Summer 2012

Optimal interpolation schemes to constrain Pm2.5 In Regional Modeling Over The United States

Sinan Dhia Jameel Sousan
University of Iowa

OPTIMAL INTERPOLATION SCHEMES TO CONSTRAIN $PM_{2.5}$ IN REGIONAL MODELING OVER THE UNITED STATES

by

Sinan Dhia Jameel Sousan

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of Philosophy degree in Chemical and Biochemical Engineering in the Graduate College of
The University of Iowa

July 2012

Thesis Supervisor: Assistant Professor Charles O. Stanier

ABSTRACT

This thesis presents the use of data assimilation with optimal interpolation (OI) to develop atmospheric aerosol concentration estimates for the United States at high spatial and temporal resolutions. Concentration estimates are highly desirable for a wide range of applications, including visibility, climate, and human health. OI is a viable data assimilation method that can be used to improve Community Multiscale Air Quality (CMAQ) model fine particulate matter ($PM_{2.5}$) estimates. $PM_{2.5}$ is the mass of solid and liquid particles with diameters less than or equal to 2.5 μm suspended in the gas phase. OI was employed by combining model estimates with satellite and surface measurements. The satellite data assimilation combined 36 x 36 km aerosol concentrations from CMAQ with aerosol optical depth (AOD) measured by MODIS and AERONET over the continental United States for 2002. Posterior model concentrations generated by the OI algorithm were compared with surface $PM_{2.5}$ measurements to evaluate a number of possible data assimilation parameters, including model error, observation error, and temporal averaging assumptions. Evaluation was conducted separately for six geographic U.S. regions in 2002. Variability in model error and MODIS biases limited the effectiveness of a single data assimilation system for the entire continental domain. The best combinations of four settings and three averaging schemes led to a domain-averaged improvement in fractional error from 1.2 to 0.97 and from 0.99 to 0.89 at respective IMPROVE and STN monitoring sites. For 38% of OI results, MODIS OI degraded the forward model skill due to biases and outliers in MODIS AOD.

Surface data assimilation combined 36 × 36 km aerosol concentrations from the CMAQ model with surface $PM_{2.5}$ measurements over the continental United States for 2002. The model error covariance matrix was constructed by using the observational method. The observation error covariance matrix included site representation that scaled the observation error by land use (i.e. urban or rural locations). In theory, urban locations should have less effect on surrounding areas than rural sites, which can be controlled

using site representation error. The annual evaluations showed substantial improvements in model performance with increases in the correlation coefficient from 0.36 (prior) to 0.76 (posterior), and decreases in the fractional error from 0.43 (prior) to 0.15 (posterior). In addition, the normalized mean error decreased from 0.36 (prior) to 0.13 (posterior), and the RMSE decreased from 5.39 μg m^{-3} (prior) to 2.32 μg m^{-3} (posterior). OI decreased model bias for both large spatial areas and point locations, and could be extended to more advanced data assimilation methods.

The current work will be applied to a five year (2000-2004) CMAQ simulation aimed at improving aerosol model estimates. The posterior model concentrations will be used to inform exposure studies over the U.S. that relate aerosol exposure to mortality and morbidity rates. Future improvements for the OI techniques used in the current study will include combining both surface and satellite data to improve posterior model estimates. Satellite data have high spatial and temporal resolutions in comparison to surface measurements, which are scarce but more accurate than model estimates. The satellite data are subject to noise affected by location and season of retrieval. The implementation of OI to combine satellite and surface data sets has the potential to improve posterior model estimates for locations that have no direct measurements.

Abstract Approved: _____

Thesis Supervisor

_____

Title and Department

_____

Date

OPTIMAL INTERPOLATION SCHEMES TO CONSTRAIN $PM_{2.5}$ IN REGIONAL MODELING OVER THE UNITED STATES

by

Sinan Dhia Jameel Sousan

A thesis submitted in partial fulfillment
of the requirements for the Doctor of Philosophy degree
in Chemical and Biochemical Engineering
in the Graduate College of
The University of Iowa

July 2012

Thesis Supervisor: Assistant Professor Charles O. Stanier

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Sinan Dhia Jameel Sousan

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Chemical and Biochemical Engineering at the
July 2012 graduation.

Thesis Committee: _____

Charles O. Stanier, Thesis Supervisor

_____

Gregory R. Carmichael

_____

Tonya Peeples

_____

Jacob Oleson

_____

Scott Spak

To My Family

ACKNOWLEDGEMENTS

I would like to thank my graduate advisor Dr. Charles Stanier for his guidance, encouragement and support during my work as a graduate student and for introducing me to atmospheric modeling. I thank Dr. Gregory Carmichael for his support during my graduate study and for recognizing me as part of CGRER. I thank Dr. Jacob Oleson for his informative statistical guidance and feedback throughout my thesis work. I thank Dr. Scott Spak for his feedback and technical support towards my research work during my graduate study. I thank Dr. Tonya Peeples and Dr. Julie Jessop for their interest, feedback, and being a part of my thesis committee. I thank the faculty members in the department of Chemical and Biochemical engineering for accepting me as a graduate student. I would like to thank Linda Wheatley, Natalie Potter, Jane Frank, Jeremie Moen, and Jill McNamara for their help and support during my graduate study.

I would like to thank Dr. Tracey Holloway and her group for their help in installing and running CMAQ. I would like to thank Dr. Jaemeen Baek, Dr. Sarika Kulkarni, Dr. Chao Wei, Dr. Sang-Rin Lee, Dr. Alicia Pettibone, Dr. Yafang Cheng, Dr. Tianfeng Chai, Dr. Min Huang, Robert Bullard, Pablo Peralta, and Pallavi Marrapu for their technical help and feedback, code sharing and paper writing. I would like to thank Fulbright Foreign Student Program for their scholarship and EPA for funding the thesis work.

I would like to thank my dear wife Nagham Abdooni and my children Marian, Yousif, and Michael for their sacrifice and support during my work as a graduate student. I would like to thank my parents Dhia Sousan and Nadia Naoum for bringing me up and giving me the opportunities and knowledge that enabled me to continue in my graduate career.

ABSTRACT

This thesis presents the use of data assimilation with optimal interpolation (OI) to develop atmospheric aerosol concentration estimates for the United States at high spatial and temporal resolutions. Concentration estimates are highly desirable for a wide range of applications, including visibility, climate, and human health. OI is a viable data assimilation method that can be used to improve Community Multiscale Air Quality (CMAQ) model fine particulate matter ($PM_{2.5}$) estimates. $PM_{2.5}$ is the mass of solid and liquid particles with diameters less than or equal to 2.5 μm suspended in the gas phase. OI was employed by combining model estimates with satellite and surface measurements. The satellite data assimilation combined 36 x 36 km aerosol concentrations from CMAQ with aerosol optical depth (AOD) measured by MODIS and AERONET over the continental United States for 2002. Posterior model concentrations generated by the OI algorithm were compared with surface $PM_{2.5}$ measurements to evaluate a number of possible data assimilation parameters, including model error, observation error, and temporal averaging assumptions. Evaluation was conducted separately for six geographic U.S. regions in 2002. Variability in model error and MODIS biases limited the effectiveness of a single data assimilation system for the entire continental domain. The best combinations of four settings and three averaging schemes led to a domain-averaged improvement in fractional error from 1.2 to 0.97 and from 0.99 to 0.89 at respective IMPROVE and STN monitoring sites. For 38% of OI results, MODIS OI degraded the forward model skill due to biases and outliers in MODIS AOD.

Surface data assimilation combined 36 × 36 km aerosol concentrations from the CMAQ model with surface $PM_{2.5}$ measurements over the continental United States for 2002. The model error covariance matrix was constructed by using the observational method. The observation error covariance matrix included site representation that scaled the observation error by land use (i.e. urban or rural locations). In theory, urban locations should have less effect on surrounding areas than rural sites, which can be controlled

using site representation error. The annual evaluations showed substantial improvements in model performance with increases in the correlation coefficient from 0.36 (prior) to 0.76 (posterior), and decreases in the fractional error from 0.43 (prior) to 0.15 (posterior). In addition, the normalized mean error decreased from 0.36 (prior) to 0.13 (posterior), and the RMSE decreased from 5.39 μg m^{-3} (prior) to 2.32 μg m^{-3} (posterior). OI decreased model bias for both large spatial areas and point locations, and could be extended to more advanced data assimilation methods.

The current work will be applied to a five year (2000-2004) CMAQ simulation aimed at improving aerosol model estimates. The posterior model concentrations will be used to inform exposure studies over the U.S. that relate aerosol exposure to mortality and morbidity rates. Future improvements for the OI techniques used in the current study will include combining both surface and satellite data to improve posterior model estimates. Satellite data have high spatial and temporal resolutions in comparison to surface measurements, which are scarce but more accurate than model estimates. The satellite data are subject to noise affected by location and season of retrieval. The implementation of OI to combine satellite and surface data sets has the potential to improve posterior model estimates for locations that have no direct measurements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF NOMENCLATURE

Abbreviations

3D-var	Three-Dimensional Variational data assimilation
4D-var	Four-Dimensional Variational data assimilation
AERONET	Aerosol Robotic Network
ANO3I	Aitken mode Nitrate
ANO3J	Accumulation mode Nitrate
AOD	Aerosol Optical Depth
AQS	Air Quality System
ASO4I	Aitken mode Sulfate
ASO4J	Accumulation mode Sulfate
b_{im}	Bias (model-observation) for a specific site i and month m
B	Background error covariance matrix
CAMx	Comprehensive Air Quality Model with Extensions
CAPs	Criteria Air Pollutants
CB05	Carbon Bond
CMAQ	Community Multiscale Air Quality model
CTM	Chemical Transport Model
dx	Difference between the column indices of two cells in the x-direction
dy	Difference between the column indices of two cells in the y-direction
EPA	U.S. Environmental Protection Agency
f_{cmaq}	Model fractional errors
f_o	Observation fractional errors
FES	Fractional Error Score
F_{urb}	Fraction of urban area in a grid cell
GOCART	Goddard Global Ozone Chemistry Aerosol Radiation and Transport model
H	Transforms the model data into the same space and variable as the observed data
HAPs	Hazards air pollutants
IMPROVE	Interagency Monitoring of Protected Visual Environments
K	Kalman gain matrix
km	kilometer

l_h	Correlation length scale in km
L_{max}	Maximum radius of influence for the observation at urban locations
L_{min}	Minimum radius of influence for the observation at urban locations
L_{repr}	Radius of influence of an observation (km)
lxy	horizontal correlation length scale for errors in the model
m	meter
MATCH	Model of Atmospheric Transport and Chemistry
MCIP	Meteorology-Chemistry Interface Processor
MFB	Mean Fractional Bias
MFE	Mean Fractional Error
MISR	Multi-angle Imaging SpectroRadiometer
MM5	Mesoscale Meteorological Model
MODIS	Moderate Resolution Imaging Spectroradiometer instrument
NEI	National Emission Inventory
NME	Normalized Mean Error
O	Observation error covariance matrix
OI	Optimal interpolation
PBL	Planetary Boundary Layer
PC	Performance Classification
$PM'_{2.5}$	Posterior model fine particulate Matter
$PM_{2.5}^{cmaq}$	Prior model fine particulate Matter
PM_m	Model fine particulate Matter
PM_o	Surface observation fine particulate Matter
PM_EC	Elemental Carbon
PM_ORG_TOT	Total Organic Carbon
PM_OTH	Unspeciated aerosols
$PM_{2.5}$	Fine particulate Matter
PMCOARS	Coarse mode particles
R_z	Correlation between the observation and prior model increment at two stations
RMSE	Root Mean Square Error
RSIG	Remote Sensing Information Gateway
SMOKE	Sparse Matrix Operating Kernel for Emissions
SSC	Coarse mode Sea Salt
SSF	Fine mode Sea Salt
STEM	Sulfur Transport dEposition Model

STN	Speciation Trends Network
WRF	Weather Research and Forecasting

Greek and non-alphabetical Symbols

τ'	Posterior AOD
τ_{anet}	AERONET AOD
τ_{cmaq}	CMAQ AOD
τ_{modis}	MODIS AOD
τ_o	Observation AOD
ε_o	Minimum observation errors
ε_{cmaq}	Minimum model errors
ε_{repr}	Observation representation error
Δx	Model grid cell size
$\mu g\ m^{-3}$	Micro gram per cubic meter
Mm^{-1}	Mega meter
σ	standard deviation
%	percent
$\pm$	plus/minus

CHAPTER 1: INTRODUCTION

1.1 Background

1.1.1 Motivation and importance of work

Atmospheric particulate matter ($PM_{2.5}$) [*Seinfeld, et al.*, 2006] is the mass of solid or liquid particles less than or equal to 2.5μm, suspended in a gas. These particles are classified as fine aerosol particles and are composed of various species that include sulfates, nitrates, organic and elemental carbon, ammonium, sea salt, metal oxides, hydrogen ions, and water. Particulate matter is categorized as either primary aerosols, which are emitted directly to the atmosphere, or secondary aerosols, which are formed from gas to particle conversion processes. Primary aerosols are emitted directly from combustion sources such as forest fires, dust, sea salt and biological debris. Secondary aerosols are formed due to atmospheric gaseous reactions from chemicals released from power plants, industry, and automobile emissions. Natural sources of secondary aerosols include dimethylsulfade from the ocean; sulfate from volcanic gases; and organic aerosols from biogenic volatile organic carbon. $PM_{2.5}$ has adverse effects on health and visibility [*McMurry et al.,* 2005*; Penner et al.,* 2001]. Since $PM_{2.5}$ can penetrate and deposit deeply within the lungs, exposure to increased levels of fine particles has adverse effects on human health, especially to the cardiopulmonary system. Aerosols exert direct effects on climate and visibility from scattering and absorption and indirectly influence climate by cloud property modification.

Constraining spatial patterns at surface-level $PM_{2.5}$ is critical in estimating aerosol impacts, and in particular, $PM_{2.5}$ exposure limits. However, $PM_{2.5}$ concentration measurements from surface monitoring networks suffer from limited spatial resolution. Temporal resolution in the monitoring network varies from high (hourly) to moderate (1 in 6 day sampling for some speciated $PM_{2.5}$ samples). Assigning point measurements to represent ambient concentrations over large areas leads to errors in exposure

classification for epidemiological studies [*Grahame,* 2009*; Adar and Kaufman,* 2007*; Jerrett et al.,* 2006*; Ito et al.,* 2004*; Zeger et al.,* 2000].

An alternate exposure estimate of speciated fine particles based on chemical transport models has been in development for the past thirty years with the advantage of high spatial and temporal coverage. The air quality model produces hourly data, at any desirable horizontal (up to 1km) and vertical resolution. However, the models are still subject to improvements and developments. Improvements in the model output can be accomplished by improving both the meteorological and emission data used by the model, improving the chemistry and physics of the model, estimating pollutants at finer model resolutions, or post analysis methods that include statistical approaches like data assimilation (common in geoscience applications).

Optimal interpolation has been shown to improve $PM_{2.5}$ concentration estimates. This method combines estimates from air quality models with available measurements such as surface measurements or satellite retrievals. Data assimilation is a scheme that should be considered in conjunction with other methods of improving model skill, like refinements to emission inventories, meteorological data, or the chemistry that is used by the air quality models. Also proven to be effective are land use regression methods [Liu, et al., 2009] that combine meteorological data, satellite data, and land use factors to construct statistical models capable of estimating $PM_{2.5}$ estimates for both spatial and temporal resolutions. In addition to the primary goal of constructing high spatial resolution $PM_{2.5}$ estimates as exposure values, the current study is also applicable to other aerosol science questions such as aerosol-climate interactions and visibility applications.

1.1.2 $PM_{2.5}$ effects on human health

Surface measurements have been used for epidemiological studies by using regression models to test associations between pollution levels and mortality and morbidity rates. Fine particles have been associated with an increase in respiratory

diseases, cardiovascular disease, lung cancer, and cardiopulmonary mortality [*Franchini, et al.*, 2009*; Pope Iii, et al.*, 2002*; O'Connor, et al.*, 2008*; Schwartz, et al.*, 2000*; Dockery, et al.*, 1993]. A recent study [*Pope, et al.*, 2009] over a 30 year period demonstrated an increase in individual average life span of about one year associated with the decrease in $PM_{2.5}$ concentrations over the three decades. Studies conducted to approach aerosol effects on human health have been beneficial to the health community in general, although these studies have a problem of misrepresenting the true atmosphere. The following section describes a few of these studies.

1.1.3 $PM_{2.5}$ exposure misclassification studies

Surface measurement studies are subject to exposure misclassification because of the scarce spatial and temporal $PM_{2.5}$ data available in the United States. Assigning point measurements to represent ambient concentrations could lead to biases in regression models. Zeger et al. [2000] performed a time series study over three months discussing the errors associated with using site measurements to represent atmospheric concentrations. Health risk estimates were influenced by differences in the average personal exposure and the atmospheric concentration where the difference between individuals and the weighted average personal exposure were minimal. Ito et el. [2004] compared speciated $PM_{2.5}$ over three monitors and showed that the correlation coefficient of source-apportioned $PM_{2.5}$ could vary between different site location measurements. Ito et el. [2004] stated that biases in $PM_{2.5}$ population exposure could be caused by the lack of monitors in certain geographic areas.

Jerrett et al. [2006] studied the relationship between particulate matter and chronic health diseases over ten years, and showed the relative risk of mortality is underestimated by two to three times due to exposure misclassification. In a study examining exposure misclassifications, Samet et al. [2007] emphasized that despite exposure misclassifications occurring due to the use of sparsely distributed $PM_{2.5}$ monitors, $PM_{2.5}$

was also found to have an impact on short term health exposure risks. Upon evaluation of multiple epidemiological studies focusing on exposure misclassifications, Grahame [2009] demonstrated that studies using deficient exposure information actually produced results that underestimated the effects of black carbon on humans. Adar et al. [2007] acknowledged the effects of using central monitors as ambient concentrations in his study of cardiovascular disease due to traffic exposure. Assigning equal exposure to people within 100 meters of the roadway could lead to misclassification. Exposure rates were equally assigned for people who lived within five meters, and exposure rates for people living in areas further than five meters exponentially declined. Another potentially valuable method of inferring $PM_{2.5}$ concentration is satellite remote sensing products explained in section 1.1.5

1.1.4 CMAQ relation to health studies

Chemical transport models (CTMs) provide alternate concentration estimates for fine particles, with more complete spatial and temporal coverage and the advantage of explicit representation of speciated PM components. Studies with the Community Multiscale Air Quality model (CMAQ) have shown that particulate matter model-measurement agreements are variable and that model estimates are associated with biases. Examples of surface measurement networks used in the current study are the Interagency Monitoring of Protected Visual Environments (IMPROVE) and Speciation Trends Network (STN) [*U.S. Environmental Protection Agency (EPA),* 2005*; Spak et al.,* 2009*; Yu et al.,* 2008]. *Marmur et al.* [2006] compared source apportionment results from CMAQ with receptor-based model results to evaluate source-resolved exposure estimates for use in health studies. *Marmur et al.* [2006] concluded that model-based estimates have higher spatial resolution but show limited skill at reproducing day-to-day observed variability. *Russell* [2008], in a review of multiple studies [*Yu et al.,* 2005*; Seigneur et al.,* 2004*; Tesche et al.,* 2006*; Boylan and Russell,* 2006], stated that typical model skill

was quantified at a normalized error of 25-50% using daily average observation-model pairs. Nitrate seasonal biases and organic carbon underestimation were identified as major sources of error. *Appel et al.* [2011] used WRF and CMAQ v4.7 to evaluate multiple species including $PM_{2.5}$ for North America (including portions of Canada and Mexico) and Europe for 2006 and showed comparable errors and biases to both previous studies and this work. Although models demonstrate skill at a variety of spatial and temporal scales, improvements in modeled concentrations are still needed to increase model performance to the level required for many applications. These improvements may come through better model mechanisms, more refined input data, or observational constraint. Tagaris et al. [2009] studied the impact of changing CMAQ $PM_{2.5}$ estimates on decreasing air quality pollutants in urban and rural areas over 50 years. The simulation was conducted over the United States, starting from 2001 at a resolution of 4° latitude by 5° longitude. Mauzerall et al. [2007] changed $PM_{2.5}$ concentrations over the continental United States for January and July of 1996 at a 36km model resolution using CMAQ. Model results showed an increase in mortality rates by a factor of ten with increases in human exposure.

Studies have shown substantial discrepancies between model output and surface measurement networks such as IMPROVE and STN [EPA, 2005, Spak, et al., 2009, Yu, et al., 2008]. Although models can generally estimate both seasonal and temporal values up to one hour, improvements are still needed to increase model performance.

1.1.5 Satellite data

Given the limited spatial coverage of $PM_{2.5}$ monitoring, researchers have been evaluating the feasibility of satellite remote sensing, geospatial analytical methods, and chemical transport models to derive high-resolution $PM_{2.5}$ estimates. $PM_{2.5}$ concentration can be inferred at high spatial resolution with the aid of satellite data. Among many sensors aboard different satellites, Moderate Resolution Imaging Spectroradiometer

(MODIS) instrument [Levy, et al., 2009] aboard Terra and Aqua has been particularly useful in developing high resolution estimates of Aerosol Optical Depth (AOD) and deriving PM concentrations [*Li et al,* 2005*; Kumar et al.* 2011]. MODIS became operational on the Terra satellite in early 2000 and on the Aqua satellite in mid-2002. Radiative forcing models are applied to retrieve AOD, a dimensionless number that quantifies extinction of light in the atmospheric column [*Levy et al.,* 2009]. AOD values higher than 1.0 correspond to low visibility and high PM levels caused by fires, sand storms, or extreme urban air pollution. Studies have shown that MODIS column AOD is correlated with surface $PM_{2.5}$ concentrations [*Kumar et al.,* 2008*; Schaap et al.,* 2009*; Engel-Cox et al.,* 2005]. Several groups have taken advantage of this correlation to produce $PM_{2.5}$ exposure estimates based solely on satellite data [van Donkelaar, et al., 2010; Al-Hamdan, et al., 2009; Wu, et al., 2006] that could relate to health studies. Land use regression methods are also employed for $PM_{2.5}$ exposure estimation [*e.g. Liu et al.,* 2009]. These studies use statistical models trained on $PM_{2.5}$ measurements and land use data.

1.1.6 Data assimilation

Data assimilation is a valuable tool for improving forward model state towards a better estimate of the prior model values. Data assimilation [*Kalnay,* 2002] combines observations with data to generate a posterior dataset sharing features of the background and the observations. The background data are hence referred to as the model data in the current work. Observational data can be ground, satellite, or aircraft measurements, in addition to remote sensing other than satellite (e.g. Lidar and Doppler data).

It is important to understand data assimilation methods and their differences in order to determine which method to apply for the current work. Optimal interpolation (OI) [*Lorenc,* 1981] is a multivariate statistical method that is computationally efficient and is applicable to atmospheric or numerical forecast models. OI is one of many data

assimilation methods, which include nudging, multivariate statistical methods, 3D-var, evolving forecast error covariance, and balancing the initial conditions. In addition to these aforementioned methods, other advanced data assimilation methods include ensemble and extended Kalman filter as well as 4D-var. The methods mentioned above vary by their computational burden and the quality of their results. Some are computationally efficient with satisfactory or even substantial results, while others are very computationally intensive and lead to better results. Following is a brief summary of the differences between the methods mentioned above and an explanation of the rationale for choosing OI as the data assimilation method in this study.

Nudging [*Hoke and Anthes,* 1976, *Kistler,* 1974], which has a low computational burden, is a data assimilation method that nudges the background towards the observations and is used with meteorological data to improve model predications. Nudging has no theoretical or statistical basis and simply pushes the models towards the observations; its advantage is its ease of use due to its low computational needs.

Optimal interpolation (OI) is more computationally expensive than nudging and includes the spatial errors (weights) for the model and the observations. The posterior value is achieved from the difference between the observations and the model data weighted by the errors and added to the initial state of the model. Thus, due to the way OI is implemented, low computational cost is achieved with high computational capability possible as long as the statistical errors are accurate or accurately constrained.

Kalman filter [*Kalman,* 1960] is similar to OI with a minor difference in the model error covariance calculation; OI calculates a constant estimate whereas Kalman filter uses the model to calculate the error covariance matrix. To achieve this, the model error is assumed to be unbiased, and the model linear; however, assuming the model error is unbiased could cause unrealistic results or add error to the model. Kalman filter, which has proven to provide similar results as OI in some cases [*Flemming et al.,* 2004], has a higher computational expense than OI.

Another method of estimating the true posterior value is the variational cost function approach [*Sasaki,* 1970], which changes the initial state of the model instead of adjusting the posterior value. This is done by adjusting a control variable with respect to which the cost function is minimized. The control variable can be any physical variable of the system, such as initial conditions or boundary conditions, or any model parameter (i.e. concentration, temperature, pressure, boundary layer, extinction coefficient, and etc.). The variational method has two parts; one minimizes the difference between the observation and prior model value, and the other minimizes the difference between the prior and posterior value. Both are weighted, respectively by the inverse of the observational and by the model error covariance. The variational cost function approach is used for 3D-var and 4D-var data assimilation. 3D-var [*e.g., Sasaki,* 1970*, Parrish and Derber,* 1992] includes a variational cost function that has a three dimensional field. Per the definition above, the distance between the analysis to the model and observations are weighted by the inverse of the error covariances for the background and observations, respectively. 3D-var has a more accurate posterior state than OI, which could be due to the fact that 3D-var uses all the observations in the domain simultaneously, whereas OI creates windows with confined sizes that restrict the observations to the windows. In other words OI restricts the effect of any observation on all the grid cells of the domain. Also, the model error covariance matrix in OI is assumed to be constant, unlike 3D-var, where the model covariances are estimated from the average of multiple analysis-model differences from short range data assimilation forecasts. Furthermore, 3D-var propagates adjustments in the control variables according to the linear or nonlinear transport and chemistry of the system, rather than through only statistical relationships as OI. 3D-var does not account for variation of observation over time, which is included in 4D-var.

3D-var (as well as some other methods) can define the model error covariance in a general approach that differs from localized OI calculation, using more global methods, which are described as follows for sparse and dense observational networks. The

National Centers for Environmental Prediction (NMC) approach, [*Parrish and Derber,* 1992] scales the magnitude of the covariance by acquiring multiple error samples over the domain, which could be defined as the model bias at each available site (model–observation) corrected by the averaged model bias over the domain. This method is appropriate for networks that have less dense observational data. For networks that have a dense network of observation through the domain, the observational method [*Hollingsworth and Lonnberg,* 1986] is used to calculate a more dependable estimate of the model error covariance. The observational method is explained in detail in the method section.

Advanced data assimilation methods are computationally expensive, but have the possibility to provide better posterior products. Assuming that the error covariance matrixes are constant over time could add additional errors to the data assimilation model; this is where more advanced methods excel. 4D-var, Ensemble or extended Kalman filter calculate the change of error over time using an adjoint model, which is a linear tangent model, to advance a perturbation backwards from the final time to the initial time. Extended Kalman filter [*Jazwinski,* 1970] uses an adjoint model to forecast the evolved error covariance matrix that is used in the analysis step. If a sufficient number of observations are available over time, extended Kalman could provide best assimilation results compared to the less computational demanding methods, even if the model starts with a poor initial guess. Unfortunately, updating the model errors over time using the adjoint model adds high computational cost to the model and in many cases is replaced with the simple constant error, due to time restrictions. Therefore, extended Kalman filter is one of the most expensive computational methods [*Kalnay,* 2002]. Ensemble Kalman filtering [*Houtekamer and Mitchell,* 1998] runs an ensemble of simultaneous data assimilation cycles independently, which reduces the extended Kalman filter computational time. 4D-var is an extension of 3D-var data assimilation over time, 4D-var includes a summation of the difference between observation and the prior model

value over time with its initial state, which uses adjoint model as described above. Finally, 4D-var [*e.g., Lewis and Derber,* 1985, *Courtier and Talagrand,* 1990] is one of the most efficient data assimilation methods. It has a higher computational expense compared to OI, but is computationally less expensive compared to Kalman filter. OI is a low computational data assimilation method that shares similar advantages as more advanced methods to inform model estimates using observation data and incorporation of error information for both model and observations.

1.1.7 Data assimilation studies

Data assimilation methods have been used to improve forecasting and model estimates using surface measurements. Nudging is mostly used to improve meteorological forecasts with surface measurements, where improving meteorology is one of the crucial steps for a good atmospheric model outcome. Multiple studies have shown that nudging has managed to improve simulation of atmospheric boundary layer [*Alapaty et al.* 2001], soil moisture [*Pleim and Xiu,* 2003], and surface wind [*Stauffer and Derber* 1990] forecasts. The Weather Research and Forecasting (WRF) model uses "observational-nudging" [*Liu et al.,* 2005] for fine grid multi scale domains. Flemming et al. [*2004*] compared OI to a Kalman filter implementation by combining surface ozone and nitrogen dioxide measurements with the REM/CALGRID [*Stern et al.,* 2003] chemical transport model, over central Europe for July 2001. Both OI and Kalman filter results were comparable, showing a similarity for the assimilation of ozone concentrations, and a better performance for Kalman filter when assimilating nitrogen dioxide.

3D-var has been used to improve both meteorological predictions [*Le Dimet et al.,* 1986*; Talagrand et al.,* 1987] and air quality model estimates. Liu et al. [*2011*] used 3D-var to assimilate MODIS AOD and modeled derived-AOD from the coupled online Weather and Research and Forecasting-Chemistry model WRF-Chem to improve PM_{10}

model estimates over East Asia for March 2010. The control variables used to minimize the cost function included 14 aerosol species for the entire horizontal and vertical domain. The aerosol species include hydrophobic and hydrophilic OC and BC, sulfate, sea salt in four particle sizes, and dust particles in five particle sizes. Comparing posterior model values with surface PM_{10} measurements showed that the assimilation gave a modest improvement over variable days of the month.

Pagowski et al. [2010] used 3D-var to improve the air quality forecast for both ozone and $PM_{2.5}$ by assimilating surface measurements using WRF-Chem in the north east of the United States during August and September 2006. The control variables used to minimize the cost function were ozone and $PM_{2.5}$ boundary conditions and multiple meteorological data (i.e. temperature, pressure, relative humidity, and velocity potential). The NMC method was used to calculate the model error covariance. The assimilation improved ozone and fine particles for a 24hr forecast. Bei et al. [2008] used 3D-var assimilation to improve ozone simulation over Mexico City, from the Comprehensive Air Quality Model with extensions (CAMx) [*ENVIRON,* 2006], by assimilating predicted wind circulation, temperature, and humidity. The control variables used to minimize the cost function were wind and ozone to improve the model estimate. The NMC method was used to calculate the model forecast errors. The research provided an insight into how 3D-var assimilation could provide improved posterior ozone over a four day episode in April 2003.

Mallet et al. [2007] described the air quality model Polyphemus that contains four data assimilation procedures, including OI, ensemble Kalman filter, reduced-rank square root Kalman filter [*Heemink et al.,* 2001], and 4D-var. The reduced-rank square root method is used to solve Kalman filter problems by reducing the covariance matrix rank, and is known as variance reduction. The ensemble Kalman filter method improves forecast and the chemical transport model concentrations. Hanea et al. [2004] used Kalman filter and the reduced-rank square root Kalman filter to assimilate ground level

ozone from the chemical transport model EUROS [*van Rheineck Leyssius et al.,* 1990*; van Pul et al.,* 1996*; Matthijsen et al.,* 2001] over Europe for June 1996. The control variables used were ozone measurements and deposition rate and NOx and VOCs emissions. Surface ozone data assimilation results were more accurate than the forward model.

Hakami et al [2005] used 4D-var to assimilate black carbon over East Asia in April 2001 using the chemical transport model STEM-2k1 [*Carmichael et al.,* 2003] and observational data. The control variables used to minimize the cost function were the anthropogenic and biomass burning black carbon emissions, the boundary conditions, and the initial conditions. Thus posterior black carbon was driven by scaling the input to the chemical transport model. Ten observation sites and platforms were available; six sites were used for data assimilation and two aircraft measurements were used for evaluation of posterior model results. 4D-var managed to significantly improve the posterior model values. Carmichael et al. [2007] gave direction for the use of 3D-var and the advanced data assimilation methods to improve meteorological model predictions and air quality model estimates, showing the advantages and challenges of using computationally expensive data assimilation methods.

1.1.8 Satellite and surface optimal interpolation studies

Optimal interpolation was first used for $PM_{2.5}$ to assimilate the Model of Atmospheric Transport and Chemistry (MATCH) and satellite AOD over the Indian Ocean [*Collins et al., 2001*]. The assimilation improved the model AOD bias, as compared with AOD from a sun photometer, from -0.12 to -0.02. *Yu et al.* [2003] assimilated AOD at a global scale in 2001, using MODIS and the Goddard Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) model. AOD estimates were compared to Aerosol Robotic Network (AERONET) AOD, and the coefficient of determination increased when assimilation was applied. A similar approach was used to

assimilate MODIS AOD with the Sulfur Transport dEposition Model (STEM) output for Asia over the period of 2001-2004 [*Adhikary et al.*, 2008]. *Chung et al.* [2010] and *Adhikary et al.* [2008] combined monthly modeled AOD and monthly MODIS and AERONET AOD over Asia. *Park et al.* [2011] performed data assimilation over East Asia using CMAQ AOD and MODIS AOD.

Satellite measurements could be affected by many factors, such as cloud contamination or surface reflectance retrieval errors due to satellite dependence on land cover and land use (e.g. snow cover and bright surfaces). So regardless of the satellite's spatial and temporal availability, the valid retrieval could add limited aerosol information that depends on the satellite's spatial location and seasonal retrieval. Surface measurements are sparse locations with limited temporal availability ranging from 3 to 7 days for daily average measurement. In spite of its limitations, it could provide valuable aerosol information at and around the site location, whether temporally averaged or not. The following cites multiple data assimilation methods that combine both surface and model data that have shown improvements in model estimates. Optimal interpolation could be a viable tool for assimilating surface and model data due its low computational cost and effective results.

Tombette et al. [2009] applied the optimal interpolation method of Collins et al. [2001] using surface PM_{10} measurements and aerosol concentration fields from the chemical transport model Polair3D. The assimilation was over Europe for January [2001] and was carried out at each time of measurement availability. The assimilated model estimates were used as initial conditions for the model. So, if the measurements occurred every three days, then data assimilation was done at the time of the measurement, and the posterior estimate was used as the initial condition to start the model, which would run for three days until the next measurement. The Balgovind approach [*Balgovind et al,* 1983] was used to calculate the model error covariance matrix. Tombette et al. [2011] used three observational networks to perform data assimilation and kept two networks for

data validation. Compared to one of the networks, the results showed an increase in correlation from 62% to 88%, and a decrease in RMSE from 12.5 to 8.4 ug.m^{-3}.

Wu et al. [2008] compared four data assimilation methods: optimal interpolation, reduced-rank square root Kalman filter, ensemble Kalman filter, and 4D variational assimilation. The study improved ozone forecasting using the surface measurements and the chemical transport model Polyphemus [*Mallet et al.,* 2007] over Western Europe for a 12 hour forecast in early July 2001. The assimilation was conducted by using two grid resolutions. The full resolution, with a 0.5 degree grid cell size, assimilating and forecasting for subsequent days between July 1st and 8th. The coarse resolution, with a 2 degree grid cell size, assimilating ozone between July 1st and 2nd for a July 2nd and 3rd forecast. The model error covariance matrix was also based on the Balgovind approach [*Balgovind et al,* 1983] for both OI and the initial data of the assimilation period for 4D var. The results show that all four data assimilation methods improved the ozone forecast when compared to the forward model. OI delivered the best ozone forecasting improvement over the four methods, which is due to the stability of the model errors calculated by the Balgovind approach [*Balgovind et al,* 1983].

The current work is motivated by the ability to conduct exposure studies over the U.S due to the availability of mortality and morbidity data. Compared to other regions or countries the dense surface $PM_{2.5}$ measurement network gives us the ability to conduct surface OI and provide a thorough model evaluation. This thesis adapts OI for data assimilation over the U.S. in 2002, for both satellite and surface measurements, motivated by the low computational needs of the method, the relatively modest requirements for model and observation error statistics, and the documented successes of the method in the literature. This work is different from previous satellite OI studies because it investigates different model-observation averaging methods for four error values. The surface OI will be different from previous studies that incorporate the observation method to calculate model error covariance and a representation error for the influence of an urban or rural

cite in the calculation of observation error covariance. Also a simple cross validation method will be used for satellite OI, compared to a more comprehensive cross validation method used for surface OI, which is considered novel for chemical transport OI evaluation. Also cross validation methods were applied for validation, which have not been used for assimilated chemical transport model validation before. These methods were adapted from regression model studies. The ability to provide accurate estimates of hourly $PM_{2.5}$ concentrations over the United States for long periods would be a significant benefit when relating $PM_{2.5}$ concentrations with epidemiological studies.

1.2 Overview of thesis

The current study includes assimilation of satellite and surface measurements with model data over the United States for the year 2002. Chapter 2 describes the detailed objectives of the current work. The methodology used in the current work has been included in the respective chapters of the thesis. Chapter 3 describes CMAQ model implementation to produce 3D $PM_{2.5}$ data speciated hourly over the United States in 2002. A description of the model-dependent emissions and meteorological data are included. Model estimates were compared and evaluated with surface $PM_{2.5}$ speciated measurements. Chapter 4 describes satellite data assimilation methods (including mapping procedures), optimal interpolation of satellite data with model AOD, annual evaluation of model values with surface $PM_{2.5}$, and comparison with previous studies. Chapter 5, which describes surface data assimilation methods and results, includes: the observational method and surface OI equations; cross validation methods; results from the observational method, OI case study for January, OI annual evaluation, and surface OI sensitivity study; and discussion and conclusions. Chapter 6 summarizes the thesis and provides recommendations for future work.

CHAPTER 2: OBJECTIVES

The purpose of this thesis was to improve fine particulate matter ($PM_{2.5}$) estimates from the Community Multiscale Air Quality (CMAQ) model that are able to simulate pollutants at high spatial and temporal resolutions using optimal interpolation (OI).

To accomplish this multiple objectives are needed:

Quantify how well OI works with MODIS retrieval over the Unites States. This is accomplished by comparison of CMAQ $PM_{2.5}$ output with $PM_{2.5}$ surface observation measurements for both rural and urban areas. CMAQ $PM_{2.5}$ output was simulated for the year 2002 to obtain hourly aerosol concentrations.

Propose a best OI strategy for U.S and explain the results. This is accomplished by comparing MODIS AOD valid retrievals with AOD surface observations from AERONET sites. The comparison is expected to give an understanding why OI works in certain regions and months based on how well MODIS is compared to AERONET, where AERONET is considered to be a more accurate measurement than MODIS.

Quantify how well OI works with MODIS retrieval at a finer spatial resolution over the United States. In this analysis MODIS 10km data is replaced with retrievals at 2km resolution. MODIS retrieval at a finer resolution eliminates errors caused by cloud contamination and misrepresentation of water-land areas. The quality assurance flag associated with MODIS retrievals reflects how well a valid retrieval can be trusted for data processing.

Quantify how well OI works with surface measurements over the United States. In this analysis MODIS 10km data is replaced with surface measurements. The use of surface $PM_{2.5}$ data will eliminate the uncertainty associated from both satellite retrievals and calculations that enable the model and satellite data to be quantified.

CHAPTER 3: PERFORMANCE OF CMAQ PM SIMULATION - FORWARD MODEL

The chapter gives an overview of CMAQ, followed by the description of model input data that includes both emissions and meteorology data. This is followed by description of the following work and model results. CMAQ 2002 aerosol concentrations within the United States domain were evaluated on a seasonal basis dividing the country into six regions (Figure 2). The seasons were defined as winter (December-February), spring (March-May), summer (June-August), and fall (September-November) and regions are according to U.S. Census divisions [*U.S. Census Bureau*, 2000]. $PM_{2.5}^{cmaq}$ performance was evaluated using MFB and MFE (equations 9 and 10).

3.1 CMAQ overview

CMAQ [*Byun, et al.*, 2006] was developed in 1999 to simulate atmospheric multi-pollutants at high time resolutions at both regional and urban scales. This atmospheric chemical transport model simulates ozone, acid deposition, visibility, and fine particulate matter throughout the troposphere. Figure 3.1 is the schematic diagram of the main models and processors utilized by the CMAQ model. Meteorology and emission data are the key elements that are translated through CMAQ's chemical transport model (CTM). CTM integrates CMAQ's governing equations, transport algorithms, gas phase chemistry, and the chemistry and dynamics of both aerosols and clouds. The Meteorology-Chemistry Interface Processor (MCIP) translates the meteorology data that is needed by the emission model and CTM. The processor also computes cloud parameters, surface and planetary boundary layer (PBL), and coordinate system transformation.

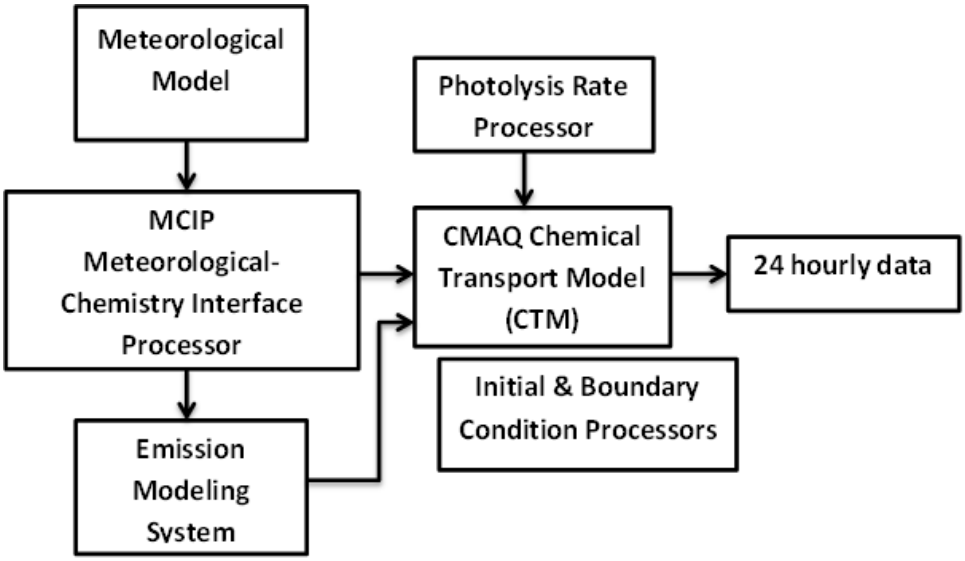

Figure 3.1 Schematic diagram of the CMAQ modeling inputs.

Initial and boundary conditions are required as individual concentrations for the model grids and surrounding domain. CMAQ is accompanied by clean-troposphere vertical profiles for use as boundary and initial conditions, but can also use previous modeling data from other chemical transport models. The photolysis processor creates lookup tables of photo-dissociation reaction rates that are required by CTM. CMAQ is also supported with surface site compare program (sitecmpV1.0) and visualization tools that can be used for post-analysis of hourly gridded pollutant concentrations.

3.2 Aerosol model in CMAQ

Binkowdki et al. [2003] described the aerosol model component in CMAQ, which is explained briefly. Aerosols are small liquid and solid particles in the atmosphere, where the particles modeled are of the size 10μm and smaller. Representation of particles in the CMAQ includes total number concentration, total mass concentration, and size distribution. The aerosol model is represented by three lognormal subdistributions called "modes". Two modes are called fine modes and represent particles with diameters less

than or equal to 2.5 μm. One of the two fine modes is the Aitken mode, which represents particles created from nucleation or from direct emissions, and another mode which is for larger particles called the accumulation mode, which represents aged particles. Particles with the size range between 2.5μm and 10μm are called coarse particles ($PM_{2.5-10}$), which is the third mode in the dynamic aerosol model. Coarse mode particles are primary composed of sea salt and wind-blown dust. The aerosol particle diameter is a function of the geometric mean diameter and the geometric standard deviation, which are 0.03μm and 1.7 for the Aitken mode, 0.3μm and 2.0 for the accumulation mode, and 6μm and 2.2 for the coarse mode respectively.

3.3 National emission inventory

The national emission inventory (NEI) [references are found from the EPA at http://www.epa.gov/ttn/chief/net/2002inventory.html] is a government sustained emission rate database for the Unites States and neighboring countries (i.e. Canada and Mexico). Only the emission rates for areas in the United States are collected and recorded by the U.S. environmental protection agency (USEPA). NEI records emission rates based on emission factors generated from source emission measurements per level of activity (e.g. annual production rates). The inventory covers all criteria air pollutants (CAPs) which include PM_{10}, $PM_{2.5}$, sulfur oxides, volatile organic carbon, nitrogen oxides, carbon monoxide and lead. Also the inventory covers hazards air pollutants (HAPs), where lead is included in both CAP and HAP lists. Emissions are categorized differently depending on source types. 2002 NEI contains point, non-point, mobile, biomass burning, and fugitive dust sources. Emissions for both Mexico and Canada are available and provided from the respective governments.

Point sources are specifically identified by sector location such as electrical generating units, forest fires, and fugitive dust. Non-point sources are identified by area sectors such as forest fires not included in point sources, agriculture burning, and fugitive

dust. Mobile sources are collected from on-road mobile emissions from vehicle emissions and non-road mobile emissions identified by area are collected from aircraft, locomotive and commercial marine emissions.

3.4 Observation data

Filter-based measurements of $PM_{2.5}$ from IMPROVE and STN (24-hour averages collected every three and six days, respectively) were used to evaluate the model and select favorable assimilation settings. IMPROVE monitors (n=133) are predominantly located in remote and rural areas, while STN monitors (n=128) are located mainly in populated areas (Figure 3.2).

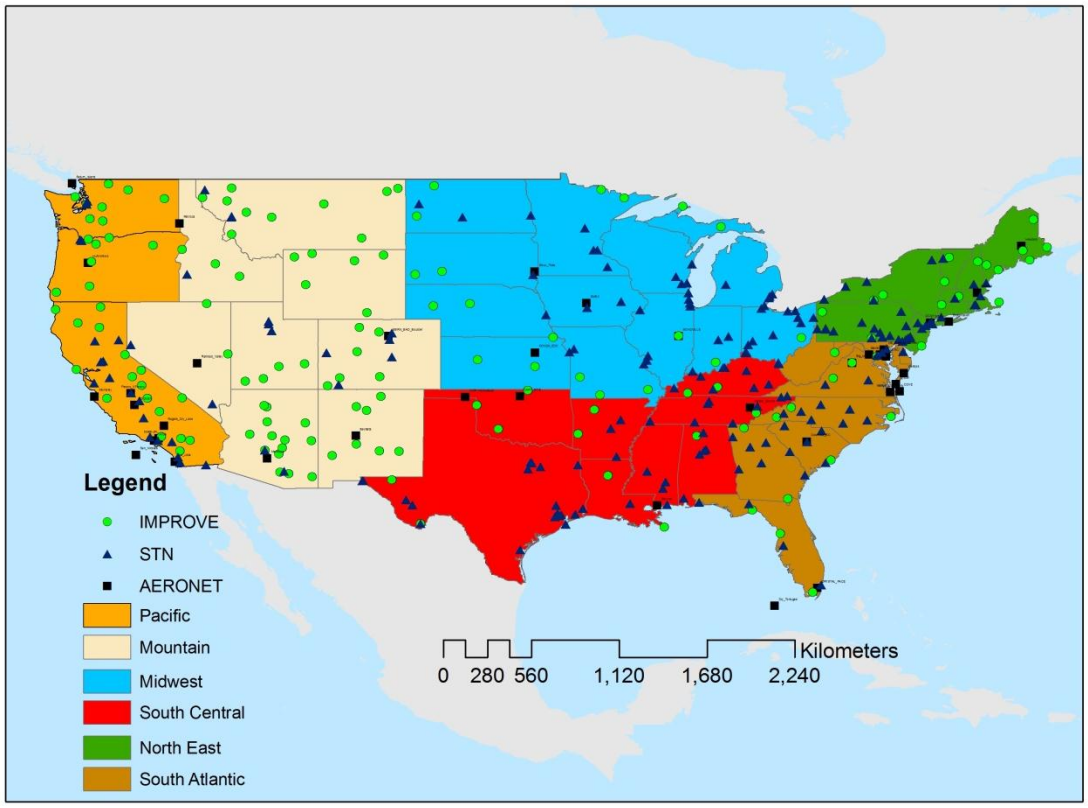

Figure 3.2 IMPROVE, STN, and AERONET monitoring sites over the continental United States for 2002 with regions used in the study shown.

3.5 Chemical transport model

CMAQ version (v4.6) [*Byun et al.,* 1999] output are hourly speciated $PM_{2.5}$ concentrations at 36 km horizontal resolution over the continental United States, with 14 vertical layers from the surface to the stratosphere. Model settings included the Carbon Bond (CB05) chemical mechanism [*Sarwar et al.,* 2008] and advection by the piecewise parabolic method. Aerosols were modeled using the CMAQ AE4 [*Binkowski and Roselle,* 2003] and ISORROPIA (v1.7) aerosol thermodynamics [*Nenes et al.,* 1998]. The simulation was carried out for all of 2002, with a ten-day spin up time from December 2001. Clean continental background initial and boundary condition vertical profiles were used [*USEPA*, 1999].

Emissions for the model were constructed using the Sparse Matrix Operating Kernel for Emissions (SMOKE) [*Houyoux et al.,* 2000] (v2.3) and the USEPA National Emission Inventory (NEI) 2002 [*U.S Environmental Protection Agency (EPA),* 2006]. Biogenic emissions were calculated from modeled meteorology using BEIS3 (v3.12). Wildfire emissions [*U.S Environmental Protection Agency (EPA),* 2008] were based on daily point locations and emission values for wildfires, managed burning, and prescribed forest burning, and non-point locations from average fire sectors based on county total emission estimates for managed burning and prescribed forest burning. When a fire is not available at spatial and temporal resolutions that are required to be included in the point fire inventory, it is categorized under non-point fires [*U.S Environmental Protection Agency (EPA),* 2008]. Meteorological data were provided by the Lake Michigan Air Directors Consortium [*Baker,* 2004] from the Mesoscale Meteorological Model (MM5) (v3.6.1) described by *Grell et al.* [1995, *available at http:// www.mmm.ucar.edu/mm5/documents/mm5-desc-doc.html*]. The Meteorology-Chemistry Interface Processor (MCIP) (v3.3) was used to prepare meteorological data for CMAQ and SMOKE.

3.6 Performance metrics

$PM_{2.5}^{cmaq}$ and $PM'_{2.5}$ were compared to surface measurements from IMPROVE and STN networks. Metrics used to characterize model skill include mean fractional bias (MFB), mean fractional error (MFE), root mean square error (RMSE), and correlation coefficient. Following *Morris et al.* [2005] four performance categories are used in this work: *excellent* [|MFB| <0.15 and MFE < 0.35]; *good* [|MFB| < 0.30 and MFE < 0.50]; *average* [|MFB| < 0.60 and MFE < 0.75]; *problematic* [|MFB| > 0.60 or MFE > 0.75].

The equations for MFB and MFE are [*Boylan et al.*, 2006]:

$$MFB = \frac{1}{N}\sum_{i=1}^{N}\frac{(C_m - C_o)}{\left(\frac{(C_m + C_o)}{2}\right)} \tag{3.1}$$

$$MFE = \frac{1}{N}\sum_{i=1}^{N}\frac{|C_m - C_o|}{\left(\frac{(C_m + C_o)}{2}\right)} \tag{3.2}$$

And Also the Normalized Mean Error (NME) was calculated for Surface OI evaluation

$$NME = \frac{\sum_{i=1}^{N}|C_m - C_o|}{\sum_{i=1}^{N}C_o} \tag{3.3}$$

where C_m and C_o are the model and observation concentration at site i respectively, and N is the number of paired model and observation concentrations. The ratio of MFE to a target MFE, $MFE_{target} = \pm[C_1 \exp(-0.5(\overline{C}_o + \overline{C}_m)/C_2) + C_3]$, with 1.5, 0.75, and 0.6 for C_1, C_2, and C_3, respectively [*Morris et al.*, 2005], is defined as the Fractional Error Score (FES) in this paper. It is used as the criteria for the selection of successful OI schemes and parameter settings. The target MFE is taken from *Boylan et al.* [*2006*]. The

target MFE is 0.6 at high concentrations, and then relaxes to larger errors at $PM_{2.5}$ concentrations below ~4 μg m^{-3} to reflect the anticipated high fraction errors under such conditions.

3.7 $PM_{2.5}$ performance evaluation

Table 3.1 lists the results of $PM_{2.5}$ performance evaluation for each season and region. Although the results of the current work apply best to the specific model configuration used (e.g. 36 km resolution, CMAQ v. 4.6, MM5 meteorology), evaluation of the model results suggests comparable prior model skill to a number of North American CMAQ implementations, and the conclusions about data assimilation should apply broadly to a wider range of model configurations. Month-region specific evaluation show lower fractional bias and fractional error at STN sites than at IMPROVE monitoring stations. At STN sites the model performance was better in the eastern U.S. than in the west. $PM_{2.5}^{cmaq}$ is underpredicted in summer in most regions, resulting in some "problematic" performance classifications relative to IMPROVE. In fall, performance generally falls into good and average categories.

Negative $PM_{2.5}$ biases were widespread in most months of the year due to organic carbon and nitrate underprediction. Organic carbon model performance in summer was characterized by negative biases for the whole United States except for IMPROVE monitors in the Mountain region. The most significant organic carbon negative biases occurred in summer in the Northeast region, and ranged from about -3.5 to -5.5 μg m^{-3}. A significant negative bias (-6.2 μg m^{-3}) occurred in the Pacific in November at STN sites. Organic carbon biases could be due to the underestimation of secondary organic aerosol, for example from biogenic isoprene oxidation [*Yu et al.*, 2007]. Nitrate is underpredicted in summer for the United States with exception of the Midwest. Nitrate underprediction in the western U.S. was also significant for majority of the year at urban locations. The most significant nitrate biases were a -6 μg m^{-3} bias in the Pacific (November, STN), and

+2.1 μg m^{-3} in the Northeast (March, IMPROVE). Differences between measured and modeled nitrate aerosol could be potentially caused by both measurement errors and model errors in total ammonia and sulfate [*Yu et al.,* 2008]. Sulfate performance was categorized as average throughout the year in all the regions. Both elemental carbon and ammonium simulations had excellent performance in various regions and seasons, although their influence on the model concentrations is mostly low.

To verify that the forward model runs used as the input for OI are representative of typical skill in regional model studies, comparisons to previously published work were performed. *Spak et al.* [2009] and the current work fall within one performance category of one another for all seasons; there are two instances of better performance in *Spak et al.,* and two instances of better performance for the current work. *Eder et al.* [2006] evaluated CMAQ's annual performance over the United States for 2001 both spatially and temporally, and reported temporal correlation coefficients of 0.5 for STN sites and 0.7 for IMPROVE sites; throughout the domain the bias ranged between ±25% for both networks. *Yu et al.* [2008] modeled at 12 km horizontal resolution summer 2004 $PM_{2.5}$ in the eastern United States (equivalent to the Northeast and South Atlantic domains in the current work) and reported $PM_{2.5}$ biases in July/August of -2.6 (STN) and -5.1 μg m^{-3} (IMPROVE), compared to -6 and -6.6 μg m^{-3}, respectively, in the current study. The studies listed above are consistent with the *Russell* [2008] review of model skill (typical model skill of normalized error of ~25-50%, using daily average observation-model pairs). More recent evaluation [*Appel et al.,* 2011], incorporating the latest version of CMAQ (v4.7), is also consistent with seasonal mean bias and mean error ranging between -0.60 to 4.00 μg m^{-3} and 4.4 to 6 μg m^{-3}, respectively.

Table 3.1 $PM_{2.5}^{cmaq}$ performance at both IMPROVE and STN networks based on Morris et al. metrics[1].

	IMPROVE								STN							
	Winter			Spring			Summer			Fall			Winter			Spring			Summer			Fall		
	FE	FB	PC	FE	FB	PC	FE	FB	PC	FE	FB	PC	FE	FB	PC	FE	FB	PC	FE	FB	PC	FE	FB	PC
Pacific	0.77	0.47	P	0.57	-0.08	A	0.62	-0.37	A	0.57	-0.08	A	0.64	-0.32	A	0.53	-0.19	A	0.57	-0.42	A	0.62	-0.36	A
Mountain	0.51	0.15	A	0.71	-0.62	A	0.91	-0.88	P	0.55	-0.26	A	0.58	-0.47	A	0.51	-0.32	A	0.61	-0.56	A	0.52	-0.37	A
Midwest	0.47	0.28	G	0.51	-0.29	A	0.73	-0.71	A	0.48	0.16	G	0.40	0.24	G	0.38	-0.04	G	0.39	-0.28	G	0.40	0.20	G
South Central	0.52	0.09	A	0.72	-0.58	A	0.89	-0.87	P	0.60	-0.07	A	0.47	0.10	G	0.60	-0.42	A	0.63	-0.53	A	0.48	-0.05	G
NorthEast	0.72	0.67	A	0.41	-0.05	G	0.60	-0.54	A	0.49	0.17	G	0.47	0.39	G	0.3	-0.05	E	0.47	-0.34	G	0.41	0.07	G
SouthAtlantic	0.50	0.12	A	0.52	-0.35	A	0.72	-0.65	A	0.49	-0.13	G	0.44	0.08	G	0.44	-0.23	G	0.60	-0.56	A	0.45	-0.16	G

[1]Fractional Error (FE), Fractional Bias (FB), and PC (Performance Classification) **E**xcellent, **G**ood, **A**verage, and **P**roblematic.

3.8 $PM_{2.5}$ speciated performance evaluation

CMAQ $PM_{2.5}$ speciation monthly performances at both IMPROVE and STN sites in 2002 are shown in Figure 3.3. Where the annual performance for both networks shows that the model is biased low compared with the observations.

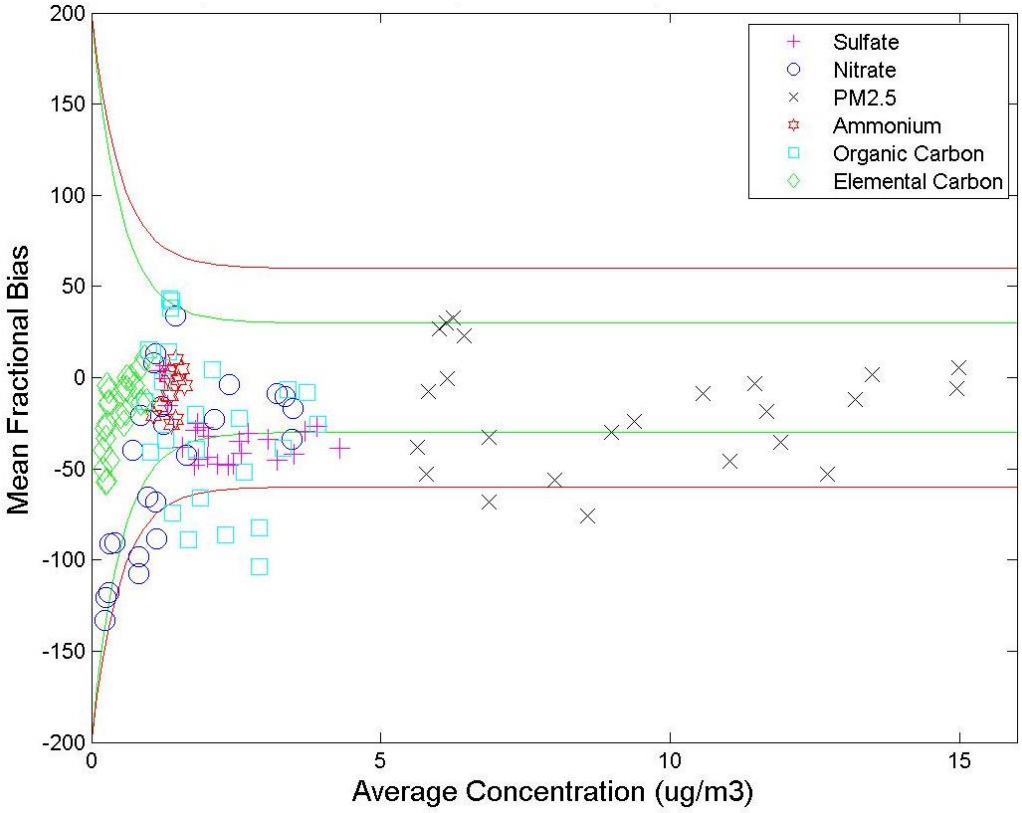

Figure 3.3 CMAQ $PM_{2.5}$ Speciated (IMPROVE and STN) CMAQ Monthly performance 2002. MFB less than or equal to ±30% are considered good, MFB larger than ±30% and less than or equal to ±60% are considered average, and MFB larger than ±60% are considered problematic.

3.8.1 Sulfate

Table 3.2 shows evaluation of sulfate based on monthly performances at both IMPROVE and STN networks for each region. Comparing model to both IMPROVE and STN networks, CMAQ has a good performance in winter and an average performance in spring and summer. In fall model performance varies from good to average. Only in summer at the Pacific, CMAQ has a problematic performance in STN sites. Sulfate has an above average performance throughout the year, where certain biases could be due to the chemical equilibrium partitioning of sulfate, aqueous sulfuric acid and gaseous SO_2 [*Spak, et al.,* 2009*; Yu, et al.,* 2005].

Table 3.2 Sulfate performance at both IMPROVE and STN networks based on Morris et al. metrics

	IMPROVE					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	Average	Good	Excellent	Good	Good	Good
Spring	Excellent	Average	Average	Average	Average	Average
Summer	Good	Average	Average	Average	Average	Good
Fall	Excellent	Good	Good	Good	Average	Good
	STN					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	Good	Average	Good	Good	Good	Good
Spring	Average	Average	Average	Average	Average	Average
Summer	Problematic	Average	Excellent	Average	Good	Good
Fall	Average	Average	Good	Good	Average	Good

3.8.2 Nitrate

Table 3.3 shows evaluation of nitrate based on monthly performances at both IMPROVE and STN networks for each region. With the exception of the Midwest

region, CMAQ is problematic in summer for the whole United States. The model performance at STN sites for the central and east parts of Unites States are excellent to average, but are problematic in the Pacific and Mountain regions. Modeled nitrate has a better performance in urban areas from rural parts. Differences between measured nitrate aerosol and modeled nitrate aerosol could be potentially caused by both measurement and model errors on total ammonia and sulfate [*Yu et al.*, 2005].

Table 3.3 Nitrate performance at both IMPROVE and STN networks based on Morris et al. metrics

	IMPROVE					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	Good	Problem	Excellent	Good	Problem	Good
Spring	Average	Average	Good	Problem	Good	Problem
Summer	Problem	Problem	Problem	Problem	Problem	Problem
Fall	Average	Average	Good	Good	Good	Average
	STN					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	Problem	Problem	Excellent	Excellent	Average	Excellent
Spring	Average	Problem	Excellent	Average	Excellent	Average
Summer	Problem	Problem	Good	Problem	Problem	Problem
Fall	Problem	Problem	Average	Good	Good	Average

3.8.3 Organic and elemental carbon

Table 3.4 shows evaluation of organic carbon based on monthly performances at both IMPROVE and STN networks for each region. Organic carbon model performance in summer is problematic for the whole United States except for the Pacific and rural areas at the Mountain region. Model estimate is excellent to average for spring and fall, but variable in winter. Table 3.5 shows evaluation of elemental carbon based on monthly

performances at both IMPROVE and STN networks for each region. The model performance is excellent at urban areas in the United States, except for the Mountain region. Compared to IMPROVE sites the model has an excellent to average performance over all regions and seasons.

Organic carbon biases could be due to the underestimation of secondary organic carbon from isoprene oxidation, which is mostly from biogenic emissions. CMAQ has a good performance for elemental carbon compared to both networks for all regions during the year but elemental carbon has small influence on $PM_{2.5}$ concentration. Model representation of elemental carbon could be explained by accurate representation of soot in the emission inventory that is measured at urban areas.

Table 3.4 Organic Carbon performance at both IMPROVE and STN networks based on Morris et al. metrics

	IMPROVE					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	Problem	Average	Average	Excellent	Problem	Excellent
Spring	Average	Excellent	Average	Problem	Good	Average
Summer	Excellent	Average	Problem	Problem	Problem	Problem
Fall	Good	Excellent	Excellent	Average	Good	Average
	STN					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	Average	Problem	Excellent	Average	Average	Good
Spring	Good	Good	Excellent	Average	Excellent	Good
Summer	Average	Problem	Problem	Problem	Problem	Problem
Fall	Problem	Average	Good	Average	Excellent	Average

Table 3.5 Elemental Carbon performance at both IMPROVE and STN networks based on Morris et al. metrics

	IMPROVE					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	Excellent	Average	Excellent	Good	Average	Excellent
Spring	Excellent	Average	Average	Average	Excellent	Average
Summer	Good	Average	Average	Problematic	Average	Average
Fall	Good	Average	Excellent	Good	Excellent	Average
	STN					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	Excellent	Problematic	Good	Excellent	Good	Excellent
Spring	Excellent	Average	Good	Good	Excellent	Excellent
Summer	Good	Average	Excellent	Excellent	Excellent	Excellent
Fall	Excellent	Problematic	Good	Excellent	Excellent	Excellent

3.8.4 Ammonium

Table 3.6 shows evaluation of ammonium based on monthly performances at STN sites for each region. The model showed excellent to average estimations of ammonium concentrations, which agrees with both studies conducted by Scott et al. [2009] and Morris et al. [2005]. Ammonium measurement is available only at STN sites.

Table 3.6 Nitrate performance at the STN network based on Morris et al. metrics

	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	Excellent	Problematic	Good	Excellent	Good	Excellent
Spring	Excellent	Average	Good	Good	Excellent	Excellent
Summer	Good	Average	Excellent	Excellent	Excellent	Excellent
Fall	Excellent	Problematic	Good	Excellent	Excellent	Excellent

3.9 Performance evaluation compared to previous studies

A Comparison of the Midwest region with the Great Lakes study performed by Spak et al. [2009] shows a number of similarities and differences. The similarities are that CMAQ estimates the highest $PM_{2.5}$ concentrations in winter, the model overpredicts $PM_{2.5}$ in fall and winter, and when compared to IMPROVE sites, the model has an underprediction of $PM_{2.5}$ in summer and an overprediction in the winter. The differences in model performance between the two studies are when CMAQ's evaluation is compared to STN sites. In the Great Lakes region the model's best estimates are from April to July, and the present work shows better performance in April through May and August through September when compared to STN sites. The modeling differences from the work of Spak et al. and the present study are the emission inventory and the modeling domain. NEI 2002 data were used in the present study compared to 2001 NEI data used for modeling CMAQ estimates over the Great Lakes. The Great Lakes domain is half the size of the Midwest region, which implies fewer sites for model-observation comparison.

Key aspects of present results are consistent with studies by Yu et al. over two domains. The first [*Yu, et al.,* 2008] evaluation of $PM_{2.5}$ concentrations was over the eastern United States for summer 2004 and used NEI 2001 data. $PM_{2.5}$ concentrations were underpredicted in comparison with both IMPROVE and STN networks. The second [*Yu, et al.,* 2007] compared IMPROVE and STN site measurements with Organic and Elemental carbon from the model output for the United States domain in 2001, which also used NEI 2001 data. The results of Yu et al. showed an underprediction of primary organic carbon due to missing sources in the emission inventory and the absence of isoprene and sesquiterpenes oxidations in the model, which led to an underestimation of secondary organic aerosols.

Recent studies incorporate the Weather Research and Forecasting (WRF) model [*Skamarock et al.,* 2008], a successor to the MM5 meteorological model used in the current study. *Appel et al.* [2010] used CMAQ v.4.7 to compare MM5 with WRF for

January and August of 2006 for the eastern United States. Both models showed similar median errors. Conclusions from the current work (based on MM5) likely apply to WRF-based simulations.

Coarse PM (i.e. $PM_{10-2.5,}$ referred to in this paper as PM_{coarse}) was underpredicted over all regions throughout the year with the most pronounced bias equal to -8 μg m^{-3} at sites in the South Central (May), Mountain (May), and the Midwest (July) regions. Several investigators have demonstrated similar PM_{coarse} underprediction [*Wang et al.*, 2002*; Tonnesen et al., 2004; Boylan et al.*, 2006].

CHAPTER 4: SATELLITE DATA ASSIMILATION

4.1 Introduction

Satellite data has proven to be effective in improving model-derived AOD over Asia and the Indian Ocean in previous studies using OI [*Collins et al.,* 2001, *Yu et al.* 2003, *Adhikary et al.,* 2008, *Park et al.* 2011]. Descriptions of those previous studies are presented in section 1.1.8. Following the methodologies used over Asia would provide improvements for model estimates over the U.S., because of the better availability of surface $PM_{2.5}$ data to evaluate model performance after OI, compared to the scarcity of surface measurements in Asia. The novel part of this work is represented by the averaging method used as data assimilation inputs for the model, in addition to the four different OI settings used for each of the averaging methods. This chapter focuses on the outcome of these methods and settings and their impact on the posterior model estimates. The methodologies for OI implementation are described in section 4.2. These include the model-observation validation methods used, the equations to calculate CMAQ-derived AOD, and the OI equations. Furthermore, the averaging methods used for CMAQ and MODIS inputs to OI, and the four different observation and model error settings used by OI are described in detail.

The results, presented in section 4.3 demonstrate an OI algorithm case study for May 2002. Detailed spatial inputs and outputs are presented as physical and mathematical evidence of OI's performance. The influence of the error covariance matrices used by OI for both the model and the observations are included in the current work giving an understanding of the effects of the four settings used for each averaging method. The annual evaluation of OI for the six regions in the continuous U.S. are described in addition to two sensitivity tests performed to evaluate the effects of both AERONET AOD and equation 4.7 on the posterior model estimates. Cross validation is performed to

validate the effectiveness of the data assimilation methods used in the current work, which is also described in section 4.3.10.

This chapter discusses OI performance over the western U.S. since data assimilation was less effective in the western parts of the U.S. than the eastern parts of the United States. Alternative input datasets that can be used by OI and possibly provide better data assimilation results are discussed and presented. An inter-comparison between the current work and previous work provides a good model AOD comparison between the current study and previous OI studies in Asia.

4.2 Methodology

4.2.1 Method overview

A graphical introduction to the methods, Figure 4.1 depicts the information flow to combine CMAQ AOD (τ_{cmaq}) estimates with aerosol optical depth data from sun photometers and satellite retrievals, producing posterior AOD values. τ_{cmaq} was calculated based on hourly CMAQ concentrations and mass extinction coefficients for each individual species (section 4.2). Posterior ($PM'_{2.5}$) and prior $PM_{2.5}$ ($PM_{2.5}^{cmaq}$) estimates were quantitatively compared with surface measurements from both the Interagency Monitoring of Protected Visual Environments (IMPROVE) [*MALM et al.*, 1994] and EPA Speciation Trends Network (STN) [*USEPA*, 2006] networks. MODIS AOD (τ_{modis}) was evaluated with ground-based AOD measurements from (τ_{anet}) (section 4.16). τ_{modis} and τ_{anet} data were regridded to the CMAQ domain, and τ_{anet} replaced τ_{modis} where available.

$PM'_{2.5}$ was calculated as the product of the $PM_{2.5}^{cmaq}$ fields and a scaling factor (the ratio of posterior to prior AOD values in OI). Different error settings, model error correlation length scales, averaging procedures, and choices for application of scaling factors were used to develop multiple realizations of $PM'_{2.5}$ via OI. Points in the algorithm with multiple choices of settings or sub-algorithms are denoted in Figure 4.1 as

A, B and C. The multiple realizations are then evaluated against surface $PM_{2.5}$ monitoring data to identify promising combinations of choices.

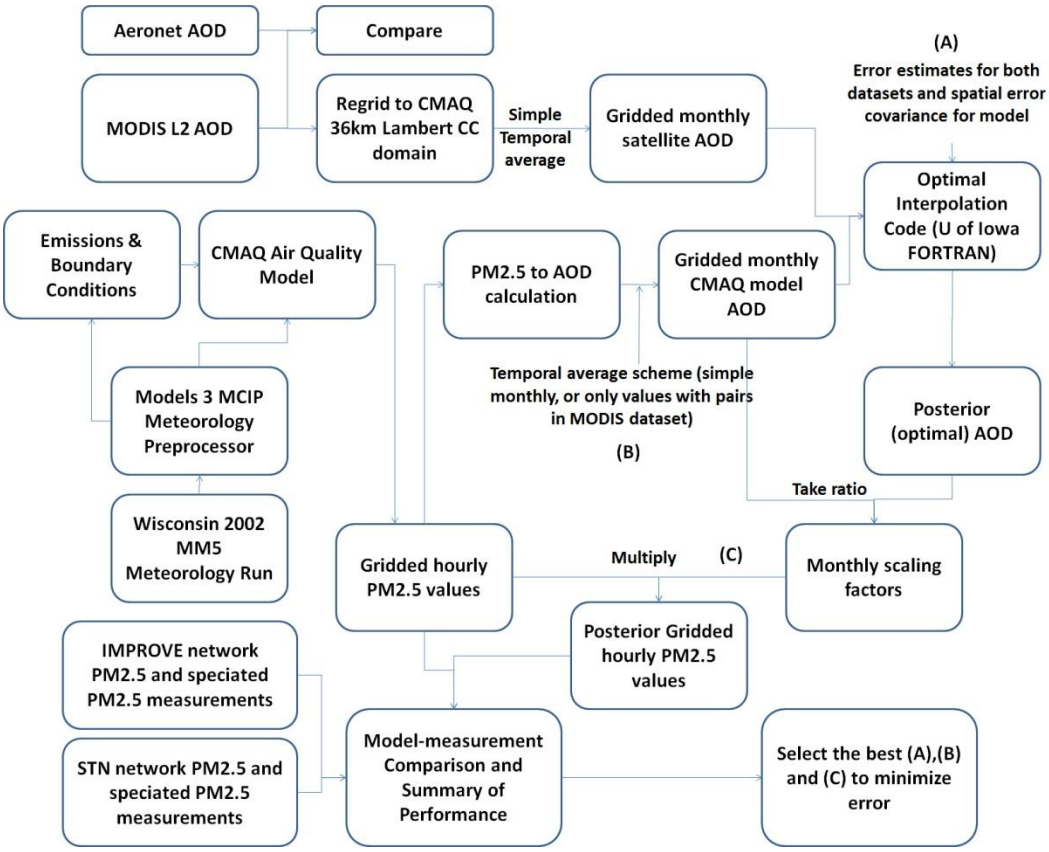

Figure 4.1 Algorithmic flow of chapter 4, (A), (B) and (C) show points where variable settings or sub-algorithms are chosen. See text for details.

4.2.2 Observation data

The same observations referred to in chapter 3 were used for evaluation and is shown Figure 3.2. τ_{modis} [*Levy et al.*, 2009] data were reprojected to the 36 km spatial resolution Lambert Conformal Conic CMAQ model grid using the Remote Sensing Information Gateway (RSIG) [Details at http://badger.epa.gov/rsig/]. RSIG averaged multiple data points for each grid cell. RSIG's AOD product is a composite of AOD from

MODIS Terra (north to south orbital transect at about 10:30 AM local time) and MODIS Aqua (south to north orbital transect about 1:30 PM local time). AOD retrievals (based on V5.2) were calculated at 0.55 μm and originally processed using MODIS data at 10 km [*Levy et al.,* 2009]. τ_{modis} is only available in a subset of observation pixels due to cloud contamination, reflective surfaces, and other factors.

4.2.3 OI implementation

Following the approach of *Adhikary et al.* [2008], AOD measurements from AERONET [*Holben et al.,* 1998] were used in the assimilation. Separately from the OI algorithm, τ_{modis} was evaluated for bias using AERONET; this is common practice since the sun photometer AOD are not sensitive to surface reflectance [*Remer et al.,* 2005]. AERONET level 1.5 was used if available; level 1.0 was used if not. Level 1.5 data is cloud screened data, but level 1.0 is not. Future work and manuscripts will exclude both level 1.0 and 1.5 datasets and use level 2.0 data which is quality assured data. For 2002, data from up to 39 AERONET sites in the domain were available (Figure 3.2) and τ_{anet} values replaced the τ_{modis} values in the respective model grid cells. If both existed at the same hour in the same grid cell, τ_{anet} was used instead of τ_{modis}. The final observational input to OI (referred to as τ_o) is a monthly mean AOD for each grid cell, and grid cells with AERONET stations may have a monthly mean of a combination of τ_{modis} and τ_{anet} data, although a large majority of τ_{anet} values is common in such cells. On average, each CMAQ grid cell had 16 available AOD retrievals per month out of a possible number of retrievals of about 60. On average all regions had more retrievals in summer and fall relative to winter and spring. During 2002, 36 AERONET stations reported at least some valid AOD measurements which are used in this work. Level 1.5 τ_{anet} data is recorded at 15-minute intervals during cloud-free daylight hours. On average, a grid cell with an AERONET station had 124 hours (out of a possible 730) of valid AOD retrievals per month.

To facilitate comparison of observed $PM_{2.5}$ and modeled $PM_{2.5}$, kriging was used to create gridded $PM_{2.5}$ values from surface observations. Kriging was performed with monthly average surface $PM_{2.5}$ values from the AQS, IMPROVE, and STN networks. Kriging was performed with GSLIB using month-specific Guassian semivarigrams [*Deutsch and Journel*, 1998]. Sill, nugget, and range parameters were optimized on a monthly basis (by nonlinear least squares) to match the experimental semivariogram.

4.2.4 CMAQ-Derived aerosol optical depth (τcmaq)

Aerosol optical depth has been calculated from CMAQ concentrations in various studies [*USEPA*, 2005; *Malm et al.*, 1994, 2007]. In this study, AOD was calculated using an empirical equation based on IMPROVE nephelometry, with the Raleigh scattering portion excluded [*USEPA*, 2005].

$$b_{ext} = [b_1 * f(RH) * (1.375 * (ASOI + ASOJ)) + b_1 * f(RH) * (1.29 * (ANO3I + ANO3J))$$

$$+ b_2 * PM_{ORG_{TOT}} + b_3 * PM_{EC} + b_4 * PM_{OTH} + b_5 * PMCOARS + b_6 * f_s(RH) * SSF$$

$$+ b_6 * f_s(RH) * SSC] \quad (4.1)$$

where b_{ext} (Mm^{-1}) is the total extinction coefficient.

Table 4.1 shows the description and dry mass extinction efficiency values for equation (4.1), which are adapted from *White* [1990] and *Malm et al.* [1994]. AOD is summed over the vertical layers to calculate total column thickness:

$$\tau_{cmaq} = \sum_{k=1..n} b_{ext_k} Z_k \quad (4.2)$$

where n is the number of layers, and z_k is the layer height of layer k. *Malm et al*. [1994] defined $f(RH)$ as $b_{ext}(RH)/b_{ext,dry}$ where $b_{ext}(RH)$ is the extinction coefficient as a function of relative humidity and $b_{ext,dry}$ is the extinction coefficient at dry conditions. The same $f(RH)$ values were used for ammonium sulfate and ammonium nitrate, taken

from the IMPROVE website [*Debell et al.*, 2006]. Sea salt $f_s(RH)$ of *Pitchford et al.* [2007] were used.

Table 4.1 Values for constants for AOD calculation in equation 4.1 [*EPA, 2005*]

CMAQ species code	Species description	Mode	Dry mass extinction efficiency value (b) (m^2 g^{-1})
ASO4I, ASO4J	Sulfate	Fine*	3
ANO3I, ANO3J	Nitrate	Fine*	3
PM_ORG_TOT	Organic Carbon	Fine*	4
PM_EC	Elemental Carbon	Fine*	10
PM_OTH	Unspeciated aerosols	Fine	1
PMCOARS	Coarse particles	Coarse	0.6
SSF, SSC	Sea Salt	Fine, Coarse	1.7

*Computed by adding both Aitken and accumulation modes

4.2.5 Satellite data assimilation via optimal interpolation

The OI method used in this study was adapted from earlier studies over Asia [*Collins et al.*, 2001; *Adhikary et al.*, 2008; *Chung et al.*, 2010]. In the governing equation

$$\tau' = \tau_{cmaq} + K(\tau_o - H\tau_{cmaq}) \quad (4.3)$$

τ' is the posterior AOD. H transforms the model data into the same space and variable as the observed data. In this work, H was the identity matrix because τ_o were mapped onto the CMAQ grid prior to the OI subroutine. K is the Kalman gain matrix, which combines the error covariance matrices for both the model and observation:

$$K = BH^T(HBH^T + O)^{-1} \quad (4.4)$$

where B and O are the error covariance matrices for the background and the observation fields respectively.

$$O = (f_o \tau_o + \varepsilon_o)^2 I \tag{4.5}$$

$$B_{ij} = (f_{cmaq} \tau_{cmaq} + \varepsilon_{cmaq})^2 exp\left[-\frac{d_x^2 + d_y^2}{2l_{xy}^2} \right] \tag{4.6}$$

f_o and f_{cmaq} are the fractional errors for the observation and the model, respectively. The parameters ε_o and ε_{cmaq} are the fixed errors in the observation and the model, respectively. The difference between the column indices of two cells is dx, while the difference between the row indices of two cells is dy. lxy is the horizontal correlation length scale for errors in the modeled AOD, which varies from one to five grid cells. OI is performed on a moving window technique that only involves 5x5 grid cells centered on the grid cell of interest.

If the posterior AOD in the grid cell, i ($\tau'(i)$) fell outside the range of the τ_{cmaq} and τ_o, then equation 4.7 was used instead, where the posterior AOD is calculated as the product of the model prior AOD with the model error covariance fraction.

$$\tau'(i) = \tau_{cmaq}(i) + \frac{B(i,i)}{B(i,i) + O(i,i)} (\tau_o(i) - \tau_{cmaq}(i)) \tag{4.7}$$

4.2.6 Satellite optimal interpolation methods and settings

The assimilation was conducted using monthly averages of τ_o. In other words, the inputs and outputs to OI (equation 4.3) represent monthly averages. Monthly averaging is supported by *Li et al.* [2009] who studied different temporal averaging windows for regression of $PM_{2.5}$ from τ_{modis} and recommended monthly averaging prior to regression. Previous OI studies [*Adhikary et al.*, 2008; *Yu et al.*, 2008] have also used monthly temporal averages.

Three OI approaches were tested (denoted as a through c in Table 4.2) which differed in the construction of monthly average τ_o and with the method of application of the OI scaling factor (i.e. the adjustment due to MODIS) to the model $PM_{2.5}$ values to produce posterior $PM_{2.5}$ values ($PM'_{2.5}$). In addition to multiple averaging approaches, the

parameters in equations 4.5 and 4.6 were varied over repeated calls to the OI algorithm to quantify sensitivity and establish best parameters. For each of the three approaches found in Table 4.2, four combinations of fractional and fixed error values (denoted 1-4 in Table 4.3) were used, generating 12 realizations of $PM'_{2.5}$. We also varied the horizontal correlation length scale for model errors by testing lxy values from 1-5 (equivalent to 36-150 km). The 60 realizations were evaluated by comparison to the surface monitoring data to identify the most suitable combinations (quantified by a fractional error score, defined below).

The three approaches in Table 4.2 have either different temporal averaging formula, or different methods of application of the scaling factors. Method a uses all τ_{cmaq} in the monthly average for each grid cell, even though there are many more values in the model than the observations since the observations occur during only 1 or 2 hours per day, and valid retrievals are not achieved every day. Method b averages hourly τ_{cmaq} values only if they are paired with valid τ_o values. $PM'_{2.5}$ is the product of prior $PM_{2.5}^{cmaq}$ by τ'/τ_{cmaq} (referred to as scaling factors in this paper). In methods a and b, scaling factors are applied equally to all hours in the month. In method c, scaling factors are applied only during and near hours with an AOD observation. The weighting of the scaling factor is adjusted using a Gaussian function, with full application of the scaling factor during the hour with the observation, and decreasing to a weight of 5% over a time window of ±12 hours. Time windows of ±8 and ±16 hours were also tested but the overall results were found to be insensitive to this selection and only the ±12 hour setting is analyzed herein.

The range of model errors in Table 4.3 was calculated by using representative model errors relative to surface $PM_{2.5}$ observations. Model error of 0.5 to 1 μg m^{-3} at conditions of 5 μg m^{-3} $PM_{2.5}$ corresponds to the fixed error term. Fraction error terms were consistent with total CMAQ errors of about ±8 μg m^{-3} at $PM_{2.5}$ levels of 20 μg m^{-3}. Absolute errors are dominant at clean conditions, while fractional error is increasing

important as $PM_{2.5}$ increases. The MODIS errors of *Adhikary et al.* [2008] and *Collins et al.* [2001] were adopted.

Table 4.2 Methods used for Optimal Interpolation

Method	Construction of MODIS Monthly Average AOD for OI Algorithm	Construction of CMAQ Monthly Average AOD for OI Algorithm	Use of scaling factor τ'/τ_{cmaq}
a	All available data in grid cell i,j	All τ_{cmaq} in grid cell i,j	$PM'_{2.5} = \tau'/\tau_{cmaq} PM_{2.5}^{cmaq}$
b		Only τ_{cmaq} with paired observed values	
c	Same as Method b		$PM'_{2.5} = PM_{2.5}^{cmaq} * \left(\left(1 - e^{-\frac{\Delta t}{2\sigma^2}}\right) + \frac{\tau'}{\tau_{cmaq}} e^{-\frac{\Delta t}{2\sigma^2}}\right)$ note*

note*. Δt (hours) is the time difference between the hour of the $PM_{2.5}$ estimate and the nearest (in time) overpass time, and σ^2 is set to 4.9 to confine the 95% influence of the scaling factor to within ±12 hours respectively.

Table 4.3 Optimal interpolation parameters (observation and model fractional error and absolute values)

Setting	f_o, e_o	f_m, e_m
1	0.2, 0.04	0.35, 0.1
2	0.2, 0.04	0.45, 0.05
3	0.4, 0.04	0.35, 0.1
4	0.4, 0.04	0.45, 0.05

4.2.7 MODIS retrieval at a finer spatial resolution

MODIS data was provided at 2km retrieval [Kumar et al., 2011], this would be different from the 10km retrieval available from the NASA website by default. The 2km

data was regridded to the 36km CMAQ domain by averaging all the available data points within one grid cell. OI was conducted using the full set of methods and settings for all regions on an annual basis.

4.3 Results

4.3.1 OI algorithm case study for May 2002

Figure 4.2 maps the AOD values of OI inputs (4.2a, 4.2c) and OI output (4.2e) for May 2002. Panels on the right are frequency distributions of the AODs from the spatial plots. Mean values for each region are listed in the inset table. Figs. 4.2a and 4.2b show τ_o, with mean of 0.22. These are higher than the counterparts τ_{cmaq} (Figs. 4.2c and 4.2d). The difference is most prominent in the west over mountain and desert regions. As expected, τ_{cmaq} was higher in more polluted locations such as the Northeast, Los Angeles, and the San Joaquin Valley. τ_o has considerable fine-scale variability while τ_{cmaq} is smoother, since the observation has significant regions of missing data.

The output of the OI algorithm (τ') is shown in Figure 4.2e, with a distribution shown in Figure 4.2f. This particular realization uses method 'a' (Table 4.2), error settings '2' (Table 4.3), and a model error correlation length parameter (lxly) of '1'; therefore, it is a 2a(1) realization. Figure 4.2e shows that the mean τ' (0.16) is higher than τ_{cmaq}. Qualitatively, τ' combines spatial features of τ_{cmaq} and the τ_o. Also, the distribution of τ' has started to take the shape of the observational histogram.

Figure 4.3 shows the logarithmic scaling factors as well as $PM_{2.5}^{cmaq}$ and $PM'_{2.5}$ for the May 2002 example. The scaling factors range from 0.3 to 16. High scaling factors can be caused by negative model biases for $PM_{2.5}^{cmaq}$ (e.g. underprediction), negative biases in the optical property calculation for τ_{cmaq}, or by positive biases in τ_o. Many of the areas with high scaling factors in the example from Figure 4.3 have high surface reflectance and dust. The model consistently had negative bias for $PM_{2.5}$ along

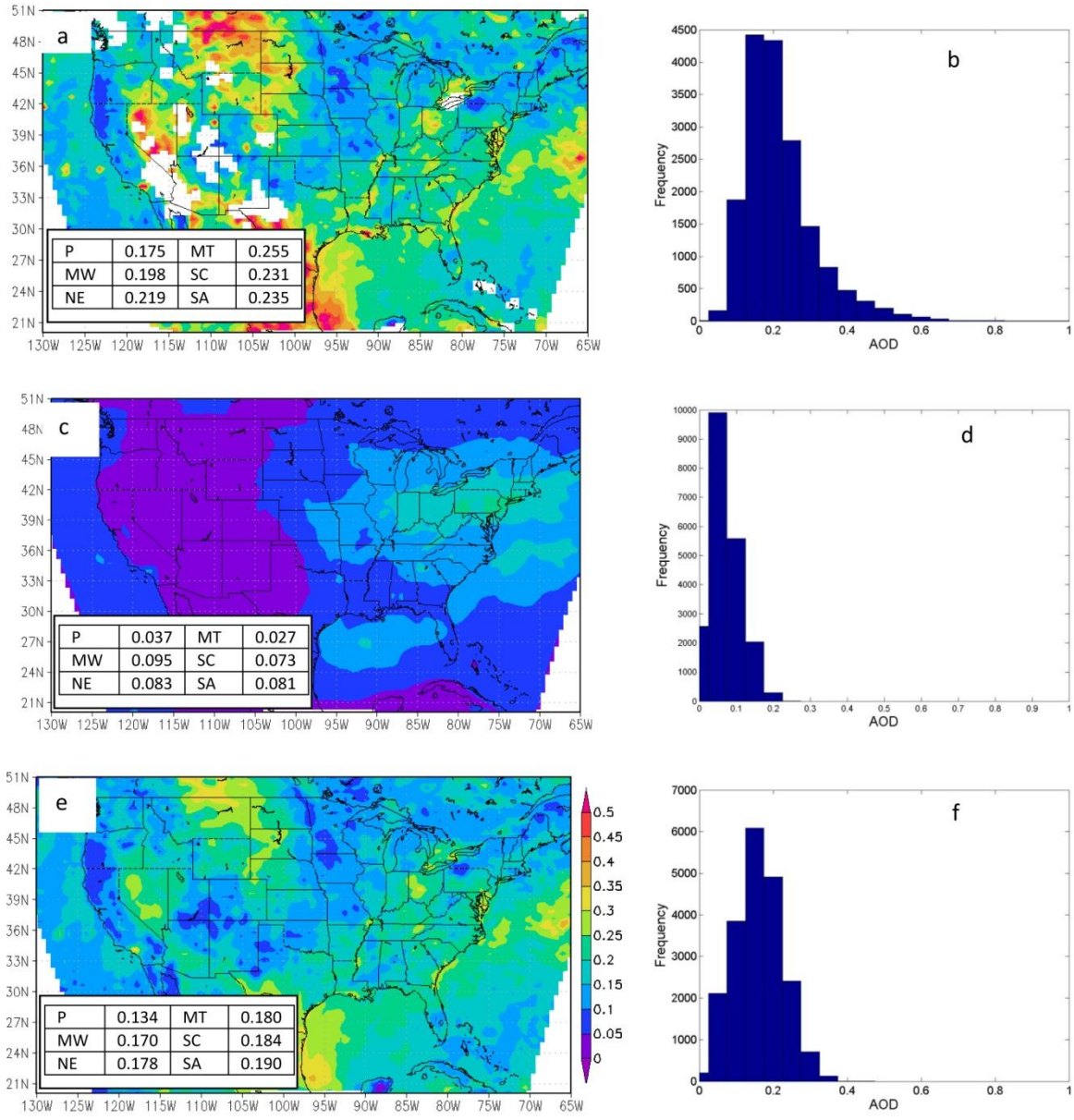

Figure 4.2 Example of the OI algorithm inputs (May 2002). (a) map and (b) histogram of monthly mean τmodis; (c) map and (d) histogram of monthly mean τ_{cmaq} ; (e) map and (f) histogram of τ' calculated using averaging method 2, error setting a, and lxly 1.

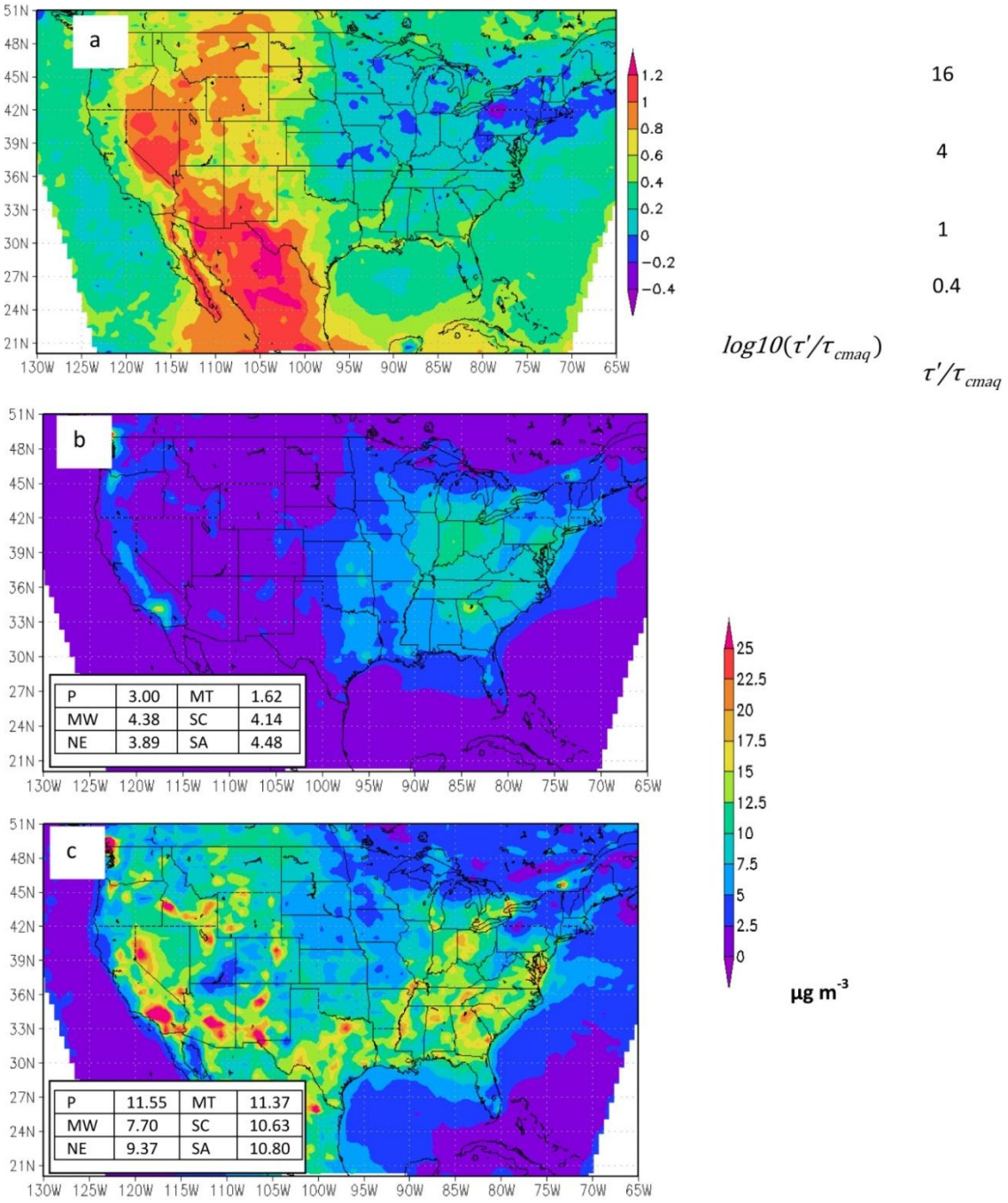

Figure 4.3 OI output (a) Log10 of CMAQ AOD scaling factors, (b) $PM_{2.5}^{cmaq}$, (c) $PM_{2.5}^{'}$ for May 2002 [settings used. method-2a, lxy1]

the U.S.-Mexico border in the southwestern U.S. (discussed in section 4.11). Scaling factors lower than one are also common, presumably resulting from positive biases in modeled aerosols, relative humidity, or potentially from negative bias in τ_o. Figures 4.3b and 4.3c map $PM_{2.5}^{cmaq}$ and $PM'_{2.5}$ average concentrations for May 2002. It is clear that $PM'_{2.5}$ is higher than $PM_{2.5}^{cmaq}$ and the high $PM_{2.5}^{cmaq}$ corresponds to locations with high τ_o (Figure 4.2a). Mean $PM_{2.5}$ concentrations in each region increased since on average, τ_o exceeds τ_{cmaq}. The Mountain region has the largest increase in mean concentration, by a factor of seven, followed by the Pacific region with a factor of about four. Across the entire domain, the mean $PM'_{2.5}$ (5.6 μg m^{-3}) was significantly higher than the mean $PM_{2.5}^{cmaq}$ (2.2 μg m^{-3}).

4.3.2 Influence of input data selection and preprocessing OI

Two major choices relative to input data selection are made in the algorithm. The first is whether to average CMAQ across all hours of the month (method a in Table 4.2), or to selectively average CMAQ during hours with AOD retrievals (methods b and c of Table 4.2). One of the distinctive features of the current work is the testing of methods b and c. Previous similar studies [*Collins et al.*, 2001; *Adhikary et al.*, 2008; *Chung et al.*, 2010] have only used method a. The second choice is whether to include τ_{anet} by replacement of τ_{modis} in pixels containing AERONET sites. A similar assessment of the impact of τ_{anet} by replacement was performed by *Adhikary et al.* [2008]. In the current work, the impact of including τ_{anet} was negligible on domain-wide performance statistics but scaling factors in the immediate neighborhoods of the AERONET sites were influenced.

Figure 4.4 shows the influence of method a (Figs. 4.4a and 4.4b) and b (Figs. 4.4c and 4.4d) on τ_{cmaq} in May 2002 as an example. The difference between τ_{cmaq} in method a and τ_{cmaq} in method b is shown in Figure 4.4e. Positive values indicate that τ_{cmaq} using method a are higher than using method b.

The pattern shown in Figure 4.4 is that τ_{cmaq} values are higher when averaging across all hours. This is likely due to the diurnal pattern of temperature and relative humidity in CMAQ, which causes higher AOD values at night and then increases τ_{cmaq} in method a but not in method b. This pattern is consistent throughout all month and all regions. Since τ_o is, for the most part, higher than τ_{cmaq} regardless of the averaging scheme, method a results in lower scaling factors than method b, because the method b τ_{cmaq} are lower, and thus require greater adjustment to approach the typically higher τ_o values. For example, in the grid cell containing Bondville, IL, the mean τ_{cmaq} value in method a is 60% higher than the value in method b; consequently, the scaling factors are much lower than those from method a.

4.3.3 Influence of the error covariance matrix for the model (B) and observation (O) on OI

The error covariance terms that quantify the errors in the model and observation values, as well as the spatial covariance in model errors, have substantial influence on τ'. Accordingly, $PM'_{2.5}$ depends on the method and error settings used for OI. The averaging method (Table 4.2) and error settings (Table 4.3) influence the amount of adjustment by OI. The variation is large enough that in some months and regions, the different error settings can lead to different signs in the $PM_{2.5}^{cmaq}$ mean fractional bias (i.e. underprediction versus overprediction).

The impact of the error settings is evident from changes between τ_{cmaq} and τ' for May 2002. As shown above, in May 2002, τ_o is significantly higher than the τ_{cmaq}; therefore, a larger increase in CMAQ AOD is indicative of the OI algorithm placing a higher weight on the observations. The increases in CMAQ AOD relative to the prior are 60%, 56%, 56%, and 50%, for settings 1a(1), 2a(1), 3a(1), and 4a(1). Therefore, error setting 1, with lower fractional errors on the observations, and higher absolute error on

the model AOD, weights the observations most heavily. Setting 4, with higher fraction error on the observations, and lower absolute error on the τ_{cmaq}, weights the model most heavily. Settings 2 and 3 are intermediate.

For both urban and rural locations, setting 4 was selected as the one with the lowest FES most frequently (37% of month/region combinations). For comparison, 26%, 7%, and 9% of month/region pairs had settings 1, 2, or 3, respectively (the remaining 37% showed no improvement under any setting). Looking at combinations of settings and averaging methods, the 4c combination has the smallest influence of the observations, produced the lowest FES values fairly frequently. The 4c combinations were particularly prevalent in the Pacific and Mountain areas where we have the least success in OI and the most evidence for bias in τ_o.

The lxly setting influences both posterior spatial patterns and error covariance, but the effects are somewhat complicated. The first effect is the spatial smoothing of τ' that is expected with higher values of lxly. The second effect is that increasing lxly increases the error covariance between model grid cells that are near to one another, and this increases the effective weight of the observations. Even in cases when the total error on the observation ($f_o\tau_o+\varepsilon_o$) is comparable to the model error ($f_{cmaq}\tau_{cmaq}+\varepsilon_{cmaq}$), the observations are weighted more heavily than the model due to the assumption of covariance in the model errors and independence in the observation errors. These basic effects of increasing lxly are further complicated by the algorithm option (equation 4.7) where τ' lying outside of the range of τ_{cmaq} and τ_o are coerced.

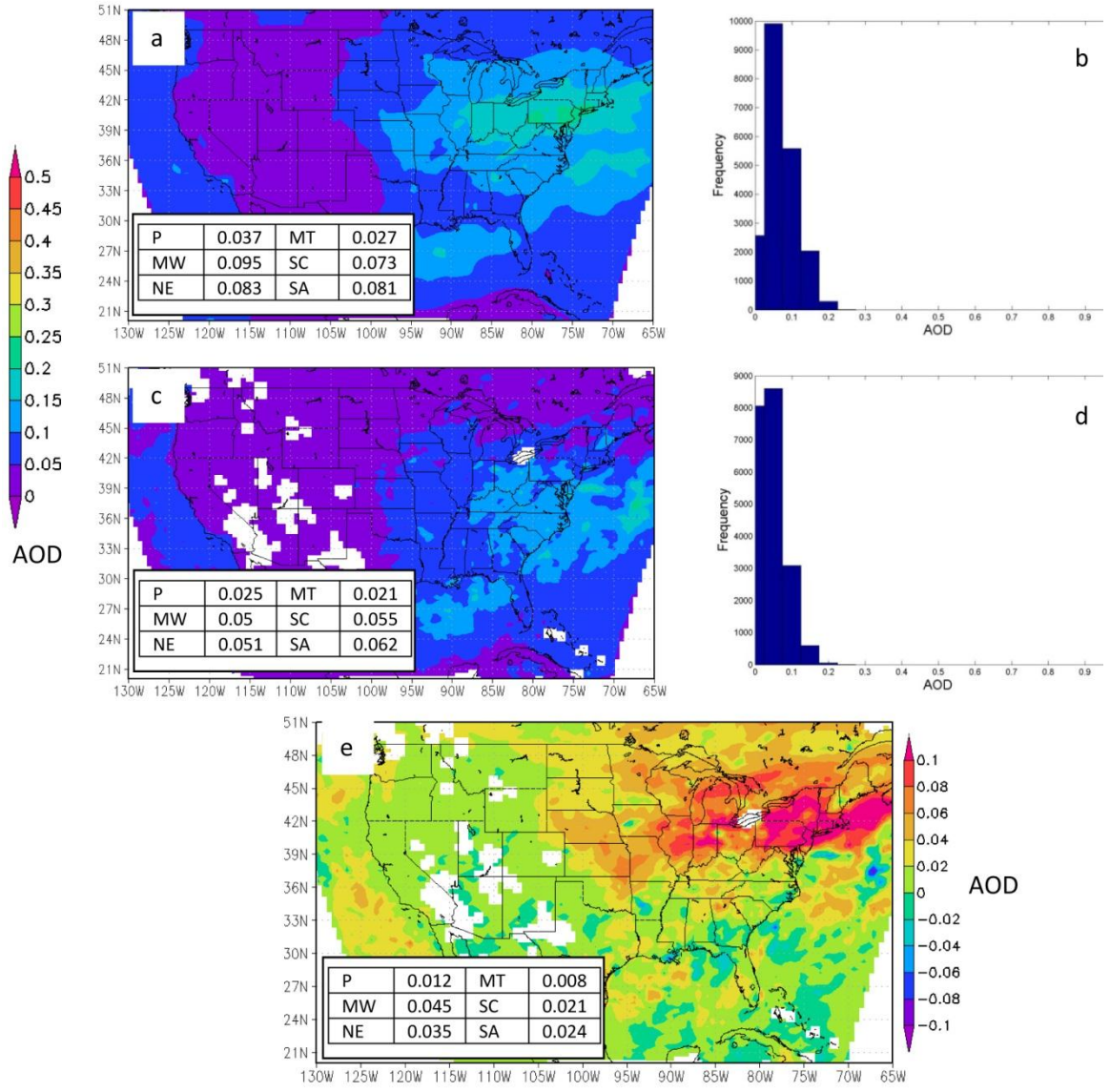

Figure 4.4 τcmaq for May 2002. (a) map and (b) histogram of τcmaq averaging all hours; (c) map and (d) histogram averaging of hours with an observation AOD pair only; and (e) difference of panels a and c. Positive values indicate that τcmaq from the all hours averaging are higher than those during hours with paired observation data. Mean values of each geographical region are shown in the inset table in each figure.

Equation 4.7 is invoked in 18 percent of grid cells at lxly 1, and this increases to 45-50 percent for lxly of 2 through lxly of 5. Equation 4.7 can be invoked in two situations. The first situation is when the weights being applied to the observation measurement difference sum to greater than one, due to numerical issues in calculating $(HBH^T + O)^{-1}$. The second potential condition where equation 4.7 is invoked is when either the observation or the model are weighted heavily, and then a smooth surface through the heavily weighted contribution will have values outside of the original range established by τ_{cmaq} and τ_o.

Comparisons between τ' maps calculated using the same settings and different lxly error covariance length scales show that from lxly of 1 to 2, some smoothing is observed and the average weighting of the observations increases. From lxly 2 to 5, there is minimal change. The reasons for this are (i) that the algorithm uses a 5x5 window to calculate the OI posterior, and this has a limited ability to do further spatial smoothing; and (ii) the alternate equation for calculating the posterior (equation 4.7) is increasingly applied, and this equation tends to not produce a smoothed surface.

4.3.4 Options for calculating the $PM'_{2.5}$

All methods use the OI-derived scaling factors on a grid cell by grid cell basis as multipliers to create the posterior modeled fine aerosol concentrations. Methods a and b apply the scaling factors uniformly to all hours. Method c (Table 4.2) applies the scaling factors at their full magnitudes only at the hour of the AOD observation as explained in the Methods section and Table 4.2.

Figure 4.5 shows the difference between method b (Figure 4.5a) and method c (Figure 4.5b). The figures show that $PM'_{2.5}$ values from method b have more influence (on average) from τ_o than those from method c. Method c tended to yield posterior values more heavily weighted to the model. The reason for this is that method c only corrects

with scaling factor's full magnitude at the time of the satellite overpass, so the posterior monthly average changes are less than with method b.

Figure 4.6 shows a time series of the $PM_{2.5}$ surface measurements (observation) at a location for a specific IMPROVE site in the Rocky Mountain National Park for May 2002. Three model-based concentration time series are also shown: $PM_{2.5}^{cmaq}$, $PM'_{2.5}$ from method 1b(1), and $PM'_{2.5}$ from 1c(1)). The scaling factor for this combination of month, grid cell, and methods was 11.9 for both 1b(1) and 1c(1). As an example to show the difference between both methods in Figure 4.6, on May 17th 2002 there was a MODIS AOD recorded at 11 AM local time. Method c had the largest adjustment only at the specific hour of the observation, and a falloff in the applied scaling factors for hours before and after. Using method c, the average $PM_{2.5}$ during this time period shifted from 1.56 to 8.24 μg m^{-3}. Using method b, the correction was applied to all hours, and the resulting mean $PM_{2.5}$ concentrations changed from 1.56 to 18.5 μg m^{-3}.

All three methods (a, b and c) increased $PM_{2.5}$ concentrations in cases where τ_o was higher than τ_{cmaq}; often, this was the desirable direction of adjustment since negative bias was common in CMAQ. For example in May in the South Atlantic region prior mean bias was -4.5 μg m^{-3} ($PM_{2.5}^{cmaq}$ - observation), and when using error settings 1, averaging method a, and lxly 1 [i.e. realization 1a(1)], the $PM'_{2.5}$ bias was 0.01 μg m^{-3}. Alternately, 1b(1) and 1c(1) gave model biases of 1.9 and -2.5 μg m^{-3}, respectively. This can be understood by the fact that original model has negative bias in this month and region and the τ_o (mean of 0.24 in the region) are higher than the τ_{cmaq}. τ_{cmaq} had a spatial mean of 0.081 using method a, and 0.062 when restricted to hours paired with observations (method b). Method c was responsible for the smallest amount of change in the model because the scaling factors derived from OI are applied selectively (only during and near hours of satellite observations). Method a was responsible for the next largest adjustment, while method b led to the largest adjustment in this case. Method a

produced a mean scaling factor of 1.02 (standard deviation 0.3) compared to 2.1 (standard deviation 0.65) for method b.

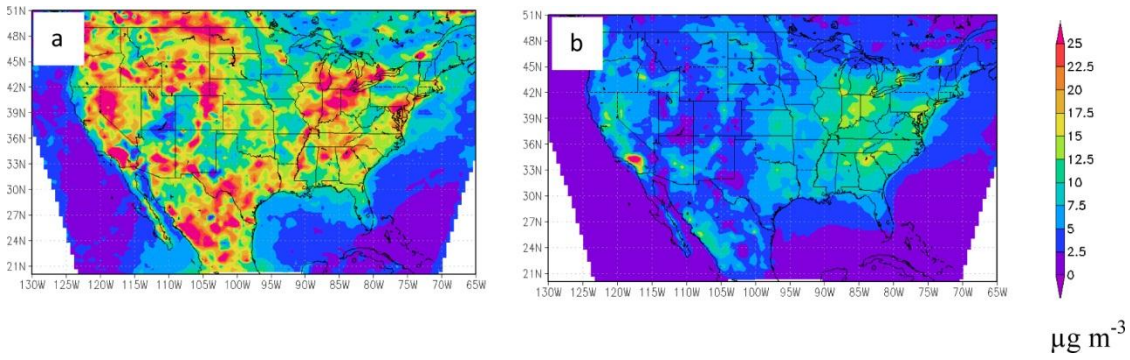

Figure 4.5 Posterior $PM_{2.5}$ ($PM'_{2.5}$) under two different ways of applying the OI scaling factors. Approach b is shown with settings (1b(1)) in panel a, while approach c is shown with settings (1c(1)) in panel b.

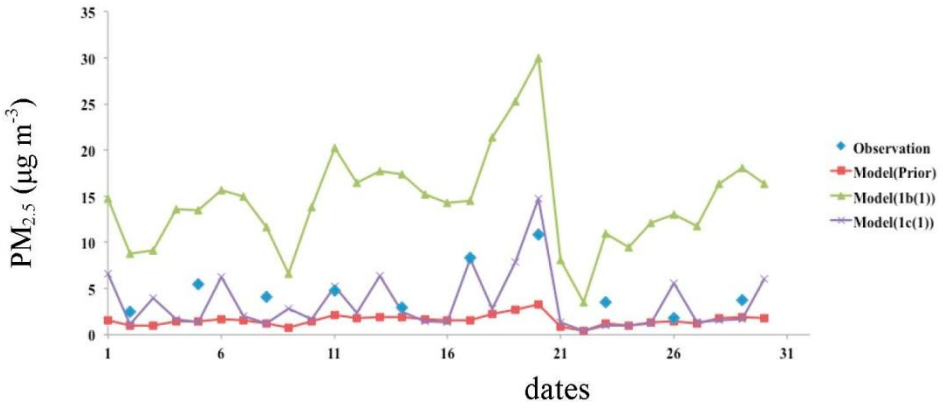

Figure 4.6 Observations (IMPROVE), model prior, model posterior (1b(1)), and model posterior (1c(1)), at the IMPROVE site in the Rocky Mountain National Park in May 2002.

4.3.5 Application of OI to all months and regions of 2002

The previously demonstrated OI algorithm was comprehensively applied to all months and regions for the year 2002. For each month, all possible combinations of methods (Table 4.2), error settings (Table 4.3), and error covariance distances (lxly settings) were applied, for a total number of 360 combinations per month. The 360 possible OI realizations for each month were evaluated by the metric fractional error score (FES). As described in the methods section, these realizations were searched for OI settings that would lead to the greatest degree of improvement in the FES.

Two important factors about the search for optimal settings should be noted. First, realizations that led to a degradation in FES relative to the prior model result had their $PM'_{2.5}$ values replaced with the $PM_{2.5}^{cmaq}$ result. In other words, where a particular OI setting is listed as achieving the best mean FES when averaged across multiple months or regions, it is important to note that instances (i.e., particular months and regions) with degradation in FES upon OI had their $PM_{2.5}^{cmaq}$ concentrations included in the final average and that OI results from those months were not used. Second, the evaluation of OI realizations was done separately versus IMPROVE and STN measurements (i.e. evaluation was done for rural and urban sites separately).

Optimal interpolation improved multiple months for all regions, but the eastern regions experienced improvement more frequently than the western regions. Table 4.4 lists settings that achieved the lowest absolute value of the FES for each month/region combination. A horizontal line (-) indicates that no setting achieved an improvement relative to the prior in the month and region in question. Although the Mountain region showed improvement from OI, method c (which limits the influence of the observations) was selected in this region more frequently than in other regions. The methods that most frequently had the superior FES were methods a and c. Method a achieved the lowest FES in 47 specific regions and months, versus 15 and 27, respectively, for methods b and c. 55 specific regions and months did not show any improvement. Method b (averaging

model AOD during only overpass hours and applying scaling factors to all hours) often yielded high scaling factors and $PM'_{2.5}$ values that were too high.

Error setting 4 (Table 4.3) was selected most frequently in both networks, where it showed up 52 times out of 89 totals (excluding 55 no improvement cases). lxly setting '1' was the most frequently used setting in both networks, where it led to the lowest FES 42 times out of 89 totals (excluding 55 no improvement cases). One distinctive feature is that both networks had improvement in the summer (June-Aug) at all regions. This is likely explained by the widespread negative bias in $PM_{2.5}^{cmaq}$ during summer, so the OI (which often yielded scaling factors > 1) was adjusting $PM_{2.5}^{cmaq}$ in the correct direction most consistently in summer. τ_o may also be most reliable in summer due to limited snow cover and presence of dark vegetation.

Figure 4.7 and Figure 4.8 plot MFB versus MFE "soccer plots" [*Morris et al.*, 2005], for IMPROVE and STN networks, respectively. The plots represent monthly (from 1-9 plus O, N and D for Oct, Nov, and Dec, respectively) values for six regions in 2002. The rectangular areas in the soccer plot correspond to performance categories, with excellent, good, average, and problematic areas radiating out from the origin. Arrow origins correspond to the performance of the prior, and arrow termini correspond to the posterior performance. Improvements in bias were larger than error improvement. Two types of behavior can be identified in the figures. The first is where (during all months with arrows) the prior model exhibited negative bias, and the satellite AOD caused an upward adjustment in concentration, often changing the bias from negative to positive. These are shown as arrows which move from left to right, and can be seen in Fig. 4.7a, b, c, and Fig. 4.8a, b and d. During these months and regions, selection of OI settings was often critical to preventing substantial overcorrection. In Figs. 4.7a and 11c there are points with substantial positive model bias, but they were not adjusted in the correct direction by OI.

Table 4.4 Settings that achieve the lowest absolute value of Fractional Error Score (FES) relative to surface measurements of (a) IMPROVE network and (b) STN for year 2002[1].

(a)	IMPROVE					
	Midwest	Mountain	Northeast	Pacific	South Atlantic	South Central
January	-	-	3a(5)	-	1a(2)	-
February	-	-	4a(1)	-	1a(5)	-
March	-	-	4a(5)	-	1a(4)	4a(1)
April	-	2c(3)	-	-	4a(1)	2a(1)
May	3a(4)	4c(1)	2a(4)	-	4b(1)	3a(1)
June	4a(1)	2a(2)	4a(1)	4c(1)	3b(2)	2b(2)
July	4b(1)	4a(1)	4a(1)	4c(1)	1b(2)	3a(1)
August	3a(2)	2a(3)	3b(4)	2c(5)	1b(2)	3a(1)
September	-	2c(2)	1b(3)	-	1b(5)	4a(3)
October	-	-	1a(5)	-	3b(5)	4c(1)
November	-	-	4a(2)	-	4c(5)	4a(3)
December	-	-	4a(5)	-	4c(2)	-
(b)	STN					
	Midwest	Mountain	Northeast	Pacific	South Atlantic	South Central
January	-	-	2a(5)	1a(2)	1a(5)	-
February	-	-	-	4a(5)	4a(1)	-
March	-	-	4c(1)	-	1c(5)	-
April	-	-	-	4a(1)	3a(2)	4b(1)
May	-	-	4a(5)	-	4a(1)	4a(1)
June	4c(1)	4c(1)	4c(1)	4c(1)	4a(1)	4b(1)
July	4a(1)	4c(1)	4a(1)	4c(1)	4a(5)	4b(5)
August	4c(5)	4c(1)	4c(1)	4c(1)	4a(1)	4a(1)
September	-	-	-	-	3a(1)	4b(5)
October	4a(5)	-	-	3c(3)	1b(3)	2c(5)
November	4a(3)	4c(1)	-	4a(5)	1c(1)	-
December	-	4c(1)	-	1a(2)	3c(4)	-

[1]Xn(Y) refers to error setting X (Table 4.3), averaging method n (Table 4.2) and lxly parameter of Y. So for example if X was setting 2 in Table 4.3, n was a in Table 4.2, and Y was 1, then the final combination will be 2a(1). '-' indicates that the prior FES is less than the FES by any OI setting.

This is consistent with persistent positive bias in τ_o. The second type of behavior was when the prior model has positive bias in some periods and negative in others, and τ_o provided (at least during several months) adjustments in the correct direction. This was less frequent and can be seen in Figs. 4.7e, 4.8c, and 4.8f. We consider the second type of behavior as a stronger indicator of valuable information content in the satellite, as the first type of behavior is consistent with a negative prior model and unbiased or a positively biased observation. A third type of behavior (a consistent positive bias in the model) with consistent reduction in concentration by the observations was not observed.

In addition to examining performance using the settings selected individually for each region and month, settings were optimized under constraints such as requiring a single monthly setting for all regions, a single regional setting for all months, etc. The constraints used are listed in Table 4.5, and these apply to the groupings found in Figure 4.9. Realizations that lead to a degradation in FES relative to the prior model result had their $PM'_{2.5}$ values replaced with the prior model result. Figure 4.9 graphs fractional error scores calculated for IMPROVE and STN averaged across multiple regions and months. Even and odd months were treated separately to allow intercomparison and cross-validation. Odd months of 2002 are shown 4.9a while 4.9b graphs even months. The optimization schemes used are listed in the figure as explained in the caption. As a means of cross validation, settings trained on odd months were applied to even months (January settings on February, March settings on April, etc.). And even month settings were applied to odd months (February on January, April on March, etc.). Figures 4.9c and 4.9d show those FES scores averaged across all regions and across 6 months. Figures 4.9c and 4.9d are discussed further below.

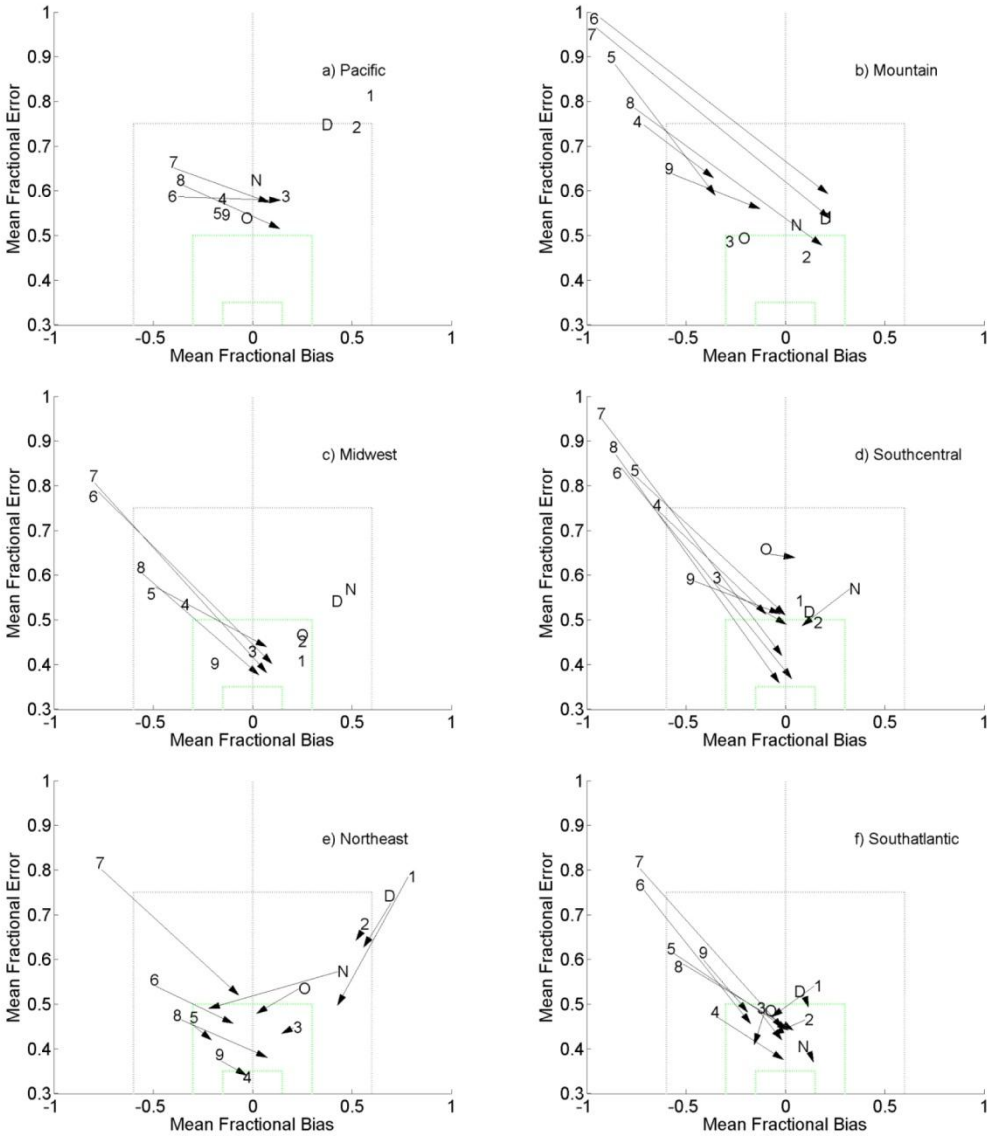

Figure 4.7 Monthly (1-9 plus O, N, and D for Oct, Nov, and Dec, respectively) prior and posterior Fractional Bias versus Fractional Error for 2002 for IMPROVE sites. The four (from excellent to problematic) performance categories described in the methods section are shown visually by the rectangular zones. The beginning and ending of the arrows represents the prior and posterior values, the arrows are only shown for months that show improvements.

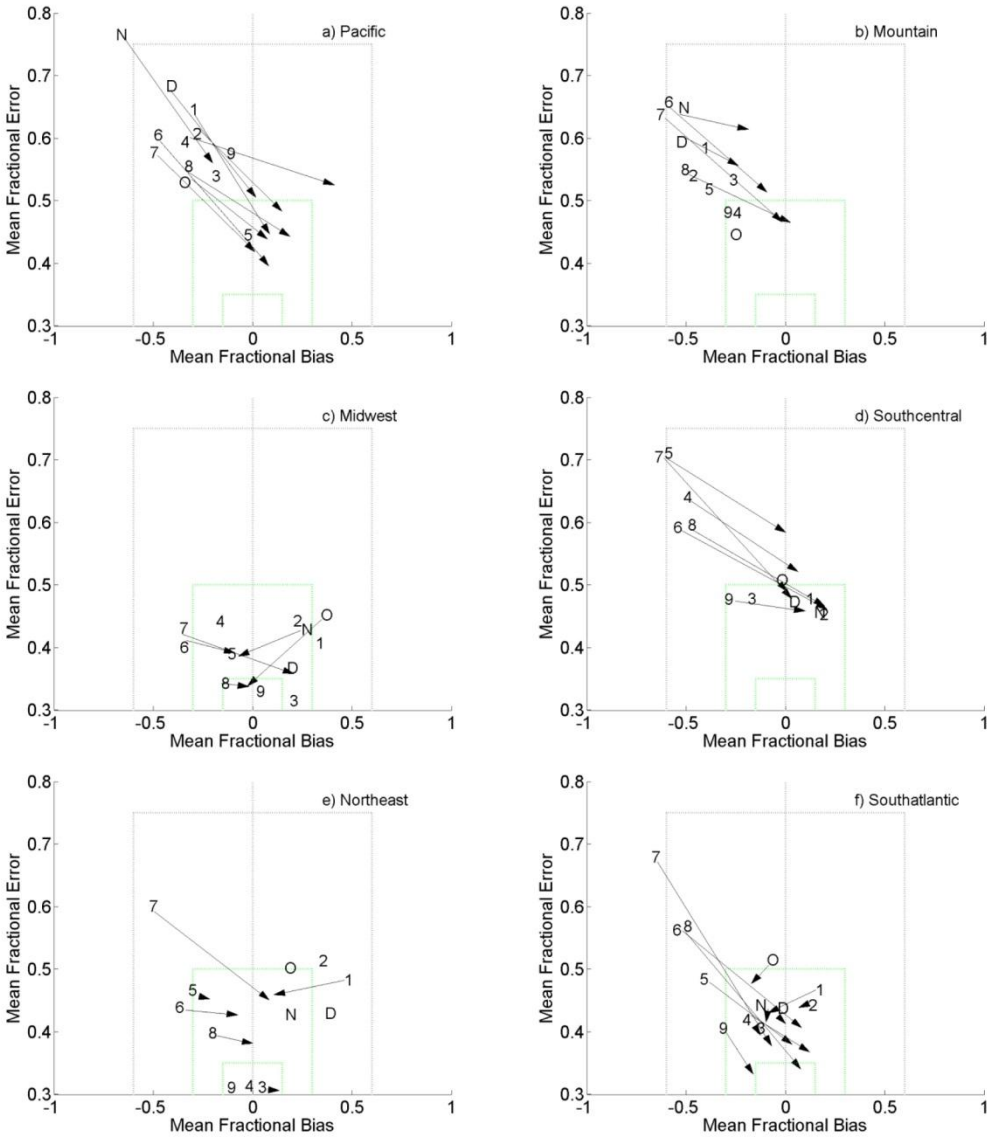

Figure 4.8 Monthly (1-9 plus O, N, and D for Oct, Nov, and Dec, respectively) prior and posterior Fractional Bias versus Fractional Error for 2002 for STN sites. The four (from excellent to problematic) performance categories described in the methods section are shown visually by the rectangular zones. The beginning and ending of the arrows represents the prior and posterior values, the arrows are only shown for months that show improvements.

Table 4.5 Categories for constraints on OI settings and methods (used in Figure 4.9)

Column	Description
1	Prior model (base case) no assimilation
2	Customized setting/method for each month and region as shown in Table 4.4.
3	One setting/method for each month, applied across entire domain.
4	One setting/ method for each region, applied across all months.
5	One setting/method applied to all regions and months.

Figure 4.9a and 4.9b show average improvements in error over the prior model from 0.99 to 0.89 (at STN sites) and from 1.2 to 0.97 (at IMPROVE sites). Since category 2 (with individual method/setting combinations for each month and region) had the highest degree of freedom for selection of settings, it naturally had the greatest improvement. The improvement stayed similar in magnitude for the more restrictive categories (3-5), with category 5 involving a single method/setting for the entire domain. Allowing variation in method/setting selection across time versus across region was very similar. With a single method/setting for 6 months and all regions, improvements over the prior was from 0.99 to 0.91 and from 1.2 to 1.00, respectively, for STN and IMPROVE monitoring locations.

Figures 4.9c and 4.9d show the result of a simple test of robustness of settings. Settings trained on odd months were applied to even months (January settings on February, March settings on April, etc.). And even month settings were applied to odd months (February on January, April on March, etc.). Figures 4.9c and 4.9d show those FES scores averaged across all regions and across 6 months. The graphs show that when applying the month's settings to opposite months the FES was similar to the improvement generated from using the settings optimized specifically for the month in question. Bar graphs in 13c and 13d use the prior in any month with degradation in the FES. The asterisks use a less forgiving procedure, where the selection of whether to use

the prior or the assimilation result is retained from the original "training" months (e.g. if the settings optimized for January in a specific region led to no improvement in FES in January, then the prior was used for February in the application of odd settings to even months). Comparison of the bar and asterisk positions show that settings are robust, but only under restriction that months with a degradation are replaced with the prior.

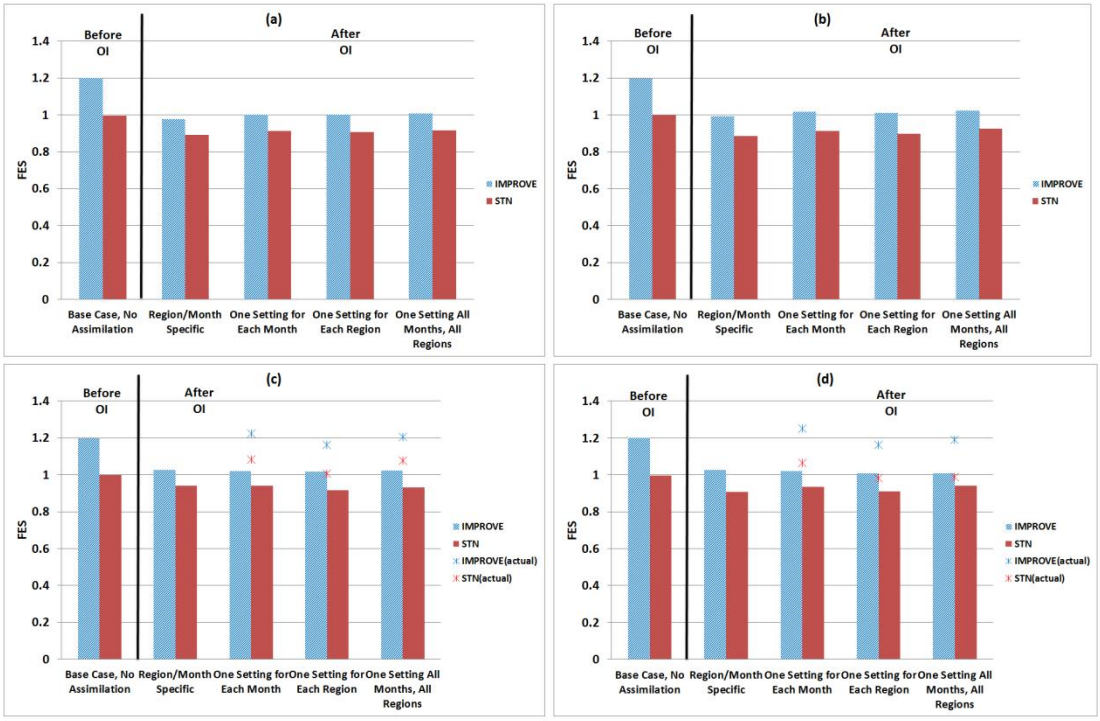

Figure 4.9 CONUS averaged fractional error score (y axis) derived from best OI settings for different constraints on the number of settings allowed (x axis categories). Left-most category is for the prior model with no OI. Categories 2-5 are described in Table 4.5. Panels a and b show results for odd and even months, respectively. Panels c and d provide a qualitative cross validation by applying settings optimized for even months to odd months (panel c) and vice versa (panel d). Lists above each bar are the specific settings used. The asterisks use a less forgiving procedure, where the selection of whether to use the prior or the assimilation result is retained from the original "training" months (e.g. if the settings optimized for January in a specific region led to no improvement in FES in January, then the prior was used for February in the application of odd settings to even months).

Figure 4.10 shows the result of OI applied to $PM_{2.5}$ concentrations for a complete year. Figure 4.10a maps 2002 annual average kriged surface observations from the AQS and IMPROVE networks. Figures 4.10b and Figure 4.10c map the 2002 annual average $PM_{2.5}^{cmaq}$ and $PM'_{2.5}$, respectively. The posterior was calculated by applying OI setting 4a(1) to odd months and 4c(1) to even months.

Bias reduction can be evaluated from $PM_{2.5}$ bias maps of $PM_{2.5}^{cmaq}$ (Figure 4.11a) and $PM'_{2.5}$ (Figure 4.11b). Bias maps are calculated as the model result minus a kriged $PM_{2.5}$ observation surface from Figure 4.10a. The plots show modest improvement in multiple locations (shifts from positive bias (red) or negative bias (blue) toward unbiased (white) colors. Areas with good performance include West Texas and Salt Lake City, the Ohio-Indiana border, around Atlanta, and eastern Pennsylvania. For example, areas along the Texas-Mexico border (Figure 4.10b) showed considerable negative bias in the forward model. Five sites were evaluated along the border with Mexico, one rural site (Big Bend National Park) and four urban sites located in El Paso and Corpus Christi, where OI reduced negative bias from 2.51-12.18 μg m^{-3} to a 0.63-7.97 μg m^{-3}. Substantial areas where model skill appears not to improve exist as well.

While Figure 4.11 shows the improvement qualitatively, the quantitative performance metrics are more reliably calculated using the original STN and IMPROVE $PM_{2.5}$ and the following were calculated using all daily paired measurements during 2002: RMSE, coefficient of determination, bias, FB, and FE were calculated. Four sets of statistics were calculated: $PM_{2.5}^{cmaq}$ vs. STN observations, $PM_{2.5}^{cmaq}$ vs. IMPROVE observations, $PM'_{2.5}$ (with settings optimized for STN) vs. STN observations, and $PM'_{2.5}$ (with settings optimized for IMPROVE) vs. IMPROVE observations. For STN, the prior metrics were RMSE of 5.34 μg m^{-3}, coefficient of determination 0.48, bias of -0.83 μg m^{-3}, FB -0.11, and FE 0.35. The OI posterior metrics were 5.25, 0.65, 1.12, 0.024, and 0.29, respectively.

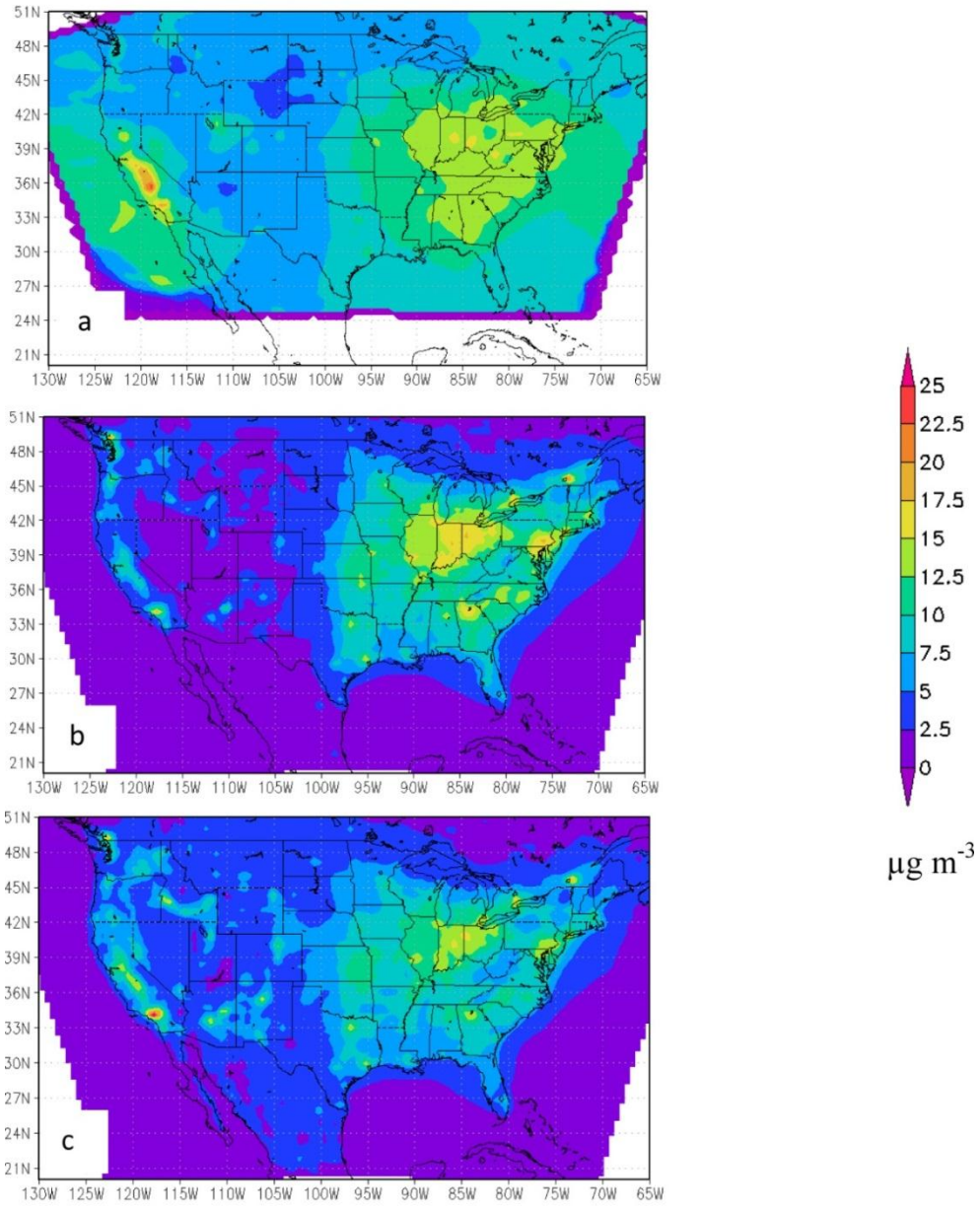

Figure 4.10 2002 annual average $PM_{2.5}^{cmaq}$ plots. (a) AQS and IMPROVE $PM_{2.5}$ values kriged to represent a combined measured $PM_{2.5}$ surface; (b) $PM_{2.5}^{cmaq}$; (c) $PM'_{2.5}$ based on 4a(1) for odd months and 4c(1) for even months.

For IMPROVE, the prior metrics were RMSE of 4.37 μg m^{-3}, coefficient of determination 0.55, bias of -0.86 μg m^{-3}, FB -0.21, and FE 0.52. The OI posterior metrics were 5.23, 0.63, 1.36, 0.09, and 0.43, respectively. Figure 4.11c shows the difference between $PM'_{2.5}$ and $PM_{2.5}^{cmaq}$ (the difference between Figure 4.11b and Figure 4.11a), and OI tends to increase $PM_{2.5}$ in the western U.S. and decrease it in the eastern U.S.

Table 4.6 shows another summary of the overall improvement from the OI techniques of this work. Classifications are based on the zones of the soccer plots (Figures 4.7 and 4.8) as proposed in *Morris et al.* [2005]. Improvements are most significant in summer. The lack of any settings that lead to OI improvement for the Pacific region (IMPROVE) is also notable.

4.3.6 Variance evaluation for the month of May

The current work relies on comparison of observation and model means; it is also reasonable to look at the variance differences. Figure 4.12 shows the temporal variance and coefficient of temporal variance for the month of May. The coefficient of variance for the observed $PM_{2.5}$ values (after averaging multiple monitors in a CMAQ grid cell) is shown in Figure 4.12a. The temporal variance is plotted for the τ_{cmaq} (Fig. 4.12b) and τ' (Fig. 4.12d) model AOD. The temporal coefficient of variance is plotted for the τ_{cmaq} (Fig. 4.12c) and τ' (Fig. 4.12e) model AOD. The coefficient of variance is the same for the τ_{cmaq} and τ' value, because the scaling factors are in both the nominator and denominator, which is constant for the whole month for each grid cell. The variance and the mean both show an increase between τ_{cmaq} and τ' values. The model AOD coefficient of variance is higher than the surface observations $PM_{2.5}$ coefficient of variance for τ_{cmaq} and τ' values.

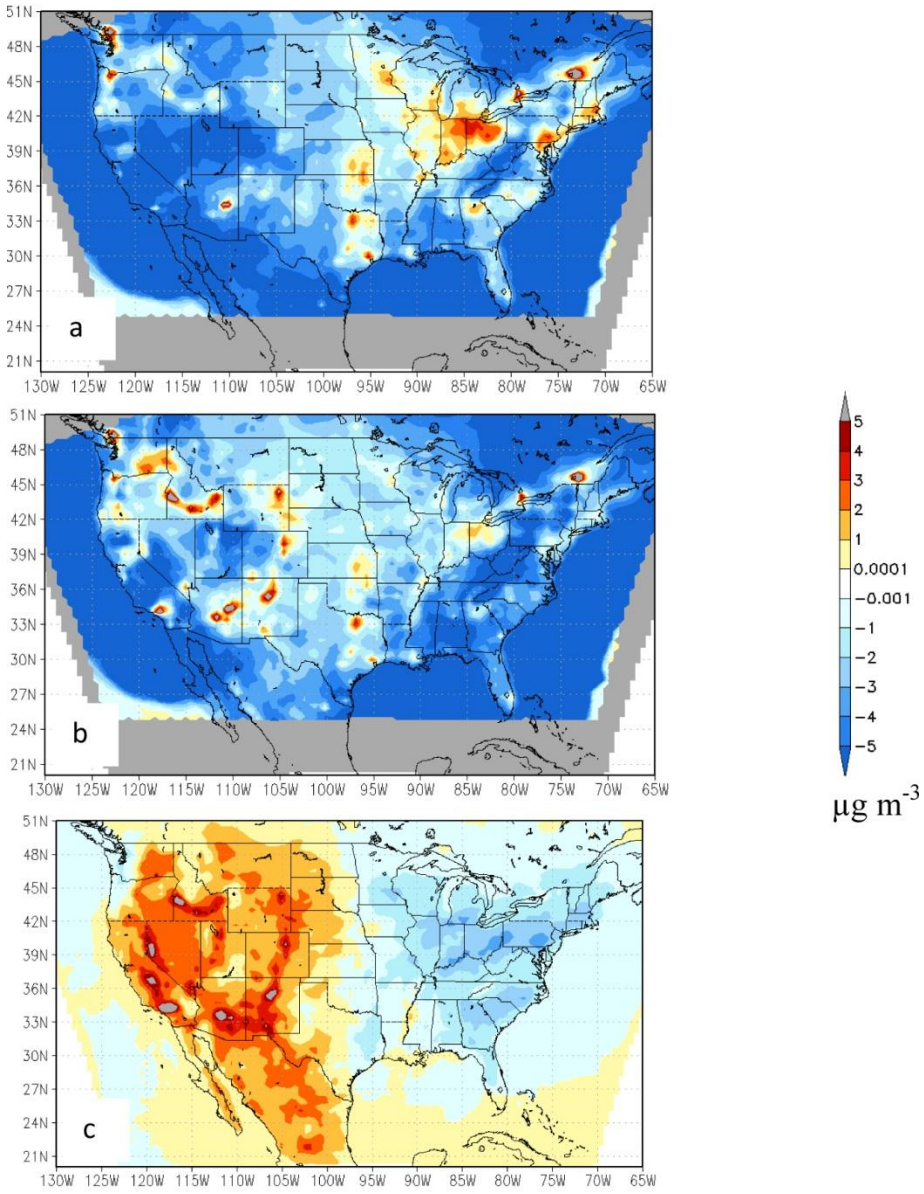

Figure 4.11 Annual average $PM'_{2.5}$ bias plots (a) before OI ($PM^{cmaq}_{2.5}$ - kriged surface obs, Fig. 11b-11a); (b) after OI ($PM'_{2.5}$ - kriged surface obs, Fig. 11c-11a); and (c) the amount of change due to OI ($PM'_{2.5} - PM^{cmaq}_{2.5}$).

Table 4.6 $PM'_{2.5}$ performance at IMPROVE and STN networks based on Morris et al. metrics[1].

	IMPROVE					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	P	A	G	A	P→A	A→G
Spring	A	A	A→G	A	G	A→G
Summer	A	P → A	A→G	P→G	A→G	A→G
Fall	A	A→G	G	A	G	G
	STN					
	Pacific	Mountain	Midwest	South Central	Northeast	South Atlantic
Winter	A→G	A	G	G	A	G
Spring	A	A	G	A	E	G
Summer	A→G	A→G	G	A→G	G	A→G
Fall	A	A	G→E	G	G	G
E=Excellent		G=Good		A=Average	P=Problematic	

[1]Arrows indicate improvements.

By definition, the coefficient of variance is the square root of the variance divided by the square root of the mean values, therefor the variance of AOD is different from the variance of $PM_{2.5}$, but the coefficient of variance is the same since the $PM_{2.5}$ to AOD conversion factor is applied to both the nominator and denominator. Spatial value of the variance and coefficient of variance for a grid cell was calculated as the mean value of a 5x5 surrounding window. Figure 4.13 shows the spatial variance and coefficient of variance for the month of May. The spatial coefficient of variance for the observed $PM_{2.5}$ values (after averaging multiple monitors in a CMAQ grid cell) is shown in Figure 4.13a. The spatial variance is plotted for the τ_{cmaq} (Fig. 4.13b) and τ' (Fig. 4.13d) model AOD. The spatial coefficient of variance is plotted for the τ_{cmaq} (Fig. 4.13c) and τ' (Fig. 4.13e) model AOD.

The coefficient of variance is different for the prior and posterior values, because the spatial covariance calculation is based on the surrounding grid cells, which all have different scaling factors. The surface coefficient of variance is higher than posterior and prior model values, which were respectively 0.47 and 0.07.

Consideration should be taken into account for the errors associated with spatial representation errors and between monitor measurements biases. An urban site should have less effect than rural sites when calculating over the spatial coverage. These variances reinforce the conclusion that MODIS is adding noise to the CMAQ because satellite OI injects high scaling factors, giving a correct regional average, but increasing the concentrations in some spatial locations more than others due to MODIS retrieval.

4.3.7 Sensitivity tests

Two sensitivity tests were conducted to understand the impact of both AERONET data and equation 4.7 on posterior model estimates. The two tests were conducted on an annual basis for all regions using the full set of methods and settings. The first test included running OI with only MODIS data. No AERONET was included in the data set; this will help understand if the AERONET sites are effective at or near AERONET measurement locations. The second test included running OI without equation 4.7, including changing the horizontal correlation length scale for errors in the modeled AOD (lxy) from 1 to 5. This test will help aid understanding of the effect of constraining the posterior value between the observation and the prior model.

4.3.7.1 AERONET

Data assimilation was repeated for the six regions on an annual basis, without using AERONET data. The results showed insignificant FES differences on a regional monthly evaluation, where the majority of the lowest FES method-settings were the same, and the few different changes in method-settings made insignificant changes from the FES values of the original OI runs that included AERONET.

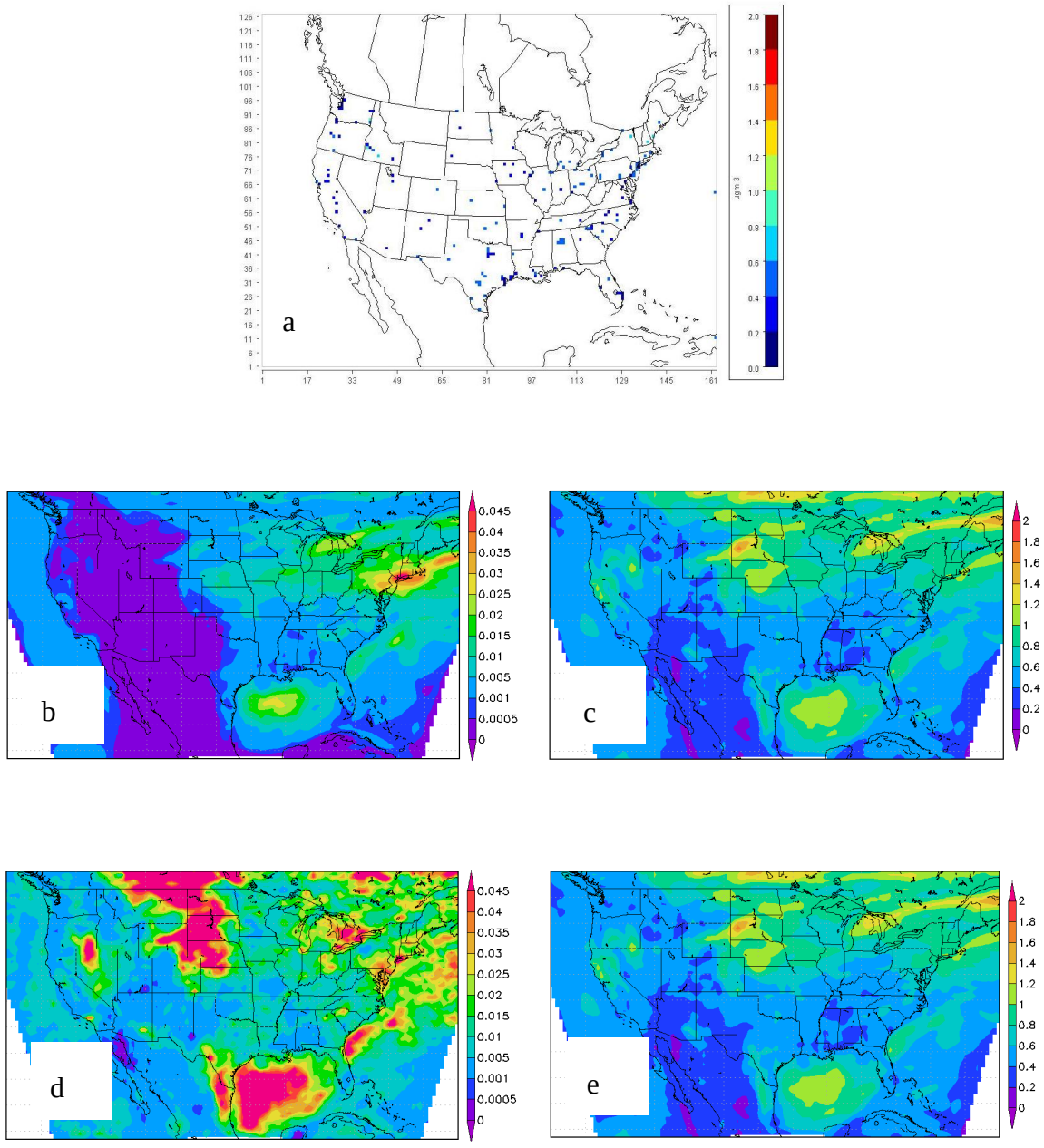

Figure 4.12 Temporal coefficient of variance and variance (May 2002). (a) surface $PM_{2.5}$ coefficient of variance; (b) τ_{cmaq} variance; (c) τ_{cmaq} coefficient of variance;(d) τ' variance; (e) τ' coefficient of variance

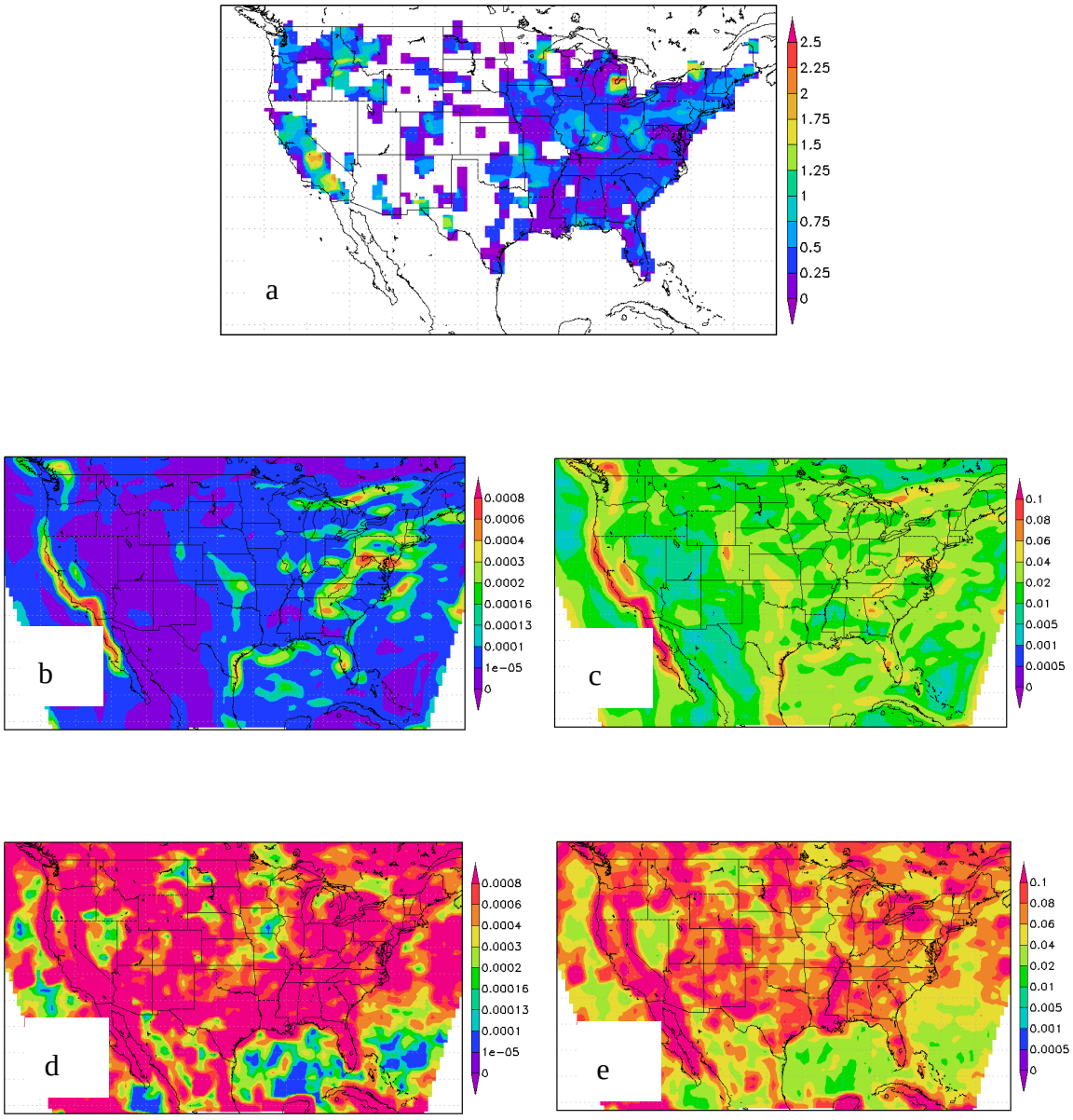

Figure 4.13 Spatial coefficient of variance and variance (May 2002). (a) surface $PM_{2.5}$ coefficient of variance; (b) τ_{cmaq} variance; (c) τ_{cmaq} coefficient of variance;(d) τ' variance; (e) τ' coefficient of variance.

To evaluate AERONET local influence (at grid cells containing or near to AERONET measurements), three sites were chosen based on the availability of collocated or adjacent

IMPROVE or STN observations. One site from IMPROVE and one site from STN were co-located (i.e. in the same grid cell) with AERONET. These sites were located at Bondvile, IL and Fresno, CA, respectively. In Columbia, SC an AERONET station and an STN monitor were in adjacent grid cells.

Figure 4.14 shows a time series of $PM_{2.5}$ concentrations (in May) for the surface observations and two $PM'_{2.5}$ concentrations (i.e. OI performed with AERONET included and OI performed without AERONET include). The $PM'_{2.5}$ values in Figures 4.14 (a through c) were chosen based on the best method-settings for the regional monthly evaluation. Both method-settings for the two posterior cases plotted (i.e. with and without AERONET) were found to be similar for these specific sites. The figures show that regional monthly method-settings are not necessarily the best for a specific site, since bias decreases are noticeable for Bondvile, IL (Figure 4.14(a)) and Columbia, SC (Figure 4.14(c)), but not for Fresno, CA (Figure 4.14(b)). In a regional monthly evaluation, the majority of sites followed the same method-settings, but this was not necessarily the case for all the sites. Figure 4.14(d) shows the same Fresno (CA) site but at the best method-settings for that site. The figure shows that posterior model $PM_{2.5}$ bias has decreased when AERONET is included with MODIS. AERONET does have an effect on the specific sites (i.e. grid cell) in optimal interpolation, but due to its scarcity, it has negligible effect on the monthly regional analysis.

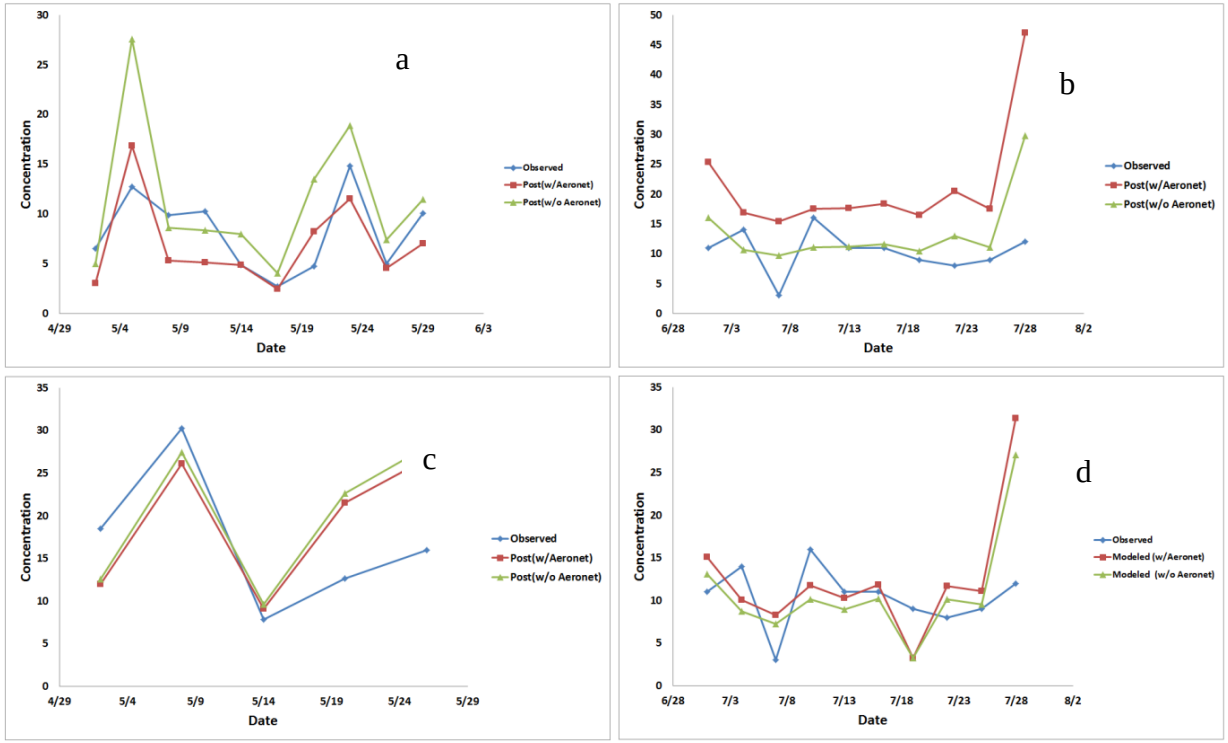

Figure 4.14 Time series for the Month of May of $PM_{2.5}$ concentrations for observations, and $PM'_{2.5}$ values (w/Aeronet and w/o Aeronet) using the same regional monthly OI method/setting, for three sites; (a) Bondville (IL); (b) Fresno (CA); (c) Columbia (SC), and the (d) Fresno (CA) using the best OI method/setting for the specific site

4.3.7.2 Equation 4.7

To constrain the $PM'_{2.5}$ value to be in-between the $PM^{cmaq}_{2.5}$ and observation values, equation 4.7 coerces values into a predominated range. The equation was included to correct numerical problems in the OI calculation, such as negative values. Removing equation 4.7 showed insignificant differences in FES values on a regional monthly evaluation, which is similar to the AERONET sensitivity test. Figure 4.15 and Figure 4.16 represent the difference between posterior and prior model estimates for lxy 1-5 with and without equation 4.7, respectively. Interesting changes are listed as follows. Setting 4 in Table 4.3 was selected by the algorithm more frequently (49% of region-month cases rather than 36% with equation 4.7). The horizontal correlation length scale lxy=1 was selected more frequently (46% of region-month cases rather than 29% with equation 4.7). The horizontal correlation length scales lxy of 3 and 4 FES differences were negligible between the four optimization schemes in Table 4.5. Figures 4.15 (with equation 4.7) show that the model concentrations have spatial smoothness compared to Figures 4.16 (without equation 4.7).

The OI working window takes into account the error covariances within the window; this allows a continuous consistency for all the posterior concentrations in that window. Equation 4.7 forces a specific grid cell to be affected by its local error covariance for, preventing the posterior value from being in continuum with the surrounding posterior values. This effect will create non-smooth spatial concentrations as can be seen in Figure 4.15. Therefore the coercion forced by equation 4.7 is physically non desirable.

The horizontal correlation length scale for errors in the modeled AOD (lxy) 1, showed no effect on the spatial plots (i.e. Figure 4.15(a) and Figure 4.16(a)). The differences start to become more evident for horizontal correlation length scale for errors in the modeled AOD (lxy) higher than 1 (i.e. 2-5). Figure 4.16 shows that AOD values are higher than AOD values in Figure 4.15, which mostly covers the Mountain, Midwest,

and South Central regions where equation 4.7 seems to localize MODIS effect. In both Figures 4.15 and 4.16, the change of lxy from 2 to 5 shows a few decreases in AOD value, as seen in the Mountain and the Midwest regions.

4.3.8 MODIS retrieval at a finer spatial resolution

Section 4.6 describes repeating OI with different MODIS input for a 2km retrieval, which is different from the 10km retrieval provided by NASA. The evaluation was only conducted on half of the year (from January to June) due to the availability of only Terra data, and since Aqua is not available until July. Results showed insignificant differences on a regional monthly evaluation, which is similar to the AERONET sensitivity test. The 2km retrieval aggregated to 36km and mapped on the CMAQ domain, is equivalent to 10km aggregated to 36km. The benefit of having 2km retrieval is to apply OI at resolutions lower than 10km (e.g. 4km model grid resolution), which is crucial for a CMAQ forward model simulation at 4km. Therefore, this test would be more meaningful for a 4km resolution model simulation.

4.3.9 Error metrics

While the selection of optimal OI settings was done using the fractional error score (calculated using paired daily $PM_{2.5}$) as a basis, alternate metrics such as RMSE could also be used. Regardless of the error metric chosen, correlation is not likely to improve significantly due to the use of monthly averaged scaling factors. Correlation (as measured by the coefficient of determination between paired model and observed daily $PM_{2.5}$ values within a region) changed by a mean amount of -0.04 (when OI was optimized to decrease the fractional error). In the settings used in Figure 4.11, temporally invariant scaling factors are used for each grid cell and month, so improved correlation based on temporal skill is not expected. Improvements (or degradation in skill) reported in this work are mainly due to changes in spatial skill. Thus the prediction of specific times for particularly clean or polluted conditions is often not improved by assimilation.

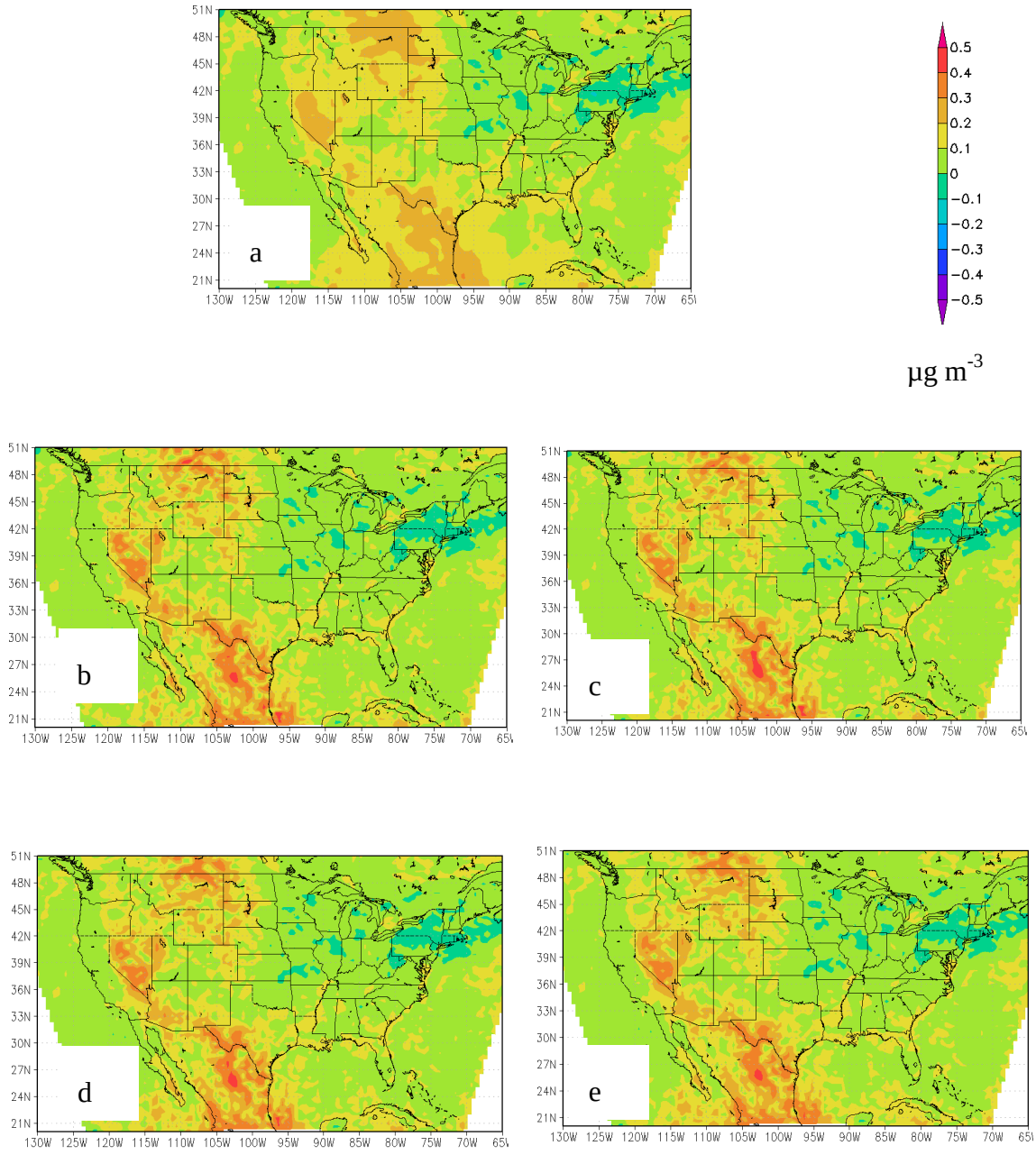

Figure 4.15 Change between $PM'_{2.5}$ and $PM^{cmaq}_{2.5}$ model concentrations for the month of May, changing the horizontal correlation length scale for errors in the modeled AOD (lxy) from 1 to 5, figure a to e respectively. With equation 4.7.

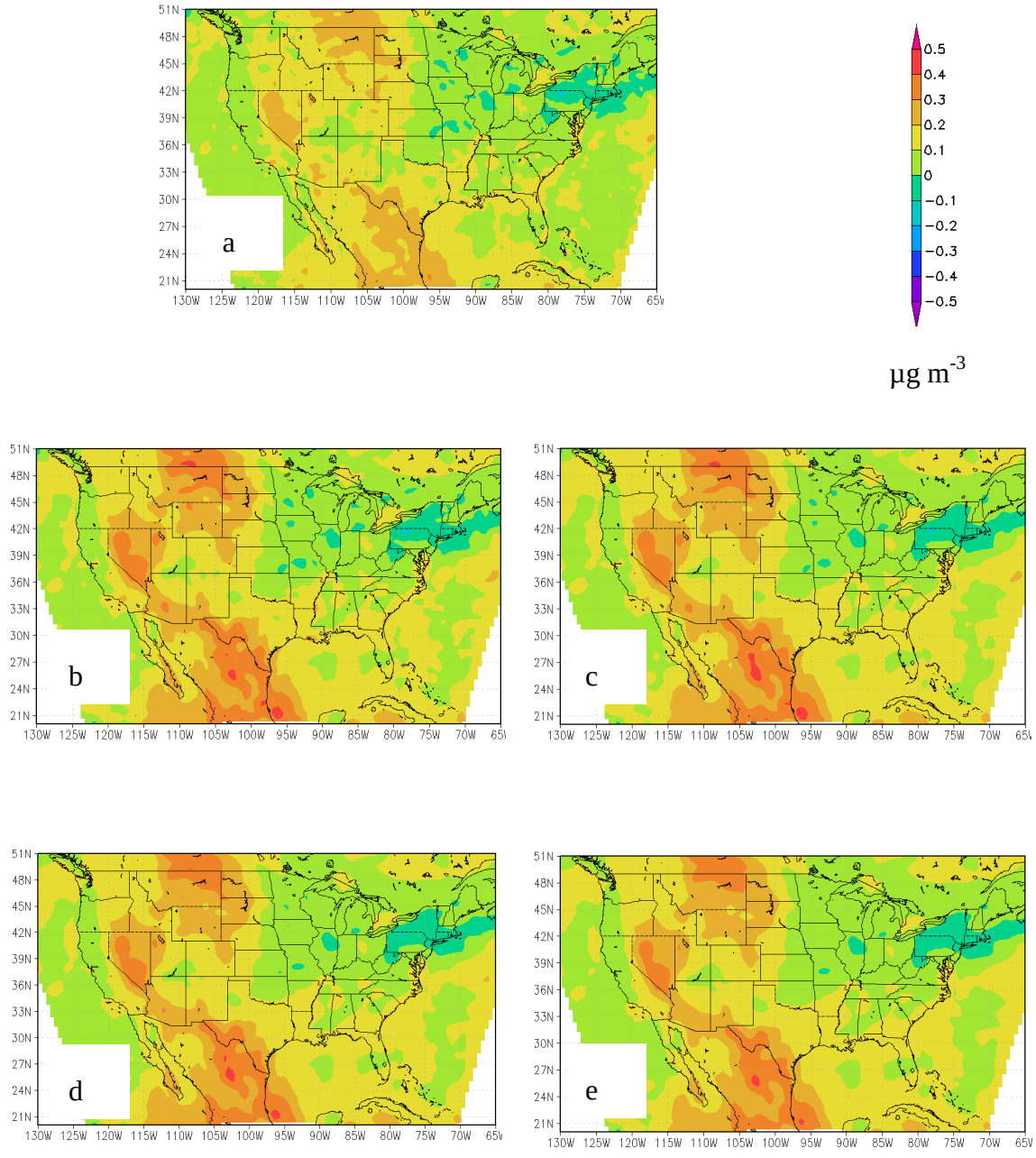

Figure 4.16 Change between $PM'_{2.5}$ and $PM^{cmaq}_{2.5}$ model concentrations for the month of May, changing the horizontal correlation length scale for errors in the modeled AOD (lxy) from 1 to 5, figure a to e respectively. Without equation 4.7.

Furthermore, improvements in spatial skill can be limited by errors in τ_o such as spurious high values (see discussion below). RMSE sometimes improves when the fractional error score is reduced, and sometimes increases. This is because RMSE and fractional error score, while both measures of error, weight errors differently.

One concern that arises in this work is whether to use settings optimized relative to IMPROVE or STN measurements to reflect both regional variability in concentrations and the urban values that affect the largest populations. Alternative approaches that blend these two datasets are also possible but have not been explored in this work. With a more urbanized group of monitoring sites, the settings based on comparison to STN will be sensitive to CMAQ errors in both urban primary and regional secondary pollutants. On the other hand, settings tuned to minimize error relative to IMPROVE will be more sensitive to CMAQ errors in regional secondary species, and in some areas, to dust and biomass burning. While future work will test simultaneous consideration of both datasets, we recommend that the application (e.g. climate forcing, human exposure, etc.) determine the weighting of the two monitoring networks or the weighting of the individual sites. In some region-month combinations, the same settings may be best for both networks, but in general this was not the case.

4.3.10 Cross validation

While the cross-validation approach of Figs. 4.9c and 4.9d gives an indication of the consistency in performance across the settings, more exhaustive cross validation approaches could be employed, such as comparing the same settings over multiple years. The limited cross-validation quantifies the effect of using settings trained on one period for use on OI of another period; it is therefore an example of withhold data from the training set.

4.4 Discussion

4.4.1 Discussion of OI performance in the western U.S.

A few alternative explanations were investigated to understand the large differences in AOD in the Western U.S. (usually $\tau_o > \tau_{cmaq}$) which often led to scaling factors in excess of 5.The alternative explanations of the CMAQ-MODIS AOD mismatch in these areas included (i) missing biomass burning emissions in the model; (ii) negative model bias in dust; (iii) positive bias in τ_o; and (iv) the use of a "clean condition" boundary in the current work.

Fires are thought to be well represented in the 2002 NEI report [*U.S Environmental Protection Agency (EPA),* 2008]. This is supported by a successful test of whether CMAQ reproduced large fire events and their impacts on $PM_{2.5}$ concentrations. A large, isolated fire outbreak occurred in southwestern Oregon on July 28th and 29th and emissions from this fire event were well represented in our CMAQ implementation [*NICC,* 2002]. On July 31st 2002 $PM_{2.5}$ observations were available for the southwestern part of Oregon showing a maximum of 64.5 µg m^{-3}, compared to the model estimate maximum of 57.5 µg m^{-3}. AOD values in the region peaked at 1.0 and 1.2 for observations and CMAQ, respectively.

Much of the limited ability of τ_o to improve $PM_{2.5}$ performance statistics relative to surface measurements is hypothesized to come from errors in τ_{modis}. The first piece of evidence of a lack of τ_o skill comes from an assessment of correlation between τ_{modis} and τ_{anet}. Figure 4.17 shows the coefficient of determination for all the sites in May 2002 from west coast to east coast of the United States. Correlation is higher in the Northeast than that in the Pacific.

PM_{coarse} exhibits negative bias as explained in section 3.1.1, and a high bias in τ_o over the Western U.S. is acknowledged in the literature and supported by an AERONET-MODIS AOD comparison. To compare the effects of these two known biases on

assimilation, we evaluated three months (April-June) at the Pacific region for AOD bias in CMAQ from missing coarse mode aerosols, relative to effects of τ_o positive bias. The CMAQ coarse AOD bias was calculated coarse PM model bias was calculated and subsequently converted to AOD units using assumed aerosol properties for the coarse aerosol. Fractions of 88% coarse PM as dust and 12% as sea salt were used, based on IMPROVE measurements averaged over the region [*Hand, et al.*, 2011]. Constants from Table 4.1 were used to calculate the sum of coarse AOD and sea salt AOD, and equation 4.2 was used assuming a 1 km boundary layer height, since the bias is based on paired surface and model concentrations. Independently, MODIS-AERONET AOD daily paired averages were compared for three months (April-June). Histograms for the coarse PM bias (Figure 4.18a) and the associated AOD values (Figure 4.18b) show skewed distributions with modes at –1.3 μg m^{-3} and -0.001 AOD units, respectively and with 5% lower percentiles at -14.76 μg m^{-3} and -0.011 AOD units, respectively. Although there is negative bias in PM_{coarse}, the extinction contribution of coarse PM in the atmosphere is low.

The MODIS-AERONET paired AOD comparison quantifies the positive τ_o bias in this period and season (mode of 0.09, and 5^{th} and 95^{th} percentiles of 0.038 and 0.34, respectively). The much larger magnitude of MODIS bias relative to that associated with coarse PM (converted to AOD units) supports that τ_{modis} biases are the main contribution towards the OI performance limitations over the Pacific.

The effect of the clean boundary condition (which led to a mean AOD in the westernmost model grid cells of ~0.03) may be comparable in magnitude to the MODIS bias. The Northeast Pacific mean AOD is 0.14-0.15 with 45-49% in the fine mode work [*Remer et al.*, 2008]. Therefore, the negative bias in the model in the Western U.S. from the boundary contribution may be of order 0.1 AOD units. Furthermore, the use of boundary $PM_{2.5}$ concentrations too low aloft may cause errors in the modeled vertical

profile of $PM_{2.5}$, and evaluation of modeled vertical profiles by in situ and remote sensing should be incorporated into future work.

Other studies have investigated τ_{modis} bias over the western U.S. *Prados et al.* [2007] compared GEOS satellite AOD with τ_{modis} and τ_{anet}. Their study showed that τ_{modis} and τ_{anet} were more highly correlated in the eastern, central, and southeastern part of the United States, and exhibited low to moderate correlation in the western United States. *Engel-Cox et al.* [2004] made a statistical analysis between $PM_{2.5}$ surface concentration (from IMPROVE and STN sites) and τ_{modis}, and found that τ_{modis} and surface $PM_{2.5}$

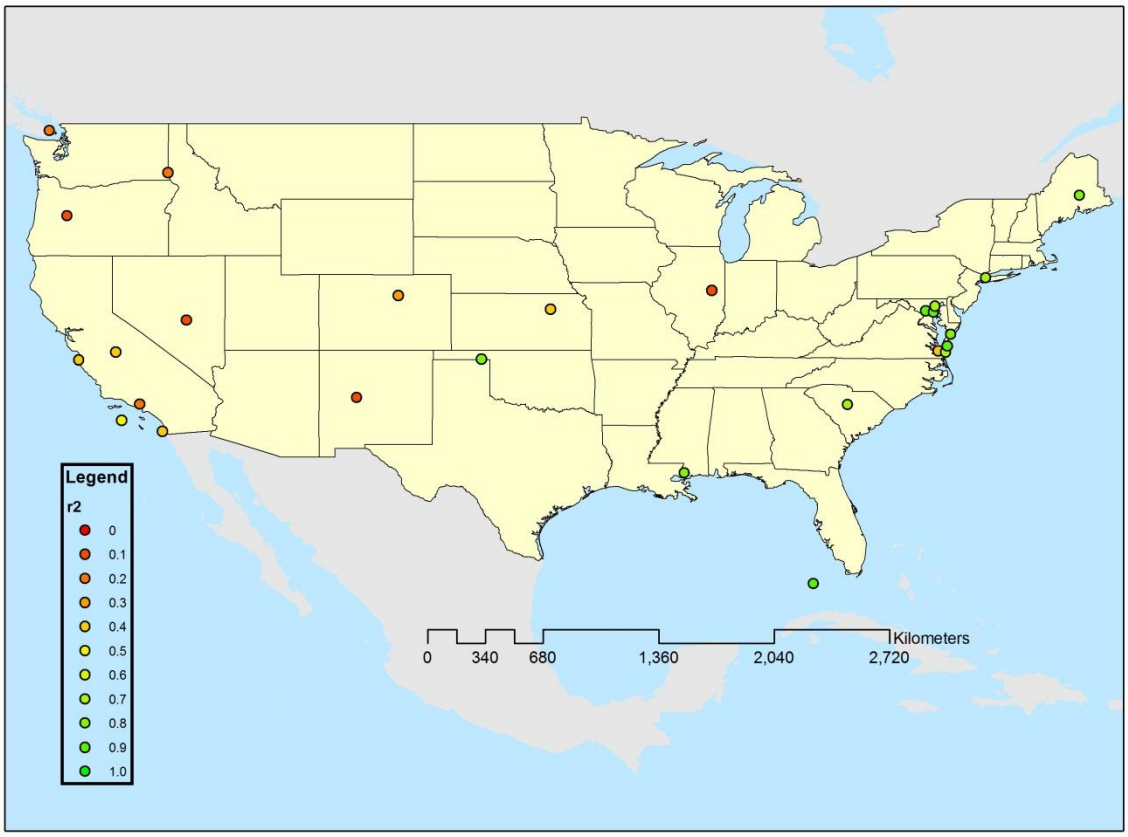

Figure 4.17 Coefficient of determination of *τanet* versus *τmodis* for May 2002

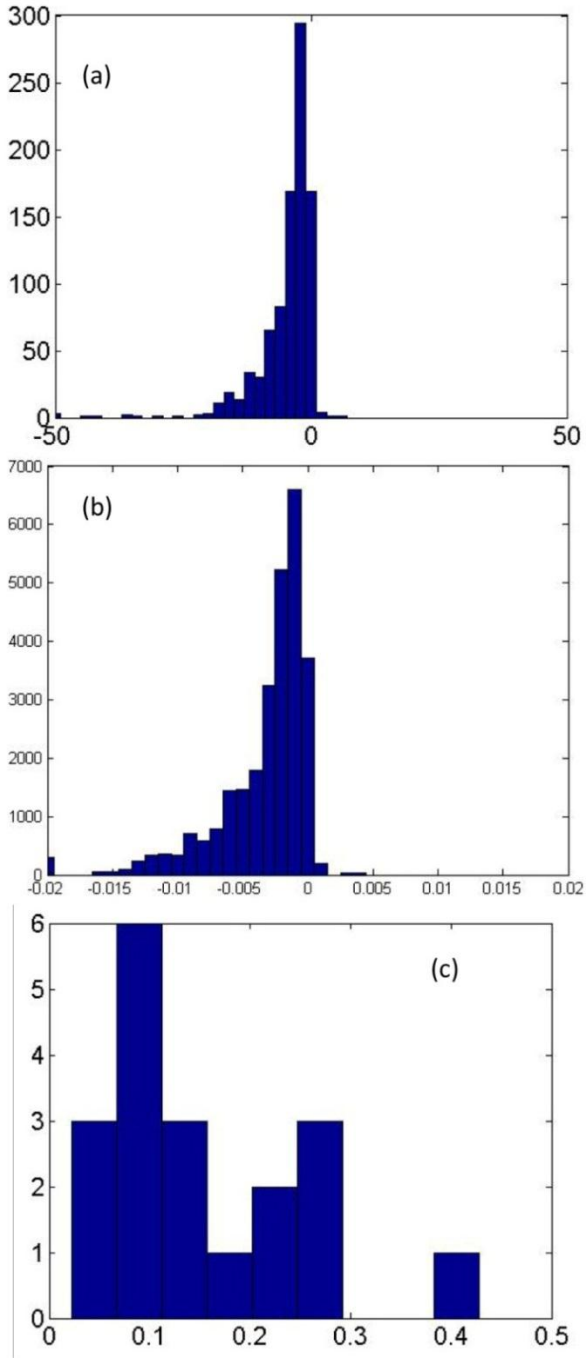

Figure 4.18 Histograms for biases in PMcoarse (model-observation) in aerosol mass (a) units of μg m-3 and after conversion to AOD units (b). Panel (c) shows *τmodis – τanet*. Figures are based on April-June in the Pacific.

measurements were correlated in eastern regions of the United States, but weakly to moderate correlated in the western regions, specifically in Los Angeles (California), Portland (Oregon), and Salt Lake City (Utah), but Denver (Colorado) showed a moderate correlation in comparison to the others specified. *Abdou et al.* [2005] attributed τ_{modis} positive bias (relative to τ_{anet}) to surface brightness over deserts and subpixel water contamination at coastal sites. *Drury et al.* [2008] reduced positive τ_{modis} bias in the western U.S. by using alternate surface reflectances in the 0.65 to 2.13 μm wavelength region. *Hu* [2009] used τ_{modis} to estimate $PM_{2.5}$ surface concentrations in locations with sparse surface $PM_{2.5}$ monitoring. Regressed $PM_{2.5}$ values (with τ_{modis} as the primary regression input) were more strongly correlated with $PM_{2.5}$ in the eastern U.S. than the western U.S.

4.4.2 Alternate input data options

Reducing observed AOD bias would likely improve the results of OI over the western U.S. Alternatives to MODIS exist, such as GOES and the Multi-angle Imaging SpectroRadiometer (MISR) [*Diner et al.*, 1998], which has the ability to capture images from nine angles, compared to for a single image at nadir viewing angle for MODIS. Several investigators have used MISR to infer $PM_{2.5}$ data over the U.S. and the world. *Liu et al.* [2004], in a comparison of MISR-derived $PM_{2.5}$ concentrations with surface observations and model predictions, reported the ratio of $PM_{2.5}$ (from MISR) to observations was 0.9-1.1 in the eastern U.S. and 0.5-0.75 in the western U.S. Simultaneous use of MISR together with MODIS is a common approach. *Chatterjee et al.* [2010] compared MODIS and MISR to AERONET in a geostatistical fusion approach; their work showed annual correlation coefficient versus AERONET of 0.63 (MISR) versus 0.3 (MODIS) over the western U.S. *Mishchenko et al.* [2009] found that MISR bias (versus AERONET) was slightly better than that of MODIS. Van *Donkelaar et al.* [2010] used MODIS and MISR AOD, coupled with AOD to $PM_{2.5}$ conversion

factors from GEOS-Chem, to derive gridded 6-year average global $PM_{2.5}$ concentration estimates. Satellite observations suspected of excessive bias (due to high surface albedo or poor comparison with AERONET) were excluded. *Van Donkelaar et al.* [2010] also found MISR more reliable than MODIS for the western U.S. These suggest that the use of a selective combination of MISR and MODIS AOD could improve OI performance over the Western U.S.

The degree of improvement from infusing MISR AOD into the current OI technique is uncertain. For example, although Van Donkelaar et al. used primarily MISR AOD in the Western U.S., but normalized bias [(model regressed data – observation)/observation] in the Western U.S. remained at about +1.0, versus a value of -0.1 to +0.1 over the eastern U.S. The increase in normalized bias is in part due to the cleaner conditions over many parts of the Western U.S., such that relatively small absolute bias (e.g. 2 μg m^{-3}) can cause large normalized bias.

The MODIS deep blue algorithm [*Hsu et al.*, 2006] may decrease errors due to ocean glint and high surface reflectance, common in regions with high dust concentrations [*Hsu et al*., 2004]. Deep blue MODIS AOD has been favorably evaluated over desert regions of Asia [*Hsu et al.*, 2006]. For the southwestern U.S., *Drury et al.* [2010] did not find any significant difference between deep blue and improved MODIS retrievals [*Drury et al.*, 2008]. However, use of MODIS deep blue product over the western U.S. should be further investigated.

Dynamic OI may also improve assimilation results relative to the static technique adopted in the current work. Static OI refers to the case where the posterior result from OI does not influence future model states. In dynamic OI, the $PM'_{2.5}$ fields influence the initial conditions for simulation of the subsequent period. *Adhikary et al.* [2008] compared static and dynamic OI over South Asia over a one month simulation. *Collins et al.* [2001] also used dynamic assimilation with OI for a two week period over South Asia. The drawback of the dynamic OI using MODIS is the sparseness of daily MODIS

retrievals, and geostationary satellites with improved temporal coverage are more appropriate for dynamic assimilation.

Improved AOD datasets, that have been subjected to enhanced quality control, bias correction, or advanced inversion, can often be more effective to use for data assimilation. One improved dataset is the MODIS level 3 product [http://gdata1.sci.gsfc.nasa.gov/daac-bin/G3/gui.cgi?instance_id=MODIS_DAILY_L3], and another has recently available from the U.S. Naval Research Laboratory [http://www.nrlmry.navy.mil/aerosol_web/]. The USNRL product has reduced systematic and random errors [*Hyer et al.,* 2010]. One potential drawback in these publically available products is their one degree spatial resolution.

4.4.3 Intercomparison of AOD OI studies

A number of studies are available for comparison to the current work. However, previous OI applications have focused on aerosol AOD and climate forcing (rather than $PM_{2.5}$), and have used AOD as the main variable for tuning OI settings and for performance evaluation. Therefore, we compare changes in AOD due to OI (rather than changes in $PM_{2.5}$) to allow comparison to studies that reported AOD performance statistics but not $PM_{2.5}$ statistics. *Collins et al.* [2001] used a dynamic OI model over the period of two weeks to improve model estimates over the Indian Ocean. The main goal of this study was to improve the model short term estimates for aerosols. *Carmichael et al.* [2010] studied the effect of static OI for the period of four years over Asia using quality assured level 3 MODIS AOD, and *Park et al.* [2011] assimilated satellite AOD to adjust model AOD over East Asia. Park et al. used season-specific observation and model errors.

In the Asian studies, mean τ_{anet} was typically between 0.1 and 0.2 over ocean and 0.3 to 0.5 over land, while the corresponding value for the CONUS domain of the current work was 0.2. All studies showed an increase in the coefficient of determination, by

~0.14±0.04 in the Asian studies and by 0.07 in the current work. All studies showed a reduction in the absolute value of bias, with a reduction of 0.25±0.15 AOD units in the Asian studies, and a reduction of 0.08 in the current work. Studies with the largest initial errors saw the largest corrections from assimilation. The magnitude of bias improvement seen over land in Asia was not seen to the same extent in the current study. We hypothesize that part of the reason for this is that we optimized settings for $PM_{2.5}$ improvement rather than for AOD improvement. The relationship between the two is not necessarily linear [*Kumar et al.,* 2008; *Schaap et al.,* 2009; *Engel-Cox et al.,* 2005], and greater improvement would be achieved if settings were optimized to reproduce τanet. Furthermore, the lower levels of AOD and aerosol burden in the CONUS domain limits the amount of improvement to be expected from assimilation of τmodis. On the other hand, the improvement in this study is dependent on the discarding of month-region specific results where the posterior degrades performance relative to surface PM2.5. Without this replacement, AOD performance relative to AERONET may have decreased in many months and regions.

4.5 Conclusion

Data assimilation was used to produce $PM_{2.5}$ estimates from a combination of $PM_{2.5}^{cmaq}$, τ_{modis}, and τ_{anet}. The specific assimilation technique was optimal interpolation (OI) of AOD performed on the CMAQ model grid. Results were compared stemming from four different sets of parameters for background (e.g. CMAQ) error covariance and observation errors. Furthermore, results were compared using two different methods for application of scaling factors calculated from OI. The methods were applied for 2002 over the continental United States, and six separate regions were used for evaluation. Our results show that OI can be implemented using two successful optimization approaches. The first optimization scheme uses spatially- and temporally-invariant settings and simple temporal (monthly) averaging of model and satellite data. The second approach uses

customized OI settings for specific regions or time periods. Averaging and correcting all model hours (case a) was the method that most frequently produced superior error scores. In both schemes, it was critical that the assimilation result was not used in cases where assimilation degraded model-measurement agreement. These "degradation" assimilation cases occurred in 38% of the possible realizations.

Multiple features in the OI results are noteworthy and are listed as follows: Setting 4 in Table 4.3 was selected by the optimization routine most often, which was the combination of errors giving the highest weight to the model and the lowest weight to observations. The model error correlation length of 36 km led to superior posterior error scores more frequently than other lxly settings (up to 150 km was investigated, but this variable is not independent). The improvement from the prior to the posterior $PM_{2.5}$ was greatest in magnitude and spatially most consistent during summer. One possible reason for this is that summer has the largest and most consistent negative bias in the prior and OI typically led to increases in $PM_{2.5}$. Alternately, the summertime AOD observations may be more accurate and noise-free due to minimized snow cover. The OI algorithm as implemented in the current work improved bias more than error. Using model AOD only during hours with observations gave significantly different scaling factors and overall results than averaging all hours of CMAQ output. Averaging and correcting the overpass hours (method c) was sometimes the method that lead to the best error score, but the interpretation of this finding is not clear because correcting the overpass hours only made a much smaller change in monthly averaged $PM_{2.5}$ than the other techniques. In some month/region combinations, a small correction exceeded the performance of larger corrections from the method c (average and correct all hours).

Investigation of the settings that produced the best assimilation results was initially done looking independently at each month and region. This led to an improvement in fractional error score from 1.2 to 0.97 at IMPROVE sites, and from 0.99 to 0.89 at STN sites. The analysis was repeated with constraints such as requiring a single

method and setting for all months in a specific region, or for all months and all regions (i.e. a single year-long nationwide method and setting). As long as months with degraded $PM_{2.5}$ performance were identified and analyzed with their prior $PM_{2.5}$ values (instead of their posterior $PM_{2.5}$ concentrations), the results were similar to the results from the approach with independent settings for each month and region. Most of the instances where OI degraded model skill occurred in the Pacific and Mountain regions.

The results show that the assimilation has different impacts in different regions as well as on different months of the year. OI achieved modest improvements in most performance metrics, but the eastern regions of the U.S. showed more improvement than the western regions. We believe the poor performance over the Western U.S. was mainly caused by MODIS random and systematic errors, although the impact of lateral boundary conditions needs more study. The possibilities of missing biomass burning emissions, negative bias in modeled coarse PM, and the use of a clean boundary condition, were evaluated. The first two potential reasons are likely not the main cause of the limited success of OI in the Western U.S.

Recommended areas for future improvement in satellite data assimilation for aerosol estimation include the use of improved MODIS retrievals as proposed in collection 6, and the integration of AOD from both MISR and MODIS.

CHAPTER 5: SURFACE DATA ASSIMILATION RESULTS

5.1 Introduction

Data assimilation using surface measurements has been shown to improve model estimates for both optimal interpolation [*Tombette et al.*, 2009, *Wu et al.*, 2008] and more advanced data assimilation methods [*Chai et al.*, 2007, *Hakami et al.*, 2005, *Pagowski et al.*, 2010, *Liu et al., 2011*]. The novel part of this work is the implementation of more advanced methods to calculate the model and observation error covariance for OI using methodologies proven to be effective in 3D-var and 4D-var data assimilation studies. In addition, three cross validation methods, usually performed in regression model studies, [*Kim et al.*, 2009, *Sheppard et al.*, 2005] are used to evaluate the OI performance based on the error covariance matrices calculation performed. The methodologies for OI implementation are described in section 5.2. This includes the optimal interpolation equations used, the observation methods used to calculate the model error covariance, and the representation error that is used to calculate the observation error covariance. Cross validation methods are viable for chemical transport model data assimilation evaluation and previous studies show the importance of implementing these methods.

The results, presented in section 5.3, illustrate first the observation method calculations followed by an OI algorithm case study for January 2002. Detailed spatial inputs and outputs are presented as physical and mathematical evidence of OI performance. The annual evaluation of OI for the continuous U.S. is described in addition to the sensitivity test performed to evaluate the minimum number of available observations required to provide an improved posterior model estimate. The chapter further discusses OI performance over the continuous U.S., presenting the positive OI performance results and comparing them to previous studies. An additional insight is provided for the utilization of the current methods to calculate the error covariance

matrices and the cross validation methods to further advance data assimilation methods such as 3D-var.

5.2 Methodology

After regridding surface $PM_{2.5}$ measurements to the model domain, OI was performed between $PM_{2.5}$ monthly averages from both surface measurements and the prior model values. The statistical analysis was based on the monthly averages for the whole domain (i.e. pairing monthly averages of the model-observations). A cross validation method was used to evaluate the surface optimal interpolation method, which included using both stratified and simple random validation.

5.2.1 Surface data assimilation via optimal interpolation

The same OI equation was implemented from section 4.2.3 but different settings, models, and observation error covariance matrices were used. The model error covariance matrix was derived by applying the observational method that assumes a network of dense surface measurements [*Hollingsworth and Lonnberg*, 1986]. First the description of the observational method is given followed by the optimal interpolation equations used.

5.2.2 The observational method

The following method was adapted from the Chai et al. [2007] assimilation of surface ozone measurements into the STEM model using 4D-var for the eastern U.S for 12 hours in July 2004. The assimilation was evaluated using surface ozone data withheld from the assimilation process. Initial O_3 concentrations were used as the control variable to minimize the cost function. The model error covariance was calculated using both the NMC method [*Parrish and Derber*, 1992] and the observational method [*Hollingsworth and Lonnberg*, 1986]. The assimilation results showed an improvement relative to prior model values.

The observational method outputs provide average correlations as a function of distance. To insure that the correlation is unbiased, the biases are then removed from the correlation using two methods. The first calculates the bias of each site individually during each month and subtracts the annual average. The second calculates the mean bias (all sites) for each month and subtracts that. This is shown in equations (5.1) and (5.2), respectively.

$$b_{im1} = b_{im} - \frac{1}{12}\sum_{m=1}^{12} b_{im} \qquad \text{m=1 to12; and a fixed site } i \qquad (5.1)$$

$$b_{im2} = b_{im} - \frac{1}{N}\sum_{i=1}^{N} b_{im} \qquad i\text{=1 to N; and a fixed month } m \qquad (5.2)$$

where

b_{im} is the model-observation difference for a specific site *i* and month *m*

b_{im1} is the model-observation difference for a specific site *i* and month *m* with bias removed using equation 5.1.

b_{im2} is the model-observation difference for a specific site *i* and month *m* with bias removed using equation 5.2.

As mentioned both bias removal procedures were tested, and both equations gave similar results, but equation 5.2 gave a slightly higher correlation, therefore b_{im2} is used in the following equations.

The observational method [*Hollingsworth and Lonnberg*, 1986] calculates the average correlation between observational increments in two model grid cells i and j [Chai et al, 2007], which is shown in equation 5.3:

$$R_{ij} = \frac{\overline{\left(PM_o^i - PM_m^i - b_{im2}^i\right)\left(PM_o^j - PM_m^j - b_{jm2}^j\right)}}{\sqrt{\overline{\left(PM_o^i - PM_m^i - b_{im2}^i\right)^2}.\overline{\left(PM_o^j - PM_m^j - b_{jm2}^j\right)^2}}} \quad (5.3)$$

where

PM_o is the $PM_{2.5}$ surface observation (μg m^{-3}),

$PM_m(c)$ is $PM_{2.5}$ model prior estimates (μg m^{-3}), and the overbar '—' denotes averaging over time.

Then parameters in equation 5.4 are fitted by the nonlinear least square method:

$$R_z.exp\left[-\frac{l_{xy}^p}{l_h^p}\right] \quad (5.4)$$

where

R_z: Correlation between the observation and prior model increment at two stations with horizontal distance (l_h), calculated at y intercept.

lxy is the distance between two model grid cells i and j.

l_h is the correlation length scale in km.

p is a fitted exponent (5.4).

5.2.3 Optimal interpolation equations

The optimal interpolation equations are similar to section 4.3 as follows:

$$M' = PM_m + K(PM_o - HPM_m) \quad (5.5)$$

Equation(4.4) is used to calculate *K*

$$B_{ij} = \left(f_{cmaq}PM_{cmaq} + \varepsilon_{cmaq}\right)^2.R_z.exp\left[-\frac{l_{xy}^p}{l_h^p}\right] \quad (5.6)$$

$$O = \left(f_oPM_o + \varepsilon_o + \varepsilon_{repr}\right)^2 I \quad (5.7)$$

$$\varepsilon_{repr} = \frac{\varepsilon_{obs}}{2} \cdot \sqrt{\left(\frac{\Delta x}{L_{repr}}\right)} \quad (5.8)$$

$$L_{repr} = L_{min} + (L_{max} - L_{min})(1 - F_{urb}) \quad (5.9)$$

where variables are defined as follows:

f_{cmaq} : model fractional error

PM_{cmaq} : $PM_{2.5}$ prior model estimate (μg m^{-3})

ε_{cmaq} : model minimum error (μg m^{-3})

f_o : Observation fractional error

PM_o : $PM_{2.5}$ surface observation (μg m^{-3})

ε_o : Surface minimum error (μg m^{-3})

ε_{repr} : is the representation error (μg m^{-3}).

ε_{obs} : is the absolute observation error.

Δx : model grid cell size (km).

L_{repr} : represents the radius of influence of an observation (km).

L_{min} and L_{max} : are the minimum and maximum radius of influence for the observation at rural and urban locations, respectively.

F_{urb} : fraction of urban area in a grid cell.

Pagowski et al. [2010] parameterized representativeness error using equation 5.8, a formula from Elbern et al. [2007]. It is important to control the observational spatial length influence (equation 5.9) based on locations. For instance urban sites should have less radius of influence while rural areas should have larger influences. The representative length here is different from Pagowski et al. [2010], which uses one minimum and one maximum fixed length. Rather, here the urban area fraction adjusts the difference between the rural and urban locations that adjusts the length from the minimum urban length (equation 5.9). Figure 5.1 graphs representation error as a

function of urban fraction in a grid cell. The representation error ranges from 0.8 μg m^{-3} for rural areas to 4.2 μg m^{-3} for urban areas.

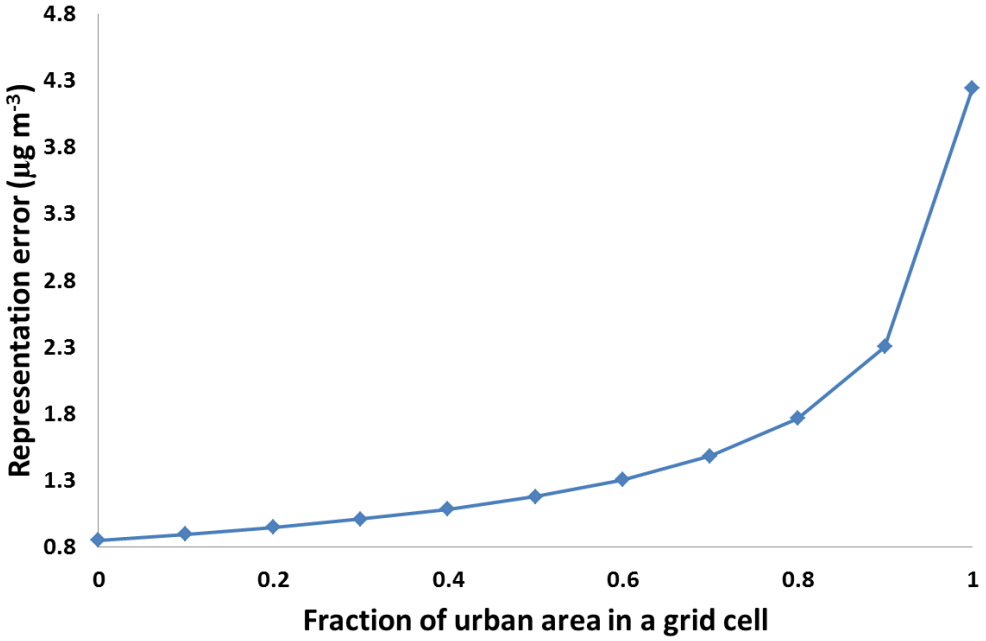

Figure 5.1 Correlation between the representation error and the fraction of urban area in a grid cell.

Parameters for equation (5.6) were calculated using the observational method, where the monthly average bias (model-observation) [equation 5.1 and 5.2] for all the sites was calculated for 12 months, obtaining 12 inputs for equation (5.3). Each one of the 12 months had more than 700 bias-corrected observational increments, which are the amount of sites after regridding to the domain. From equation 5.3 the average correlations were calculated over distance and the parameters for equation (5.6) were fit. The model fractional error and minimum error were constant values of 0.35 and 1μg m^{-3}, respectively. The observation fractional error (0.05) and minimum error (0.4 μg m^{-3}) were adapted from USEPA [2012], a 3-year quality assurance report for 1999 through

2001. The errors were based upon the differences between two networks that measure surface concentration and their median absolute error, shown in Figure 3.2 of the report. For absolute observation error and for the minimum and maximum radius of influence for the observation at rural and urban locations, we adopt values from Pagowski et al. [2010], with values of 2μg m^{-3}, L_{max} = 50km, and L_{min} = 2km, respectively.

5.2.4 Cross validation methods

Cross validation [*Browne*, 2000] is a statistical method used to validate performance of a forecast model, regression model, or other model. The method helps inform the validity of a certain procedure towards a sample of data. This is accomplished by dividing a dataset into two samples; the first sample is used towards weight calibration in a regression model; the calibrated weights are then used to validate the criteria of the regression model proposed using the second set.

Typically [*Browne*, 2000], the data is divided into two parts. One is used for training the regression model parameters, while the other is used to validate the statistical accuracy of the regression model. It is important however to leave 70%-90% of the sample data for the regression model to accurately represent the state of the system. Therefore multiple methods have been proposed for removing data for validation, such as the method of leaving out one at a time for a sample size of N and repeating the cross validation N number of times, but this is computationally expensive. Another method is to simplify this by leaving out X (some) at a time, where validation samples have to be greater than one. The validation [*Gilbert*, 1987] data has to be withheld randomly from the sample data each time, which can be accomplished by the simple random method and the stratified method.

Cross validation has been used in $PM_{2.5}$ exposure models, $PM_{2.5}$-AOD correlation, and land use regression modeling. In two different studies, *Paciorek et al.* [2009, 2007] withheld 10% of the training observations for land-use regression model cross validation.

Ross et al. [2007] and *Henderson et al.* [2007] used a leave-one-out cross-validation method. Finally, *Kumar et al.* [2011] used both the leave-one-out cross-validation method and division of the data set randomly into a training data set and a withheld dataset. Kim et al. [2009] validated a $PM_{2.5}$ exposure model for Los Angeles and surrounding areas using average area-wide monitors. Exposure was based on monitors closest to the subject or by kriging the observation monitors over the domain. The correlation was implemented by selecting 2000 residential locations and stratified for up to 3 random selections for each stratum. Sheppard et al. [2005] used a personal exposure model to study the effect of $PM_{2.5}$ on areas with increased air pollution in Seattle. The cross validation included both simple and stratified random sampling in addition to averaging of all individuals in the population.

5.2.5 Surface optimal Interpolation methods and settings

The surface assimilation was conducted similarly to the satellite assimilation, using monthly averages of PM_o and PM_m. Over 1500 $PM_{2.5}$ surface measurement sites were available in the US for 2002 [*U.S Environmental Protection Agency (EPA), 2002*]. All hourly data was averaged to 24 hour averages. Cross validation was performed 3 times using slightly different procedures for selecting the monitor values for cross validation. In all three cases, the cross validation was performed 10 times, with 10% data withholding in each repetition. Withheld sites were sampled randomly without replacement. Therefore, each monitor value was used in one (and no more than one) of the 10 repetitions. Comparing ten repetitions informed the validity of withholding different random datasets towards the same outcome. Validation of the current OI implementation is achieved by comparison between all three methods. A conclusion that the methodology was correctly implemented would be drawn if all cross validation methods managed to improve the model estimates. In contrast, if one or more cross

validation methods disagree in the final conclusion, then more investigation is required to understand the discrepancies.

The three cross validation methods are summarized in Table 5.1. To compare the prior and posterior model results from equation 5.5, we followed Sheppard et al. [2005] for stratified and unstratified cross validation methods. It is important to mention that all three cross validation methods in Table 5.1 validated the withheld data after averaging the observations. For the first two methods, the data is averaged before performing cross validation. For the third method, the withheld data is averaged (if more than one data point is withheld) before validation.

Table 5.1 Cross validation methods

	Cross validation method	Description
1	Stratified-gridded	Using the stratified random sampling after regridding observations to the domain
2	Unstratified-gridded	Using the unstratified random sampling after regridding observations to the domain
3	Unstratified-ungridded	Using the unstratified random sampling before regridding observations to the domain

The stratified-gridded cross validation method assumed that each stratum should have the same sampling amounts of data withheld, proportional to the other stratums (proportionate stratification). This is accomplished by using the same fraction withheld for each stratum, regardless of the amount of data points. As an example for three stratums with 4, 8, and 12 data points and a 50% sampling fraction, the amount of data points withheld would be 2, 4, and 6, respectively. The amount of strata was calculated based on a 9x9 grid cells. For validation, the monthly average posterior and prior model

concentrations were compared with the withheld monthly average surface measurements. All statistics in this chapter are based on comparison to withheld data; data to inform the OI is never used in evaluation.

A separate set of repetitions were formulated based on the stratified-gridded cross validation method, but the strata have unequal numbers of data points (disproportionate stratification). The method randomly leaves one observation point out from all the strata at a time, until the polygon with the maximum number of observations is empty. The maximum number of observation in one stratum is 32 or 33 depending on the month. The stratum that removes its observations remains empty until the stratum with the maximum number of observations is empty. The method is repeated 10 times to check the effect of choosing random observations each time. The number of repetitions for January is 32 runs, each withholding data for 10 random times yielding 320 files. This method will inform the amount of observation data to be withheld, which will give a sufficient model improvement.

In contrast with the satellite assimilation, the moving window did not use a 5x5 grid cell size; rather, a sensitivity test was performed to accommodate the correlation length scale. These sensitivity tests included window sizes 5x5, 9x9, 13x13, 19x19, 25x25, and 31x31 to choose the appropriate window size. Tests were compared by model skill evaluation at withheld monitors; the same monitors were withheld for each window size for valid intercomparison. Model skill metrics included correlation coefficient, root mean square error (RMSE), factional error (FE), and normalized mean error (NME).

5.3 Results

5.3.1 Observational method results

Figure 5.2 shows the average correlation coefficient as a function of distance between observations (explained in section 5.2). Equation 5.4 had a 0.7 intercept

(i. e. the zero intercept of the curve $R_z = lim_{\Delta h \to 0} R(\Delta h) \approx 0.7$), a 260km correlation length scale, and a 1.5 power. Equation 5.4 can be written as:

$$0.7.\, exp\left[-\frac{l_{xy}^{1.5}}{260^{1.5}}\right]$$

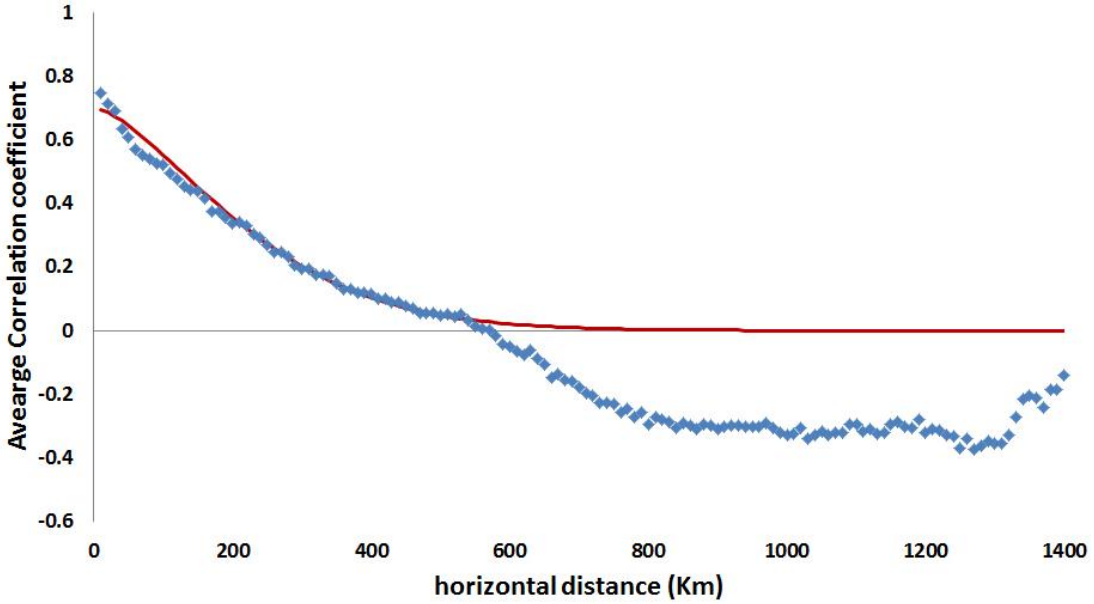

Figure 5.2 Average correlation coefficient over distance between observational increments and two models grid.

The fitted line does not fall below zero despite the negative data because equation 5.4 cannot give a negative value.

5.3.2 OI working or moving window.

Optimal interpolation is typically implemented over a limited spatial window. In chapter 4, a 5x5 window was used. In this chapter, windows of variable size are explored. Varying the size of windows is motivated by the fact that error-covariance in the model extends out to large distances (e.g. Figure 5.2). Figure 5.3 compares prior and

posterior error metrics and computational effort as a function of the OI window sizes. Specifically, performance metrics (i.e. FE, R^2, and NME) are found in Fig. 5.3a, RMSE in Fig. 5.3b, and log of machine run time in Fig. 5.3c. The error results are very similar for the 9x9 to 31x31 windows, while the run times increase dramatically as the window size increases. Taking into account both metric values and OI run times, the 9x9 window was chosen as an optimal value for all the runs, and is used in the remainder of this chapter.

5.3.3 OI algorithm case study for January 2002

The example for the Satellite OI was taken for the month of May because of the high AOD spatial coverage compared to January (due to snow and cloud cover). For the surface OI, all the months could provide substantial information, so the example focused on the first month of the year. Figure 5.4 shows the fractional urban percent of grid cells according to the meteorology data by MCIP. Figure 5.5 maps $PM_{2.5}$ concentrations for OI inputs (5.5a, 5.5c) and OI output (5.5e) for January 2002. The output example shown is calculated using 90% of the available grid cells with available data, with the withheld grid cells chosen according to the stratified-gridded cross validation method (described in section 5.5). Panels on the right are the frequency distributions of surface $PM_{2.5}$ and model-surface paired $PM_{2.5}$ from the spatial plots. Model values (prior and posterior) are paired with observational values, since the observational measurements were relatively few (only about 700 grid cells) compared to the model grid cells (n=20412).

Mean values for the whole domain for Figs. 5.5a, 5.5b, and 5.5c are 11.2, 3.6, and 3.4 μg m^{-3}, respectively. The paired model observation value for Figs. 5.5b and 5.5c are 12.25 and 10.64 μg m^{-3}, respectively. Figure 5.6a and 5.6b show that CMAQ posterior values are decreased mostly in the eastern parts of the U.S and increased mostly in the western parts of the U.S.

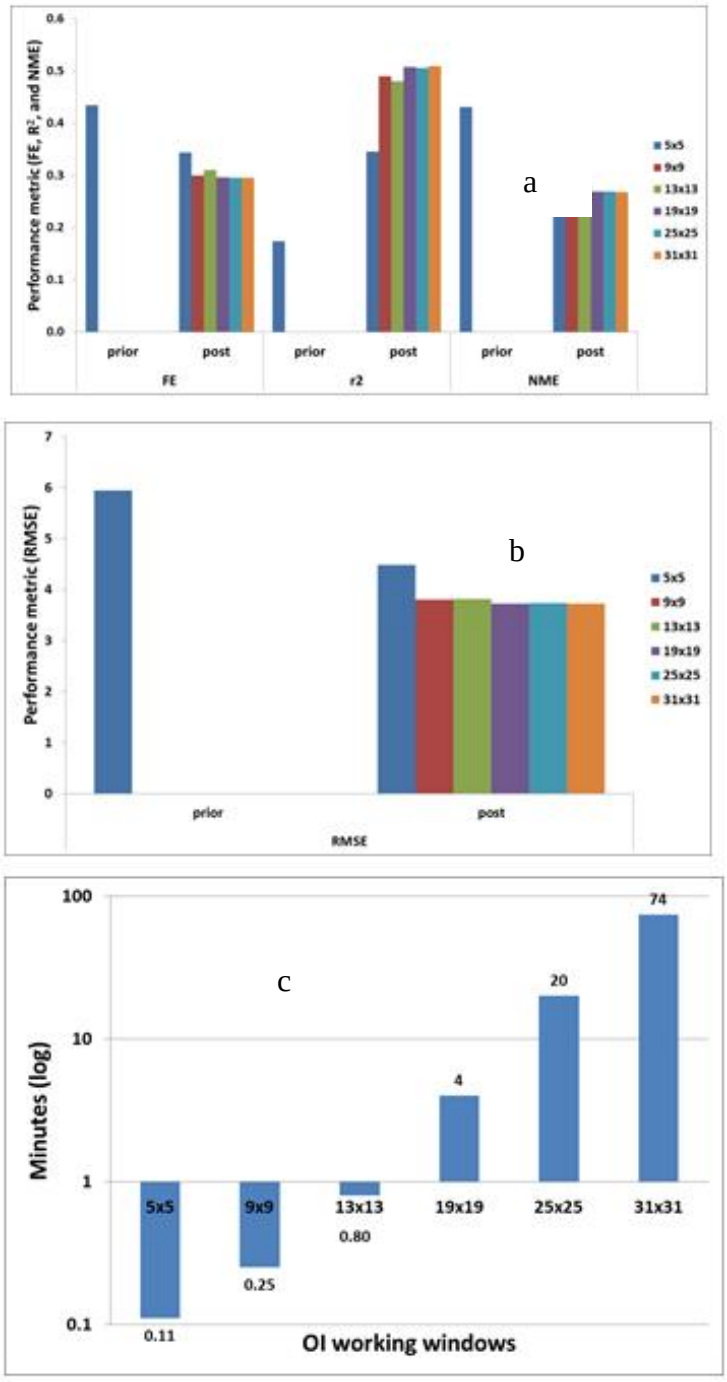

Figure 5.3 Sensitivity test Metric for different OI working windows (Figures (a) and (b)) and the OI duration time for each window (figure c, by fixing the random withheld input data, calculated using stratified-gridded cross validation in section 5.5.

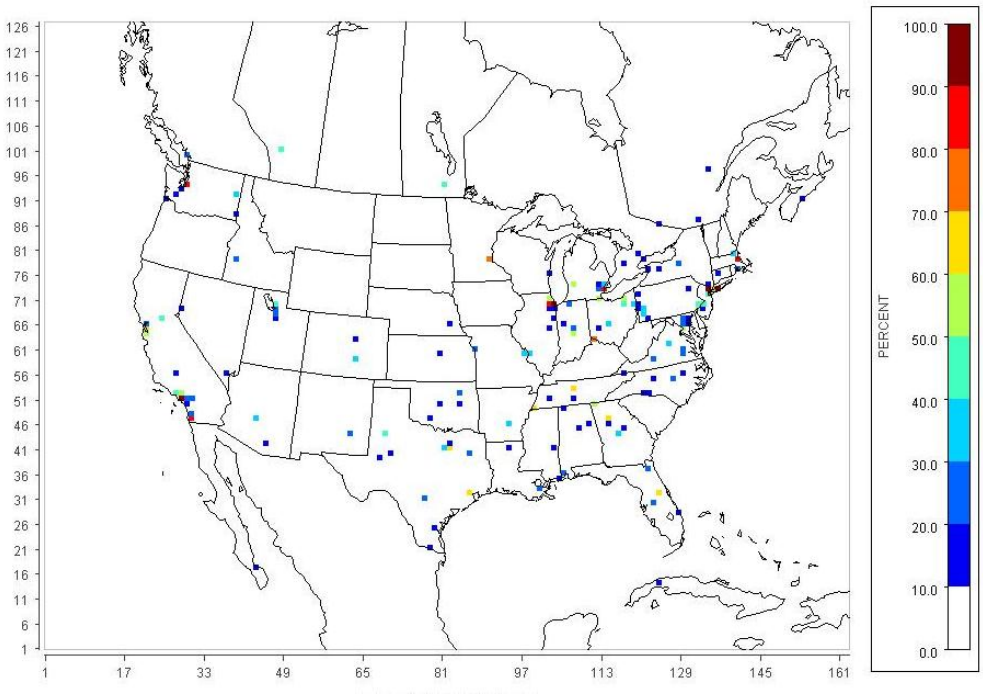

Figure 5.4 Fractional urban percent of cell based on land

OI improved the forward model estimates and decrease the amount of model-observation bias towards zero for the month of January. The highest decrease in $PM_{2.5}$ for the model is shown around the Boston, MA, area, and the highest increases are shown around Salt Lake City, UT. The histograms in Figures 5.5 and 5.6 show that the posterior model starts to take the shape of the surface measurements, and that the average $PM_{2.5}$ decreases over the domain. Although the increases are more substantial (up to 25 μg m^{-3}) compared to decreases (as low as -17 μg m^{-3}) the overall frequencies of the decreases are higher.

OI predominantly corrects CMAQ positive bias (i.e. overprediction) in the eastern U.S. and CMAQ negative bias (i.e. underprediction) in the in the western U.S. The model had a negative bias for $PM_{2.5}$ along the U.S.-Mexico border in the southwestern U.S.

(discussed in section 4.17). Overall the model changes are spatially spread between -5 to 5 µg m^{-3}, where Figure 5.7 shows a scatter plot for the withheld model-observation pairs for prior and posterior concentrations, Figure 5.5c and 5.5e, respectively. Figure 5.7 shows the posterior values shifted towards the one to one line, with an improvement in the correlation coefficient from 0.05 (prior) to 0.27 (posterior). The values reflect only one random test out of ten repetitions, where the maximum improvement for the month of January was 0.3 to 0.9 for prior and posterior values for one of the tests. The mean correlation for the ten random runs, in addition to other metric values, will be shown in the following section.

5.3.4 Application of surface OI to all months of 2002

The previously demonstrated OI algorithm was comprehensively applied to all months for the year 2002 for three cross validation methods described in section 5.5. Figure 5.8 demonstrates the annual (2002) metric evaluation for the three cross validation methods; stratified-gridded (a) and (b); unstratified-gridded (c) and (d); and unstratified-ungridded (e) and (f). Correlation coefficient, RMSE, FE, and NME are calculated for each month as the metric average of the ten random runs. The results show that the annual metric improvement is for all the months of the year, and this is true for all the random cases done for each month. The improvement over all months of the year was random and the pattern of improvements would change if cross-validation was repeated. Table 5.2 contains the annual average metric improvement for the three cross validation methods. The correlation improved from 0.36 to 0.76, the FE decreased from 43% to 15%, the NME decreased from 36% to 13%, and RMSE decreased from 5.39 to 2.32 µg m^{-3} for prior and posterior values respectively. The results are similar between all cross validation methods and could slightly change for a set of random runs. The three cross validation methods show similar results leading to a conclusion, as stated in section 5.5,

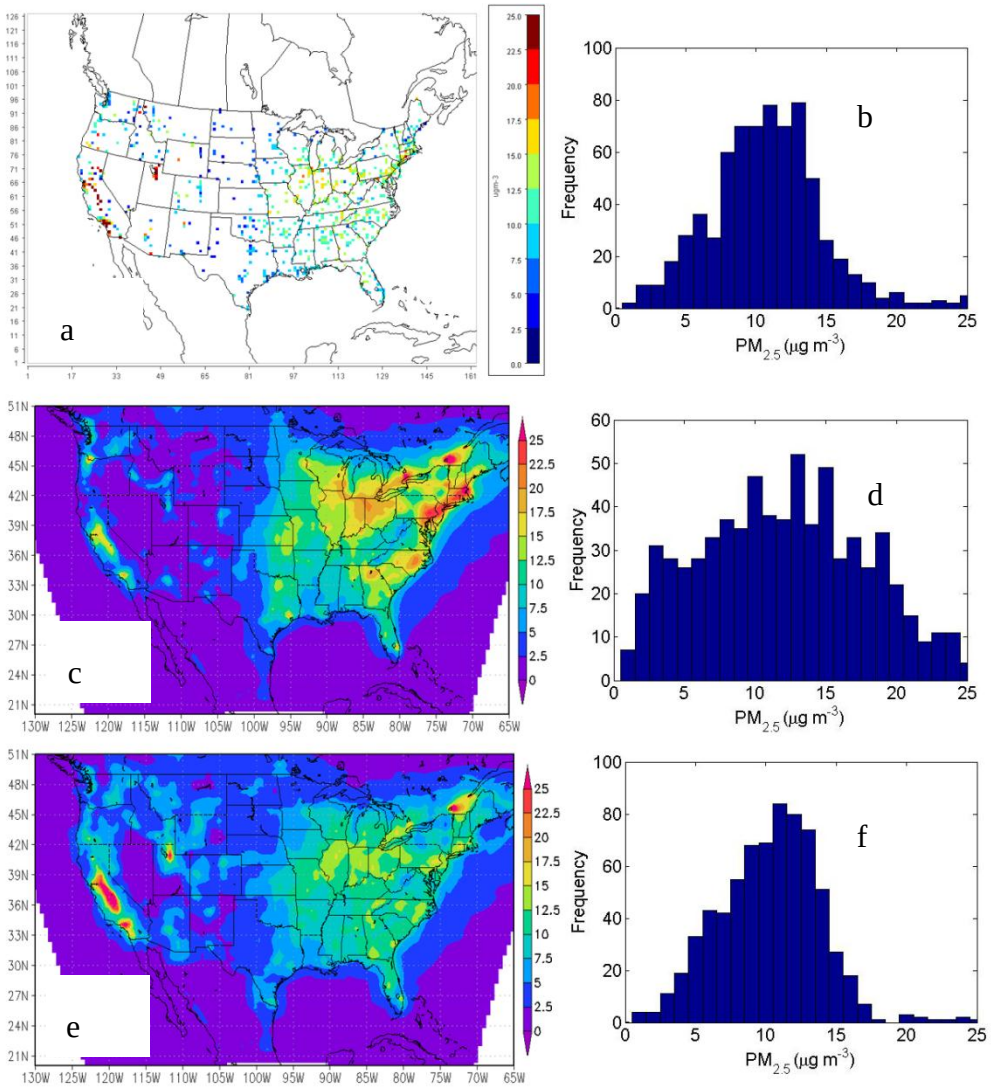

Figure 5.5 Example of the OI algorithm inputs and output (January 2002). (a) map and (b) histogram of monthly mean surface $PM_{2.5}$; (c) map prior model $PM_{2.5}$; (d) histogram of monthly mean of observation-prior model $PM_{2.5}$ paired; (e) map posterior model $PM_{2.5}$; and (f) histogram of monthly mean of observation-posterior model $PM_{2.5}$ paired, both e and f were calculated using stratified-gridded cross validation

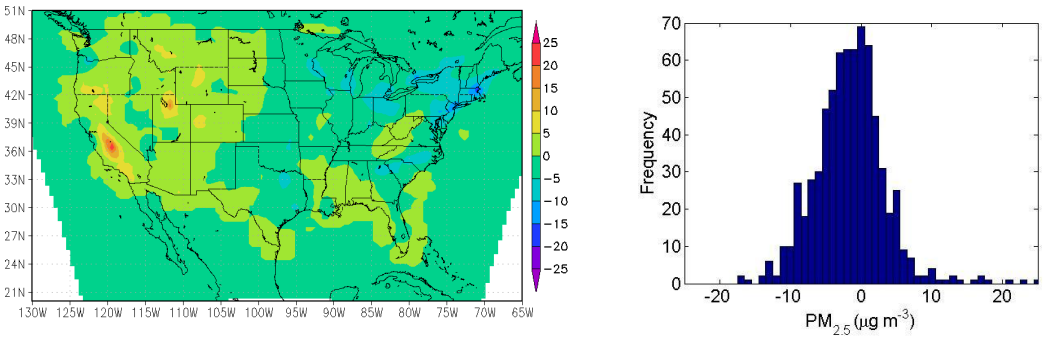

Figure 5.6 Example of the OI algorithm change (January 2002). (a) map and (b) histogram of monthly mean model $PM_{2.5}$ concentration, between posterior and prior values (i.e. Figure 5.5(e) – Figure 5.5(c)). Using stratified-gridded cross validation in section 5.5.

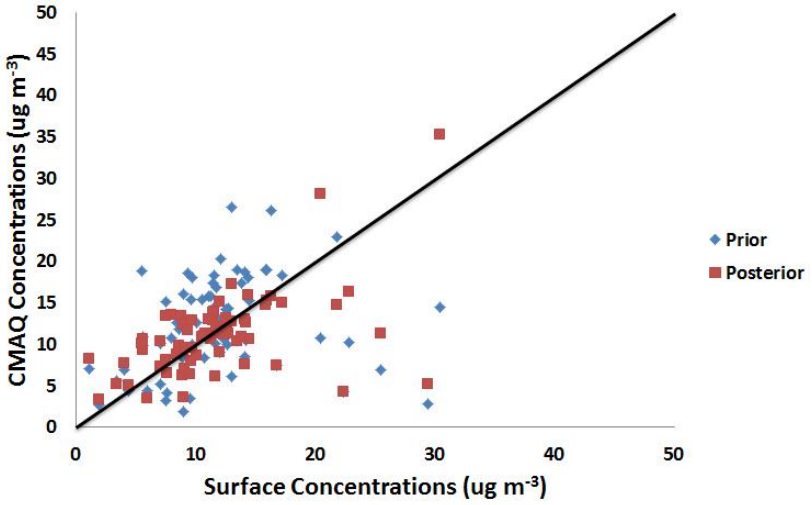

Figure 5.7 Scatter plot for the model-observation pairs (January 2002) of the prior and posterior concentrations, for Figure 5.5c and 5.5e respectively. Using stratified-gridded cross validation in section 5.5

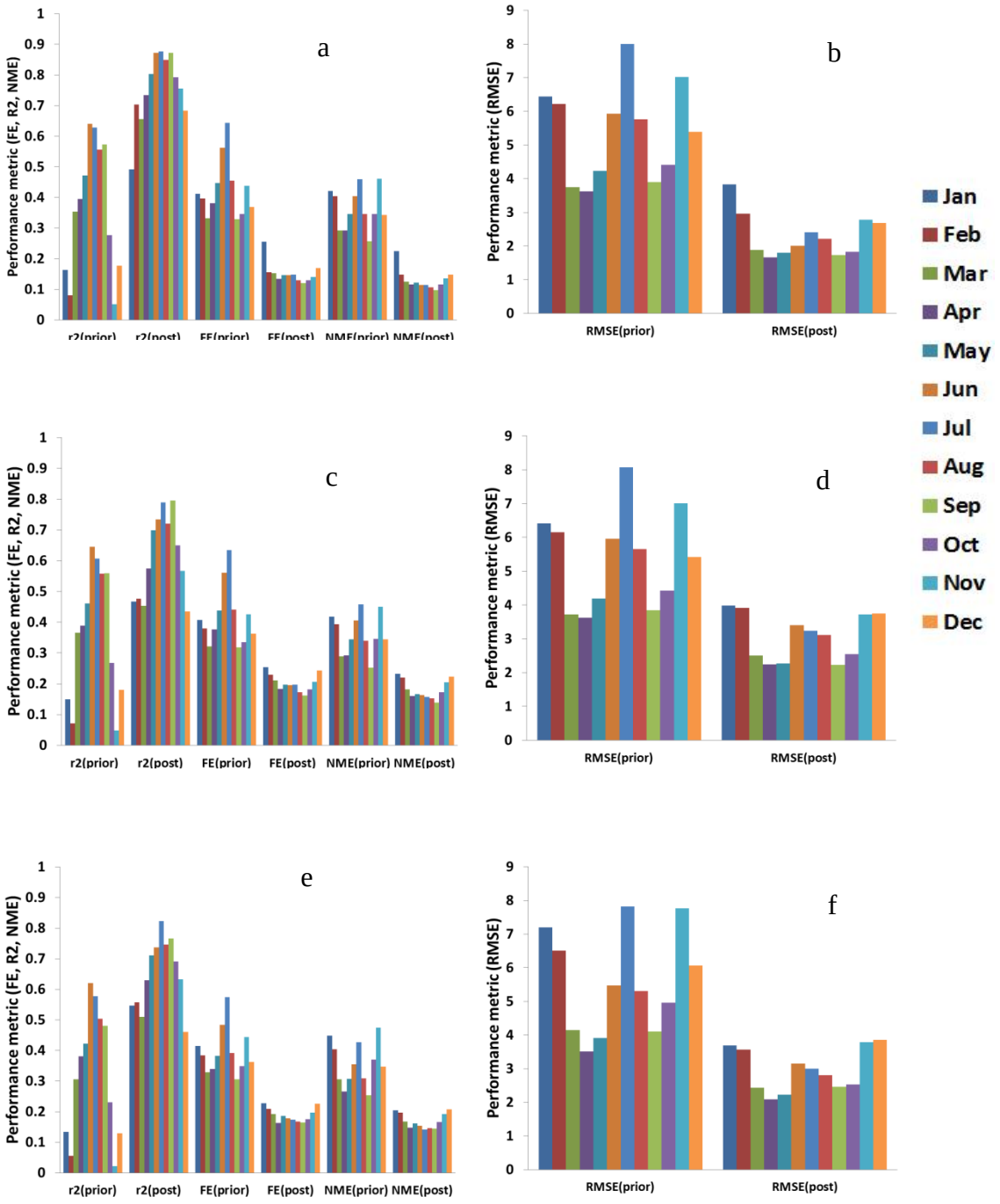

Figure 5.8 Annual (2002) metric evaluation for cross validation methods; stratified-gridded (a) and (b); unstratified-gridded (c) and (d); unstratified-ungridded (e) and (f), explained in section 3.9.

Table 5.2 Annual average metric improvement for the three cross validation methods.

	Strata_Gridded		Unstrata_Gridded		Unstrata_Ungridded	
	Prior	Posterior	Prior	Posterior	Prior	Posterior
r^2	0.36	0.76	0.36	0.61	0.32	0.65
FE	43%	15%	42%	20%	40%	19%
NME	36%	13%	36%	18%	36%	17%
RMSE $\mu g\ m^{-3}$	5.39	2.32	5.37	3.08	5.57	2.97

that specific aspects of our methodology were correct, such as the observational method and the representative error.

MFB versus MFE "soccer plots" [*Morris et al.,* 2005] were calculated for each month as the average of the ten random runs and are graphed in Figure 5.9. Figures 5.9 a, b, and c represent the stratified-gridded, unstratified-gridded, and unstratified-ungridded cross validation methods, respectively. The plots represent monthly (from 1-9 plus O, N and D for Oct, Nov, and Dec, respectively) values for 2002. The rectangular areas in the soccer plot correspond to performance categories, with excellent, good, average, and problematic areas radiating out from the origin. Arrow origins correspond to the performance of the prior, and arrow termini correspond to the posterior performance. Improvements in both bias and error were substantial for all of the three cross validation methods. The fractional errors in all months were lower than 20% in value, and the fractional bias was close to zero, or had a low negative value.

Figure 5.10 shows the result of OI applied to $PM_{2.5}$ concentrations for a complete year. Figure 5.10 maps 2002 annual average kriged surface observations from the AQS and IMPROVE networks. Figures 5.10b and Figure 5.10c map the 2002 annual average prior and posterior values, respectively. Posterior values were based on the first random stratified-gridded cross validation method for each month, described in section 5.5.

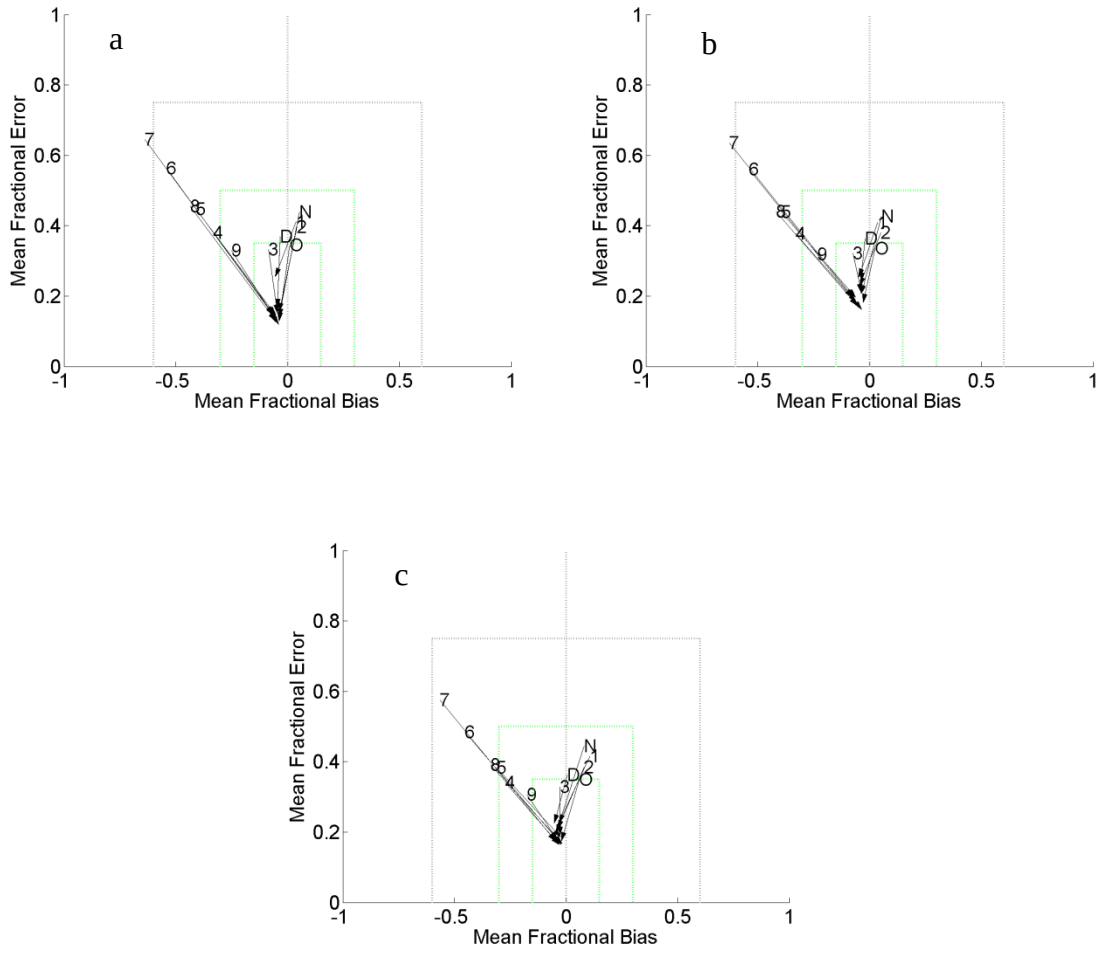

Figure 5.9 Monthly (1-9 plus O, N, and D for Oct, Nov, and Dec, respectively) prior and posterior Fractional Bias versus Fractional Error for three cross validation methods described in section 5.5; (a) stratified-gridded; unstratified-gridded; and (c) unstratified-ungridded. The four (from excellent to problematic) performance categories described in the methods section are shown visually by the rectangular zones. The beginning and ending of the arrows represents the prior and posterior values, the arrows are only shown for months that show improvements.

Bias reduction can be evaluated from $PM_{2.5}$ bias maps of prior (Figure 5.11a) and posterior (Figure 5.11b) values. Bias maps are calculated as the model result minus a kriged $PM_{2.5}$ observation surface from Figure 5.10a. The data used to perform kriging

was IMPROVE $PM_{2.5}$ downloaded from The Visibility Information Exchange Web System (VIEWS) [2002] and USEPA AQS $PM_{2.5}$ measurements [2002a, 2002b]. The comparison is based on an assumption that the kriged maps are accurate. The plots show high improvements in the eastern parts of the U.S, where large spatial locations have values that are approaching zero (white, light blue, or yellow areas) the positive biases (red) have been reduced considerably, with a few remaining locations. The western parts of the U.S also have considerable model bias reduction (white areas), but also have areas of increased model bias, where the biases shifted from unbiased (white, light blue, or yellow) to positively biased (red). Areas of improved performance include New York, Washington, D.C., Philadelphia, Atlanta, Chicago, Dallas, Salt Lake City, and Portland. In contrast, the surrounding spatial area around Miami and Los Angeles has shifted to a positive bias from a low bias and negative bias, respectively. The annual mean bias reduction for the whole domain was decreased from -4.37 to -3.66 μg m^{-3}, and mean RMSE reduced from 4.3 to 3.8 μg m^{-3}. Figure 5.11c shows the difference between posterior and prior (the difference between Figure 5.10b and Figure 5.10a). OI tends to increase $PM_{2.5}$ in the eastern and western U.S. with a slight decrease in some spatial areas of the eastern U.S. The mean $PM_{2.5}$ prior value increased from 4.14 μg m^{-3} to a posterior value of 4.84 μg m^{-3}.

5.3.5 Sensitivity test of depleting surface measurements gradually

Section 5.3.5 describes a sensitivity test that gradually removes observations from all strata until the observations were completely removed, to understand the effect of OI on observation-rich and observation-poor locations. Figure 5.12 shows number of observations left out versus prior and posterior mean FE for the whole domain (repeated10 times) for the month of January. Posterior FEs (upper group of lines) are higher than posterior FEs (lower group of lines).

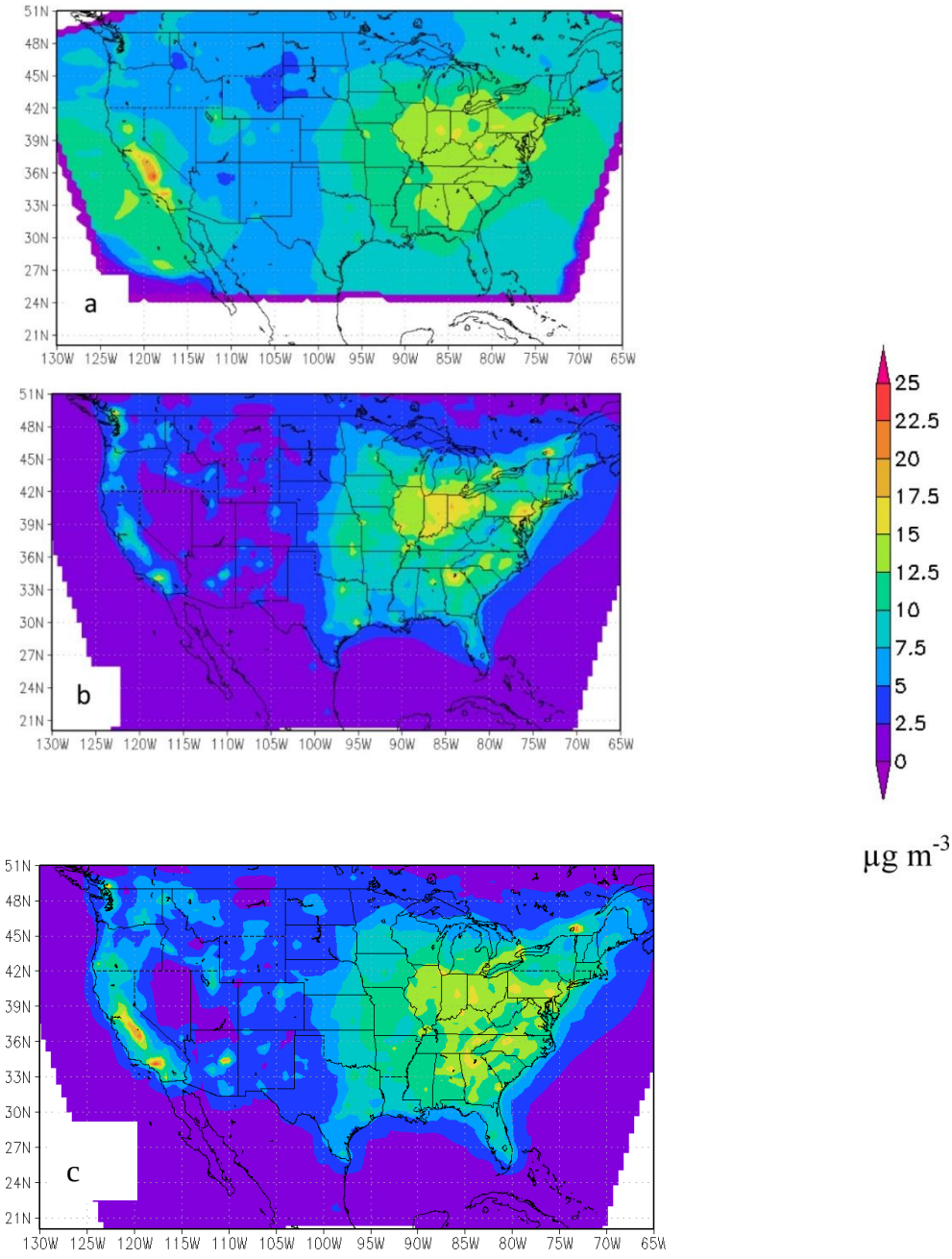

Figure 5.10 Annual average $PM_{2.5}$ plots (2002). (a) AQS and IMPROVE $PM_{2.5}$ values kriged to represent a combined measured $PM_{2.5}$ surface; (b) Prior Model; (c) Posterior based on the first random stratified-gridded cross validation method for each month, described in section 5.5.

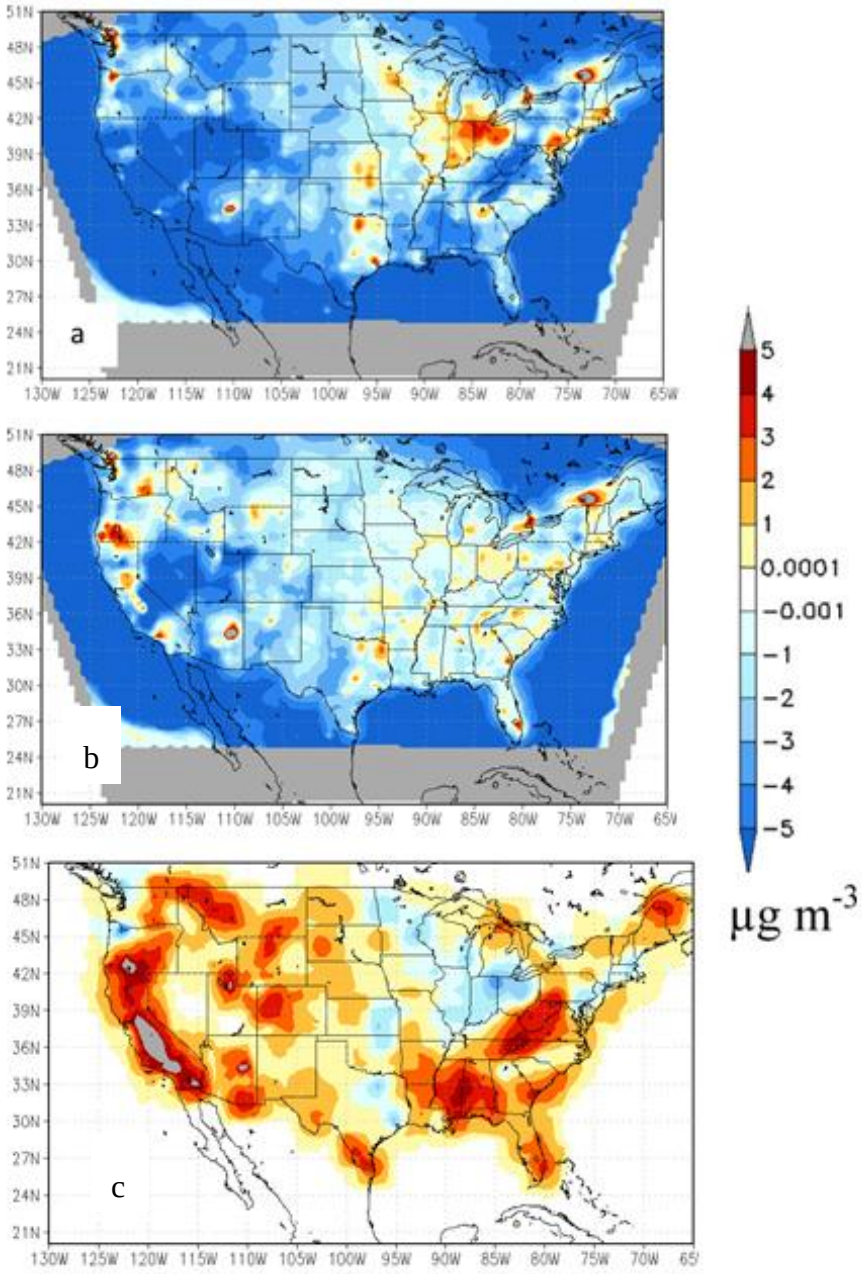

Figure 5.11 Annual average posterior bias plots (a) before OI (prior - kriged surface obs, Fig. 5.10b-5.10a); (b) after OI (posterior - kriged surface obs, Fig. 5.10c-5.10a); and (c) the amount of change due to OI (Posterior - Prior).

OI improves performance statistics for all of the cross-validation repetitions, even when almost all data is withheld. This could prove to be useful when data assimilation is required over a specific area such as Iowa City, IA, or Bondvile, IL, with limited observational data.

Figure 5.12 shows that the prior model FE decreases with the number of available measurements for evaluation. To understand this Figure 5.13a shows FE calculated from strata that have only a few measurements (i.e. 1-5) while Figure 5.13b shows FE calculated from strata that have considerable amounts of measurements (i.e. 12-16). The mean Fractional Error for Strata that only have a few measurements is higher than the mean Fractional Error Strata that have a lot of measurements.

The sensitivity test shows that surface OI is applicable for any site with only a few measurements around the area of interest to considerable amount of measurements for large spatial locations.

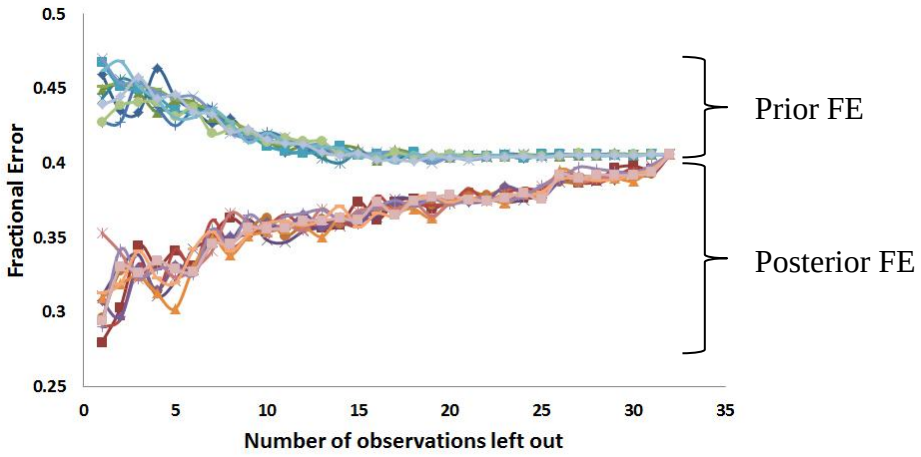

Figure 5.12 Number of observations left out versus prior and posterior mean FE for the whole domain (repeated 10 times). The test for the month of January assumed stratified-gridded observations withheld from equal amount of strata

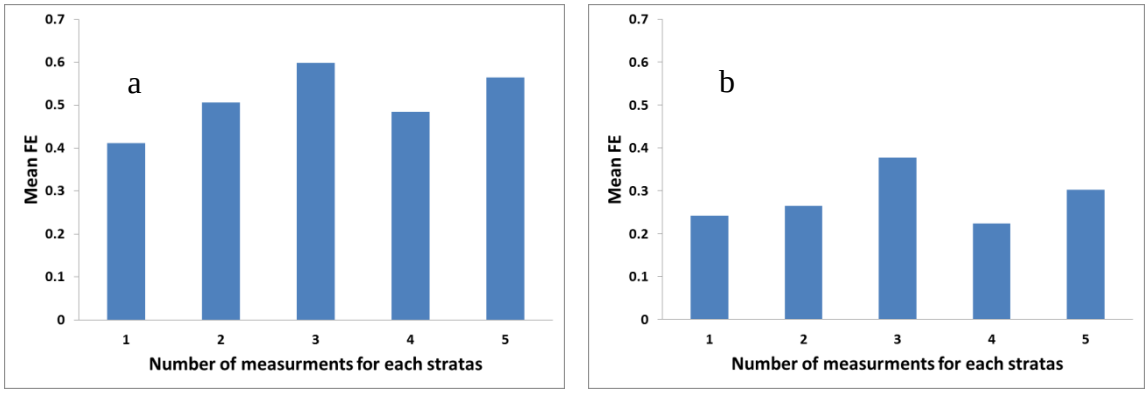

Figure 5.13 The mean FE for (a) stratum that have 1-5 measurements and (b) stratum that have 12-16 measurements.

5.4 Discussion

OI results show practical implementation of surface data assimilation towards an improved model estimate. Despite OI being a modest data assimilation technique (i.e. it provides less accuracy) compared to the more advanced data assimilation methods, it achieves substantial performance increases. The question arises how the methods in this chapter will compare to more advance data assimilation methods.

Previous studies have also shown substantial improvements using 3D-var, with which Pagowski et al. [2010] showed substantial improvements over the eastern U.S. for correlation coefficients, RMSE, and bias. Correlation coefficients improved from 0.57 to 0.8, RMSE decreased from 9 to 6 or 3 μg m^{-3}, and bias changed from values higher than -4 μg m^{-3} to values lower than -4 μg m^{-3}. Implementing the observational method with 3D-var could prove to increase model performance after data assimilation. One of the limitations of OI is its working (or moving) window, which does not include all the observations in the domain simultaneously, which is one of the features of 3D-var. The

observational method is calculated based on the observational increments for the whole domain, which could be fully or accurately exploited in 3D-var.

Figs 5.11a and 5.11b, representing the observation biases for posterior and prior model estimates, show a substantial decrease from the forward model. Large spatial areas show that the bias is approaching zero, but this is not true for a few spatial locations, such as Los Angeles and Miami, where we can see an increase in model bias between the posterior and observation, which could be an artifact of kriging. The representation error decreased the effect of urban location to influence its surroundings; this was successful for most spatial areas. Another error term is the influence or change in elevation for the surrounding areas around the site measurement. If a measurement site is within a basin, then its influence should be limited for the surrounding areas. This can be recognized by using a slope parameter that indicates the degree of elevated topography around the measurement site.

Based on Pagowski et al. [2010], a cutoff was used to remove any data measurements recorded at or higher than 150 μg m^{-3}. A more developed method would remove high concentrations based on their seasonal and spatial locations (i.e. a site in Chicago should not be treated the same as a site in Iowa City). This can be accomplished by an outlier detection method that identifies peak concentrations at different sites. Although a general approach as described to build a seasonal and spatial database for outliers would be time efficient, investigating high values using remote sensing data would prove valuable, although time consuming.

5.5 Conclusion

The assimilation technique was OI of $PM_{2.5}$ performed on the CMAQ model grid. Data assimilation was used to produce monthly average $PM_{2.5}$ estimates by combining monthly averages from prior model and surface $PM_{2.5}$ values. Results were compared based on the three cross validation methods applied over the domain. Results show that

OI can improve model estimates using any of the cross validation methods. Stratified cross validation was the first cross validation method applied after regridding the observations to CMAQ domain. The observations were aggregated in a single cell when regridding the observations to the model domain. The second cross validation method is based on simple random systems applied after regridding the observations to CMAQ domain. The third cross validation method is based on simple random systems applied before regridding the observations to CMAQ domain. In each method 10% of the data was withheld randomly for model evaluation after data assimilation. This was repeated ten times with different 10% withheld data each time. Also a different set of repetitions was achieved using stratified-gridded systems, by removing all the data gradually (i.e. 10%, 20% …100%), which improved model posterior estimates throughout all the withheld data.

Multiple features in the OI results are noteworthy and are listed as follows: all three cross validation methods showed substantial posterior model improvements, which are very similar, indicating that the methodology applied for data assimilation is implemented correctly; the random repetitions in some cases increased the correlations coefficient from 0.3 to 0.9; all four evaluation metrics used showed a constant bias reduction (posterior-observation) over the forward model in all repetitions for all months of the year; soccer plots showed that all posterior values had an MFE lower than 20% and an MFB approaching zero; the annual bias evaluation of posterior and prior values compared to kriged surface measurements, showed overall bias reduction; removing all surface measurements gradually showed that all repetition managed to improve FE.

OI substantially improved performance statistics, and improvement was nearly independent of the cross validation (i.e. sample withholding) strategy used to assess performance. Cross validation shows that the results are dependable on the data withheld and the data forced through data assimilation (i.e. is site dependent). Whereas there is a difference between each repetition, the final outcome is the same. The results, although

random, substantially enhance model performance on average. Data assimilation captures the spatial pattern of the surrounding OI areas and is not limited by the number of monitoring sites included. The spatial improvements include major urban cities that are of most interest in a variety of studies (e.g. exposure studies). The OI model could be used for any data assimilation method and would prove to be a viable evaluation and model performance improvement tool.

CHAPTER 6: SUMMARY AND RECOMMENDATIONS

6.1 Summary

The current work was focused on improving air quality model estimates for fine particulate matter. This was achieved by combining model estimates with observations from both satellite and surface measurements using data assimilation. The availability of fine particles at high spatial and temporal resolutions is crucial for epidemiology studies, and current studies must either rely on scarce surface measurements or model concentrations. The current work emphasizes the importance of using both satellite data, which has high spatial temporal retrieval, and surface measurements, which are scarce and temporally limited, in order to improve model estimates through data assimilation. MODIS AOD datasets are relatively biased depending on the topography and the seasonal retrieval. Surface measurements are considered accurate, from a modeling perspective, and are used for model assessment and validation. The objectives of the current work are discussed explaining the outcomes of each.

The first objective was to quantify how well OI works with MODIS retrieval over the Unites States. This was accomplished in two steps, the first step was establishing a 24 hour daily data base for $PM_{2.5}$ model estimates using CMAQ for the year 2002, this was covered in chapter 3. The model estimates were evaluated and compared with previous studies for the U.S. The evaluation was done for both rural and urban sites using observations from IMPROVE and STN networks, respectively. A detailed model evaluation of $PM_{2.5}$ and its species was conducted for the whole U.S for 2002. The evaluation was based on dividing the U.S into six regions, based on the U.S census division, for 12 months of the year. This was crucial for the satellite OI evaluation, since the MODIS retrieval is dependent on topography and seasonal changes. For surface OI, the whole continental U.S. was used for evaluation over 12 months for 2002. $PM_{2.5}$ performance was affected by organic carbon and nitrate, which were both underpredicted

(i.e. negatively biased). Organic carbon biases reached a maximum of -3.5 to -5.5 μg m^{-3} in the North East region in summer, and a -6.2 μg m^{-3} bias was estimated at the Pacific in November for urban locations. Nitrate biases reached a maximum of -6 μg m^{-3} in the Pacific for November at urban sites, and +2.1 μg m^{-3} in the North East in March for rural sites. Overall the forward model results showed an agreement with previous studies, thereby setting a $PM_{2.5}$ database for OI with both satellite and surface measurements.

The second step was applying satellite data assimilation with optimal interpolation, which is covered in chapter 4. To perform OI, AEONET and MODIS were re-gridded on CMAQ domain, and CMAQ-derived AOD was calculated from CMAQ $PM_{2.5}$. The inputs for the data assimilation procedure are both monthly averages of CMAQ-derived AOD and monthly averages of the observation (i.e. AERONET replaced MODIS AOD wherever existed). Three methods were applied for averaging the model-observation inputs and for correcting posterior $PM_{2.5}$ estimates using posterior AOD values. The posterior model correction was the product of the prior estimate and the ratio of posterior to prior AOD (referred to as the scaling factor). The first method used monthly averages from CMAQ and MODIS, correcting the posterior values using the scaling factors. The second method is similar to the first, except the monthly averages from CMAQ were based on the satellite overpass, and only the same hours of the satellite retrieval over a specific location were averaged. The third method has the same averaging procedure as the second method, but the scaling factors are not used at full magnitude, only at the time of the satellite overpass, while the ±12 hours before and after that have a decreasing value for the scaling factor. Four sets of error settings were used for OI, two based on low MODIS errors and two based on high MODIS errors.

The evaluation of model posterior estimates was similar to the evaluation of the CMAQ model output without data assimilation found in Chapter 3. The results showed that the first method was the most effective, followed by third method, with the second being least effective. As a combination of regions, months, and site location, a total of

144 metric values were calculated, 72 for each rural and each urban location. The first method decreased FES values in 47 specific regions and months, versus 27 and 15 for the third and second methods, respectively. The remaining 55 region and months from both rural and urban locations were not improved by OI. The best improvement (FES reduction) was noticeable in the eastern parts of the US, compared to a poor OI result in the Pacific, especially at rural locations. OI error settings varied based on region, season and methods/settings applied. An annual evaluation was based on four categorizations for constraints on OI settings and methods, compared to the base case (i.e. no data assimilation). The four categories are as follows: the first constraints the methods and settings for each month and region; the second has one method/setting for each month, applied for the whole domain; the third has one method/setting for each region, applied for the whole year; the last has one method/setting applied for all regions and months. All of the four categories showed that OI managed to decrease FES values, but only offered unsubstantial improvements over the forward model. The best combinations led to a domain average improvement in fractional error from 1.2 to 0.97 at rural locations, and from 0.99 to 0.89 at urban locations. The annual bias reduction based on the posterior and prior model estimates with kriged surface measurements, showed a bias reduction in areas such as West Texas and Salt Lake City, the Ohio-Indiana border, around Atlanta, and eastern Pennsylvania. But results also showed an increase in negative bias compared to the forward model estimate.

The second objective proposed a best OI strategy for U.S explaining the results. This was accomplished by comparing MODIS AOD with AERONET AOD, where AERONET is considered to be a more accurate measurement than MODIS. The coefficient of determination for all AERONET sites in May 2002 was calculated, from west coast to east coast of the United States. Correlations were higher in the Northeast than that in the Pacific, giving an indication of OI week performance in the Pacific.

Two sensitivity tests were conducted based on the same regional/monthly evaluation conducted above. The first included removing AERONET data from the observational input, and running the data assimilation as before. The results showed insignificant differences based on a monthly/regional evaluation, but comparing model to surface measurements showed differences based on the method/settings chosen. The second was removing equation 4.7 effects, where the equation forced posterior values to be in between the observations and prior estimates. The results also showed no significant difference based on a monthly/regional evaluation.

The third objective quantifies how well OI works with MODIS retrieval at a finer spatial resolution over the United States. To accomplish this, model runs were conducted based on a 2km MODIS retrieval, which is different from the 10km data available from NASA. The 2km data was averaged to 36km to match the model grid resolution. CMAQ simulations were only conducted for the first six months of the year, and the results showed no significant difference based on a monthly/regional evaluation. The benefit of having a 2km resolution is for model simulations that have finer model resolution (e.g. 4km), and aggregating 2km to 36km showed no benefit from the new dataset.

The final objective quantified how well OI works with surface measurements over the United States, which is shown in chapter five. The novel part of the surface OI was the calculation of the error covariance matrices for the model and observation. For model error covariance values (i.e. B matrix), the observational method was applied. The observation error covariance calculation in OI was different from the satellite calculation because representation error was calculated for each surface $PM_{2.5}$ observation based on urban fraction of the measurement location. Three cross validation methods were applied to the data assimilation method to validate the current OI implementation using both the observational method and representation error to calculate the error covariance matrices for the model and observations, respectively. The three cross validation methods included using stratified random sampling after regridding observations to the model domain,

using the simple random sampling after regridding observations to the model domain, and using the simple random sampling before regridding observations to the model domain. Each of the cross validation methods was repeated ten times to assess variability. The methods withheld 10% of the measurements at a time for each repetition, cycling through all the measurements without replacement. The withheld data was used for validation and model evaluation.

The annual surface OI evaluation was based on the coterminous U.S. for all three cross validation methods. Significant model bias reduction was noticeable for all cross validation methods independent of month. OI improved the annual metric values of the correlation coefficient from 0.36 prior value to 0.76 posterior value. OI decreased the fractional error, normalized mean error, and RMSE from 43% (prior) to 15% (posterior), 36% (prior) to 13% (posterior), and 5.39 (prior) to 2.32 μg m^{-3} (posterior), respectively. All three cross validation methods were similar and gave a conclusion that surface OI implementation is correctly constructed. The monthly average mean fractional error and mean fractional bias were significantly decreased from values exceeding 60% and -50% to values lower than 20% and -2%, respectively. The annual bias reduction based on posterior and prior model estimates with kriged surface measurements showed a bias reduction in most of the eastern areas and in the western parts of the U.S. It is important for the method to be applied in locations that have abundant epidemiological data and where bias reductions were noticeable, such as in cities that included New York, Washington, D.C., Philadelphia, Atlanta, Chicago, Dallas, Salt Lake City, and Portland. It is also important to note that an increase in bias was noticeable in a few areas such as Los Angeles and Miami.

Another set of repetitions were done using stratified random sampling to inform the amount of observation data to be withheld while still providing a sufficient model improvement. This was accomplished by removing all the observational data gradually using different repetitions. Each repetition removed a percent of the observational data

until all the data were removed. This test was used to test the effectiveness of OI in situations of an abundance of data compared to situations with only a few data points. The results show a fractional error decrease over all prior model estimates for all repetitions. This indicates that surface OI can be applied to areas with minimal observations and still achieve posterior model bias lower than the forward model.

From the study, we concluded that optimal interpolation schemes to constrain $PM_{2.5}$ over the United States have proven to be successful by running data assimilation with satellite or surface measurements. Although surface OI has proven to be far superior to satellite OI, the spatial and temporal limitations of the surface measurement indicate a need for the use of satellite retrievals. MODIS retrieval (level 2 data) at 10km is subjected to biases and noise, which depend on the topography and seasonal retrieval. The results over the eastern parts of the continental U.S. showed improvements in posterior model estimates, but can be described as modest improvements in most performance metrics. Satellite OI is recommended for the eastern parts of the western region in the U.S. Surface OI showed substantial improvements in the posterior values over the prior model. This is credited to the implementation of the observational method and the representative error in the calculation of the error covariance matrices for the model and observations, respectively. The three cross validation methods proved that the implementation described improved model performance in all the methods. The surface OI can be applied towards the EPA five year study to improve model performance for generating $PM_{2.5}$ concentrations that can be used in epidemiological studies.

6.2 Future work

The ability to combine both satellite and surface OI, if possible, could provide substantial improvements for areas in between scarce surface-measurement locations. Problems combining both measurements are due to MODIS errors. Surface measurements are considered accurate; therefore site locations and their surrounding

areas are influenced by the surface OI. Satellite data will add noise to the system due to the random errors in the satellite retrievals, which could increase the model bias shown in the surface OI implementation. To combine both measurements, the random errors have to be removed from MODIS retrieval. This is accomplished by using quality-assured data (referred to in section 4.18). The noise-free data could substantially decrease the MODIS AOD spatial retrieval by 50% or more, but could still provide valuable spatial coverage for the locations without nearby surface monitors. Also MODIS data can be improved by using land regression models [Kumar, et al., 2008, Schaap, et al., 2009, Engel-Cox, et al., 2005] with AERONET data, land use and land cover, snow cover, and meteorological data (e.g. temperature, pressure, relative humidity, wind speed, and wind direction). Based on this thesis research, I recommend two procedures. The first takes into account that the surface measurements are accurate, giving surface priority over satellite OI. Therefore OI is first established with the surface measurements and then the satellite data assimilation covers areas that did not have any effect from surface OI, or had an increase in model bias. This procedure can be used for the raw MODIS data (without quality assurance) and areas of model bias reduction can be used to cover areas with no bias reduction. The second procedure includes combining both AOD and surface $PM_{2.5}$ with respect to the transformation between model and observation space. Surface measurements should have priority over satellite retrieval, and MODIS AOD should be replaced by surface $PM_{2.5}$ at site locations. Both respective covariance matrices that include the observation representation error should be accounted for. The two procedures could prove to be an improvement over the current performance shown in the current work, and this suggestion could serve as a goal of future research.

In future work, the EPA study of surface OI methods and application will be applied over a five-year run simulation using WRF-CMAQ over the continental U.S. The domain has three nested areas, 36km, 12km, and 4km. The 4km nested area only applies

for one location over Chicago. Also satellite OI can be applied for the 4km resolution area, using 2km retrieval data described in section 4.6.

REFERENCES

Abdou, W. A., D. J. Diner, J. V. Martonchik, C. J. Bruegge, R. A. Kahn, B. J. Gaitley, K. A. Crean, L. A. Remer, and B. Holben (2005), Comparison of coincident Multiangle Imaging Spectroradiometer and Moderate Resolution Imaging Spectroradiometer aerosol optical depths over land and ocean scenes containing Aerosol Robotic Network sites, *J. Geophys. Res., 110*, D10S07, doi:10.1029/2004JD004693

Adar, S. D. and J. D. Kaufman (2007), Cardiovascular Disease and Air Pollutants: Evaluating and Improving Epidemiological Data Implicating Traffic Exposure, *Inhal. Toxicol., 19*, 135-149.

Adhikary, B., S. Kulkarni, A. Dallura, Y. Tang, T. Chai, L. R. Leung, Y. Qian, C. E. Chung, V. Ramanathan, and G. R. Carmichael (2008), A regional scale chemical transport modeling of Asian aerosols with data assimilation of AOD observations using optimal interpolation technique, *Atmos. Environ., 42*, 8600-8615.

Alapaty, K., N. Seaman, D. Niyogi, and A. Hanna (2001), Assimilating surface data to improve the accuracy of atmospheric boundary layer simulations, J. Appl. Meteorol., 40, 2068-2082.

Al-Hamdan, M. Z. et al. (2009), Methods for Characterizing Fine Particulate Matter Using Ground Observations and Remotely Sensed Data: Potential Use for Environmental Public Health Surveillance, *J. Air Waste Manage. Assoc., 59*, 865-881.

Appel, K. W., C. Chemel, S. J. Roselle, X. V. Francis, R. M. Hu, R. S. Sokhi, S. Rao, and S. Galmarini (2011), Examination of the Community Multiscale Air Quality (CMAQ) model performance over the North American and European domains, *Atmos. Environ.*, doi.org/10.1016/j.atmosenv.2011.11.016, in press.

Appel, K. W., S. J. Roselle, R. C. Gilliam, and J. E. Pleim (2010), Sensitivity of the Community Multiscale Air Quality (CMAQ) model v4.7 results for the eastern United States to MM5 and WRF meteorological drivers, *Geosci. Model Dev., 3*, 169-188.

Baker, K. (2004), Midwest Regional Planning Organization Modeling Protocol, Lake Mich. Air Dir. Consort., Midwest Reg. Plann. Org., Des Plaines, Ill.

Balgovind, R., Dalcher, A., Ghil, M., and Kalnay, E. (1983), A Stochastic-Dynamic Model for the Spatial Structure of Forecast Error Statistics, Mon. Weather Rev., 111, 701–722.

Bei, N., B. de Foy, W. Lei, M. Zavala, and L. T. Molina (2008), Using 3DVAR data assimilation system to improve ozone simulations in the Mexico City basin, Atmos. Chem. Phys., 8, 7353-7366.

Binkowski, F. S., and S. J. Roselle (2003), Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 1. Model description, *J. Geophys. Res., 108*, doi:10.1029/2001JD001409.

Boylan, J. W., and A. G. Russell (2006), PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models, *Atmos. Environ., 40*, 4946-4959.

Browne, M. W. (2000), Cross-validation methods, J. Math. Psychol., 44, 108-132.

Byun, D. W., and J. K. S. Ching (1999), Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) modeling system, EPA/600/R-99/030, Off. of Res. and Dev., U.S. Environ. Prot. Agency, Washington, D. C.

Carmichael, G. R., S. Kulkarni, C. E. Chung, and V. Ramanathan, (2010), Constraining Aerosol Distribution in Asia by Integrating Models with Multi-sensor Observations, paper presented at American Geophysical Union, Fall Meeting 2010, San Francisco, CA.

Carmichael, G. R., A. Sandu, T. Chai, D. N. Daescu, E. M. Constantinescu, and Y. Tang, (2008), Predicting air quality: Improvements through advanced methods to integrate models and measurements, J. Comp. Phys., 227, 3540 - 3571, doi:10.1016/j.jcp.2007.02.024.

Carmichael, G. R., et al. (2003), Regional-scale chemical transport modeling in support of the analysis of observations obtained during theTRACE-P experiment, J. Geophys. Res., 108(D21), 8823, doi:10.1029/2002JD003117.

Chatterjee, A., A. M. Michalak, R. A. Kahn, S. R. Paradise, A. J. Braverman, and C. E. Miller (2010), A geostatistical data fusion technique for merging remote sensing and ground-based observations of aerosol optical thickness, *J. Geophys. Res., 115*, D20207, doi:10.1029/2009JD013765.

Chung, C. E., V. Ramanathan, G. Carmichael, S. Kulkarni, Y. Tang, B. Adhikary, L. R. Leung, and Y. Qian (2010), Anthropogenic aerosol radiative forcing in Asia derived from regional models with atmospheric and aerosol data assimilation, *Atmos. Chem. Phys., 10*, 6007-6024.

Collins, W. D., P. J. Rasch, B. E. Eaton, B. V. Khattatov, J. Lamarque, and C. S. Zender (2001), Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX, *J. Geophys. Res., 106*, 7313-7336, doi:10.1029/2000JD900507.

Courtier, P., and O. Talagrand, 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow water equations. Tellus 42A, 531-549.

Debell, L. J., et al. (2006), Spatial and Seasonal Patterns and Temporal Variability of Haze and its Constituents in the United States (Report IV), CIRA, Colorado State University, Fort Collins, CO.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Library and User's Guide*, Oxford Univ. Press, New York, N.Y.

Diner, D. et al. (1998), Multi-angle Imaging SpectroRadiometer (MISR) - Instrument description and experiment overview, *IEEE Trans. Geosci. Remote Sens., 36*, 1072-1087.

Drury, E., D. J. Jacob, J. Wang, R. J. D. Spurr, and K. Chance (2008), Improved algorithm for MODIS satellite retrievals of aerosol optical depths over western North America, *J. Geophys. Res., 113*, D16204, doi:10.1029/2007JD009573.

Drury, E., D. J. Jacob, R. J. D. Spurr, J. Wang, Y. Shinozuka, B. E. Anderson, A. D. Clarke, J. Dibb, C. McNaughton, and R. Weber (2010), Synthesis of satellite (MODIS), aircraft (ICARTT), and surface (IMPROVE, EPA-AQS, AERONET) aerosol observations over eastern North America to improve MODIS aerosol retrievals and constrain surface aerosol concentrations and sources, *J. Geophys. Res., 115*, D14204, doi:10.1029/2009JD012629.

Eder, B., and S. Yu (2006), A performance evaluation of the 2004 release of Models-3 CMAQ, Atmos. Environ., 40, 4811-4824.

Elbern H, Strunk A, Schmidt H, Talagrand O., (2007), Emission rate and chemical state estimation by 4-dimensional variational inversion. Atmos. Chem. Phys. 7: 3749–3769.

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

Engel-Cox, J. A., G. S. Young, and R. M. Hoff (2005), Application of satellite remote-sensing data for source analysis of fine particulate matter transport events, *J. Air Waste Manage. Assoc., 55*, 1389-1397.

ENVIRON, (2006), User's Guide Comprehensive Air Quality Model with Extension (CAMx) Version 4.31, ENVIRON International Corporation, Novato, online: http://www.camx.com, California.

Flemming, J., M. van Loon, R. Stern, (2004), Data assimilation for CTM based on optimum interpolation and KALMAN filter; C. Borrego, S. Incecik (Eds.), Air Pollution Modeling and its Application, vol. XVI Kluwer Academic/Plenum Publishers, New York.

Gandin, L. S. 1963, Objective analysis of meteorological fields. Translated from Russian by the Israeli Program for Scientific Translations (1965).

Gilbert RO, 1987, Statistical methods for environmental pollution monitoring. New York Van Norstrand Reinhold Company.

Grahame, T. J. (2009), Does improved exposure information for $PM_{2.5}$ constituents explain differing results among epidemiological studies?, *Inhal. Toxicol., 21*, 381-393.

Grell GA, Peckham SE, Schmitz R, McKeen SA, Frost GJ, Skamarock WC, Eder B., (2005), Fully-coupled 'online' chemistry within theWRFmodel. Atmos. Environ. 39: 6957–6975.

Gupta, P., and S. Christopher (2009), Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2, A neural network approach, *J. Geophys. Res.*, 114, D20205, doi:10.1029/2008JD011497.

Hakami A., D.K. Henze, J.H. Seinfeld, T. Chai, Y. Tang, G.R. Carmichael, A. Sandu, (2005), Adjoint inverse modeling of black carbon during the Asian pacific regional aerosol characterization experiment, J. Geophys. Res. 110 (D14). Art. No.D14301.

Hand, J. L., et al. (2011), Spatial and Seasonal Patterns and Temporal Variability of Haze and its Constituents in the United States (Report V), CIRA, Colorado State University, Fort Collins, CO.

Hanea, R., G. Velders, and A. Heemink (2004), Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model, J. Geophys. Res. - Atmos., 109, D10302.

Heemink, A. W., Verlaan, M., and Segers, A. J., (2001), Variance reduced ensemble Kalman filtering, Mon. Wea. Rev., 129, 1718–1728.

Henderson, S. B., B. Beckerman, M. Jerrett, and M. Brauer (2007), Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter, *Environ. Sci. Technol.*, *41*, 2422-2428.

Henze, D.K., Seinfeld, J.H., Shindell, D.T., (2008), Inverse modeling and mapping US air quality influences of inorganic $PM_{2.5}$ precursor emissions using the adjoint of GEOS-Chem. Atmos. Chem. Phys. Discuss. 8, 15031–15099.

Hoke, J., and R. Anthes, 1976: The initialization of numerical models by a dynamic relaxation technique. Mon. Wea.Rev. 104, 1551-1556.

Holben, B. N. et al. (1998), AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, *Remote Sens. Environ.*, *66*, 1-16.

Hollingsworth, A., and P. Lonnberg (1986), The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, Tellus, Ser. A, 38, 111– 136.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev. 126, 796-811.

Houyoux, M. R., J. M. Vukovich, C. C. J. Jr, N. J. M. Wheeler, and P. S. Kasibhatla (2000), Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project, *J. Geophys. Res.*, *105*, 9079-9090, doi:10.1029/1999JD900975.

Hsu, N. C., S. C. Tsay, M. D. King, and J. R. Herman (2004), Aerosol properties over bright-reflecting source regions, *IEEE Trans. Geosci. Remote Sens.*, *42*, 557-569.

Hsu, N. C., S. C. Tsay, M. D. King, and J. R. Herman (2006), Deep blue retrievals of Asian aerosol properties during ACE-Asia, *IEEE Trans. Geosci. Remote Sens.*, *44*, 3180-3195.

Hu, Z. (2009), Spatial analysis of MODIS aerosol optical depth, $PM_{2.5}$, and chronic coronary heart disease, *Int. J. Health Geogr.*, *8*, 27.

Hyer, E., J. Reid, and J. Zhang (2010), An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals, *Meas. Tech. Discuss.*, *3*, 4091-4167.

Ito, K., N. Xue, and G. Thurston (2004), Spatial variation of $PM_{2.5}$ chemical species and source-apportioned mass concentrations in New York City, *Atmos. Environ.*, *38*, 5269-5282.

Jazwinski, A. H., 1970: Stochastic processes and filtering theory. Academic Press, New York.

Jerrett, M. et al. (2006), Spatial analysis of air pollution and mortality in Los Angeles, *Epidemiology*, *17*, S69-S69.

Kalman, R. E. 1960. "A New Approach to Linear Filtering and Prediction Problems," Transaction of the ASME—Journal of Basic Engineering, pp. 35-45 (March 1960).

Kalnay, E. (2002), *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, United Kingdom.

Kim, S., L. Sheppard, and H. Kim (2009), Health Effects of Long-term Air Pollution Influence of Exposure Prediction Methods, Epidemiology, 20, 442-450.

Kistler, R. E., 1974: A study of data assimilation techniques in an autobarotropic primitive equation channel model. MS Thesis, Dept of Meteorology, Penn State University.

Kumar, N., A. D. Chu, A. D. Foster, T. Peters, and R. Willis (2011), Satellite Remote Sensing for Developing Time and Space Resolved Estimates of Ambient Particulate in Cleveland, OH, Aerosol. Sci. Technol., 45, 1090-1108.

Kumar, N., A. Chu, and A. Foster (2008), Remote sensing of ambient particles in Delhi and its environs: estimation and validation, *Int. J. Remote Sens.*, *29*, 3383-3405.

Kumar, N., A. D. Chu, A. D. Foster, T. Peters, and R. Willis (2011), Satellite Remote Sensing for Developing Time and Space Resolved Estimates of Ambient Particulate in Cleveland, OH, *Aerosol. Sci. Technol.*, *45*, 1090-1108.

Le Dimet, F.X., O. Talagrand, (1986), Variational algorithms for analysis and assimilation of meteorological observations – theoretical aspects, Tellus Series – Dyn. Meteorol. Oceanogr. 38 (2) 97–110.

Levy, R. C., L. A. Remer, S. Mattoo, E. F. Vermote, and Y. J. Kaufman (2007), Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance, *J. Geophys. Res.*, *112*, D13211, doi:10.1029/2006JD007811.

Lewis, J., and J. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraint. Tellus 37A, 309-322.

Li, C., A. K. H. Lau, J. Mao, and D. A. Chu (2005), Retrieval, validation, and application of the 1-km aerosol optical depth from MODIS measurements over Hong Kong, *IEEE Trans. Geosci. Remote Sens.*, *43*, 2650-2658.

Li, H., F. Faruque, W. Williams, M. Al-Hamdan, J. Luvall, W. Crosson, D. Rickman, and A. Limaye (2009), Optimal temporal scale for the correlation of AOD and ground measurements of $PM_{2.5}$ in a real-time air quality estimation system, *Atmos. Environ.*, *43*, 4303-4310.

Liu, Z., Q. Liu, H. Lin, C. S. Schwartz, Y. Lee, and T. Wang (2011), Three-dimensional variational assimilation of MODIS aerosol optical depth: Implementation and application to a dust storm over East Asia, J. Geophys. Res. -Atmos., 116, D23206.

Liu, Y., C. J. Paciorek, and P. Koutrakis (2009), Estimating Regional Spatial and Temporal Variability of $PM_{2.5}$ Concentrations Using Satellite Data, Meteorology, and Land Use Information, *Environ. Health Perspect., 117*, 886-892.

Liu, Y., A. Bourgeois, T. Warner, S. Swerdlin and J. Hacker, (2005), An implementation of obs nudging-based FDDA into WRF for supporting ATEC test operations. 2005 WRF user workshop. Paper 10.7.

Liu, Y., R. Park, D. Jacob, Q. Li, V. Kilaru, and J. Sarnat (2004), 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, doi:10.1029/2004JD005025.

Lorenc, A. C. (1986), Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc., 112*, 1177-1194.

Lorenc, A., 1981: A global three-dimensional multivariate statistical interpolation scheme. Mon. Wea. Rev. 109, 701-721.

Malm, W. C., J. F. Sisler, D. Huffman, R. A. Eldred, and T. A. Cahill (1994), Spatial and seasonal trends in particle concentration and optical extinction in the United States, *J. Geophys. Res., 99*, 1347-1370, doi:10.1029/93JD02916.

Malm, W. C., and J. L. Hand (2007), An examination of the physical and optical properties of aerosols collected in the IMPROVE program, *Atmos. Environ., 41*, 3407-3427.

Mallet, V.,Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I.,Wu, L., Roustan,Y., Sartelet, K., Tombette, M., Foudhil, H., (2007), Technical Note: The air quality modeling system Polyphemus. Atmos. Chem. Phys. 7, 5479–5487.

Marmur, A., S. Park, J. Mulholland, P. Tolbert, and A. Russell (2006), Source apportionment of $PM_{2.5}$ in the southeastern United States using receptor and emissions-based models: Conceptual differences and implications for time-series health studies, *Atmos. Environ., 40*, 2533-2551.

Matthijsen, J., L. Delobbe, F. Sauter, and L. de Waal (2001), Changes of surface ozone over Europe upon the Gothenburg protocol abatement of 1990 reference emissions, in Transport and Chemical Transformation in the Troposphere: Proceedings of EUROTRAC Symposium 2000, Garmisch-Partenkirchen, Germany, 27-31 March 2000, edited by P. M. Midgley, M. Reuther, and M. Williams, pp. 1384– 1388, Springer- Verlag, New York.

McMurry, P. H., M. F. Shepherd, and J. S. Vickery (Eds.) (2004), *Particulate Matter Science for Policy Makers: A NARSTO Assessment*, Cambridge Univ. Press, New York, N.Y.

Mishchenko, M. I., L. Liu, I. V. Geogdzhayev, L. D. Travis, B. Cairns, and A. A. Lacis (2010), Toward unified satellite climatology of aerosol properties. 3. MODIS versus MISR versus AERONET, *J. Quant. Spectrosc. Radiat. Transfer, 111*, 540-552.

Morris, R. E., D. E. McNally, T. W. Tesche, G. Tonnesen, J. W. Boylan, and P. Brewer (2005), Preliminary evaluation of the community multiscale air, quality model for 2002 over the southeastern United States, *J. Air Waste Manage. Assoc.*, *55*, 1694-1708.

Nenes, A., S. N. Pandis, and C. Pilinis (1998), ISORROPIA: A New Thermodynamic Equilibrium Model for Multiphase Multicomponent Inorganic Aerosols, *Aquatic Geochemistry*, *4*, 123-152.

National Interagency Coordination Center (NICC) (2002), 2002 Statistics and Summary, National Interagency Fire Center, Boise, ID.

Paciorek C. J., Y. Liu, H. Moreno-Macias, and S. Kondragunta (2008), Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level $PM_{2.5}$, *Environ. Sci. Technol.*, *42*(15), 5800-5806.

Paciorek, C. J., J. D. Yanosky, R. C. Puett, F. Laden, and H. H. Suh (2009), Practical Large-Scale Spatio-Temporal Modeling of Particulate Matter Concentrations, *Ann. Appl. Stat.*, *3*, 370-397.

Pagowski M, Grell GA, McKeen SA et al (2010) Threedimensional variational data assimilation of ozone and fine particulate matter observations. Some results using the Weather Research and Forecasting – Chemistry model and Gridpoint Statistical Interpolation. Q. J. R. Meteorol. Soc. 136: 2013–2024. DOI:10.1002/qj.700.

Park, R., C. Song, K. Han, M. Park, S. Lee, S. Kim, and A. Shimizu (2011), A study on the aerosol optical properties over East Asia using a combination of CMAQ-simulated aerosol optical properties and remote-sensing data via a data assimilation technique, *Atmos. Chem. Phys.*, *11*, 12275-12296.

Parrish, D., and J. Derber (1992), The National Meteorological Center's spectral statistical-interpolation analysis system, Mon. Weather Rev., 120(8), 1747–1763.

Penner, J. E., et al., Climate Change 2001: The Scientific Basis. Contribution of working group I to the third assessment report of the Intergovernmental Panel on Climate Change, Chap. Aerosols, their Direct and Indirect Effects, p. 881, Cambridge Univ. Press, New York, 2001.

Pitchford, M., W. Malm, B. Schichtel, N. Kumar, D. Lowenthal, and J. Hand (2007), Revised algorithm for estimating light extinction from IMPROVE particle speciation data, *J. Air Waste Manage. Assoc.*, *57*, 1326-1336.

Pleim, J. and A. Xiu (2003), Development of a land surface model. Part II: Data assimilation, J. Appl. Meteorol., 42, 1811-1822.

Prados, A. I., S. Kondragunta, P. Ciren, and K. R. Knapp (2007), GOES Aerosol/Smoke product (GASP) over North America: Comparisons to AERONET and MODIS observations, *J. Geophys. Res.*, *112*, D15201, doi:10.1029/2006JD007968.

Remer, L. A. et al. (2005), The MODIS Aerosol Algorithm, Products, and Validation, *J. Atmos. Sci.*, *62*, 947-973.

Remer, L. A., R. G. Kleidman, R. C. Levy, Y. J. Kaufman, D. Tanré, S. Mattoo, J. V. Martins, C. Ichoku, I. Koren, and H. Yu (2008), Global aerosol climatology from the MODIS satellite sensors, *J. Geophys. Res.*, *113*, D14S07, doi:10.1029/2007JD009661.

Ross, Z., M. Jerrett, K. Ito, B. Tempalski, and G. D. Thurston (2007), A land use regression for predicting fine particulate matter concentrations in the New York City region, *Atmos. Environ.*, *41*, 2255-2269.

Russell, A. G. (2008), EPA Supersites Program-related emissions-based particulate matter modeling: Initial applications and advances, *J. Air Waste Manage. Assoc.*, *58*, 289-302.

Sasaki, Y. 1970. Some basic formalisms in numerical variational analysis. Mon. Wen. Rev. 98, 875-883.

Sarwar, G., D. Luecken, G. Yarwood, G. Z. Whitten, and W. P. L. Carter (2008), Impact of an updated carbon bond mechanism on predictions from the CMAQ modeling system: Preliminary assessment, *J. Appl. Meteorol. Climatol.*, *47*, 3–14.

Schaap, M., A. Apituley, R. M. A. Timmermans, R. B. A. Koelemeijer, and G. de Leeuw (2009), Exploring the relation between aerosol optical depth and $PM_{2.5}$ at Cabauw, the Netherlands, *Atmos. Chem. Phys.*, *9*, 909-925.

Seigneur, C., B. Pun, S.-Y. Chen, and K. Lohman (2004), Performance Evaluation of Four Air Quality Models Applied for an Annual Simulation of PM over the Western United States, Atmospheric & Environmental Research, Inc., San Ramon, CA.

Sheppard, L., J. Slaughter, J. Schildcrout, L. Liu, and T. Lumley (2005), Exposure and measurement contributions to estimates of acute air pollution effects, J. Expo. Anal. Environ. Epidemiol., 15, 366-376.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers (2005), A description of the Advanced Research WRF Version 2, NCAR Tech. Note 468+STR, Natl. Cent. for Atmos. Res., Boulder, CO.

Spak, S. N., and T. Holloway (2009), Seasonality of speciated aerosol transport over the Great Lakes region, *J. Geophys. Res.*, *114*, D08302, doi:10.1029/2008JD010598.

Stauffer, D.R. and N.L. Seaman, (1990), Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. Mon. Wea. Rev., 118, 1250-1277.

Stern, R., Yamartino, R., Graff, A., (2003). Dispersion Modelling within the European Community's air quality directives: Long term modelling of O3, PM10 and NO2. 26th ITM on Air Pollution Modelling and its Application. May 26–30, 2003, Istanbul, Turkey.

Talagrand , O., P. Courtier, (1987), Variational assimilation of meteorological observations with the adjoint vorticity equation. 1. Theory, Quart. J. Roy. Meteor. Soc. 113 (478) 1311–1328.

Tesche, T. W., R. Morris, G. Tonnesen, D. McNally, J. Boylan, and P. Brewer (2006), CMAQ/CAMx annual 2002 performance evaluation over the eastern US, *Atmos. Environ.*, *40*, 4906-4919.

The Visibility Information Exchange Web System (VIEWS), Download Data available at http://views.cira.colostate.edu/web/About/AboutVIEWS2.aspx (accessed May, 2012)

Tombette, M., V. Mallet, and B. Sportisse (2009), PM_{10} data assimilation over Europe with the optimal interpolation method, Atmos. Chem. Phys., 9, 57-70.

Tonnesen, G., B. Wang, C. J. Chien, Z. Wang, M. Omary, Z. Adelman, A. Holland, and R. Morris (2004), WRAP 2002 Visibility Modeling: Annual CMAQ Performance Evaluation using Preliminary 2002 version C Emissions. Prepared for Western Regional Air Partnership, EPA: Research Triangle Park, NC.

U.S. Census Bureau (2000), Census Regions and Divisions of the United States, U.S. Department of Commerce Economics and Statistics Administration, Washington, D.C.

U.S Environmental Protection Agency (EPA) (2008), Technical Support Document: Preparation of Emissions Inventories For the 2002-based Platform, Version 3, Criteria Air Pollutants, EPA, Research Triangle Park, NC.

U.S. Environmental Protection Agency (EPA) (2007), 3-Year Quality Assurance Report Calendar Years 2002, 2003 and 2004: The SLAMS $PM_{2.5}$ Ambient Air Monitoring Program; available at http://www.epa.gov/ttn/amtic/files/ambient/pm25/PM2.5%202002_2004%20final%20report.pdf (accessed February, 2012).

U.S Environmental Protection Agency (EPA) (2006), Documentation for the Final 2002 Point Source National Emissions Inventory, Emission Inventory and Analysis Group, Air Quality and Analysis Division, EPA, Research Triangle Park, NC.

U.S. Environmental Protection Agency (EPA) (2005), CMAQ Model performance evaluation for 2001, Docket OAR-2003-0053-1716, Research Triangle Park, N. C.

U.S Environmental Protection Agency (EPA) (2002a): Download Detailed AQS Data available at http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm (accessed May, 2012)

U.S Environmental Protection Agency (EPA) (2002b): Download Detailed AQS Data available at http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm (accessed May, 2012)

U.S. Environmental Protection Agency (EPA) (1999), The Initial Concentration and Boundary Condition Processors, Gipson EPA/600/R-99/030, Research Triangle Park, N. C.

van Donkelaar, A., R. V. Martin, M. Brauer, R. Kahn, R. Levy, C. Verduzco, and P. J. Villeneuve (2010), Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application, *Environ. Health Perspect.*, *118*, 847.

van Pul, W. A. J., J. A. van Jaarsveld, and C. M. J. Jacobs (1996), Deposition of persistent organic pollutants over Europe, in Air Pollution Modeling and Its Application XI, edited by S. E. Gryning and F. A. Schiermeier, pp. 203– 212, Plenum, New York.

van Rheineck Leyssius, H. J., F. A. A. M. de Leeuw, and B. H. Kessenboom (1990), A regional scale model for the calculation of episodic concentrations and depositions of acidifying components, Water Air Soil Pollut., 51, 327– 344.

Wang, Z., and G. Tonnesen (2002), WRAP Regional Haze CMAQ 1996 Model Performance Evaluation, WRAP Regional Modeling Center, Denver, CO.

White, W. H. (1990), The components of atmospheric light extinction: A survey of ground-level budgets, *Atmos. Environ., Part A., 24*, 2673-2679.

Wu, J., A. M. Winer, and R. J. Delfino (2006), Exposure assessment of particulate matter air pollution before, during, and after the 2003 Southern California wildfires, *Atmos. Environ., 40*, 3333-3348.

Wu, L., V. Mallet, M. Bocquet, and B. Sportisse (2008), A comparison study of data assimilation algorithms for ozone forecasts RID E-1966-2011, J. Geophys. Res. - Atmos., 113, D20310.

Yu, H., R. Dickinson, M. Chin, Y. Kaufman, B. Holben, I. Geogdzhayev, and M. Mishchenko (2003), Annual cycle of global distributions of aerosol optical depth from integration of MODIS retrievals and GOCART model simulations, *J. Geophys. Res., 108*, 4128, doi:10.1029/2002JD002717.

Yu, S., R. Dennis, S. Roselle, A. Nenes, J. Walker, B. Eder, K. Schere, J. Swall, and W. Robarge (2005), An assessment of the ability of three-dimensional air quality models with current thermodynamic equilibrium models to predict aerosol NO3, *J. Geophys. Res., 110*, D07S13, doi:10.1029/2004JD004718.

Yu, S., P. V. Bhave, R. L. Dennis, and R. Mathur (2007), Seasonal and regional variations of primary and secondary organic aerosols over the Continental United States: Semi-empirical estimates and model evaluation, *Environ. Sci. Technol., 41*, 4690-4697.

Yu, S., R. Mathur, K. Schere, D. Kang, J. Pleim, J. Young, D. Tong, G. Pouliot, S. A. McKeen, and S. T. Rao (2008), Evaluation of real-time $PM_{2.5}$ forecasts and process analysis for $PM_{2.5}$ formation over the eastern United States using the Eta-CMAQ forecast model during the 2004 ICARTT study, *J. Geophys. Res., 113*, D06204, doi:10.1029/2007JD009226.

Zeger, S. L., D. Thomas, F. Dominici, J. M. Samet, J. Schwartz, D. Dockery, and A. Cohen (2000), Exposure measurement error in time-series studies of air pollution: concepts and consequences, *Environ. Health Perspect., 108*, 419-426.