Methods for Characterizing Fine Particulate Matter Using Satellite Remote-Sensing Data and Ground Observations: Potential Use for Environmental Public Health Surveillance

Mohammad Z. Al-Hamdan [a], William L. Crosson [a], Ashutosh S. Limaye [a], Douglas L. Rickman [b], Dale A. Quattrochi [b], Maurice G. Estes, Jr. [a], Judith R. Qualters [c], Amanda S. Niskar [d], Amber H. Sinclair [e], Dennis D. Tolsma [e], Kafayat A. Adeniyi [f]

[a] Universities Space Research Association at NASA Marshall Space Flight Center, National Space Science and Technology Center, NASA Global Hydrology and Climate Center, Huntsville, AL

[b] Earth Science Office at NASA Marshall Space Flight Center, National Space Science and Technology Center, NASA Global Hydrology and Climate Center, Huntsville, AL

[c] Centers for Disease Control and Prevention (CDC), Atlanta, GA

[d] Israel Center for Disease Control, Tel Hashomer, Israel

[e] Kaiser Permanente, Atlanta, GA

[f] Quantell, Inc., Atlanta, GA

1

ABSTRACT

This study describes and demonstrates different techniques for surfacing daily environmental hazards data of particulate matter with aerodynamic diameter less than or equal to 2.5 micrometers ($PM_{2.5}$) for the purpose of integrating respiratory health and environmental data for the Centers for Disease Control and Prevention (CDC's) pilot study of Health and Environment Linked for Information Exchange (HELIX)-Atlanta. It described a methodology for estimating ground-level continuous $PM_{2.5}$ concentrations using B-Spline and inverse distance weighting (IDW) surfacing techniques and leveraging National Aeronautics and Space Administration (NASA) Moderate Resolution Imaging Spectrometer (MODIS) data to complement The Environmental Protection Agency (EPA) ground observation data. The study used measurements of ambient $PM_{2.5}$ from the EPA database for the year 2003 as well as $PM_{2.5}$ estimates derived from NASA's satellite data. Hazard data have been processed to derive the surrogate exposure $PM_{2.5}$ estimates. The paper has shown that merging MODIS remote sensing data with surface observations of $PM_{2.5}$ not only provides a more complete daily representation of $PM_{2.5}$ than either data set alone would allow, but it also reduces the errors in the $PM_{2.5}$ estimated surfaces. The results of this paper have shown that the daily IDW $PM_{2.5}$ surfaces had smaller errors, with respect to observations, than those of the B-Spline surfaces in the year studied. However the IDW mean annual composite surface had more numerical artifacts, which could be due to the interpolating nature of the IDW that assumes that the maxima and minima can occur only at the observation points. Finally, the methods discussed in this paper improve temporal and spatial resolutions and establish a foundation for environmental public health linkage and association studies for which determining the concentrations of an environmental hazard such as $PM_{2.5}$ with good accuracy levels is critical.

IMPLICATIONS

The described method of estimating fine particulate matter whose aerodynamic diameter is less than or equal to 2.5 micrometers ($PM_{2.5}$) concentrations by merging Moderate Resolution Imaging Spectrometer (MODIS) remote sensing data with surface observations of $PM_{2.5}$ not only provides a more complete daily representation of $PM_{2.5}$ than either data set alone would allow,

but it also reduces the errors in the $PM_{2.5}$ estimated surfaces with respect to observations. This information would facilitate more effective research into the association of selected health events with environmental air quality levels. These new data products have the potential to serve as a tool for environmental public health surveillance to monitor trends and serve as an early warning system for prevention of human exposure to potential hazards.

INTRODUCTION

A major challenge in studying the relationship between air quality and human health outcomes such as asthma is characterization of population-level or individual-level exposures. Human exposure measurements are typically unavailable and are estimated using a variety of techniques that rely on environmental measures available from the existing ambient air-monitoring network. While monitoring data provide the best characterization of pollutant concentrations levels at a particular place and time, temporal and spatial gaps in this data can limit their applicability for exposure assessment in health studies. Available fixed-site air quality monitoring stations tend to be located strategically in areas where high levels of pollutants are expected and/or where there is high population density (Watkins and Boothe; Bell, 2006). The purpose of these monitors is to provide data to measure regulatory compliance, not personal exposure information. Thus, many epidemiology studies examining the association between particulate matter and asthma have had to rely on measurements from stationary ambient monitoring sites located substantial distances from where many individuals actually lived or worked (Liu et al, 2004; Ito et al., 2001) to develop surrogates from human exposures. Moreover, the frequency of monitoring for particular pollutants varies from hourly to one every several days.

Researchers have used a number of modeling techniques to address issues in estimating exposure concentrations. (Jerret et al, 2005; Bell, 2006; Wong et al, 2003). These include proximity to air monitor models, statistical interpolation, land use regression, dispersion models, integrated emission-meteorological models, and hybrid models. A comparative analysis by Jerrett et al. (2005) outlines the strengths and limitations of each. For example, proximity models can provide a straightforward and cost-effective approach for characterizing air pollution exposure. However, they are best used for exploratory analyses since they are more likely to misclassify exposure due to lack of consideration of covariates that confound the relationship

between air pollution and health. Jerret et al (2005) reported that the best way to measure concentrations of air pollutants that an individual may be exposed to is to use personal air monitors; however, this method is expensive and its use in large population-based studies or ongoing public health tracking is cost prohibitive.

One promising method for characterizing $PM_{2.5}$ exposure for public health practice and epidemiologic research is integration of remote sensing satellite systems data with air monitoring network data (Engle-Cox et al., 2004). Remote sensing data have been used to detect and track particulate matter plumes from major events such as dust storms, volcanic emissions, and fires (EPA, 2002). However, the aerosol optical properties retrieved by space-borne sensors may also be useful in filling the temporal and spatial gaps found with monitoring ground level data. Satellite data cover large geographic areas at moderate spatial resolution for multiple years and with reliable repeated measurements (Liu et al, 2004). National Aeronautics and Space Administration (NASA)'s MODIS satellite provides a measure of Aerosol Optical Depth (AOD) - the measure of the degree to which sunlight is scattered and absorbed by aerosols of various sizes throughout the entire atmospheric column. The MODIS AOD product is available for any area up to two times each day and can be used to estimate the amount of aerosols present in the atmosphere. Research has shown that AOD is indirectly related to ground level $PM_{2.5}$, with the correlation between the two being strongest on days with low cloud cover, low relative humidity, and good vertical mixing within the atmospheric column. (Gupta and Christopher, 2006; Gupta et al., 2005; Rush et al, 2004; Engel-Cox et al, 2004; Wang and Christopher, 2003; Chu et al., 2003). In addition to developing a model for estimating $PM_{2.5}$ from MODIS AOD, this paper develops methods and algorithms for MODIS estimated $PM_{2.5}$ bias adjustment, Air Quality System (AQS) $PM_{2.5}$ quality control, as well as merging AQS and MODIS estimated $PM_{2.5}$ to generate continuous $PM_{2.5}$ spatial surfaces.

The use of remote sensing data with ground level monitoring has not had broad public health application beyond those relating to infectious disease (Morain et al. 2005; Patz, 2005; Beck et al., 2000). This paper explores methods for utilizing AOD to enhance $PM_{2.5}$ exposure estimation for a study of asthma exacerbations and air quality in the five-county Atlanta metropolitan area (Clayton, Cobb, DeKalb, Fulton, and Gwinnett). It describes and demonstrates different surfacing techniques for estimating daily ambient concentrations of $PM_{2.5}$ that can be linked to health outcomes data. Measurements of ambient $PM_{2.5}$ from the Environmental Protection

Agency (EPA) AQS database for the year 2003 as well as $PM_{2.5}$ estimates derived from MODIS AOD data are used. This project is part of the Centers for Disease Control and Prevention (CDC) Health and Environment Linked for Information Exchange (HELIX)-Atlanta project. HELIX-Atlanta's goal is to explore and pilot methodologies that could be used in the National Environmental Public Health Tracking Network, a CDC-led initiative to build a nationwide information system that integrates environmental hazard, exposure and health effect data for use in improving public health.

The objective of this paper is to describe and demonstrate different techniques for surfacing daily environmental hazards data of $PM_{2.5}$ for the purpose of integrating respiratory health and environmental data for HELIX-Atlanta. The study will use measurements of ambient $PM_{2.5}$ from the EPA AQS database for the year 2003 as well as $PM_{2.5}$ estimates derived from MODIS AOD data. Hazard data have been processed to derive the surrogate exposure $PM_{2.5}$ estimates.

SURFACING TECHNIQUES

Two spatial surfacing techniques, the inverse distance weighted (IDW) and B-Spline, were used to generate daily $PM_{2.5}$ surfaces and the results were compared in this study. In the IDW technique, observational points are weighted during interpolation such that the influence of one point relative to another declines with distance from the given point. Weights are assigned to observational points through the use of a power function, which controls how weighting factors decrease as the distance from the given point increases.

A B-Spline fits a polynomial equation to the data between a set of user-defined dividing points, termed knots. The number and position of the knots determines how well the fitted polynomial models the data. The B-Spline first fits a global estimator through all points. This model is used to estimate a coarse grid of values. The study area is then recursively subdivided into smaller sub-areas and at each iteration a B-Spline fit is made to the actual data along with the model values made by the previous iteration. The subdivision process is done in such a way that a single actual data value in a sub-area has the same significance as all of the model values in the sub-area. If there is more than one observation within a sub-area the model values have relatively little significance in that area.

IDENTIFYING DATA AND DATA SOURCES

Two sources of $PM_{2.5}$ data were identified for use in this project. The first is the EPA AQS, which measures several key air pollutants at a network of ground monitoring stations across the U.S. The AQS network provides $PM_{2.5}$ measurements from monitoring stations concentrated around metropolitan areas and a few monitors in rural areas. The AQS measurements are direct ground level concentrations and are well calibrated. In the Atlanta area, AQS $PM_{2.5}$ data are available from only five AQS sites, although the sites are well distributed across the five-county study area. AQS ground observations are made at time frequencies ranging from hourly and daily to every sixth day leaving some temporal gaps in their coverage area including the five-county of Atlanta metropolitan area. The AQS database is updated nearly every day by state and local environmental agencies that operate the monitoring station.

The second $PM_{2.5}$ data source is NASA's MODIS satellite which provides measurements of AOD. The MODIS AOD observations are at an approximate 10 km spatial resolution and are available for each day of the year for clear sky areas. Two NASA MODIS sensors are currently in orbit on the Terra and Aqua satellites, which, in sun-synchronous orbit, observe any location on the Earth's surface at about 10:30 AM and 1:30 PM local standard time, respectively, each day.

AQS DATA PROCESSING

Quality Control (QC) Procedure

AQS $PM_{2.5}$ data for five states – Georgia, Alabama, North Carolina, South Carolina and Tennessee – were obtained from the U.S. EPA for the 2003-2004 period. This region was chosen to provide a regional perspective and so that daily spatial surfaces could be created using a large number of $PM_{2.5}$ observations. A QC procedure for eliminating anomalous AQS $PM_{2.5}$ measurements was developed. The procedure utilizes observations from surrounding sites to determine whether a given measurement is acceptable or is considered anomalous and thus eliminated from further analysis. The QC procedure is based on a non-parametric (rank-order) spatial analysis. Before the observation values are used for generating spatial surfaces, a Corroborative Neighbors Statistic (CNS), predetermined based on a rank-order spatial analysis of the monitoring values, was used to filter the raw data using the following criteria: The test station was dropped out of the data set if all five closest neighbors have values larger than the CNS

times the test observation, or all five closest neighbors have values smaller than the inverse of CNS times the test observation. The value of CNS used in this procedure was 1.4, which was the 95th percentile of all CNS values.

We also compared the anomalous results identified in our CNS analyses with data flagged in the AQS for QC issues. In most cases the AQS flags indicating suspicious data confirmed anomalies identified in our QC procedure. Where an AQS datum was not flagged but our QC algorithm indicated a spurious value, this datum was excluded from analysis. EPA flags some data to indicate a specific cause of the apparently anomalous value, such as a forest fire or construction work in the vicinity of the site. In these cases, although the measurements may correctly reflect the local conditions, we made the decision to be conservative and eliminate these data if our QC algorithm identified them as anomalous. The reason for this was that the effect of these observations on the spatial surface generated with the surfacing algorithm was over a much larger spatial extent than the local phenomenon warranted, especially in regions where observations are sparse.

The results showed that out of 19403 AQS data points in the year 2003, 450 anomalous data points (1-2 data points a day on average) were identified and eliminated. Figures 1 and 2 show examples of how using the QC procedure enhanced the output $PM_{2.5}$ B-Spline and IDW surfaces by preventing any unrealistic ripples to be formed within the surface on October 9, 2003. The arrow indicates the location of the anomalous value and its effect was clearly over a much larger spatial extent than the local phenomenon warranted.
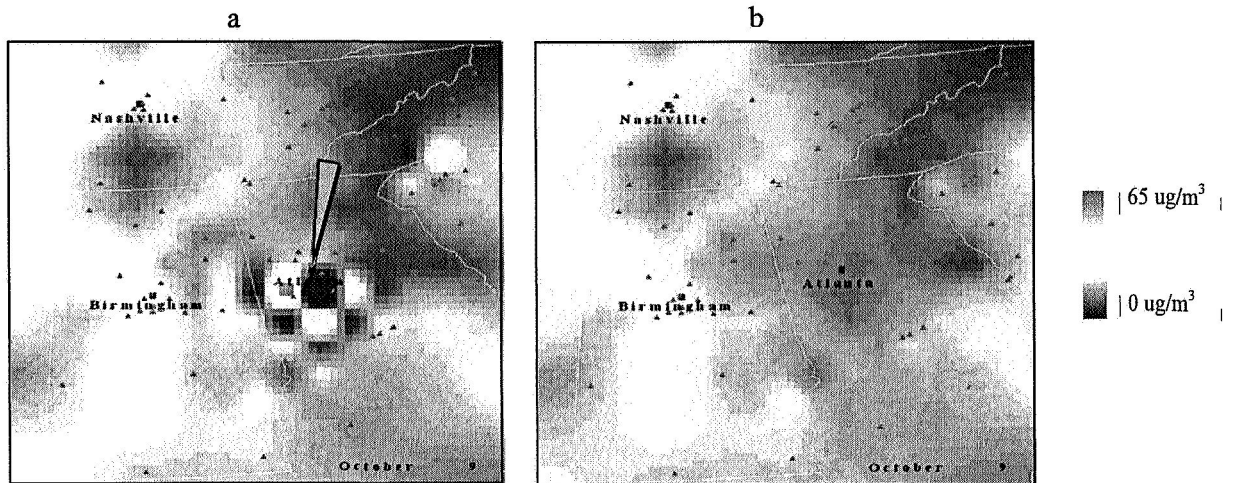
Figure 1. PM$_{2.5}$ B-Spline surfaces (a) without and (b) with quality control procedures using data from the EPA AQS network for October 9, 2003. The green arrow indicates the location of a value believed to be anomalous.
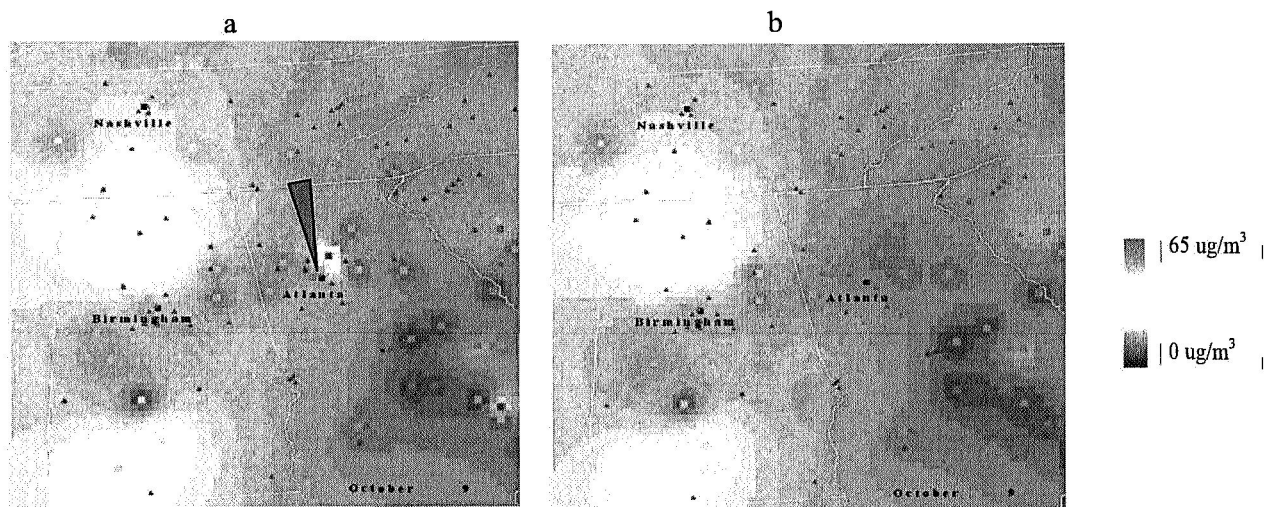

Figure 2. PM$_{2.5}$ IDW surfaces without (left) and with (right) quality control procedures using data from the EPA AQS network for October 9, 2003.

Adjustment of non-Federal Reference Method (FRM) AQS Measurements

The AQS ground measurements comprise three measurement types: FRM, 'Continuous' and 'Speciation' measurements. In the Atlanta five-county area, all but one of the AQS stations is an FRM system. FRM data are recognized as the standard but have temporal resolution of one day or longer. Furthermore, FRM data require several weeks processing time and are thus not available for near real-time analysis. Due to the different measurement types, there is the potential for non-FRM observations to be biased with respect to FRM observations, particularly in specific environmental conditions, and indeed some studies have revealed systematic differences (Gillespie, 2005; Kaldy et al., 2003; Eberly, 2002). These differences seem to be site- and season-specific. In order to utilize both types of measurements together in our algorithms, we evaluated this issue. Toward this end, we examined 28 sets of co-located FRM and non-FRM observations within the 5-state area and performed a regression-based adjustment to the non-FRM observations (Figure 3) As shown in Figure 4, the slope of this regression equation was 0.944 and the correlation coefficient was 0.96. We applied this equation to adjust all non-FRM measurements to the FRM standard.
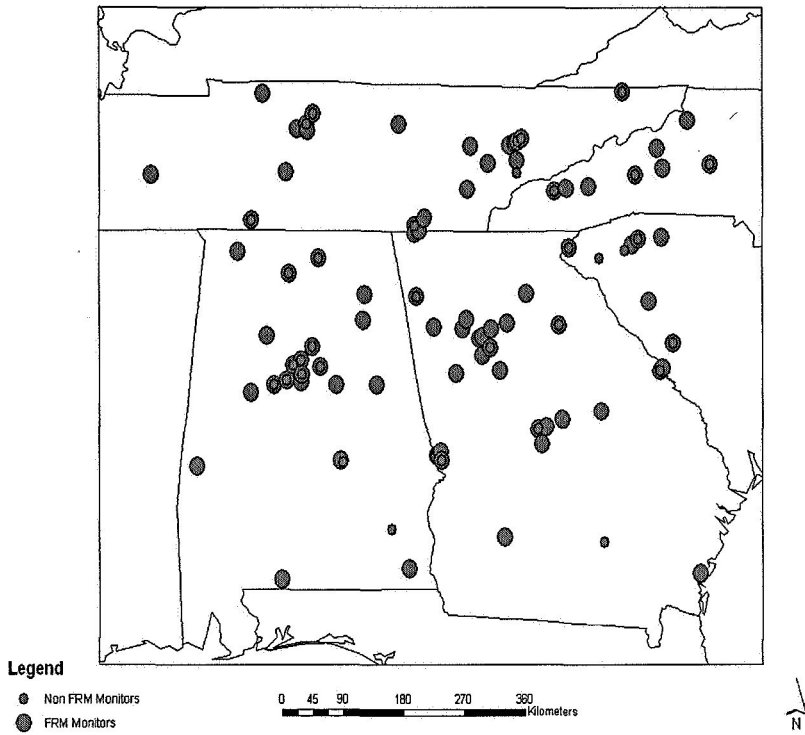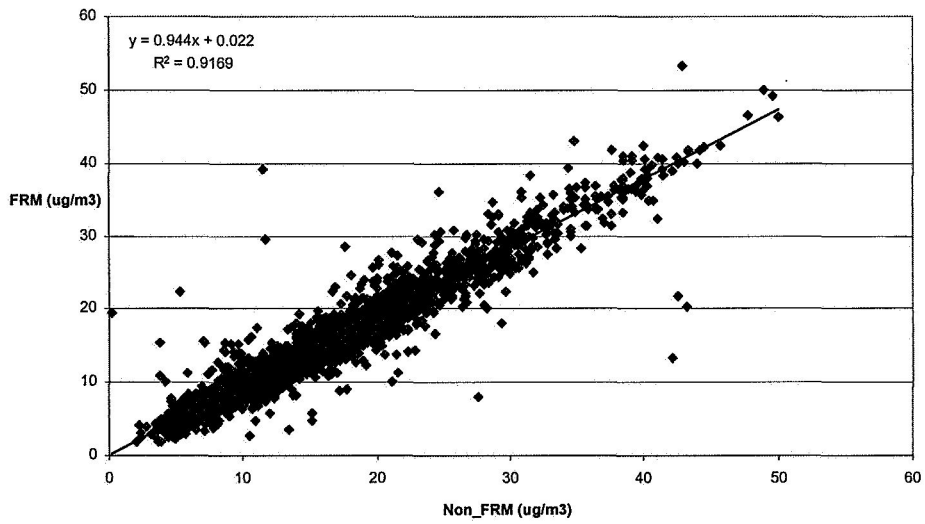
Figure 3. FRM and Non-FRM sites.

Figure 4. Linear regression model for adjusting the non-FRM AQS measurements.

MODIS DATA PROCESSING

AOD-PM$_{2.5}$ Regression Models

To estimate PM$_{2.5}$ from MODIS AOD observations, regression models were established separately for the Terra and Aqua MODIS data. First, MODIS AOD data from both Terra and Aqua satellites were obtained for the year 2003. AOD data corresponding to the locations of the AQS sites were extracted from the MODIS data files by selecting any AOD observations located within a 10 x 10 km box centered at the site location. If more than one MODIS observation fell within the box, the values were averaged to give the AOD value for the site. Linear correlation coefficients were then calculated on a monthly basis for each satellite sensor, using all of the paired daily AOD - PM$_{2.5}$ observations for the month.

Table 1 summarizes the linear correlation coefficients by month for both satellite sensors using all of the daily paired AOD - PM$_{2.5}$ observations for the years 2000-2003. The regression analysis between MODIS AOD and AQS PM$_{2.5}$ observations revealed that the relationship is generally weak during the cool season (October – March) and relatively strong during the warm season (April - September). This is consistent with previous research results shown in Rush et al. (2004) and has been attributed to weaker boundary layer mixing or differences in PM$_{2.5}$ speciation between summer and winter. Consequently, we grouped the data for April through September for each year (50 % of the daily available AOD data sets) and determined correlation coefficients and regression equations for each satellite sensor, which are also shown in Table 1. The warm season regression equations were applied to MODIS AOD observations to estimate ground-level PM$_{2.5}$.

Table 1. Linear correlation coefficients by month and sensor, and regression coefficients for April-September for each year and sensor. $PM_{2.5}$ is the dependent variable and AOD is the dependent variable: $PM_{2.5} = Slope*AOD + Intercept$

	2000 Terra	2001 Terra	2002 Terra	2002 Aqua	2003 Terra	2003 Aqua
January		0.062	0.121		0.036	0.432
February		0.553	0.475		0.728	0.198
March		0.038	-0.015		0.457	-0.133
April		0.511	0.326		0.806	0.397
May		0.431	0.269		0.328	0.469
June	-0.140		0.420		0.452	0.759
July	0.676	0.709	0.732	0.705	0.481	0.812
August	0.756	0.640	0.446	0.058	0.409	0.824
September	0.758	0.707	0.415	0.341	0.652	0.591
October	0.741	0.295	0.171	0.658	0.225	0.359
November	0.927	0.372	0.100	-0.077	0.052	0.164
December	0.181	0.234	0.224	-0.466	-0.411	0.151
April - September	0.579	0.643	0.559	0.401	0.661	0.727
	(June - Sept.)		(July - Sept.)			
Regression coefficients, April - September:						
Intercept	11.29	11.69	8.88	11.40	8.85	6.47
Slope	15.88	19.33	17.16	9.41	18.57	18.39

Bias Removal Procedure

An assumption made in developing the merged AQS-MODIS $PM_{2.5}$ product was that the AQS observations are unbiased with respect to the local value of $PM_{2.5}$, but there could be biases in the MODIS $PM_{2.5}$ estimates due to the indirect nature of the observation and the imperfect relationship between AOD and $PM_{2.5}$. To account for this potential bias, we determined, on a daily basis, a spatial MODIS bias field, which was then used to adjust the MODIS $PM_{2.5}$ estimates to match the AQS observations in a mean sense. The bias field was calculated on the 10 x 10 km grid as the difference between a highly smoothed MODIS-estimated $PM_{2.5}$ (Figures 5c and 6c) and a similarly smoothed AQS field (Figures 5d and 6d). Two different techniques were tested to perform this smoothing. The first was a two-step B-Spline algorithm as illustrated in Figure 5 for June 24, 2003, a date characterized by excellent MODIS data coverage. The second technique was IDW as shown in Figure 6. By inspection, the B-Spline algorithm resulted in a much smoother bias field as clear in Figures 5e and 6e. However, the reason a two-step B-Spline was used instead of a one-step B-Spline is to make sure that the algorithm is not over

smoothing. Thus, a two-step B-Spline technique was used to remove the bias in the AQS observations with respect to ground observations on a daily basis even if the IDW were subsequently used to create the PM$_{2.5}$ surfaces.
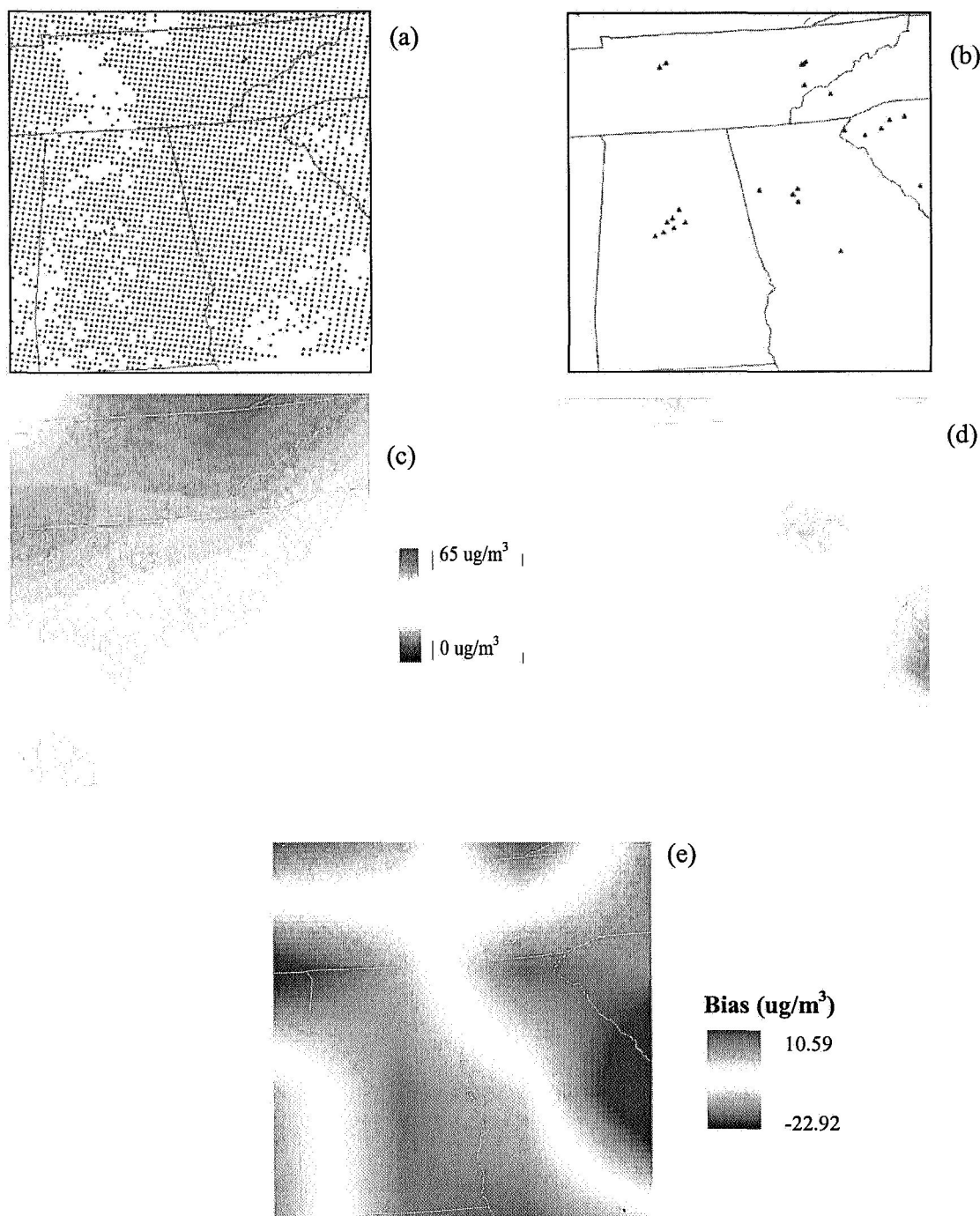

Figure 5. The bias determination procedure using B-Spline for June 24, 2003: (a) MODIS coverage, (b) AQS coverage, (c) Smooth MODIS: The 2nd iteration of the B-Spline algorithm with 2 knots on X and Y, (d) Smooth AQS: The 2nd iteration of the B-Spline algorithm with 2 knots on X and Y, (e) Difference between smooth MODIS and AQS fields (bias).

(a)

(b)

(c)

| 65 ug/m^3 |

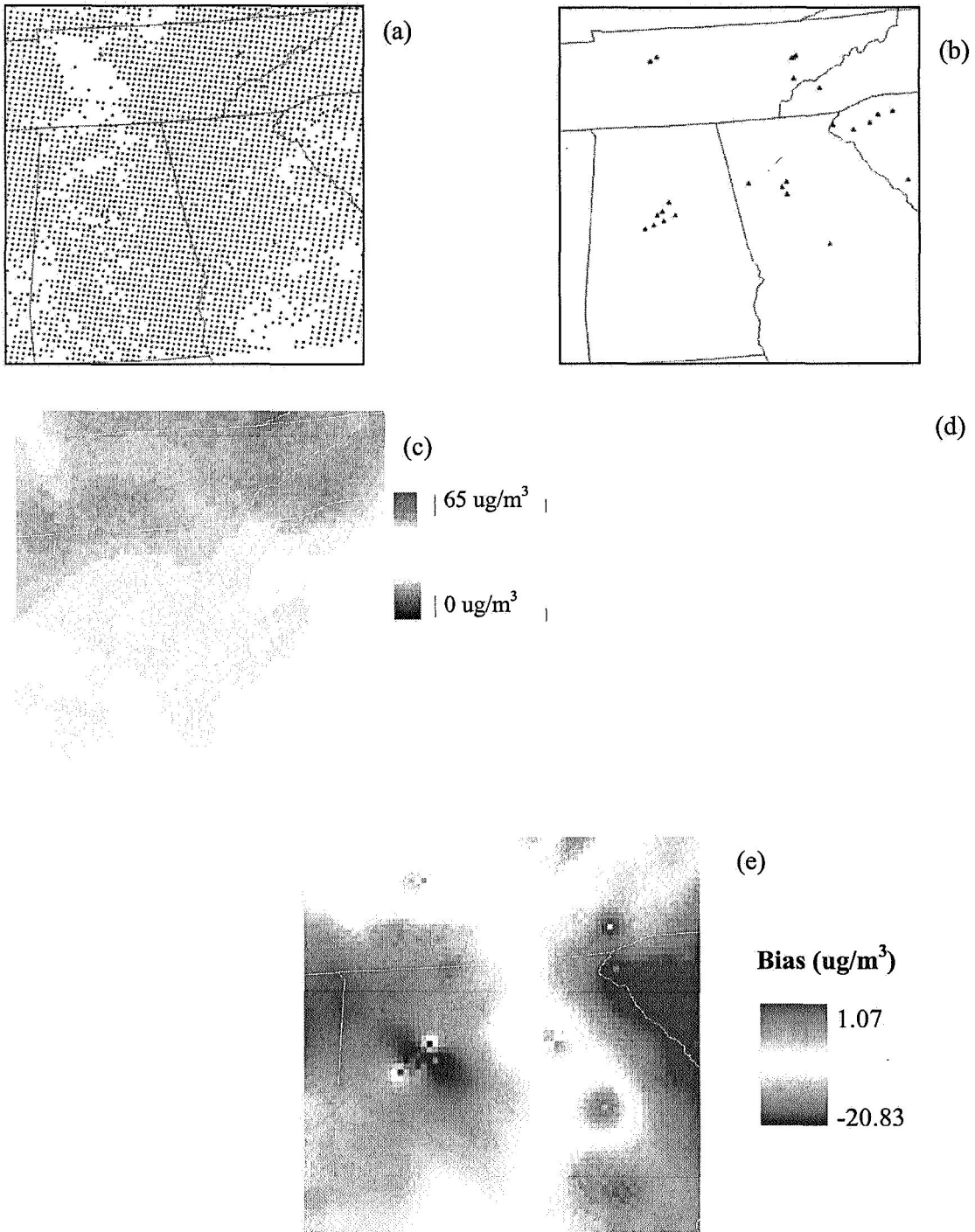| 0 ug/m^3 |

(d)

(e)

Bias (ug/m^3)

1.07

-20.83

Figure 6. The bias determination procedure using IDW for June 24, 2003: (a) MODIS coverage, (b) AQS coverage, (c) Smooth MODIS: IDW Surface (d) Smooth AQS: IDW Surface, (e) Difference between smooth MODIS and AQS fields (bias).

14

AQS AND MODIS DATA MERGER

After applying the quality control procedure to eliminate anomalous ground observations and the bias removal algorithm to remove biases in the satellite observations with respect to ground observations on a daily basis, the AQS ground-based $PM_{2.5}$ data were then merged with the Terra MODIS $PM_{2.5}$ estimates for the period of April 1- September 30, 2003 to produce a spatial surface of estimated $PM_{2.5}$ for each day using the B-Spline and IDW surfacing techniques. Figures 7a and 7b show the $PM_{2.5}$ B-Spline surface for June 24, 2003 using the MODIS-derived data set and the AQS data set separately, and Figure 7c demonstrates the $PM_{2.5}$ B-Spline surface using the merged data set. In merging the MODIS and AQS data, separate weights were applied to the two data sets to reflect their relative uncertainties in the B-Spline surfacing algorithm. Using a simplified Kalman Filter approach we determined the appropriate weighting for the MODIS data to be approximately 0.1. Thus, each MODIS observation was weighted by this factor, with AQS values weighted by a factor of 1.0, in the B-Spline surfacing algorithm. Figure 8 shows the same results using the IDW surfacing technique.
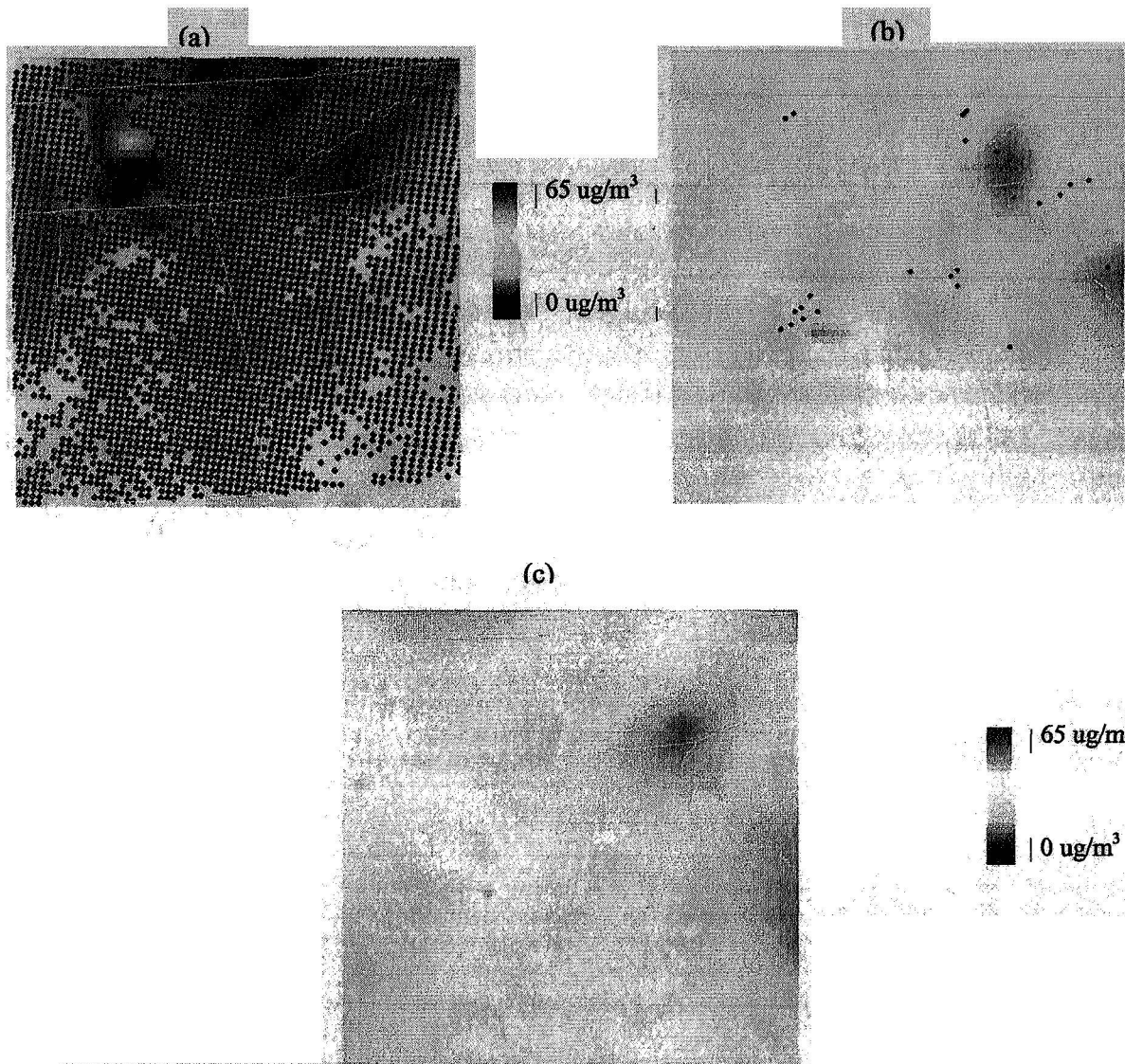
Figure 7. AQS-MODIS merging process for June 24, 2003: (a) PM$_{2.5}$ B-Spline surface using MODIS data only (b) PM$_{2.5}$ B-Spline surface using AQS data only (c) PM$_{2.5}$ B-Spline surface using a merged data set with a weight of 1 for AQS and a weight of 0.1 for MODIS.
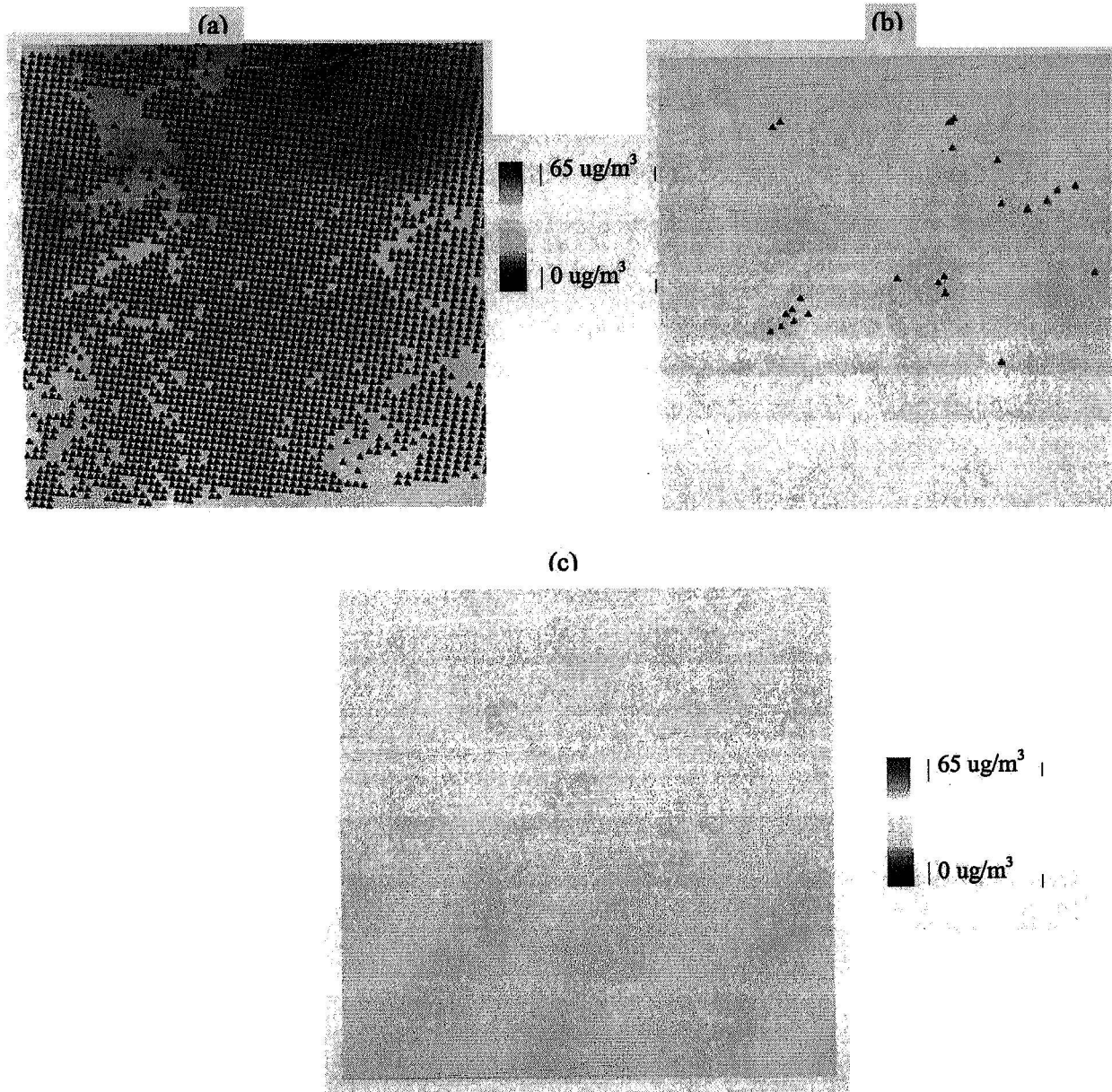
Figure 8. AQS-MODIS merging process for June 24, 2003: (a) $PM_{2.5}$ IDW surface using MODIS data only (b) $PM_{2.5}$ IDW surface using AQS data only (c) $PM_{2.5}$ IDW surface using a merged data set.

17

CROSS VALIDATION ANALYSIS (BOOTSTRAP ANALYSIS)

Cross-validation analysis enables error statistics to be generated for the estimated $PM_{2.5}$ surfaces. In cross-validation analysis, each observed value in the AQS $PM_{2.5}$ data set is individually removed from the set, the surface generated with the other points, and the value of the surface at the location of the omitted observation compared to the observation. Root mean square difference (RMSD) statistics have been compiled for each day to provide estimates of the expected errors in the daily surfaces, and by measurement site to identify sites where the surface is most uncertain.

The daily time series of RMSD between the B-Spline and observed AQS $PM_{2.5}$ values are shown in Figure 9. There is a slight tendency for higher RMSD during the summer, when mean $PM_{2.5}$ values are higher. The range of values is approximately 1-9 ug/m^3, with the highest RMSD occurring on days with fewer observations. The mean RMSD for the entire time period is 2.7 ug/m^3. The daily time series of the RMSD's between the IDW and observed AQS $PM_{2.5}$ values are shown in Figure 10. As in the B-Spline case, there is a slight tendency for higher RMSD's during the summer, when mean $PM_{2.5}$ values are higher. The range of vales is approximately 1-7 ug/m^3 with the highest RMSD occurring on days with fewer observations. The mean RMSD for the entire time period is 2.1 ug/m^3 which is 22% lower than for B-Spline.

When sorted by site, the RMSD between the bootstrap and observed $PM_{2.5}$ values indicate geographic locations where the $PM_{2.5}$ surfaces are more uncertain. Figures 11and 12 show RMSD values by AQS site, averaged over all days for the two surfacing techniques. Values range from about 1-6 ug/m^3, with results for the IDW technique being slightly lower than for the B-Spline.
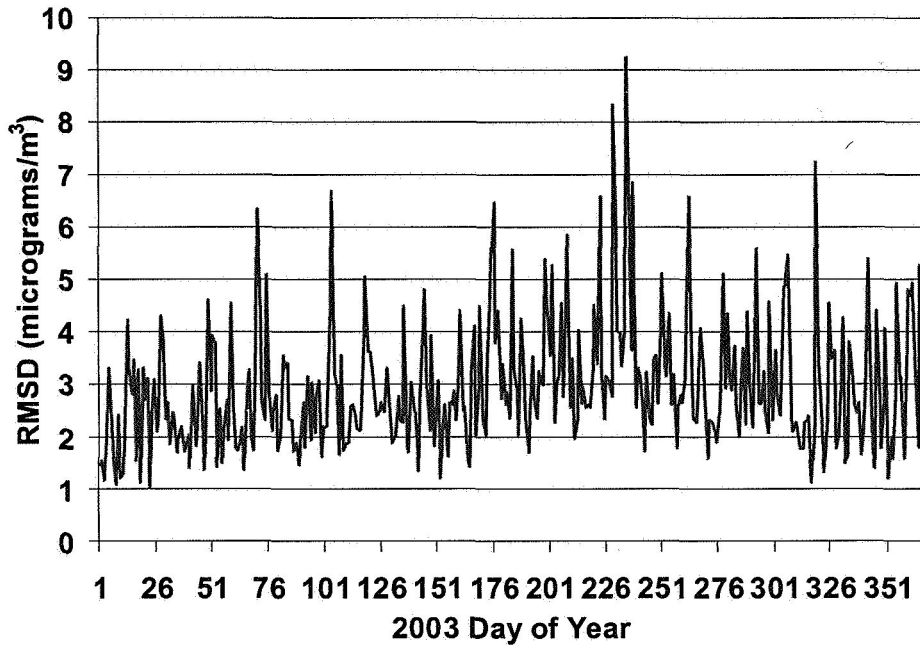
Figure 9. Daily time series of Root Mean Square Differences between B-Spline surface estimates and observed $PM_{2.5}$ values from the AQS data set, estimated by the bootstrap analysis.
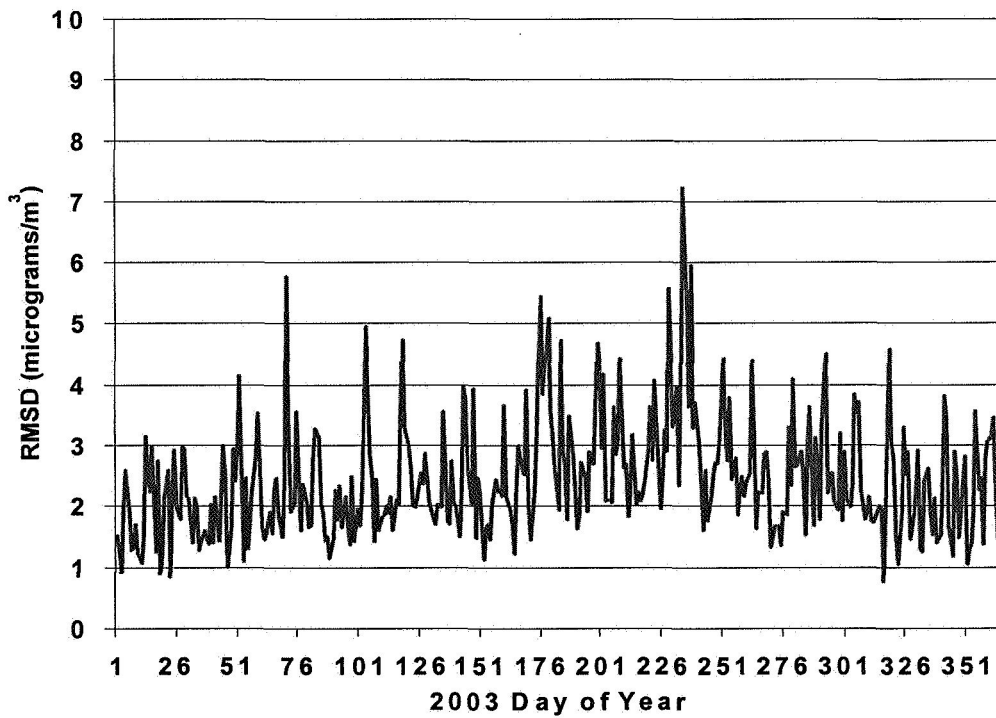

Figure 10. Daily time series of Root Mean Square Differences between IDW surface estimates and observed $PM_{2.5}$ values from the AQS data set, estimated by the bootstrap analysis.

19

Legend
RSMD
BOOT_OBS
- ● 0.61 - 1.70
- ◎ 1.71 - 2.79
- ○ 2.80 - 3.88
- ○ 3.89 - 4.97
- ◉ 4.98 - 6.09

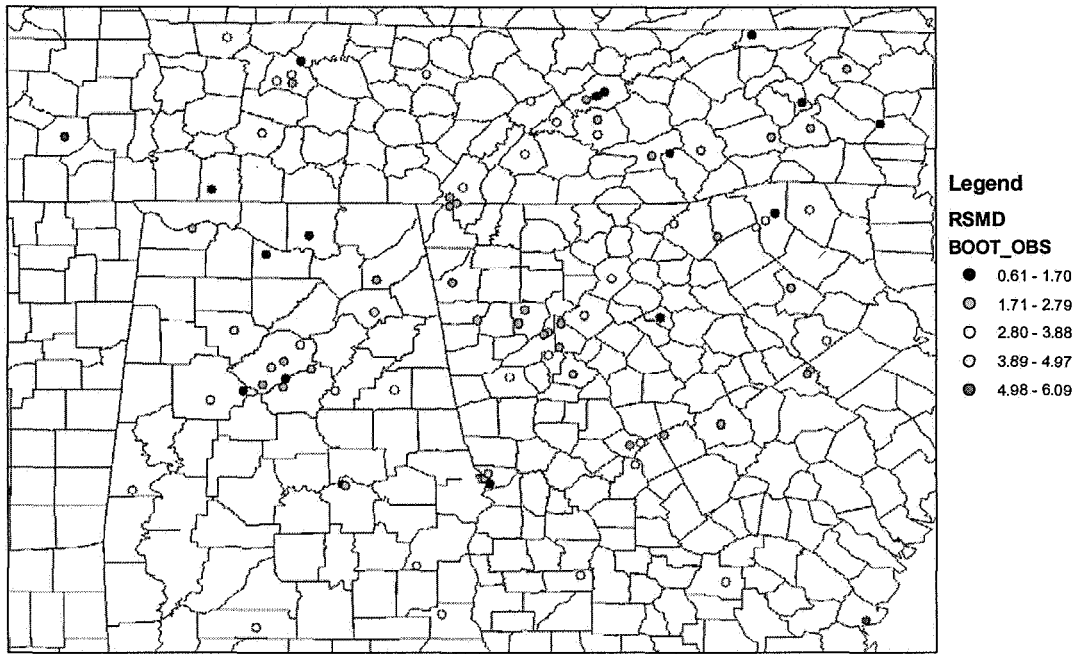Figure 11. Root Mean Square Differences between B-Spline and observed $PM_{2.5}$ values, estimated by the bootstrap analysis for each AQS site.

Legend
RMSD
Boot_OBS
- ● 0.6 - 1.70
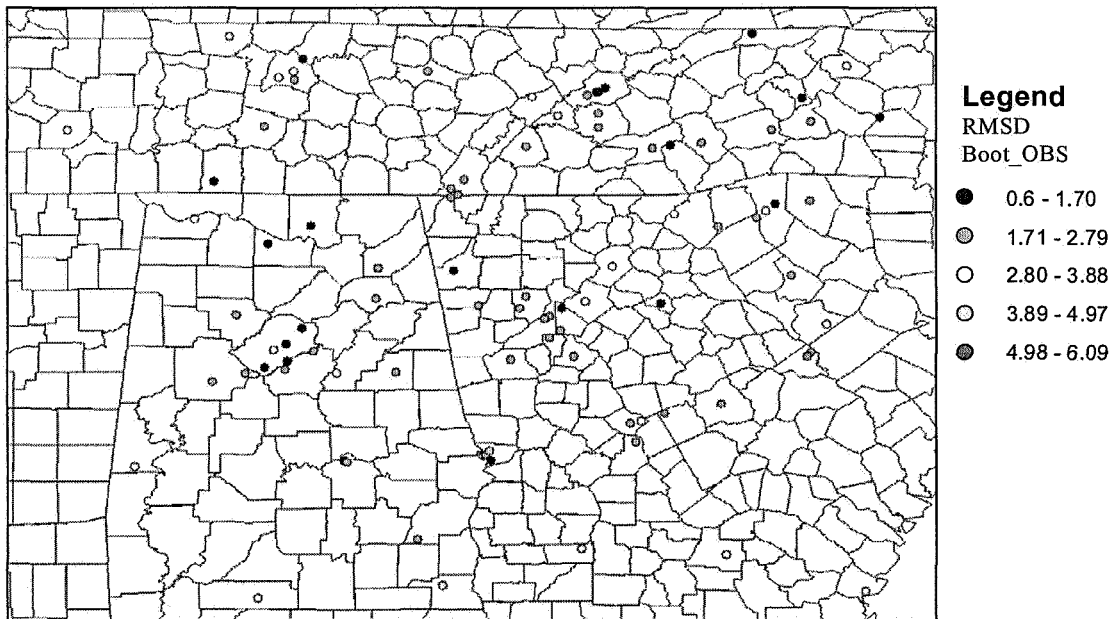- ◉ 1.71 - 2.79
- ○ 2.80 - 3.88
- ○ 3.89 - 4.97
- ◉ 4.98 - 6.09

Figure 12. Root Mean Square Differences between IDW and observed $PM_{2.5}$ values, estimated by the bootstrap analysis for each AQS site.

QC vs. No QC Comparison

In order to statistically evaluate improvements in the $PM_{2.5}$ estimates obtained by applying the QC procedure, daily B-Spline surfaces of $PM_{2.5}$ were generated for the year 2003 using only AQS data, once using the QC filtered dataset and once with the raw dataset. The cross validation results results showed that the QC reduced the mean RMSD between the bootstrap and observed AQS $PM_{2.5}$ values averaged over all 365 days from 3.3 to 2.9 for an improvement of 12% over the raw dataset. Also, correlation coefficients obtained using the QC filtered dataset increased to 0.91, compared with 0.88 in the data with no QC.

AQS Only vs. Merged MODIS-AQS Comparison

Improvements in $PM_{2.5}$ estimates obtained by merging the MODIS-derived $PM_{2.5}$ data with the AQS ground data were quantified using cross validation analysis performed on the daily surfaces of $PM_{2.5}$ that were generated for the warm season of 2003 (April 1-September 30), once using the AQS data only and once with the merged MODIS-AQS data set. The results showed that adding the MODIS data reduced the mean RMSD between the B-Spline and observed values averaged over the 182 days from 3.2 to 2.7 for a 16% improvement over the AQS-only data set. Also, as presented in Figure 13, which shows estimated versus actual values in both cases, the coefficient of determination increased from 0.840 to 0.874 corresponding to an increase of the correlation coefficient from 0.917 to 0.935.

For the IDW case (Figure 14), adding the MODIS data reduced the mean RMSD between the IDW and observed values averaged over the 182 days from 2.7 to 1.6 for a 40% improvement over the AQS-only data set. The correlation coefficient increased from 0.94 to 0.97. These statistics are summarized in Tables 2 and 3.
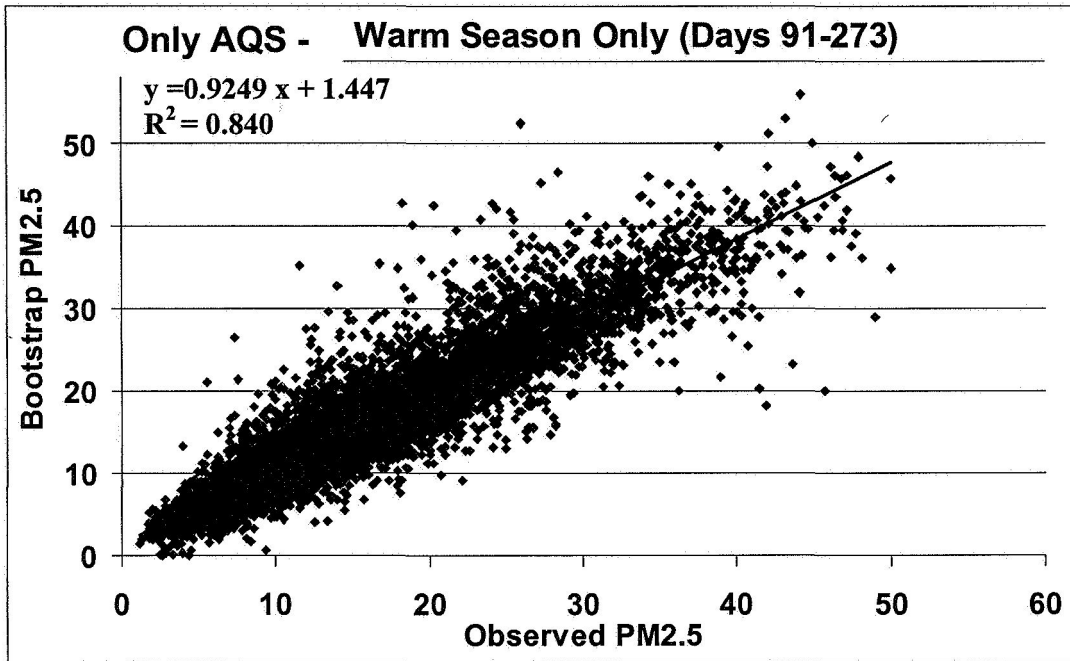
Table 2: Root mean square differences and regression statistics for different surfacing techniques and data sources.

Surfacing Technique and Data Source	RMSD All Days	RMSD Warm Season (Days 91-273)	R^2	Slope	Intercept
Bspline, AQS only, no QC	3.302	3.556	0.795	0.895	1.970
Bspline, AQS only, with QC	2.927	3.164	0.840	0.925	1.447
IDW, AQS only	2.450	2.686	0.878	0.899	2.088
B-Spline, merged AQS/MODIS	N/A	2.756	0.870	0.925	1.390
IDW, merged AQS/MODIS	N/A	1.613	0.949	0.924	1.356

Table 3: Improvements ofRoot Mean Square Differences for different surfacing techniques and data sources.

Surfacing Technique and Data Source	Improvement
Bspline: QC vs. No QC	12 %
Bspline: AQS only vs. merged AQS/MODIS	16 %
IDW: AQS only vs. merged AQS/MODIS	40 %

(a)

Only AQS - Warm Season Only (Days 91-273)

$y = 0.9249 x + 1.447$
$R^2 = 0.840$

Bootstrap PM2.5 (y-axis)
Observed PM2.5 (x-axis)

(b)

Merged Data - Warm Season Only (days 91-273)

$y = 0.9255 x + 1.391$
$R^2 = 0.874$
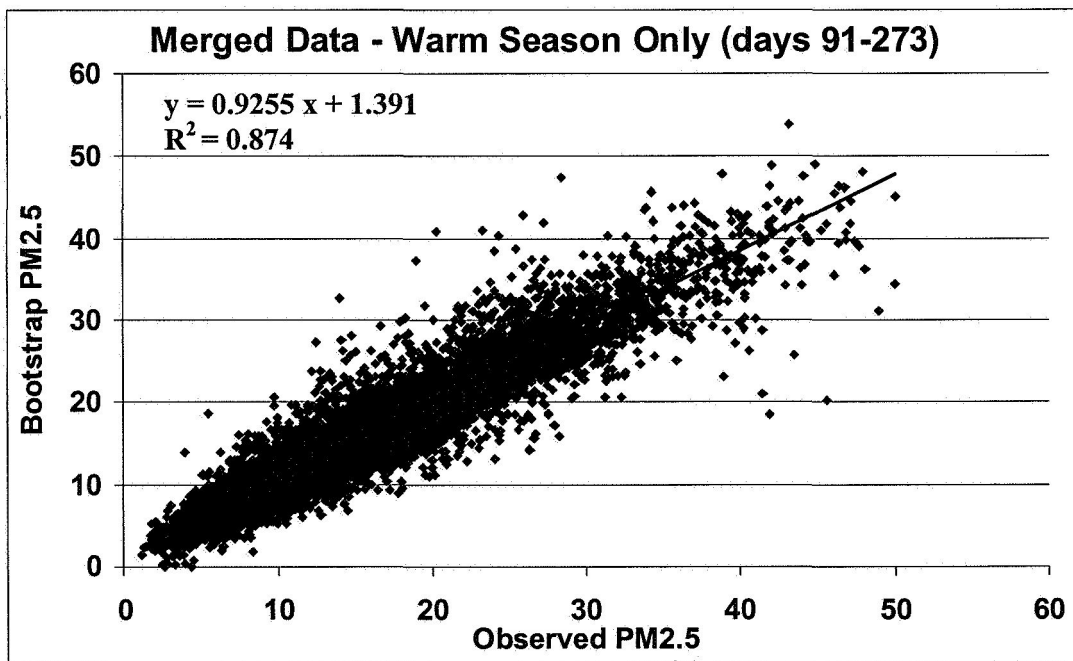
Bootstrap PM2.5 (y-axis)
Observed PM2.5 (x-axis)

Figure 13. Cross-validation results for the daily B-Spline surfaces of year 2003 warm season (April 1-September 30) a) Using AQS only b) Using Merged MODIS-AQS data set.
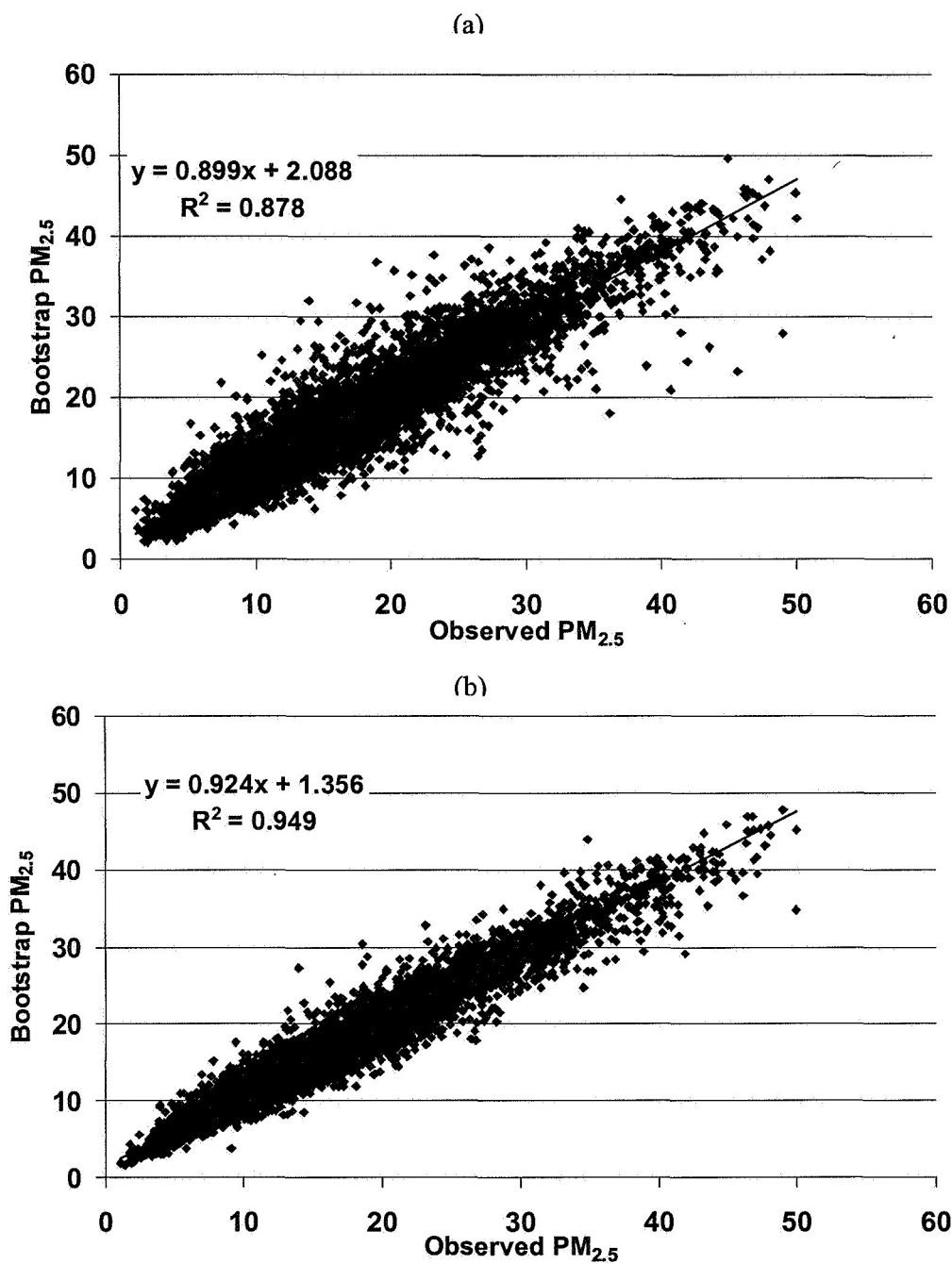
(a)

(b)

Figure 14. Cross-validation results for the daily IDW surfaces of year 2003 warm season (April 1-September 30) a) Using AQS only b) Using Merged MODIS-AQS data set

IDW vs. B-Spline Comparison

The results showed that, using only AQS data, the mean RMSD for the IDW technique was 15% lower than that for the B-Spline. Using the merged AQS and MODIS data, the mean RMSD was 41% lower than that of the B-Spline. Those results are also shown in Table 2. The regression analysis also showed that the IDW case had higher coefficients of determination in both cases. Figures 15 and 16 show the composite (mean) surface of the IDW and B-Spline daily surfaces respectively for year 2003. It can be noted that the composite surfaces from both techniques have similar structures in general but the IDW can introduce numerical artifacts like those indicated by the arrows in Figure 16, which could be due to the interpolating nature of the IDW, which assumes that the maxima and minima can occur only at the observation points.
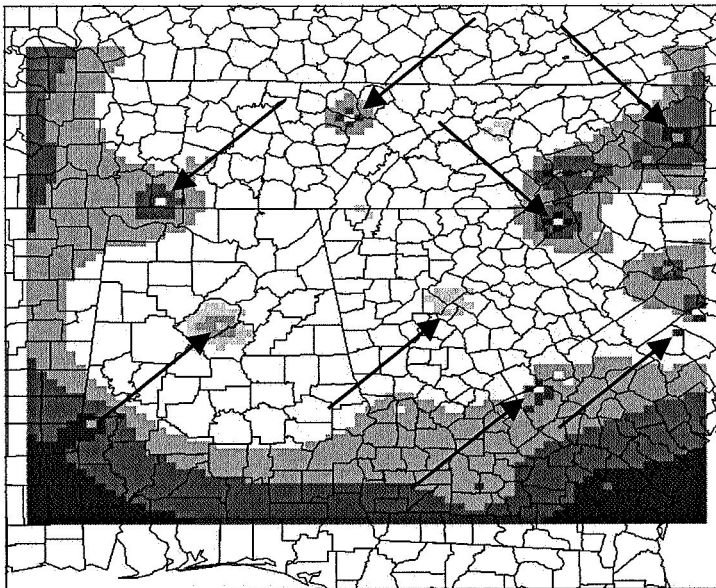
PM2.5 B-Spline Surfaces Year 2003 Composite
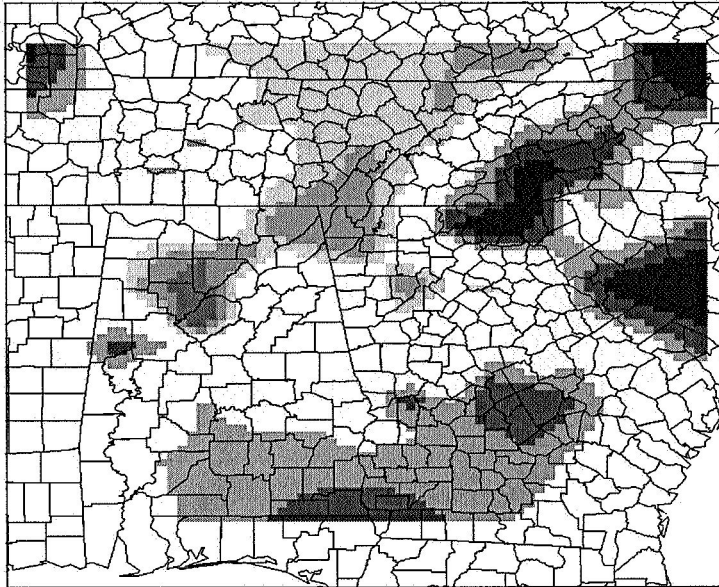
PM2.5 (ug/m^3)

■	10.24 - 11.97
■	11.98 - 12.57
▦	12.58 - 12.97
	12.98 - 13.35
	13.36 - 13.74
	13.75 - 14.23
	14.24 - 14.85
	14.86 - 15.54
▦	15.55 - 16.50
■	16.51 - 18.36

Figure 15. 2003 annual mean of all PM$_{2.5}$ B-Spline surfaces

PM2.5 IDW Surfaces Year 2003 Composite

PM2.5 (ug/m^3)

■	10.24 - 11.97
■	11.98 - 12.57
▦	12.58 - 12.97
	12.98 - 13.35
	13.36 - 13.74
	13.75 - 14.23
	14.24 - 14.85
	14.86 - 15.54
▦	15.55 - 16.50
■	16.51 - 18.36

Figure 16. 2003 annual mean of all PM$_{2.5}$ IDW Surfaces

CONCLUSIONS

This paper has described and demonstrated a methodology for estimating ground-level continuous $PM_{2.5}$ concentrations using B-Spline and IDW surfacing techniques and leveraging NASA MODIS data to complement EPA AQS data. The paper has shown that merging MODIS remote sensing data with surface observations of $PM_{2.5}$ not only provides a more complete daily representation of $PM_{2.5}$ than either data set alone would allow, but it also reduces the errors in the $PM_{2.5}$ estimated surfaces.

The IDW technique's strengths are the simplicity of the underlying principle and the speed of calculation. However, it is recognized that this technique can easily be affected by an uneven distribution of observational data points since an equal weight will be assigned to each of the data points even if it is in a cluster. In addition, maxima and minima in the IDW surface can only occur at data points since IDW is an interpolating technique. On the other hand, recursion of the B-Spline technique provides a robust methodology for data sets with mixtures of data sparse and data rich regions, which is a common condition with many environmental and health datasets. The B-Spline technique is able to produce maximum and minimum values at locations away from point observations, but it does not handle discontinuities in the assumed surface without advanced programming logic. The paper has also shown that the daily IDW $PM_{2.5}$ surfaces had smaller errors, with respect to observations, than those of the B-Spline surfaces in the year studied. However the IDW surfaces had more numerical artifacts as was clear in the annual composite surface, which could be due to the interpolating nature of the IDW that assumes that the maxima and minima can occur only at the observation points.

Finally, the methods discussed in this paper increase temporal and spatial resolution of fine particulate estimates and have the potential to provide public health practitioners with more tools to describe the public health impact of air pollutants such as $PM_{2.5}$, ozone, and other pollutants. This paper establishes a foundation for environmental public health linkage and association studies for which determining the concentrations of an environmental hazard such as $PM_{2.5}$ with good accuracy levels is critical.

Major Study Contributions

Globally, environmental epidemiologists have been trying for decades to develop valid estimates of dose and duration of human exposure to air pollutants. Where we do have air quality

information, we lack spatial and temporal resolution because of the limited resources of ground monitoring and global information on air quality. To compensate for this, we use models to estimate levels of pollutants in areas without monitoring. Epidemiological studies suggest that there is an association between incidence and exacerbation of adverse respiratory and cardiovascular health effects and air pollution. Studies also suggest an association between cancer and air pollution and birth defects and air pollution. However, the findings are inconsistent and controversial due to the weak exposure data products.

The methods discussed in this paper increase temporal and spatial resolution of fine particulate estimates and have the potential to provide public health practitioners with more tools to describe the public health impact of multiple air pollutants including $PM_{2.5}$, and ozone. This information would facilitate more effective research into the association of selected health events with environmental air quality levels. In addition these new data products would serve as a tool for environmental public health surveillance to monitor trends and changes over time, to serve as an early warning system for prevention of exposure of humans to potential hazards, and provide information for decision-making and program planning and evaluation. If an algorithm could be developed to estimate air quality with satellite data in locations where there is no ground monitoring (much of the world) then we would have more information to prevent and control public health problems globally. There is also a potential cost-benefit to reduce our dependence on ground monitoring for air quality information.

FUTURE WORK

The estimated $PM_{2.5}$ results will be linked with Health Maintenance Organization (HMO) asthma visits in Metro-Atlanta counties on the grid aggregated level as well as the individual level to demonstrate the feasibility of linking environmental data with health outcomes data for association studies. Having an accurate continuous representation or a spatial surface of the environmental hazard facilitates such linkage.

REFERENCES

1. Beck, L., Lobitz, B., Wood, B. (2000). Remote Sensing and Human Health: New Sensors and New Opportunities, *Emerging Infectious Diseases Journal, 6(3), 217-226.*

2. Bell,

3. Chu, A., Kaufman, Y. J., Zibordi, G., Chern, J. D., Mao, J., Li, C., Holben, B. N. (2003). Global monitoring of air pollution over land from the Earth Observing System-Terra Moderate Resolution Imaging Spectroradiometer (MODIS) *Journal of Geophysical Research,* 108, D21.

4. Eberly, S., Fitz-Simons, T., Hanley, T., Weinstock, L., Tamanini, T., Denniston, G., Lambeth, B., Michel, E., Bortnick, S. (2002). Data Quality Objectives (DQOs) for Relating Federal Reference Method (FRM) and Continuous PM2.5 Measurements to Report an Air Quality Index (AQI), EPA-454/B-02-002.EPA

5. Engel-Cox, J.A., Holloman, C., Coutant, B., and Hoff, R. (2004). Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality, *Atmos. Environ.,* 38 (16), 2495-2509.

6. Gillespie. (2005). Correlating Federal Reference Method and Continuous PM2.5 Monitors in the MARAMA Region available at
   http://www.marama.org/reports/MARAMA_Correlation_Report-090105.pdf

7. Gupta, P., Christopher, S., Wang, J., Gehrig, R., Lee, Y., and Kumar, R. (2005). Satellite Remote Sensing of Particulate Matter and Air Quality over Global Cities, *Atmospheric Environment,* in press.

8. Gupta, P., and Christopher, S. (2006). Multi year satellite remote sensing of particulate matter air quality over Sydney, Australia, *International J. Remote Sensing, submitted.*

9. Ito, K., Thurston, G., Na'das, A., and Lippmann, M. (2001), Monitor-tomonitor temporal correlation of air pollution and weather variables in the north-central U.S., *J. Exposure Anal. Environ. Epid.,* 11(1), 21–32.

10. Jerret et al, 2005

11. Kaldy, W.; Rizzo, M.; Scheff, P. A.; Adjusting Tapered Element Oscillating Microbalance Data for Comparison with Federal Reference Method PM2.5 Measurements in Region 5. (2003). *Journal of the Air & Waste Management Association,* 53, 596-607.

12. Liu, Y., Park, R. J., Jacob, D. J., Li, Q., Kilaru, V. and Sarnat, J. A. (2004). Mapping annual mean ground-level $PM_{2.5}$ concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States; *J. Geophysical Research*, 109, D22206.

13. Morain S., Budge A., Budge T., Baros K., Hudspeth W., Bales C., Sanchez G., Sprigg W., Yin D., Barbaris B., Chandy B., Nickvic S., Caskey S., Speer J., Bradbury J. (2005). Modeling atmospheric dust for a public health decision support system. The ISPRS 31st International Symposium on Remote Sensing of Environment, June 20-24, St. Petersburg, Russia.

14. Patz, J. (2005). Satellite Remote Sensing Can Improve Chances of Achieving Sustainable Health, *Environ Health Perspect.,* 113(2): A84–A85

15. Rush, A., Dougherty, J; Engel-Cox, J. (2004). Correlating seasonal averaged in situ monitoring of fine PM with satellite remote sensing data using geographic information system (GIS). Presented at the 49th annual meeting of the International Society for Optical Engineering, August 2-6, 2004, Denver, CO.

16. Wang, J., and Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: implications for air quality studies. Geophysical Research Letters, 30, 21, 2095, doi:10.1029/2003/GL018174.

17. Wong et al. 2003

18. Watkins and Boothe

19. World Health Organization. (2000).
    Available at: http://www.who.int/mediacentre/factsheets/fs206/en/

ABOUT THE AUTHORS

The corresponding author, Dr. Mohammad Z. Al-Hamdan, is a research scientist at the Universities Space Research Association at NASA Marshall Space Flight Center, National Space Science and Technology Center, NASA Global Hydrology and Climate Center, Huntsville, Alabama 35816. Email: mohammad.alhamdan@nasa.gov; Tel.: (256) 961-7465; Fax: (256) 961-7377. Dr. William L. Crosson, Mr. Maurice G. Estes, Jr. and Dr. Ashutosh S. Limaye are also research scientists at the Universities Space Research Association at NASA/MSFC, National Space Science and Technology Center, NASA Global Hydrology and Climate Center. Drs. Dale

A. Quattrochi and Douglas L. Rickman are reasearch scientists at the Earth Sciences Office at NASA Marshall Space Flight Center, National Space Science and Technology Center, NASA Global Hydrology and Climate Center. Ms. Kafayat A. Adeniyi is a public health advisor at Quantell, Inc. Dr. Judith R. Qualters is the Chief of the Environmental Health Tracking Branch in the Division of Environmental Hazards and Health Effects, National Center for Environmental Health, Centers for Disease Control and Prevention (CDC). Dr. Amanda S. Niskar is a research scientist at Israel Center for Disease Control Gertner Institute. Mr. Dennis D. Tolsma and Ms. Amber H. Sinclair are research scientists at Kaiser Permanente.