of such approaches does not necessarily lead to attenuation toward the null. We expect that progress on this issue can be made but that it will rely on characterizing the error structure in exposure-estimation procedures. For proxies, this involves the challenge of validating or calibrating against a gold standard. This would be particularly difficult in areas with little monitoring, even though these are precisely the areas where the proxies are needed. However, targeted monitoring campaigns may provide useful calibration data.

Our efforts pushed the limits of current satellite technology for the remote sensing of aerosol. It has only been 10 years since the launch of the MISR and MODIS instruments, which are the first generation of dedicated aerosol sensors. Current sensors are designed to study global radiative forcing of aerosol, and validation efforts have generally focused on larger scales than considered here. Current AOD products generally have large relative retrieval errors at low aerosol levels such as those generally present in the United States, in part due to difficulty in distinguishing aerosols from the background surface reflectance, a critical issue in terms of our interest in spatial variability in $PM_{2.5}$. As a result, urban-scale air pollution monitoring and public health research using satellite data are for the time being highly exploratory and far from mature. While our results suggest caution in using satellite AOD and CMAQ output to infer urban-scale variation of $PM_{2.5}$ in epidemiologic contexts, much current research focuses on improving remote-sensing retrievals and CTMs, including the integration of satellite and in situ measurements into CTMs, so continued consideration of these tools for public health research is worthwhile.

ACKNOWLEDGMENTS

We thank Steve Melly at the Harvard School of Public Health for his extensive GIS and data-management work in support of this research, without which this work would not have been possible. The GASP AOD team at NOAA, led by Shobha Kondragunta, with support from Pubu Ciren and Chuanyu Xu, provided access to GASP AOD and related data. We thank them for their comments and suggestions about the assessment of GASP AOD as a proxy for $PM_{2.5}$ and about new methods for making use of GOES channel 1 reflectance. We thank Shobha and Hortensia Moreno-Macias (at the Harvard School of Public Health) for their contributions to section 3, published in *Environmental Science & Technology*, and Doug Dockery for his assistance in initiating the Harvard–NOAA collaboration. CMAQ output for 2001 (funded by the EPRI as part of a separate project) was graciously provided by AER based on their CMAQ–Model of Aerosol Dynamics, Reaction, Ionization and Dissolution–Advanced Plume Treatment (MADRID–APT) version, with assistance from Christian Seigneur, Betty Pun, and Krish Vijayaraghavan, at AER, and Eladio Knipping and Leonard Lavin, at EPRI. The authors thank Louise Ryan, Jeff Yanosky, the Environmental Statistics group in the Biostatistics Department at the Harvard School of Public Health,

Helen Suh, Francine Laden, Charles Stanier, and Montserrat Fuentes for feedback. Data collection for the Brooklyn data used in Appendix D was funded by the Gilbert and Ildiko Butler Foundation. We thank Jonathan Levy of the Boston University School of Public Health and Leonard Zwack for access to the data and feedback on the buffer methodology of Appendix D. NARR data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, from their Web site at *www.cdc.noaa.gov*.

REFERENCES

Al-Saadi J, Szykman J, Pierce RB, Kittaka C, Neil D, Chu DA, Remer L, Gumley L, Prins E, Weinstock L, MacDonald C, Wayland R, Dimmick F, Fishman J. 2005. Improving national air quality forecasts with satellite aerosol observations. Bul Am Met Soc 86:1249–1261.

Banerjee S, Gelfand AE, Finley AO, Sang H. 2008. Gaussian predictive process models for large spatial data sets. J R Stat Soc Series B Stat Methodol 70:825–848.

Beelen R, Hoek G, Fischer P, Brandt PA, Brunekreef B. 2007. Estimated long-term outdoor air pollution concentrations in a cohort study. Atmos Environ 41:1343–1358.

Berrocal VJ, Gelfand AE, Holland DM. 2009. A spatio-temporal downscaler for output from numerical models. J Agric Biol Environ Stat 15:176–197.

Besag J, Mondal D. 2005. First-order intrinsic autoregressions and the de Wijs process. Biometrika 92:909–920.

Burnett R, Ma R, Jerrett M, Goldberg MS, Cakmak S, Pope CA 3rd, Krewski D. 2001. The spatial association between community air pollution and mortality: A new method of analyzing correlated geographic cohort data. Environ Health Perspect 109:375–380.

Cakmak S, Burnett RT, Jerrett M, Goldberg MS, Pope CA 3rd, Ma R, Gultekin T, Thun MJ, Krewski D. 2003. Spatial regression models for large-cohort studies linking community air pollution and health. J Toxicol Environ Health Part A 66:1811–1823.

Campbell JB. 1996. Introduction to Remote Sensing. Second edition. The Guilford Press, New York.

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. Measurement Error in Nonlinear Models: A Modern Perspective. Chapman & Hall CRC, Boca Raton, Florida.

Chin M, Chu A, Levy R, Remer L, Kaufman Y, Holben B, Eck T, Ginoux P, Gao Q. 2004. Aerosol distribution in the Northern Hemisphere during ACE-Asia: Results from global model, satellite observations, and sun photometer measurements. J Geophys Res 109:D23S90.

Chin M, Diehl T, Ginoux P, Malm W. 2007. Intercontinental transport of pollution and dust aerosols: Implications for regional air quality. Atmos Chem Phys 7:5501–5517.

Chin M, Ginoux P, Kinne S, Torres O, Holben BN, Duncan BN, Martin RV, Logan JA, Higurashi A, Nakajima T. 2002. Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements. J Atmos Sci 59:461–483.

Crainiceanu CM, Ruppert D, Wand MP. 2005. Bayesian analysis for penalized spline regression using WinBUGS. J Stat Soft 14:1–24.

Diggle PJ, Menezes R, Su TL. 2010. Geostatistical inference under preferential sampling. J R Stat Soc Series C 59:191–232.

Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG, Speizer FE. 1993. An association between air pollution and mortality in six U.S. cities. N Engl J Med 329:1753–1759.

Dominici F, Peng R, Bell M, Pham L, McDermott A, Zeger S, Samet J. 2006. Fine particles, air pollution and hospital admission for cardiovascular and respiratory diseases. J Am Med Assoc 295:1127–1135.

Draper D, Krnjacic M. 2006. Bayesian Model Specification. Technical Report. Department of Applied Mathematics and Statistics, University of California Santa Cruz, Santa Cruz, CA.

Drury E, Jacob DJ, Wang J, Spurr RJD, Chance K. 2008. Improved algorithm for MODIS satellite retrievals of aerosol optical depths over western North America. J Geophys Res 113:D16204.

Engel-Cox JA, Holloman CH, Coutant BW, Hoff RM. 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. Atmos Environ 38:2495–2509.

Fuentes M, Raftery AE. 2005. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. Biometrics 61:36–45.

Fuentes M, Reich B, Lee G. 2008. Spatial-temporal mesoscale modelling of rainfall intensity using gauge and radar data. Ann App Stat 2:1148–1169.

Gelfand AE, Sahu SK. 2009. Combining monitoring data and computer model output in assessing environmental exposure. In: The Handbook of Bayesian Analysis

(O'Hagan A, West M, eds.). Oxford University Press, Oxford, UK.

Goldstein M, Rougier J. 2009. Reified Bayesian modelling and inference for physical systems. J Stat Plan Infer 139:1221–1239.

Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. 2009. Measurement error caused by spatial misalignment in environmental epidemiology. Biostatistics 10:258–274.

Hoff RM, Christopher SA. 2009. Remote sensing of particulate pollution from space: Have we reached the Promised Land? J Air Waste Manag Assoc 59:645–675.

Jiang X, Liu Y, Yu B, Jiang M. 2007. Comparison of MISR aerosol optical thickness with AERONET measurements in Beijing metropolitan area. Rem Sens Environ 107:45–53.

Jun M, Stein ML. 2004. Statistical comparison of observed and CMAQ modeled daily sulfate levels. Atmos Environ 38:4427–4436.

Kahn R, Banerjee P, McDonald D, Diner D. 1998. Sensitivity of multiangle imaging to aerosol optical depth and to pure-particle size distribution and composition over ocean. Geophys Res 103:32195–32213.

Kammann EE, Wand MP. 2003. Geoadditive models. Appl Stat 52:1–18.

Karamchandani P, Vijayaraghavan K, Chen S-Y, Seigneur C, Edgerton E. 2006. Plume-in-grid modeling for particulate matter. Atmos Environ 40:7280–7297.

Kennedy MC, O'Hagan A. 2001. Bayesian calibration of computer models. J R Stat Soc Series B 63:425–464.

Knapp KR, Vonder Haar TH, Kaufman YJ. 2002. Aerosol optical depth retrieval from GOES-8: Uncertainty study and retrieval validation over South America. J Geophys Res 107:D4055.

Koelemeijer RBA, Homan CD, Matthijsen J. 2006. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. Atmos Environ 40:5304–5315.

Kumar N, Chu A, Foster A. 2008. Remote sensing of ambient particles in Delhi and its environs: Estimation and validation. Int J Rem Sens 29:3383–3405.

Laden F, Schwartz J, Speizer FE, Dockery DW. 2006. Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities Study. Am J Res Crit Care Med 173:667–672.

Liu Y, Franklin M, Kahn R, Koutrakis P. 2007a. Using aerosol optical thickness to predict ground-level $PM_{2.5}$ concentrations in the St. Louis area: A comparison between MISR and MODIS. Rem Sens Environ 107:33–44.

Liu Y, Kahn R, Turquety S, Yantosca RM, Koutrakis P. 2007b. Estimating $PM_{2.5}$ component concentrations and size distributions using satellite retrieved fractional aerosol optical depth: Part II — A case study. J Air Waste Manag Assoc 57:1360–1369.

Liu Y, Koutrakis P, Kahn R. 2007c. Estimating $PM_{2.5}$ component concentrations and size distributions using satellite retrieved fractional aerosol optical depth: Part I — Method development. J Air Waste Manag Assoc 57:1351–1359.

Liu Y, Park RJ, Jacob DJ, Li Q, Kilaru V, Sarnat JA. 2004a. Mapping annual mean ground-level $PM_{2.5}$ concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. J Geophys Res 109:D22206.

Liu Y, Sarnat JA, Coull BA, Koutrakis P, Jacob DJ. 2004b. Validation of Multiangle Imaging Spectroradiometer (MISR) aerosol optical thickness measurements using aerosol robotic network (AERONET) observations over the contiguous United States. J Geophys Res 109:D06205.

Liu Y, Sarnat JA, Kilaru V, Jacob DJ, Koutrakis P. 2005. Estimating ground-level $PM_{2.5}$ in the eastern United States using satellite remote sensing. Environ Sci Technol 39:3269–3278.

Liu Y, Schichtel BA, Koutrakis P. 2009. Estimating particle sulfate concentrations using MISR retrieved aerosol properties. IEEE J Sel Top Appl Earth Observ Remote Sens 2:176–184.

McMillan NJ, Holland DM, Morara M, Feng J. 2010. Combining numerical model output and particulate data using Bayesian space-time modeling. Environmetrics 21:48–65.

Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jović D, Woollen J, Rogers E, Berbery EH, Ek MB, Fan Y, Grumbine R, Higgins W, Li H, Lin Y, Manikin G, Parrish D, Shi W. 2006. North American regional reanalysis. Bul Am Met Soc 87:343–360.

Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson GL, Kaufman JD. 2007. Long-term exposure to air pollution and incidence of cardiovascular events in women. N Engl J Med 356:447–459.

Mugglin AS, Carlin BP, Gelfand AE. 2000. Fully model-based approaches for spatially misaligned data. J Am Stat Assoc 95:877–887.

Paciorek CJ. 2007. Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP package. J Stat Soft 19:2.

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

Paciorek CJ, Liu Y, Moreno-Macias H, Kondragunta S. 2008. Spatio-temporal associations between GOES aerosol optical depth retrievals and ground-level $PM_{2.5}$. Environ Sci Technol 42:5800–5806.

Paciorek CJ, Yanosky JD, Puett RC, Laden F, Suh HH. 2009. Practical large-scale spatio-temporal modeling of particulate matter concentrations. An Appl Stat 3:369–396.

Pelletier B, Santer R, Vidot J. 2007. Retrieving of particulate matter from optical measurements: A semiparametric approach. J Geophys Res 112:D06208.

Pope CA 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston G. 2002. Lung cancer, cardiopulmonary mortality and long-term exposure to fine particulate air pollution. J Am Med Assoc 287:1132–1141.

Pope CA 3rd, Burnett RT, Thurston GD, Thun MJ, Calle EE, Krewski D, Godleski JJ. 2004. Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease. Circulation 109:71–77.

Prados AI, Kondragunta S, Ciren P, Knapp KR. 2007. GOES aerosol/smoke product (GASP) over North America: Comparisons to AERONET and MODIS observations. J Geophys Res 112:D15201.

Puett RC, Hart JE, Yanosky JD, Paciorek CJ, Schwartz J, Suh HH, Speizer FE, Laden F. 2009. Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. Environ Health Perspect 117:1697–1701.

Remer LA, Kaufman YJ, Tanre D, Mattoo S, Chu DA, Martins JV, Li RR, Ichoku C, Levy RC, Kleidman RG, Eck TF, Vermote E, Holben BN. 2005. The MODIS aerosol algorithm, products, and validation. J Atmos Sci 62: 947–973.

Robinson IS. 2004. Measuring the Oceans from Space: The Principles and Methods of Satellite Oceanography. Springer, Berlin, Germany.

Rogers R, Kondragunta S, McQueen J, Moshary F, Gross B, Hoff R. 2006. Validation of ETA-CMAQ-Modeled Planetary Boundary Layer Height with Elastic Lidar. Technical Report. 7th ISTP Symposium, Boulder, CO.

Rue H, Held L. 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall, Boca Raton, FL.

Ruppert D, Wand MP, Carroll RJ. 2003. Semiparametric Regression. Cambridge University Press, Cambridge, UK.

Rush AC, Dougherty JJ, Engel-Cox JA. 2004. Correlating seasonal averaged in-situ monitoring of fine PM with satellite remote sensing data using geographic information system (GIS). In: Proceedings of SPIE, Volume 5547 (Chu A, Szykman J, eds.). SPIE, Bellingham, WA.

Sahu SK, Gelfand AE, Holland DM. 2009. Fusing point and areal level space-time data with application to wet deposition. J R Stat Soc Ser C Appl Stat 59:77–103.

Samet JM, Dominici F, Curriero F, Coursac I, Zeger SL. 2000. Particulate air pollution and mortality: Findings from 20 U.S. cities. N Engl J Med 343:1742–1757.

Silverman BW. 1984. Spline smoothing: The equivalent variable kernel method. Ann Stat 12:898–916.

Smith LA. 2002. What might we learn from climate forecasts? Proc Natl Acad Sci U S A 99:2487–2492.

Smith RL, Kolenikov S, Cox LH. 2003. Spatiotemporal modeling of $PM_{2.5}$ data with missing values. J Geophys Res 108:D9004.

Sollich P, Williams CKI. 2005. Using the equivalent kernel to understand Gaussian Process regression. In: Advances in Neural Information Processing Systems 17. MIT Press, Cambridge, MA.

Stein ML, Fang D. 1997. Discussion of ozone exposure and population density in Harris County, Texas, by R.J. Carroll et al. J Am Stat Assoc 92:408–411.

Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar SD, Kaufman JD. 2010. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. Environmetrics 21:606-631.

U.S. Census Bureau. United States — County by State, and for Puerto Rico. GCT-PH1. Population, Housing Units, Area, and Density. Census 2000. Accessed on December 14, 2007, at *www.census.gov/geo/www/gazeteer/places2k.html#counties*. (This site is no longer active. Refer to *http://factfinder.census.gov/servlet/GCTTable?_bm=y&-ds_name=DEC_2000_SF1_U&-CONTEXT=gct&-mt_name=DEC_2000_SF1_U_GCTPH1_US9&-redoLog=false&-_caller=geoselect&-geo_id=&-format=US-25|US-25S&-_lang=en*.)

van Donkelaar A, Martin RV, Brauer M, Kahn R, Levy R, Verduzo C, Villeneuve P. 2010. Global estimates of exposure

to fine particulate matter concentrations from satellite-based aerosol optical depth. Environ Health Perspect 118:847–855.

van Donkelaar A, Martin RV, Park RJ. 2006. Estimating ground-level $PM_{2.5}$ using aerosol optical depth determined from satellite remote sensing. J Geophys Res 111: 21201.

Wang J, Christopher SA. 2003. Intercomparison between satellite-derived aerosol optical thickness and $PM_{2.5}$ mass: Implications for air quality studies. Geophys Res Lett 30:2095.

Wood SN. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall, Boca Raton, FL.

Wu J, Winer AM, Delfino RJ. 2006. Exposure assessment of particulate matter air pollution before, during, and after the 2003 southern California wildfires. Atmos Environ 40:3333–3348.

Yanosky JD, Paciorek CJ, Schwartz J, Laden F, Puett R, Suh HH. 2008. Spatio-temporal modeling of chronic PM_{10} exposure for the Nurses' Health Study. Atmos Environ 42:4047–4062.

Yanosky JD, Paciorek CJ, Suh HH. 2009. Predicting chronic fine and coarse particulate exposures using spatio-temporal models for the northeastern and midwestern U.S. Environ Health Perspect 117:522–529.

Yue Y, Speckman PL. 2010. Nonstationary spatial Gaussian Markov random fields. J Comput Graphl Stat 19:96–116.

Zeger SL, Dominici F, McDermott A, Samet JM. 2007. Mortality in the Medicare Population and Chronic Exposure to Fine Particulate Air Pollution. Technical Report 133. Johns Hopkins University Department of Biostatistics, Baltimore, MD.

Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, Cohen A. 2000. Exposure measurement error in time series studies of air pollution: Concepts and consequences. Environ Health Perspect 108:419–426.

Zhang Y, Pun B, Vijayaraghavan K, Wu S-Y, Seigneur C, Pandis S, Jacobson M, Nenes A, Seinfeld JH. 2004. Development and application of the Model of Aerosol Dynamics, Reaction, Ionization and Dissolution. J Geophys Res 109:D01202.

Zhou Y, Levy JI. 2007. Factors influencing the spatial extent of mobile source air pollution impacts: A meta-analysis. BMC Public Health 7:89.

APPENDIX A. Data Description

GIS PROCEDURES

Software and Background

We used ArcGIS 9.2 (Esri, Redlands, CA) to perform spatial analyses on various datasets in order to link together data with different spatial resolutions and to create variables that could be used in exploratory statistical analysis and statistical modeling. "Spatial dataset" is used here to refer to a dataset that has features, including points, polygons, or lines, that are associated with both spatial information and nonspatial attributes. Spatial datasets supported by ArcGIS used in this project included personal geodatabases, file geodatabases, and shapefiles. ESRI software and data were used under a site license agreement with Harvard University.

Projection

The primary spatial extent of this project was the United States east of 100° W longitude (the "study area"). Because of the size of the study area, a continental projection was used, the USA Contiguous Albers Equal Area Conic projection, U.S. Geological Survey version. This is referred to as the study projection. The projection details are as follows: standard parallels = 29.5, 45.5; longitude of central meridian = −96; latitude of projection origin = 23; distance is in meters; horizontal datum = North American Datum of 1983 (NAD1983); ellipsoid = geodetic reference system (GRS) 80; semimajor axis = 6378137; denominator of flattening ratio = 298.257222.

Base Grid of 4 km

We used ET GeoWizards (*www.ian-ko.com*) to create a 4-km grid that included south Texas, with a starting point at 100° W longitude. The size of the grid was 669 × 677, with the southwest-corner cell centroid at (−417082.6, 310832.93) and the northeast-corner cell centroid at (2254917.39,3014832.93). Each grid cell was assigned a unique ID number. Grid cells were intersected with Esri Data & Maps 8.3 for U.S. counties, Mexico, and Canada to determine if cells intersected the United States or land surface. Grid cells that intersected the United States were coded as inUS = 1. Grid cells that intersected land were coded as landMask = 1. Grid cells with inUS = 1 that also intersected the area east of 100° W longitude were coded as dataMask = 1. In general, we did not plot predictions for cells with dataMask = 0.

For our eastern United States analyses, we used the entire 669 × 677 grid. For our mid-Atlantic analyses, we

used a 175 × 100 subset of the grid with the southwest-corner cell centroid at (1202917.39,1902832.93) and the northeast-corner cell centroid at (1898917.39,2298832.93).

Grid cells were associated with state and county Federal Information Processing Standards codes. Cells with centroids that fell inside a county were assigned to that county. Cells that intersected one or more counties but with centroids that fell outside all counties were assigned to the county that intersected the grid cell the most. Detailed county boundaries were obtained from Esri StreetMap USA 9.2 (streetmap9_2\census\dtl_cnty.sdc).

DATA SOURCES AND MANIPULATIONS

$PM_{2.5}$

We downloaded 24-hour averages of $PM_{2.5}$ data from the U.S. EPA's Technology Transfer Network Web site, *www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm*. We downloaded AQS data (parameter code 88101) and IMPROVE data (parameter code 88502) for the years 2000 through 2006. We downloaded AQS data for 2000 through 2005 on September 29, 2006, and for 2006 on June 9, 2009. We downloaded IMPROVE data for 2000 through 2004 on November 10, 2006, and for 2005 through 2006 on June 9, 2009. For some initial analyses, we also downloaded hourly $PM_{2.5}$ data (parameter code 88101) on September 28, 2006. We included all data regardless of qualifiers based on a conversation with Michael Papp of the U.S. EPA (November 13, 2006), who said that data were only supposed to be submitted if the monitoring agency believed the data were valid for attainment analysis. We modified one outlier, at site 29-125-0001 (just southeast of Belle, Missouri, about 50 miles southwest of St. Louis), which reported a daily value of 640.6 on September 13, 2001. Given the presence of high concentrations on other days and values of 50 $\mu g/m^3$ on days just before this one, we assumed the true value was 64.6. Data were read from flat text files into the statistical software, SAS (SAS Institute Inc., Cary, NC), used under license to the Harvard School of Public Health.

In some analyses, we used only the primary monitor at a site with two monitors, to avoid double-counting. In our statistical modeling described in sections 5 and 6, we treated each monitor as contributing individual data points, and the model errors were structured so that only instrument error contributed to differences in the model representation of the colocated data points.

Information on the monitoring sites was obtained from the appropriate rows in the flat text files and converted to a spatial dataset (personal geodatabase feature class) in the study projection. The x- and y-coordinates in the study projection were added as attributes. Separate datasets were created for the contiguous United States and for the study area. A spatial join was performed with the 4-km grid, to obtain the ID of the grid cell where each monitor was located.

NARR

Data for a number of meteorologic variables were downloaded from the NARR (Mesinger et al. 2006) Web site (*www.esrl.noaa.gov/psd/data/gridded/data.narr.html*). The NARR provides estimates of fields based on data assimilated through a fixed, state-of-the-art meteorologic model on a 32-km grid with 3-hour sampling intervals. From the "monolevel" data grouping, we obtained the RH at 2 m above the surface (file names beginning with rhum.2m), planetary boundary layer (hpbl), u and v wind speeds at 10 m above the surface (uwnd.10m, vwnd.10m), accumulated precipitation (apcp), temperature at 2 m above the surface (air.2m), and pressure at mean sea level (prmsl). The RH, winds, and PBL were obtained on November 18, 2006, and the remaining variables were obtained on June 17, 2009. We read the netcdf files into the statistical software R using the ncdf package.

NARR grid points were on an approximate 32-km grid in the Lambert North American conformal conic projection with these projection details: false easting = 5632642.22547; false northing = 4612545.65137; central meridian = −107; standard parallels = 50, 50; latitude of origin = 50; datum = NAD1983; and prime meridian = 0. NARR grid points were converted to a spatial dataset in the study projection. The ArcGIS point distance tool was used to calculate distance between monitoring sites and NARR points for all NARR points within 40 km of monitoring sites. Microsoft Access (Microsoft, Redmond, WA) was used to join the result to the NARR file on the feature ID (FID) and to identify the four NARR points nearest to each monitoring site. The same procedure was followed to find the four NARR points nearest to the centroid of each cell in the 4-km grid.

To estimate a meteorologic field value for a 4-km-grid cell or a monitor location, we generally calculated the IDW average of field values at the four NARR grid points nearest to the location of interest, for a given 3-hour time period, and then calculated a 24-hour average based on the eight 3-hour values (in UTC) coinciding with a calendar day in the eastern United States.

National Emissions Inventory

We downloaded data from the U.S. EPA's 2002 inventory on November 29, 2007, from *www.epa.gov/ttn/chief/net/2002inventory.html*. We made use of the county area Tier 1

$PM_{2.5}$ emissions and of $PM_{2.5}$ primary point emissions. Point emissions locations were converted to a GIS spatial dataset in the study projection. Two obvious spatial errors were corrected (Norfolk Naval Shipyard in Portsmouth, Virginia, and Gulf Coast Recycling in Tampa) by geocoding their addresses provided with Esri StreetMap 9.2 streets. To normalize the county area emissions, we divided the emissions by the land area of each county, obtained from the Census 2000 Summary File 1 (U.S. Census 2000) on December 14, 2007.

CMAQ Model

We obtained CMAQ output for 2001 from a 36-km model run over the entire United States, provided by AER and funded by the EPRI. The model version was the CMAQ–Model of Aerosol Dynamics, Reaction, Ionization and Dissolution–Advanced Plume Treatment (MADRID–APT) version developed by AER. MADRID is a $PM_{2.5}$ module developed at AER, and APT offers a more accurate treatment of large-point-source emissions. AER provided hourly three-dimensional fields of $PM_{2.5}$ mass and components in 2001 at a 36-km horizontal resolution with 14 vertical layers over the United States in CMAQ netcdf format. Three-dimensional fields of model input (MM5-derived) pressure, temperature, and layer heights were also provided, to facilitate calculation of AOD, which we did not end up doing.

The MADRID module is described in Zhang and colleagues (2004), and the APT module is described in Karamchandani and colleagues (2006).

CMAQ raw data files in netCDF format were read using Interactive Data Language (IDL). Flat text files for $PM_{2.5}$ species and meteorologic parameters were created, with individual files containing information about one species or parameter for all vertical layers and all hours (UTC) in a single day.

CMAQ output was provided on a 36-km grid with 6036 pixels in the United States Lambert conformal conic projection with the southwest-corner pixel at (−2736000, −2088000) (i.e., this was the southwest corner of the southwest-most pixel). For our analyses in section 6, we used a 19×11 subset of the grid over our mid-Atlantic region, with the southwest-corner pixel centered at (1314000,54000), the northeast-corner pixel at (1962000, 414000), and a 73×77 grid over the eastern United States, with the southwest-corner pixel centered at (−270000, −1566000) and the northeast-corner pixel at (2322000, 1170000). The projection details were as follows: standard parallels = 33, 45; central meridian = −97; latitude of origin = 40; and datum = GRS 80.

We calculated the overlap between CMAQ pixels and 4-km-grid cells as follows: Because projecting the CMAQ pixels into the study projection resulted in some distortion, with distances between pixel centroids greater than 36 km at higher latitudes, the 4-km grid was converted to the Lambert projection. The 4-km grid was then intersected with the CMAQ grid using the ArcGIS intersect tool (ArcToolbox/Analysis tools/Overlay), which created polygons for each CMAQ–4-km-grid cell combination and calculated the area in square meters for each polygon. We used Microsoft Access for each CMAQ–4-km-grid cell combination, to calculate the percent of the cell in the 4-km grid covered by the intersection. The result could be used to calculate a weighted average of variables calculated at the 4-km-grid resolution for each CMAQ pixel. We also calculated in R the nearest pixel for each grid cell, by minimizing the distance between pixels and cell centroids.

We calculated the CMAQ pixel nearest to each monitoring site as follows: Monitoring sites were projected in the Lambert projection. We used the ArcGIS 9.2 Near tool to identify the CMAQ pixel nearest to each monitoring site and the distance in meters between the two.

MISR

We downloaded MISR data products for 2000 through 2006 from the NASA Langley Research Center (LARC) Atmospheric Science Data Center (*http://eosweb.larc.nasa.gov*). For section 4, we used version 22, downloaded in June 2009. For section 5, we used version 15, downloaded in 2008. The parameters used in this analysis, including best-estimate AOD in the green band (558 nm), Retrieval Success Flag, Regional Class, and Type flag, were extracted from the raw data in hierarchical data format (HDF) using the EOS_GD_READFIELD function in the IDL software. The original ancillary geographic product (AGP) data has a resolution of 1.1 km. For each 17.6-km AOD pixel, a $16 \times 16 = 256$ set of 1.1-km pixels was averaged. Pixels were nominally 17.6-km squares.

MISR paths during an orbit around the globe repeat in precisely the same locations every 16 days, with identical pixel locations over time for all orbits of the earth on a given path. A spatial dataset of pixel centroids for MISR paths 8 through 36 was obtained from the MISR AGP, also downloaded from LARC, and projected in the study projection.

We calculated the MISR pixel nearest to each monitoring site as follows: Each path was processed separately. Monitoring sites within 12.4 km (where 12.4 km is the distance from the center of a 17.6-km square to a corner) of a path were considered to be associated with the nearest MISR pixel in that path. The ArcGIS Near tool was used to identify the FID of the MISR pixel centroid nearest each

monitoring site. The MISR pixel ID was obtained by joining to the MISR path spatial dataset on the FID.

We calculated the MISR pixel nearest to each centroid of a cell in the 4-km grid as follows: A point spatial dataset containing centroids of 4-km-grid cells was created. Each path was processed separately. The 4-km-grid cells were intersected with 12.4-km buffers around each path and saved as separate spatial datasets. The ArcGIS Near tool was run for each path to identify the FID of the MISR pixel nearest to each 4-km-grid cell. The MISR pixel ID was obtained by joining to the MISR path spatial dataset on the FID. Microsoft Access was used to combine the results for each path into one file. Since paths overlapped, each 4-km-grid cell was associated with three to six MISR pixels, representing multiple paths.

MODIS

We downloaded MODIS collection 5 aerosol data for 2000 through 2006 from the Goddard Space Flight Center MODIS Level 1 and Atmosphere Archive and Distribution System (*http://ladsweb.nascom.nasa.gov*) in February 2007. The parameters used in this analysis, including Optical_Depth_Land_And_Ocean, Longitude, and Latitude, were extracted from the raw MODIS data in HDF format in IDL using the EOS_SW_READFIELD function. MODIS pixels had a nominal resolution of 10 km.

MODIS does not having repeating paths, so pixels do not occur at the same locations over time and differ from orbit to orbit. This makes working with the MODIS output difficult. MODIS data were extracted from HDF files into CSV files that included latitude and longitude. These files were added as XY data and converted to file geodatabase feature classes in the study projection. Each year from 2000 through 2006 was split into 11 to 14 different time periods.

We calculated the MODIS pixels nearest to the centroids of the 4-km-grid cells at each point in time when MODIS data were available from that location. After some trial and error, a procedure was developed for handling the large MODIS files. The 4-km-grid cells were broken into two files (north and south). Results were processed using SAS running on the Harvard School of Public Health high-performance computer cluster ("the cluster"). The SAS code broke the north results into two separate files. Microsoft Access was used to import the MODIS CSV files with latitude and longitude into a personal geodatabase table using Access Import/Export Specifications to specify the field names. The geodatabase table was added as XY data and exported to a geodatabase feature class in the study projection. The FID (ObjectID) associated with each feature was used as the MODIS pixel ID within each MODIS file. This ID was renamed MODIS*n*_ID where *n* was a file number from 1 through 14. The point distance tool was used to calculate the distance between MODIS pixel centroids and 4-km-grid centroids for pixels within 25 km of the 4-km-grid centroid. This step was done separately for the north and the south, and each portion of the country took approximately 20 minutes to complete. The results were saved as .dbf files, which could be imported directly into SAS datasets. The MODIS feature classes that included the MODIS IDs were imported to SAS directly from personal geodatabases (*.mdb files) on a desktop computer, or were first exported to .dbf files and then imported to SAS datasets on the cluster. SAS was used on the cluster to sort the distance files by MODIS ID; join the distance files to MODIS files; sort by 4-km-grid FID; join the result to 4-km-grid files containing the 4-km-grid IDs; sort the result by 4-km-grid ID, orbit, and distance; and finally create a dataset that identified the MODIS pixel in each orbit nearest to each 4-km-grid cell. This SAS processing took several minutes for each distance file. These results could be linked back to the original extracted MODIS files by year, file name, and MODIS*n*_ID.

For an analysis of 2004 data, we calculated the MODIS pixels for 2004 within 7.1 km of each monitoring site as follows: The ArcGIS point distance tool was used to calculate distances between MODIS pixel centroids and monitoring sites for MODIS centroids within 7.1 km of each monitoring site. Microsoft Access was used to join to MODIS data on the FID.

GOES

Via an FTP site, we obtained from NOAA, courtesy of Shobha Kondragunta and the NOAA GASP team, GOES-8 (East) data (for the years 2001 through 2002) and GOES-12 (East) data (for the period April 24, 2003, through November 30, 2007). The data included GASP AOD and fields used to calculate AOD, such as channel 1 reflectance and the cloud mask information. GOES pixels have a nominal spatial resolution of 4 km. GOES is a geostationary satellite, and measurements are made every half hour during daytime.

The NOAA GASP team updates the AOD retrieval algorithm occasionally; our values were from algorithms current at the time of downloading (2001 data: downloaded March 2006; 2002 data: May 2007 and September 2008; 2003 data: January 2007; 2004 data: September 2006; 2005 data: September 2006; 2006 data: May 2006 and February 2007; 2007 data: December 2007 and June 2008). In 2009, the team updated their algorithm and applied the algorithm to the period of June through August 2004 for our use; we downloaded these retrievals in April 2009. GASP AOD core data processing codes for GOES-8 and GOES-12 data files were provided by the GASP team and adapted for IDL. GOES pixel centroids provided by the GASP team in

ASCII format were first imported into ArcGIS and clipped according to this study's spatial domain. The selected GOES pixel coordinates were merged with aerosol data in the IDL code.

We followed the same procedures for both GOES-8 and GOES-12. Since GOES is a geostationary satellite, the pixels for GOES values are at the same locations over time. GOES pixel centroid coordinates were converted to a spatial dataset in the study projection, with x- and y-coordinates in the study projection added as attributes.

We calculated the GOES pixel nearest to each monitoring site using the ArcGIS 9.2 Near tool. We calculated the overlap between GOES pixels and 4-km-grid cells as follows: In the study projection, the distance between the GOES pixel centroids varied across the study area. Using the ArcGIS Thiessen polygons tool (ArcToolbox/Analysis Tools/Proximity/Create Thiessen Polygons), Thiessen polygons were created from the GOES pixel centroids, to approximate the area covered by each pixel. GOES pixels were limited to those within 200 km of the eastern United States (east of 100° W longitude). The Thiessen polygons were intersected with the 4-km-grid cells using the ArcGIS intersect tool (ArcToolbox/Analysis tools/Overlay), and the area of intersection for each pixel–grid-cell combination was saved. In addition, Microsoft Access was used to identify the 4-km-grid cell that intersected the largest portion of the GOES Thiessen polygon.

GIS-DERIVED VARIABLES

Values of various land-use and related variables were calculated as follows for each cell in the 4-km grid:

Population Density

Census block boundaries were obtained from the Harvard Geospatial Library (*http://dixon.hul.harvard.edu:8080/HGL/hgl.jsp*) and from *www.geographynetwork.com* (a few counties were missing from the Harvard Geospatial Library data). Census block boundaries were intersected with 4-km-grid cells. Population was assumed to be uniformly distributed throughout a census block. Census-block areas were obtained from the U.S. Census. These areas may be slightly different from areas calculated using GIS software and census polygons. Using SAS, the fraction of a census block area that fell inside a 4-km-grid cell was calculated and multiplied by the total population of the census block. These results were summed up for each 4-km-grid cell to provide an estimated population for each cell. Note that no population data for Canada or Mexico were obtained, so population estimates for 4-km-grid cells that crossed the border represented only the U.S. population.

Elevation

The North American Digital Elevation Model — from the U.S. Geological Survey, EROS Data Center Distributed Active Archive Center (series name ESRI Data & Maps) — was converted to the study projection. ArcGIS Spatial Analyst Extract to Points, with the interpolation option selected, was used to extract the elevation in meters of the centroids of the 4-km grids. Elevations were also extracted for monitoring sites.

Road Density and Distance to Major Roads

Esri StreetMap 9.2 roads for the United States east of 100° W longitude were used. The following types of roads were exported as three separate feature classes to a file geodatabase in the study projection: from StreetMap USA\streets\mjr_hwy\, Class 1 (primary roads with limited access) and Class 2 (primary roads without limited access), and from StreetMap USA\streets\highways\, Class 3 (secondary and connecting roads, state and county highways). Roads in each of the three road-feature classes were intersected separately with the 4-km grid. The resulting attribute tables, which contain the length in meters of each intersected street segment, were exported to a Microsoft Access file. Microsoft Access queries were used to sum up the road-segment lengths by 4-km-grid ID and to combine the results into one table for export to a comma-delimited text file. In addition, roads for the entire contiguous United States from the same source were used to calculate the distances to the nearest Class 1, Class 2, and Class 3 roads from each monitoring site.

Land Use and Land Cover

We obtained the 2001 National Land Cover Database from the Multi-Resolution Land Characteristics Consortium (*www.mrlc.gov*). The database classifies 30-m pixels using the following categories: open water; developed, open space; developed, low intensity; developed, medium intensity; developed, high intensity; barren land; deciduous forest; mixed forest; evergreen forest; shrub/scrub; grassland/herbaceous; pasture/hay; cultivated crops; woody wetlands; and emergent herbaceous wetlands. Each pixel was assigned a single category. We used the ArcGIS Spatial Analyst Combine command to intersect the land-cover pixels and 4-km-grid cells. In Microsoft Access we then calculated the proportion of each 4-km-grid cell described by its land-cover category. For cells along the Canadian and Mexican borders, the proportions were based on 30-m pixels inside the United States. At the time of downloading, working with the data was time-consuming, so in the process of dividing the region for downloading, some cells along the western border of the 4-km-grid region were excluded. In addition, a small

number of pixels along the Missouri–Arkansas border and the eastern edge of Nantucket were excluded accidentally. These areas were treated as having missing land cover and were not used in the analysis in section 4.

CURRENT DATA AVAILABILITY

Although we cannot post the satellite data online because of the expense associated with hosting such large data files, data used in this project can be obtained from Dr. Paciorek. Provision of GOES output may require permission from Dr. Shobha Kondragunta at NOAA. The same holds for CMAQ output provided by AER and EPRI. Most of our other data are publicly available online, but our cleaned and formatted versions can be made available on request.

APPENDICES AVAILABLE ON THE WEB

Appendices B, C, D, and E contain supplemental material not included in the printed report. They are available on the HEI Web site *http://pubs.healtheffects.org*.

Appendix B. Additional Analysis of Spatiotemporal Associations between GOES Aerosol Optical Depth Retrievals and Ground-level $PM_{2.5}$

Appendix C. Statistical Details for Flexible Spatial Latent Variable Modeling

Appendix D. Flexible Buffer Modeling Using Penalized Splines

Appendix E. R Code for Flexible Buffer Modeling

ABOUT THE AUTHORS

Christopher J. Paciorek was an assistant professor in the Department of Biostatistics at Harvard School of Public Health while this research was conducted. He is currently a researcher, lecturer, and statistical computing consultant at the University of California–Berkeley. He has a Ph.D. and a master's degree in statistics from Carnegie Mellon University and a master's degree in botany (ecology) from Duke University. Dr. Paciorek's research interests are in the areas of Bayesian statistics and statistics for spatial and spatiotemporal data. The applied focus of his work is in environmental applications, including environmental health, ecology, and climate. Much of his recent work has been applied to improving exposure assessment and understanding statistical issues in environmental epidemiology.

Yang Liu graduated from Harvard School of Engineering and Applied Sciences in 2004 with a Ph.D. degree in environmental sciences and engineering. He joined the Rollins School of Public Health of Emory University as Assistant Professor in the Department of Environmental and Occupational Health in 2009. Dr. Liu is an environmental scientist, with primary experience in the applications of satellite remote sensing, atmospheric model simulations, spatial statistics in air pollution exposure modeling, and the health impacts of climate change. His ongoing work includes (1) PM exposure modeling in the northeastern United States in relation to birth outcomes; (2) an investigation of improving the accuracy of satellite-derived PM concentrations using space-borne and ground-based LIDAR profiles; (3) coupling satellite observations with ground monitoring to study the associations between cardiovascular hospital admissions and air pollution in Beijing, China; and (4) developing environmental public health indicators related to particulate matter pollution in metro Atlanta for the CDC Public Health Tracking Network.

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

Paciorek CJ. 2012. Combining spatial information sources while accounting for systematic errors in proxies. J R Stat Soc Ser C (Appl Stat) 61:429–451.

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

Paciorek CJ, Liu Y, Moreno-Macias H, Kondragunta S. 2008. Spatio-temporal associations between GOES aerosol optical depth retrievals and ground-level $PM_{2.5}$. Environ Sci Technol 42:5800–5806.

ABBREVIATIONS AND OTHER TERMS

AER	Atmospheric and Environmental Research
AERONET	Aerosol Robotic Network
AGP	ancillary geographic product
AOD	aerosol optical depth
AQS	Air Quality System
AR	autoregressive
CALIPSO	Cloud-Aerosol LIDAR and Infrared Pathfinder Satellite Observations
CAR	conditional autoregressive
CI	confidence interval
CMAQ	Community Multi-Scale Air Quality model
CTM	chemistry and transport model

EPRI	Electric Power Research Institute
FID	feature ID
FRM	federal reference method
GASP	GOES East Aerosol/Smoke Product
GCV	generalized cross-validation
GIS	geographic information system
GOES	Geostationary Operational Environmental Satellite
GRS	geodetic reference system
HDF	hierarchical data format
hPa	hectopascal
IDL	Interactive Data Language
IDW	inverse distance-weighted
IMPROVE	Interagency Monitoring of Protected Visual Environments
LARC	NASA Langley Research Center
LIDAR	light detection and ranging
MADRID–APT	Model of Aerosol Dynamics, Reaction, Ionization and Dissolution–Advanced Plume Treatment
MCMC	Markov chain Monte Carlo
MISR	multiangle imaging spectroradiometer
MODIS	moderate resolution imaging spectroradiometer
MRF	Markov random field
NARR	North American Regional Reanalysis
NASA	U.S. National Aeronautics and Space Administration
NEI	National Emissions Inventory
NOAA	U.S. National Oceanic and Atmospheric Administration
PBL	planetary boundary layer
PM	particulate matter
$PM_{2.5}$	PM with an aerodynamic diameter ≤ 2.5 μm
RH	relative humidity
RMSPE	root mean squared prediction error
SH	specific humidity
TPS	thin plate spline
U.S. EPA	U.S. Environmental Protection Agency
UTC	Coordinated Universal Time, i.e., Greenwich Mean Time

Research Report 167, *Assessment and Statistical Modeling of the Relationship Between Remotely Sensed Aerosol Optical Depth and $PM_{2.5}$ in the Eastern United States*, C.J. Paciorek and Y. Liu

INTRODUCTION

Over the past decade, satellite-based estimates of ground-level pollution have emerged as a potentially important source of information on human exposure to health-damaging pollutants such as fine particulate matter (*$PM_{2.5}$; PM with an aerodynamic diameter of 2.5 µm or smaller) and nitrogen dioxide (NO_2). Health effects researchers have now begun to apply satellite-based estimates in both epidemiologic research and risk assessment; however, their ultimate utility remains the subject of debate regarding the accuracy and precision of estimates provided by different satellite-based estimators and the circumstances in which such estimates might make the most important contributions.

Dr. Christopher J. Paciorek of the Harvard School of Public Health submitted an application under Request for Applications 05-2, the "Walter A. Rosenblith New Investigator Award," which was established to provide support for an outstanding investigator beginning his or her independent research career. In his proposed study, "Integrating Monitoring and Satellite Data to Estimate $PM_{2.5}$ Exposure and Its Chronic Health Effects in the Nurses' Health Study," Paciorek planned to reanalyze data on the chronic health effects of $PM_{2.5}$ in the Nurses' Health Study, a large epidemiologic cohort study, by integrating satellite data with ground monitoring data to improve the exposure-assessment modeling. The HEI Health Research Committee urged Paciorek to focus his study on the estimates of exposure, rather than the epidemiologic study. Ultimately, HEI funded the current study, "Integrating Monitoring and Satellite Data to Retrospectively Estimate Monthly $PM_{2.5}$ Concentrations in the Eastern United States," which began in 2006.

This Commentary is intended to aid the sponsors of HEI and the public by highlighting both the strengths and limitations of the study and by placing the Investigators' Report into scientific and regulatory perspective.

SCIENTIFIC BACKGROUND

Epidemiologic studies of the health effects of outdoor air pollution most often rely on measurements of air pollution at fixed monitoring sites. Air pollution concentrations have been measured routinely in North America and Western Europe since the 1970s to ensure compliance with air quality standards and directives; currently, these networks are extensive. In epidemiologic studies, the monitored levels are typically used either alone or as part of more complex spatiotemporal models that may provide more refined estimates of exposure among study participants (Jerrett et al. 2005; Szpiro et al. 2011). Nonetheless, sparse spatial and temporal coverage by existing monitoring networks in some areas of Europe and North America still limits the scope and size of potential health effects studies; the large American Cancer Society (ACS) cohort study of particulate air pollution and mortality could include less than half of the members of the ACS cohort because of limited monitoring data. In addition, many Canadians live in areas with no ground-level monitoring (Krewski et al. 2009; Crouse et al. 2012). Moreover, there is limited information on exposure to particulate air pollution in most of the world, specifically including regions thought to have the highest ground-level concentrations and the largest burdens of disease attributable to air pollution. These regions include large parts of Asia, Africa, and the Middle East (HEI International Scientific Oversight Committee 2010; Brauer et al. 2012).

Satellite observations of aerosol optical depth (AOD) — a measure of light extinction by aerosols in the total atmospheric column calculated from measurements of light scattering at various wavelengths — are of great interest for estimating ground-level concentrations of $PM_{2.5}$. The AOD data, such as those provided by the National Aeronautics and Space Administration (NASA) from two satellites, indicate how aerosols modify the radiation exiting the top

Dr. Paciorek's 3-year study, "Integrating Monitoring and Satellite Data to Retrospectively Estimate Monthly $PM_{2.5}$ Concentrations in the Eastern United States," began in August 2006. Total expenditures were $297,239. The draft Investigators' Report from Drs. Paciorek and Liu was received for review in December 2009. A revised report, received in August 2010, was accepted for publication in October 2010. During the review process, the HEI Health Review Committee and the investigators had the opportunity to exchange comments and to clarify issues in both the Investigators' Report and the Review Committee's Commentary.

* A list of abbreviations and other terms appears at the end of the Investigators' Report.

of the atmosphere after being scattered by the Earth's atmosphere and surface (Hoff and Christopher 2009).

Remote sensing of ground-level air pollution concentrations using satellite-based measurements is increasingly viewed by atmospheric scientists and health effects researchers as a valuable source of information on human exposure, with potential applications in epidemiologic research and risk assessment (Liu et al. 2005; Martin 2008). However, AOD estimates of ground-level $PM_{2.5}$ are derived from measurements of the total atmospheric column, and may need to be combined with other data, such as data from chemical-transport models, to provide accurate estimates of ground-level concentrations. (Chemical-transport models also provide local and temporal estimates of aerosol composition, which also affects light scattering.) AOD estimates are affected by factors such as the reflectance of the Earth's surface (e.g., from snow- and water-covered surfaces) and missing observations due to heavy cloud cover (Hoff and Christopher 2009). In addition, although the spatial resolution of satellite-based AOD estimates is improving, the most widely available data are at a resolution of 0.1×0.1 degrees (~10 km $\times$ 10 km at the equator). These factors may contribute to measurement error in exposure estimates, which could lead to either overestimates or underestimates of the effects of air pollution on adverse health outcomes (Szpiro et al. 2011).

STUDY SUMMARY

STUDY OBJECTIVES

The general objective of this study was to assess the ability of approaches that use satellite-based measurements of AOD to fill spatial and temporal gaps in existing monitoring networks in the eastern United States by providing estimates of spatial patterns in ambient $PM_{2.5}$ concentrations at monthly and longer time scales.

Specifically, the investigators proposed to do the following:

1. Develop Bayesian statistical models for integrating monitoring, satellite, and geographic information system (GIS) data to estimate monthly ambient $PM_{2.5}$ concentrations at high spatial resolution.
2. Estimate monthly $PM_{2.5}$ concentrations across the eastern United States for the period 2000 to 2006 at a fine spatial resolution (10 km $\times$ 10 km or finer) by combining $PM_{2.5}$ monitoring data with satellite measurements of AOD from the moderate resolution imaging spectroradiometer (MODIS) and multiangle imaging spectroradiometer (MISR) satellites.
3. Develop an understanding of temporal and spatial heterogeneity in $PM_{2.5}$ and the ability to characterize it based on satellite, monitoring, and GIS data.
4. Use a measurement-error framework to quantify the reduction in exposure uncertainty and the potential reduction in uncertainty in health effects estimates arising from using both satellite and monitoring data relative to the uncertainty arising from using monitoring data only.
5. Assess the promise of two additional sources of information, the Geostationary Operational Environmental Satellite (GOES) and the Community Multiscale Air Quality (CMAQ) atmospheric chemistry model (for vertical-profile information), to improve exposure estimates for 1995 through 1999, years for which satellite measurements are unavailable.
6. Develop a general statistical framework to account for any systematic discrepancies between a proxy measure, such as AOD, and $PM_{2.5}$ measurements.
7. Analyze reflectance data from the GOES satellite and develop methods to screen and calibrate reflectance as a proxy measure for $PM_{2.5}$.
8. Assess the degree to which CMAQ vertical-profile information can be used to improve calibration of AOD to $PM_{2.5}$.

METHODS

Sources and Compilation of Data

The analyses presented in this report utilized data from the eastern United States obtained from the following:

- U.S. Environmental Protection Agency (EPA) Interagency Monitoring of Protected Visual Environments $PM_{2.5}$ monitoring network for ground-level estimates of $PM_{2.5}$ (2000–2006);
- U.S. National Oceanic and Atmospheric Administration's (NOAA's) North American Regional Reanalysis database for meteorologic factors, including temperature and barometric pressure (2000–2006);
- U.S. EPA National Emissions Inventory of $PM_{2.5}$ emissions (2002–2006);
- U.S. EPA CMAQ model for emissions-based estimates of $PM_{2.5}$ (2001);
- NASA's MODIS and MISR satellites for AOD measurements (2000–2006);
- NOAA's GOES satellite for AOD estimates and reflectance measurements (2001–2007); and

- Multiple sources for GIS-derived variables, including population density (Harvard Geospatial Library), elevation (U.S. Geological Survey), road density and distance to major roads (ESRI StreetMap 9.2), and land use and land cover (Multi-Resolution Land Characteristics Consortium).

Detailed descriptions of each data source and the procedures used to compile the data from them at the appropriate spatial and temporal scales for analysis are provided in Appendix A of the Investigators' Report.

Statistical Analyses

The overall objective of the statistical analysis was to analyze the relationship between $PM_{2.5}$ and AOD in both space and time. In initial analyses, the investigators estimated the correlation between $PM_{2.5}$ and AOD both before and after calibration with meteorologic factors, and after performing large-scale spatial and temporal calibrations, to account for discrepancies between AOD and $PM_{2.5}$.

In more complex analyses, the investigators used both raw and calibrated AOD data as variables in statistical models in order to predict $PM_{2.5}$ concentrations in two ways: as a separate data source contributing a second likelihood to a Bayesian statistical model, and as a data source on which to regress monitor-based $PM_{2.5}$ measurements. In additional analyses, they modeled the discrepancies between proxies, such as AOD, and $PM_{2.5}$ and explored the scales of the spatial relationship between the proxy and $PM_{2.5}$ using methods they had developed for this purpose.

KEY FINDINGS AND CONCLUSIONS

The investigators report that satellite-based AOD estimates did not improve predictions of $PM_{2.5}$ concentrations for the eastern United States as compared with predictions resulting from the use of other geospatial models. Although AOD was temporally correlated with $PM_{2.5}$, correlations of long-term spatial averages were relatively weak unless they were adjusted statistically for the discrepancy between AOD and $PM_{2.5}$ (see sections 3, 4, and 5 of the Investigators' Report; later references to sections are also to the Investigators' Report).

The investigators split the $PM_{2.5}$ data, using part of the data to develop multivariable predictive models and using the remaining data to test the predictive ability of the models they developed. Although multivariable models that combined AOD, $PM_{2.5}$ observations, and land-use and meteorologic variables were highly predictive of $PM_{2.5}$ concentrations, AOD contributed little to the predictive power of those models over and above the other variables (see sections 5 and 6).

Further, the investigators report that substituting $PM_{2.5}$ estimates from the U.S. EPA's CMAQ model for AOD estimates also did not improve the ability of the multivariable models to predict the measured $PM_{2.5}$ data. This they attributed to CMAQ's smoothing of large-scale spatial patterns in $PM_{2.5}$ monitor values (see section 6).

Using statistical models that accounted for potential discrepancies between AOD and $PM_{2.5}$ at both large and small spatial scales was an important determinant of predictive ability. Models that did not account for discrepancies at small spatial scales had poor predictive ability for measured $PM_{2.5}$, a result the investigators attribute to the fact that their analysis did not account for spatial variation in the vertical profile of the aerosol (see section 4).

The investigators' results suggest that, to some extent, correlations between raw AOD estimates and measured $PM_{2.5}$ concentrations may reflect a correlation with land-surface brightness (as a function of land use) rather than with actual aerosol levels, because the AOD algorithms estimate poorly the background surface reflectance (see sections 4 and 7). Their attempts to use data from the GOES satellite to correct for surface reflectance were largely unsuccessful (see section 7).

The investigators conclude the following:

- The inability of AOD to improve the spatial prediction of monthly and yearly average $PM_{2.5}$ concentrations in the eastern United States is a result of the spatial discrepancy between AOD and measured $PM_{2.5}$, particularly at smaller spatial scales. They emphasize the importance of accounting explicitly for such discrepancies in statistical models that use proxy estimates such as AOD. The objective should be to distinguish the "signal" in the proxy measure (AOD) with respect to the process of interest ($PM_{2.5}$) from the "noise" contributed by the discrepancy between the proxy and the process of interest. They note that achieving this objective will be particularly challenging when measured data are sparse.
- There is little evidence that current satellite-based AOD estimates can improve the prediction of ground-level $PM_{2.5}$ at small-to-moderate scales in the eastern United States. They argue that, until more evidence regarding the reliability of satellite-derived AOD data is available, it is premature to use these data in epidemiologic studies as a proxy for $PM_{2.5}$.
- Future research on the application of AOD estimates in epidemiologic research on the chronic effects of long-term $PM_{2.5}$ exposure should attempt to distinguish

temporal and spatial correlations, so that the spatial scale of the correlation and the discrepancies between AOD and $PM_{2.5}$, may be better understood.

- More promising results might be obtained in areas that have higher levels of ambient $PM_{2.5}$ than the eastern United States, because the difference between the AOD signal in such locales and the "noise" contributed by background surface reflectance would be greater. They note that in developing countries that have high levels of ambient $PM_{2.5}$, but may lack surface measurements and land-use data, satellite-based estimates may add more incremental information.

THE HEI HEALTH REVIEW COMMITTEE'S EVALUATION AND INTERPRETATION OF THE RESULTS

This report presents a careful analysis of specific applications of satellite-based AOD to estimate ground-level concentrations of $PM_{2.5}$. It is an outstanding example of the type of detailed work that is required, and should be continued, to evaluate the performance of satellite-based estimates for application in epidemiologic research and risk assessment. The statistical approaches used to evaluate the performance of the AOD estimator were well conceived and well executed from both air quality science and epidemiologic perspectives, and the report reflects a deep understanding of the operation of the quantitative models that were applied, as well as the impacts of simplifying assumptions.

The overall objective of the study was to evaluate the potential of remotely sensed AOD estimates to improve estimates of $PM_{2.5}$ in epidemiologic studies, particularly those relying on spatial variation in pollution and on longer time scales. However, the investigators found little evidence that the AOD estimator, as conceived and applied in this research, improved the estimates of exposure to $PM_{2.5}$. They reached this conclusion after carefully considering alternative approaches, and they were able to provide some insights regarding the reasons for their admittedly disappointing results.

Having observed that part of the correlation between AOD and surface $PM_{2.5}$ may be due to a correlation between land use and AOD (attributable to spatial variation in surface brightness), rather than to a direct relationship between AOD and $PM_{2.5}$, the investigators argue that their results do not depend on the specific instrumentation used in the study, which comprised a variety of sensors that have been widely used by others. This was a strength of their study. However, although Paciorek and Liu demonstrate convincingly that the use of raw AOD estimates as a proxy for ground-level concentrations of $PM_{2.5}$ air pollution is unwarranted and potentially misleading, they suggest that alternative applications of AOD under different conditions may have merit. For example, other researchers have combined data from different sensors (MODIS and MISR) and have developed approaches in which AOD or other satellite estimates are linked with chemical-transport models. These approaches (Liu et al. 2004; van Donkelaar et al. 2006, 2010) provide spatially (and temporally) resolved information on atmospheric vertical profiles and aerosol composition, and generally show stronger relationships with surface $PM_{2.5}$, in the United States and elsewhere, than were reported by Paciorek and Liu. The authors acknowledge this: The analysis in this report speaks primarily to the shortcomings of using raw AOD estimates, which may not apply to the general applicability of satellite-based approaches, where ongoing efforts are underway to improve the technology.

The focus of the study on a single geographic area, the eastern United States, is another limitation of the study that the investigators acknowledge. From a continental perspective, their decision to focus on the eastern United States was entirely reasonable, given that the satellite-based AOD–$PM_{2.5}$ relationship is more poorly correlated in the western United States, the population density and air pollution levels are higher in the eastern United States, and a relatively rich database of ground-level measurements in the eastern United States afforded the opportunity for comparing AOD estimates and $PM_{2.5}$ measurements. However, the focus on the eastern United States may have produced a too-limited picture of the potential utility of satellite-based estimates in other settings. The eastern United States is characterized by constrained spatial variability and generally lower levels of $PM_{2.5}$ from a global perspective, and it affords an extensive surface-monitoring network and a rich database of other predictors such as those related to land use. In a setting such as this, it is perhaps not surprising that satellite-based estimates added little information. That said, it is not entirely clear in retrospect that the choice of location was a major limitation. Other investigators have reported a close relationship of AOD-based estimates of $PM_{2.5}$ with monitored levels across the United States and in the eastern United States in particular (van Donkelaar et al. 2006; Kloog et al. 2011; Lee et al. 2011). The relationship with monitored levels in the eastern United States was improved when a daily AOD–$PM_{2.5}$ calibration was used and when AOD estimates were combined with land-use data (Kloog et al. 2011; Lee et al. 2011).

This report offers important insights into potential sources of inaccuracy in the AOD estimator and sounds appropriate notes of caution concerning the naïve application of this technology. However, it does not address what

might be the most important issue from a global perspective: Can satellite-based estimates be used for health effects research instead of monitoring data in locations currently without extensive $PM_{2.5}$ surface monitors, such as many parts of Asia, Africa, and the Middle East, where there is evidence of very high levels of particulate air pollution (World Health Organization 2006; HEI International Scientific Oversight Committee 2010; Brauer et al. 2012)? A systematic effort is needed to reduce the uncertainties surrounding the global use of satellite-based estimates, and the methods employed in this report should be useful in future applications of this kind.

SUMMARY AND CONCLUSIONS

Looking to the future, several conclusions seem warranted regarding the application of satellite-based estimates of $PM_{2.5}$ air pollution in health effects research.

- The use of raw AOD estimates as a proxy for measured $PM_{2.5}$ in health effects research should be avoided in general. Rather, approaches that combine information from multiple sources — remote sensing, model-based estimates, and ground-level measurements — may offer the most promise (Brauer et al. 2012). For example, van Donkelaar and colleagues (2010) derive a scaling factor from the GOES-Chem model that is then used as input in a land-use regression model. This approach is similar to that of Yanosky and colleagues (2009) and is now being applied in cohort studies in the United States and Canada (e.g., Crouse et al. 2012). Recent results suggest that it explains a large proportion of between-city variation in $PM_{2.5}$ but a much smaller proportion of within-city variation (Hystad et al. 2011). This difference is perhaps to be expected — $PM_{2.5}$ displays little within-city variation. For NO_2, which exhibits considerable within-city spatial variation, two recent studies show that models that use satellite-based estimates alone explain only a small proportion of the variation in NO_2 relative to the amount that can be explained when land-use and other data are also included in the models (Hystad et al. 2011; Novotny et al. 2011). It would appear that satellite-based estimates can explain a fair amount of between-city variation in $PM_{2.5}$ (less so for NO_2), but that they need to be combined with other data to explain within-city variation.
- Applications in health effects research should include evaluations of the relationship between satellite-based estimates and monitoring data, and should quantify, to the extent possible, the contribution of satellite-based estimates to total exposure measurement error in epidemiologic effect estimates. Some recent epidemiologic studies have made efforts to do so (see, for example, Anderson et al. 2012; Crouse et al. 2012).
- Satellite-based estimates can play — and have played — an important role in the evaluation of exposure to, and health effects of, short-term episodes of high levels of air pollution from burning vegetation or dust events in areas with limited or no monitoring. In such settings, satellite-based estimates can help define the spatial extent of the exposure — something that cannot be done with monitoring data if the network is sparse — and could be used to quantify ground-level concentrations. Recent papers illustrate the potential and the considerable challenges of such applications (Wu et al. 2006; Henderson et al. 2011; van Donkelaar et al. 2011).
- Global ground-based measurements of long-term exposure to particulate air pollution are likely to be insufficient to address the needs of epidemiologic research and public health–based risk assessment in global locales where such measurements are not likely to be collected for the foreseeable future. Satellite remote sensing may offer promise for providing information on exposure to $PM_{2.5}$ at regional-to-global scales, especially in places with the highest levels of pollution and the greatest estimated burden of disease attributable to it. However, as this report makes clear, there are limitations to, and outstanding questions about, the accuracy and precision with which ground-level aerosol mass concentrations can be inferred from satellite remote sensing. A key source of uncertainty is the level of global variation that exists in the relationship of annual average $PM_{2.5}$ with columnar AOD at specific satellite overpass times during cloud-free conditions. These issues need to be addressed if the promise of satellite-based technology is to be realized. To do so, a systematic effort including ground-level measurements of $PM_{2.5}$ in selected global regions will be necessary to assess the factors that most affect the relationship between satellite-based and ground-level estimates.

ACKNOWLEDGMENTS

The Health Review Committee thanks the ad hoc reviewers for their help in evaluating the scientific merit of the Investigators' Report. The Committee is also grateful to Dr. Annemoon van Erp for her oversight of the study, to Dr. Aaron Cohen for his assistance in preparing its Commentary, to Leah Shriro for science editing this Report and Commentary, and to Suzanne Gabriel, Barbara Gale, Hope Green, Virgi Hepner, Fred Howe, Flannery McDermott, and Ruth Shaw for their roles in preparing this Research Report for publication.

REFERENCES

Anderson HR, Butland B, van Donkelaar A, Brauer M, Strachan DP, Clayton T, van Dingenen R, Amann M, Brunekreef B, Cohen A, Dentener F, Lai C, Lamsal LN, Martin R, ISAAC Phase One and Phase Three Study Groups. 2012. Satellite-based estimates of ambient air pollution and global variations in childhood asthma prevalence. Environ Health Perspect. Available from doi:10.1289/ehp.1104724.

Brauer M, Amann M, Burnett RT, Cohen A, Dentener F, Ezzati M, Henderson SB, Krzyzanowski M, Martin RV, Van Dingenen R, van Donkelaar A, Thurston GD. 2012. Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. Environ Sci Technol 46:652–660.

Crouse DL, Peters PA, Donkelaar A, Goldberg MS, Villeneuve PJ, Brion O, Khan S, Atari DO, Jerrett M, Pope CA III, Brauer M, Brook JR, Martin RV, Stieb D, Burnett RT. 2012. Risk of non-accidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A Canadian national-level cohort study. Environ Health Perspect. Available from *http://dx.doi.org/10.1289/ehp.1104049.*

HEI International Scientific Oversight Committee. 2010. Outdoor Air Pollution and Health in the Developing Countries of Asia: A Comprehensive Review. Special Report 18. Health Effects Institute, Boston, MA.

Henderson SB, Brauer M, MacNab YC, Kennedy SM. 2011. Three measures of forest fire smoke exposure and their associations with respiratory and cardiovascular health outcomes in a population-based cohort. Environ Health Perspect 119:1266–1271.

Hoff RM, Christopher SA. 2009. Remote sensing of particulate pollution from space: Have we reached the Promised Land? J Air Waste Manag Assoc 59:645–675.

Hystad P, Setton E, Cervantes A, Poplawski K, Deschenes S, Brauer M, van Donkelaar A, Lamsal L, Martin R, Jerrett M, Demers P. 2011. Creating national air pollution models for population exposure assessment in Canada. Environ Health Perspect 119:1123–1129.

Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, Morrison J, Giovis C. 2005. A review and evaluation of intraurban air pollution exposure models. J Expos Anal Environ Epidemiol 15:185–204.

Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J. 2011. Assessing temporally and spatially resolved $PM_{2.5}$ exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmos Environ 45:6267–6275.

Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope CA III, Thurston G, Calle EE, Thun MJ. 2009. Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality. Research Report 140. Health Effects Institute, Boston, MA.

Lee HJ, Liu Y, Coull BA, Schwartz J, Koutrakis P. 2011. A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. Atmos Chem Phys 11:7991–8002.

Liu Y, Park RJ, Jacob DJ, Li Q, Kilaru V, Sarnat JA. 2004. Mapping surface concentrations of fine particulate matter using MISR satellite observations of aerosol optical thickness. J Geophys Res 109:1–10.

Liu Y, Sarnat JA, Kilaru V, Jacob DJ, Koutrakis P. 2005. Estimating ground-level $PM_{2.5}$ in the eastern United States using satellite remote sensing. Environ Sci Technol 39:3269–3278.

Martin RV. 2008. Satellite remote sensing of surface air quality. Atmos Environ 42:7823–7843.

Novotny EV, Bechle M, Millet DB, Marshall JD. 2011. National satellite-based land-use regression: NO_2 in the United States. Environ Sci Technol 45:4407–4414.

Szpiro AA, Paciorek CJ, Sheppard L. 2011. Does more accurate exposure prediction necessarily improve health effect estimates? Epidemiology 22:1–6.

van Donkelaar A, Martin RV, Levy RC, da Silva AM, Krzyzanowski M, Chubarova NE, Semutnikova E, Cohen AJ. 2011. Satellite-based estimates of ground-level fine particulate matter during extreme events: A case study of the Moscow fires in 2010. Atmos Environ 45:6225–6232.

van Donkelaar A, Martin RV, Brauer M, Kahn R, Levy R, Verduzco C, Villeneuve PJ. 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. Environ Health Perspect 118:847–855. Available from *http://ehp03.niehs.nih.gov/article/info:doi/10.1289/ehp.0901623.*

van Donkelaar A, Martin RV, Park RJ. 2006. Estimating ground-level $PM_{2.5}$ using aerosol optical depth determined from satellite remote sensing. J Geophys Res 111:1–10. Available from *http://fizz.phys.dal.ca/~atmos/publications/vanDonkelaar_2006_JGR.pdf.*

World Health Organization. 2006. Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide. WHO, Copenhagen, Denmark.

Wu J, Winer A, Delfino R. 2006. Exposure assessment of particulate matter pollution before, during, and after the 2003 Southern California wildfires. Atmos Environ 40:3333–3348.

Yanosky JD, Paciorek CJ, Suh HH. 2009. Predicting chronic fine and coarse particulate exposures using spatio-temporal models for the northeastern and midwestern U.S. Environ Health Perspect 117:522–529.

RELATED HEI PUBLICATIONS: ASSESSING EXPOSURE TO AIR POLLUTION

Number	Title	Principal Investigator	Date*
Research Reports			
160	Personal and Ambient Exposures to Air Toxics in Camden, New Jersey	P.J. Lioy	2011
158	Air Toxics Exposure from Vehicle Emissions at a U.S. Border Crossing: Buffalo Peace Bridge Study	J.D. Spengler	2011
152	Evaluating Heterogeneity in Indoor and Outdoor Air Pollution Using Land-Use Regression and Constrained Factor Analysis	J.I. Levy	2010
140	Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality	D. Krewski	2009
139	Effects of Long-Term Exposure to Traffic-Related Air Pollution on Respiratory and Cardiovascular Mortality in the Netherlands: The NLCS-AIR Study	B. Brunekreef	2009
130	Relationships of Indoor, Outdoor, and Personal Air (RIOPA)		
	Part I. Collection Methods and Descriptive Analyses	C.P. Weisel	2005
	Part II. Analyses of Concentrations of Particulate Matter Species	B.J. Turpin	2007
HEI Communications			
10	Improving Estimates of Diesel and Other Emissions for Epidemiologic Studies		2003
HEI Special Reports			
17	Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects	HEI Panel on the Health Effects of Traffic-Related Air Pollution	2010
	Research Directions to Improve Estimates of Human Exposure and Risk from Diesel Exhaust	HEI Diesel Epidemiology Working Group	2002
	Part I. Report of the Diesel Epidemiology Working Group		
	Part II. Investigators' Reports		
	Cancer Risk from Diesel Emissions Exposure in Central and Eastern Europe: A Feasibility Study	P. Boffetta	
	Cancer Risk from Diesel Exhaust Exposure in the Canadian Railroad Industry: A Feasibility Study	M.M. Finkelstein	
	Quantitative Assessment of Lung Cancer Risk from Diesel Exhaust Exposure in the US Trucking Industry: A Feasibility Study	E. Garshick	
	Measurement of Diesel Aerosol Exposure: A Feasibility Study	D.B. Kittelson	
	Measuring Diesel Emissions Exposure in Underground Mines: A Feasibility Study	B. Zielinska	

** Reports published since 2000.*
Copies of these reports can be obtained from the Health Effects Institute and many are available at pubs.healtheffects.org.

HEI BOARD, COMMITTEES, and STAFF

Board of Directors

Richard F. Celeste, Chair *President Emeritus, Colorado College*

Sherwood Boehlert *Of Counsel, Accord Group; Former Chair, U.S. House of Representatives Science Committee*

Enriqueta Bond *President Emeritus, Burroughs Wellcome Fund*

Purnell W. Choppin *President Emeritus, Howard Hughes Medical Institute*

Michael T. Clegg *Professor of Biological Sciences, University of California–Irvine*

Jared L. Cohon *President, Carnegie Mellon University*

Stephen Corman *President, Corman Enterprises*

Gowher Rizvi *Vice Provost of International Programs, University of Virginia*

Linda Rosenstock *Dean, School of Public Health, University of California–Los Angeles*

Henry Schacht *Managing Director, Warburg Pincus; Former Chairman and Chief Executive Officer, Lucent Technologies*

Warren M. Washington *Senior Scientist, National Center for Atmospheric Research; Former Chair, National Science Board*

Archibald Cox, Founding Chair *1980–2001*

Donald Kennedy, Vice Chair Emeritus *Editor-in-Chief Emeritus,* Science*; President Emeritus and Bing Professor of Biological Sciences, Stanford University*

Health Research Committee

David L. Eaton, Chair *Associate Vice Provost for Research and Director, Center for Ecogenetics and Environmental Health, School of Public Health, University of Washington–Seattle*

David T. Allen *Gertz Regents Professor in Chemical Engineering; Director, Center for Energy and Environmental Resources, University of Texas–Austin*

David Christiani *Elkan Blout Professor of Environmental Genetics, Harvard School of Public Health*

David E. Foster *Phil and Jean Myers Professor, Department of Mechanical Engineering, Engine Research Center, University of Wisconsin–Madison*

Uwe Heinrich *Professor, Medical School Hannover, Executive Director, Fraunhofer Institute for Toxicology and Experimental Medicine, Hanover, Germany*

Grace LeMasters *Professor of Epidemiology and Environmental Health, University of Cincinnati College of Medicine*

Sylvia Richardson *Professor of Biostatistics, Department of Epidemiology and Public Health, Imperial College School of Medicine, London, United Kingdom*

Richard L. Smith *Director, Statistical and Applied Mathematical Sciences Institute, University of North Carolina–Chapel Hill*

James A. Swenberg *Kenan Distinguished Professor of Environmental Sciences, Department of Environmental Sciences and Engineering, University of North Carolina–Chapel Hill*

HEI BOARD, COMMITTEES, and STAFF

Health Review Committee

Homer A. Boushey, Chair *Professor of Medicine, Department of Medicine, University of California–San Francisco*

Ben Armstrong *Reader in Epidemiological Statistics, Public and Environmental Health Research Unit, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, United Kingdom*

Michael Brauer *Professor, School of Environmental Health, University of British Columbia, Canada*

Bert Brunekreef *Professor of Environmental Epidemiology, Institute of Risk Assessment Sciences, University of Utrecht, the Netherlands*

Mark W. Frampton *Professor of Medicine and Environmental Medicine, University of Rochester Medical Center*

Stephanie London *Senior Investigator, Epidemiology Branch, National Institute of Environmental Health Sciences*

Armistead Russell *Georgia Power Distinguished Professor of Environmental Engineering, School of Civil and Environmental Engineering, Georgia Institute of Technology*

Lianne Sheppard *Professor of Biostatistics, School of Public Health, University of Washington–Seattle*

Officers and Staff

Daniel S. Greenbaum *President*
Robert M. O'Keefe *Vice President*
Rashid Shaikh *Director of Science*
Barbara Gale *Director of Publications*
Jacqueline C. Rutledge *Director of Finance and Administration*
Helen I. Dooley *Corporate Secretary*

Kate Adams *Senior Scientist*
Aaron J. Cohen *Principal Scientist*
Maria G. Costantini *Principal Scientist*
Philip J. DeMarco *Compliance Manager*
Suzanne Gabriel *Editorial Assistant*
Hope Green *Editorial Assistant (part time)*
L. Virgi Hepner *Senior Science Editor*
Anny Luu *Administrative Assistant*
Francine Marmenout *Senior Executive Assistant*
Nicholas Moustakas *Policy Associate*
Hilary Selby Polk *Senior Science Editor*
Sarah Rakow *Science Administrative Assistant*
Evan Rosenberg *Staff Accountant*
Robert A. Shavers *Operations Manager*
Geoffrey H. Sunshine *Senior Scientist*
Annemoon M.M. van Erp *Senior Scientist*
Katherine Walker *Senior Scientist*
Morgan Younkin *Research Assistant*

HEALTH
EFFECTS
INSTITUTE

101 Federal Street, Suite 500
Boston, MA 02110, USA
+1-617-488-2300
www.healtheffects.org

RESEARCH
REPORT

Number 167
May 2012