

May 2013

Improved Estimation of PM_{2.5} Using Lagrangian Satellite-Measured Aerosol Optical Depth

Rolando Olivas Saunders
University of Wisconsin-Milwaukee

Follow this and additional works at: <https://dc.uwm.edu/etd>

 Part of the [Atmospheric Sciences Commons](#), and the [Environmental Sciences Commons](#)

Recommended Citation

Olivas Saunders, Rolando, "Improved Estimation of PM_{2.5} Using Lagrangian Satellite-Measured Aerosol Optical Depth" (2013).
Theses and Dissertations. 144.
<https://dc.uwm.edu/etd/144>

This Dissertation is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UWM Digital Commons. For more information, please contact open-access@uwm.edu.

IMPROVED ESTIMATION OF $PM_{2.5}$ USING LAGRANGIAN
SATELLITE-MEASURED AEROSOL OPTICAL DEPTH

by

Rolando Olivas Saunders

A Thesis Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
in Mathematics

at

The University of Wisconsin-Milwaukee

May 2013

ABSTRACT
IMPROVED ESTIMATION OF PM_{2.5} USING LAGRANGIAN
SATELLITE-MEASURED AEROSOL OPTICAL DEPTH

by

Rolando Olivas Saunders

The University of Wisconsin-Milwaukee, 2013
Under the Supervision of Professor Jonathan D.W. Kahl

Suspended particulate matter (aerosols) with aerodynamic diameters less than 2.5 μm (PM_{2.5}) has negative effects on human health, plays an important role in climate change and also causes the corrosion of structures by acid deposition. Accurate estimates of PM_{2.5} concentrations are thus relevant in air quality, epidemiology, cloud microphysics and climate forcing studies. Aerosol optical depth (AOD) retrieved by the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite instrument has been used as an empirical predictor to estimate ground-level concentrations of PM_{2.5}. These estimates usually have large uncertainties and errors. The main objective of this work is to assess the value of using upwind (Lagrangian) MODIS-AOD as predictors in empirical models of PM_{2.5}.

The upwind locations of the Lagrangian AOD were estimated using modeled backward air trajectories. Since the specification of an arrival elevation is somewhat arbitrary, trajectories were calculated to arrive at four different elevations at ten measurement sites within the continental United States. A systematic examination revealed trajectory model calculations to be sensitive to starting elevation. With a 500 m difference in starting elevation, the 48-hr mean horizontal separation of trajectory endpoints was 326 km. When the difference in starting elevation was doubled and tripled to 1000 m and 1500m, the mean horizontal separation of trajectory endpoints approximately doubled and tripled to 627 km and 886 km, respectively. A seasonal dependence of this sensitivity was also found: the smallest mean horizontal separation of trajectory endpoints was exhibited during the summer and the largest separations during the winter.

A daily average AOD product was generated and coupled to the trajectory model in order to determine AOD values upwind of the measurement sites during the period 2003-2007. Empirical models that included *in situ* AOD and upwind AOD as predictors of PM_{2.5} were generated by multivariate linear regressions using the least squares method. The multivariate models showed improved performance over the single variable regression (PM_{2.5} and *in situ* AOD) models. The statistical significance of the improvement of the multivariate models over the single variable regression models was tested using the extra sum of squares principle. In many cases, even when the R² was high for the multivariate models, the improvement over the single models was not statistically significant.

The R² of these multivariate models varied with respect to seasons, with the best performance occurring during the summer months. A set of seasonal categorical variables was included in the regressions to exploit this variability. The multivariate regression models that included these categorical seasonal variables performed better than the models that didn't account for seasonal variability. Furthermore, 71% of these regressions exhibited improvement over the single variable models that was statistically significant at a 95% confidence level.

I dedicate this work to my friends Tera and Jon without whom
this endeavor would have been impossible.

It also goes to my mom without whom I would not even have had a chance.

All have their worth and each contributes to the worth of the others.

-J.R.R. Tolkien

TABLE OF CONTENTS

	Page
1. Introduction.....	1
1.1 Particulate material.....	3
1.2 Remote sensing of particulate material	4
1.3 Estimating PM _{2.5} using remotely sensed aerosol optical depth.....	8
1.4 The role of transport.....	12
2. Objectives.....	14
3. Data.....	16
3.1 Study locations.....	16
3.2 Respirable particulate material	18
3.3 Air trajectory models	19
3.4 Aerosol optical depth.....	20
3.5 Relative Humidity.....	21
4. Trajectory Model Sensitivity to Starting Elevation	23
4.1 Trajectory calculations	25
4.2 Horizontal separation and directional difference.....	28
4.3 Conclusions.....	39
5. AOD Spatial Correlation.....	42
5.1 AOD pixel correlation.....	45
5.2 Distance weighted average.....	49
6. Estimation of PM _{2.5} Using Lagrangian MODIS-AOD	52
6.1 Linear regressions of PM _{2.5} and Lagrangian MODIS-AOD.....	54
6.2 Statistical significance of the regression models.....	56

6.3 Multicollinearity	74
6.4 Sample size and seasonal variability.....	77
6.5 Estimates of PM2.5 in cases of missing in situ AOD	86
6.6 Summary	88
7. Summary, Conclusions and Future Work.....	90
Bibliography	95

LIST OF TABLES

	Page
Table 1.1 MODIS channels and primary uses for aerosol determinations (Guenther 1996; Levy et al., 2009).	6
Table 3.1 Study sites and their geophysical characteristics.	17
Table 3.2 Study sites, codes and elevations.	18
Table 4.1 Mean horizontal separation (km) of 48-hr trajectories with different starting elevations, January 1st, 2000 to December 12th, 2007. Standard deviation in parenthesis.	36
Table 6.1 Data set for Milwaukee, Wisconsin with the location of the Lagrangian AOD determined by the trajectories arriving at 500 m agl for the combined winters of 2003-2007. Included are predicted PM _{2.5} using the single and multivariate regressions. AOD subscripts refer to the number of hours upwind.	55
Table 6.2 Regression coefficients for single and multivariate linear regressions for Milwaukee, Wisconsin with the location of the Lagrangian AOD determined by the trajectories arriving at 500 m agl for the combined winters of 2003-2007.	55
Table 6.3 Goodness of fit and statistical significance test for yearly regressions for the period 2003-2007. (Bold p-values indicate values less than 0.05.)	60
Table 6.4 Goodness of fit and statistical significance test for seasonal regressions for the period 2003-2007. (Bold p-values indicate values less than 0.05.)	65
Table 6.2 Correlation matrix of Lagrangian AOD predictors. AOD subscripts refer to the number of hours upwind.	76
Table 6.6 Comparison of the percentage of statistically significant “multivariate+seasons” regressions with respect to single model and median R ² to the properties of the study sites.	84
Table 6.7 Sample test data for Milwaukee, Wisconsin on August 7 and 19, 2007 and August 7, 2007. The data corresponds to Lagrangian AOD determined by trajectories arriving at 500 m agl.	87
Table 6.8 Regression coefficients for single and multivariate+seasons linear models for Milwaukee, Wisconsin with Lagrangian AOD determined by trajectories arriving at 500 m agl for 2007.	87

LIST OF FIGURES

	Page
Figure 3.1 Study sites (details listed in Tables 3.1 and 3.2).	16
Figure 4.1 (top) Sample 48-hr back-trajectories for Milwaukee, Wisconsin arriving at 1800 UTC on January 5th, 2000. The filled circles along each trajectory are drawn at 3, 12, 24, 36 and 48-hr upwind. (bottom) Trajectory altitude as a function of travel time.	29
Figure 4.2 Horizontal separation analysis for sample back-trajectories arriving at Milwaukee, Wisconsin at 1800 UTC on January 5th, 2000.	30
Figure 4.3 Directional difference analysis for sample back-trajectories arriving at Milwaukee, Wisconsin at 1800 UTC on January 5th, 2000.	31
Figure 4.4 Length of trajectories arriving at Milwaukee, Wisconsin during January 1st, 2000 - December 31st, 2007. Results are shown for trajectories with starting elevations of 500 m, 1000 m, 1500 m, and 2000 m agl. The box-plots show the median (horizontal line within the box), the 25th and 75th percentiles (vertical extent of box) and the 5th and 95th percentiles (vertical extent of dashed lines).	32
Figure 4.5 Horizontal separation analysis for trajectories arriving at Milwaukee, Wisconsin during January 1st, 2000 - December 31st, 2007. Results are shown for pairs of trajectories with starting elevations of 500 m, 1000 m, 1500 m, and 2000 m agl. The box-plots show the median (horizontal line within the box), the 25th and 75th percentiles (vertical extent of box) and the 5th and 95th percentiles (vertical extent of dashed lines).	33
Figure 4.6 Directional difference analysis of 3-hr endpoints for trajectories arriving at Milwaukee, Wisconsin during January 1st, 2000 - December 31st, 2007. Results are shown for pairs of trajectories with starting elevations of 500 m, 1000 m, 1500 m, and 2000 m agl. The box-plots show the median (horizontal line within the box), the 25th and 75th percentiles (vertical extent of box) and the 5th and 95th percentiles (vertical extent of dashed lines).	34
Figure 4.7 Directional difference analysis for trajectories arriving at Milwaukee, Wisconsin during January 1st, 2000 - December 31st, 2007. The histogram shows directional differences of 3-hr endpoints for trajectories calculated with 500 m and 2000 m agl starting elevations.	35
Figure 4.8 Mean horizontal separation as a function of the difference in starting elevation separation for all sites combined during January 1st, 2000 - December 31st, 2007.	38
Figure 5.1(a) Sample MODIS granule retrieved by Aqua on July 31, 2007 at 1850 UTC in the visible band (images generated by MODIS LAADS Granule Browser).	43

Figure 5.1(b) Sample MODIS granule retrieved by Aqua on July 31, 2007 at 1850 UTC in Optical Depth Land and Ocean (images generated by MODIS LAADS Granule Browser).	43
Figure 5.2 MODIS-AOD granules for July 31, 2007 retrieved by Terra and Aqua overlapped on the 10°N to 60°N and 50°W to 140°W region (unprocessed AOD).	44
Figure 5.3 Correlation of MODIS-AOD pixels (10km distance bin) for a sample granule for the Aqua - MODIS granule for July 31st, 2007 at 1850 UTC in Figure 5.1(b).	46
Figure 5.4 Correlation of MODIS-AOD pixels (10km distance bin) for twelve random granules.	47
Figure 5.5 Mean correlation of MODIS-AOD with respect to distance (1 standard deviation bars) from 100 randomly selected granules.	48
Figure 5.6 Sample AOD fields for July 31st, 2007 (distance weighted average).	50
Figure 6.1 Scatter plots of observed vs. predicted PM _{2.5} mass concentration for single (a) and multivariate (b) linear regression for Milwaukee, Wisconsin with the location of the Lagrangian AOD determined by the trajectories arriving at 500 m agl for the combined winters of 2003-2007. Red line indicates perfect fit between observed and predicted PM _{2.5}	56
Figure 6.2 Comparison of the distribution of R ² for yearly (a) and seasonal (b) data aggregates for each site for the period 2003-2007. Single, multivariate and control regressions are shown. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.	69
Figure 6.3 Goodness of fit distribution of yearly linear regressions for individual sites for the period 2003-2007. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.	71
Figure 6.4 Goodness of fit distribution of seasonal linear regressions for individual sites for the period 2003-2007. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.	72
Figure 6.5 Goodness of fit distribution for yearly linear regressions grouped into arrival elevations for the period 2003-2007. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.	74
Figure 6.6 Correlation of upwind AOD to in situ AOD (AOD ₀) as a function of upwind time.	75
Figure 6.7 Improvement over single regressions of PM _{2.5} - Lagrangian AOD - Seasonal Categorical Variables (multivariate+seasons) (a) and yearly multivariate regressions of PM _{2.5} - Lagrangian AOD (multivariate) (b) for individual study sites at each of the four trajectory arrival elevations for the period 2003-2007. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.	79

Figure 6.8 Goodness of fit distribution for four types of yearly linear regressions for the period 2003-2007 (individual sites and elevations). Multivariate regression with control sample AOD predictors is shown as a control. All cases considered regardless of statistical significance of improvement over the single model.80

Figure 6.9 R^2 distribution of the yearly $PM_{2.5}$ - Lagrangian AOD - Seasonal Categorical Variable linear regression for the period 2003-2007 as a function of number of AOD predictors. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions. Percentage of statistically significant cases with respect to single variable regression without seasonal variables also shown (solid black line, axis on the right).82

Figure 6.10 R^2 distribution of the yearly $PM_{2.5}$ - Lagrangian AOD - Seasonal Categorical Variable linear regression for the period 2003-2007 grouped into individual sites. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions. Percentage of statistically significant cases with respect to single variable regression without seasonal variables also shown (solid black line, axis on the right).83

Figure 6.11 Goodness of fit distribution for three types of yearly linear regressions for the period 2003-2007 (individual sites and elevations). Multivariate, multivariate+seasons and a full model including relative humidity. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.85

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Jonathan Kahl, who taught me the true nature of scientific research and how to effectively communicate my results. He taught me not to be so worried about the future and to focus on the present. He also taught me to work in a smarter way and not just do things because I can, which is a lot of fun, but not always the right thing to do. During these five years in Milwaukee he was more than an academic advisor, he always made me feel as family. He also provided me with monetary support during my last summer in Milwaukee and this work resulted in an article. He also allowed me to participate in the study abroad course as an academic co-director; that was an invaluable learning experience and hopefully in the future I will be able to organize study abroad courses of my own.

I would like to thank Dr. Paul Roebber, Dr. Sergey Kravtsov, Dr. Kyle Swanson and Dr. Vincent Larson. Never once during my time at the University of Wisconsin - Milwaukee did I feel like a stranger or a misfit for having a different scientific background. They patiently answered all of my questions and always gave me great advice, not only concerning my work at UWM, but also my continued professional development. I would also like to thank Dr. Jugal Ghorai whose invaluable advice made this work take a completely different perspective.

I would also like to acknowledge, from very far away, my professors from the Universidad Autónoma de Nuevo León who believed in me and encouraged me to pursue my doctoral studies.

I also would like to acknowledge the support of the Mexican National Council of Science and Technology (CONACyT) that made this work possible.

1. Introduction

Atmospheric aerosols are particles, solid or liquid, small enough to remain suspended in the atmosphere. Exposure to suspended particulate matter with aerodynamic diameters less than 2.5 μm (PM_{2.5}) has negative effects on human health and may induce respiratory problems, cardiovascular and lung diseases, and additional health problems (Pope et al., 2002; Kappos et al., 2004). Aerosols and their chemical reaction products also erode the materials that constitute sites of historical and cultural interest (Reyes et al., 2009). Furthermore, aerosols play an important role in the climate system by altering the Earth's energy budget as they can scatter and absorb radiation (Charlson et al., 1992; Li, 1998) and modify the microphysical and radiative properties of clouds (Twomey et al., 1984; Hansen et al., 1997). Any changes in the atmospheric aerosol load can result in an associated altering of the global climate. The measurement of ground-level PM_{2.5} provides valuable information for an effective forecasting of air quality (Al-Saadi et al., 2005). Air quality monitoring networks have been established in many industrialized countries to make measurements of pollutant concentrations at different locations, on a daily or hourly basis.

Even though ground-based measurements are generally accurate, they are representative of only relatively small areas around the monitoring stations. Often, the limited spatial coverage and irregular distribution of these stations restricts the study of air pollution and its effects on human health and the environment.

Satellite remote sensing has been employed to supplement the prediction and measurement of ground-level PM_{2.5} concentrations. Satellites are able to monitor vast spaces, especially areas where ground-based monitoring stations are not available. For aerosol studies, the launch of the Moderate Resolution Imaging Spectroradiometer

(MODIS) installed on the Terra and Aqua satellites has enabled the retrieval of aerosol optical depth (AOD) data globally from the satellites' spectral observations (Kaufman and Boucher, 2002). AOD is a measure of light extinction by aerosols in an atmospheric column during the satellite overpass.

There is a direct correlation, although it's not perfect, between the aerosol load in the atmosphere and AOD. With the evolution of the retrieval algorithms which process the irradiance measured by the instrument, MODIS AOD has become increasingly important in the role of producing more accurate estimations of ground-level $PM_{2.5}$ concentration (Kaufman et al., 1997; Tanré et al., 1997; Remer et al., 2005; Levy et al., 2005).

Single regression models between AOD and measured $PM_{2.5}$ have been used to estimate the ground-level $PM_{2.5}$ concentrations in several major cities worldwide by using the satellite-derived MODIS-AOD (Gupta et al. 2006). However, MODIS measures the light extinction within the total integrated atmospheric column, while the aerosol concentration is measured only at surface level. For this reason the correlation between them is influenced by the vertical distribution of aerosols and other meteorological factors that affect the aerosol extinction coefficient such as relative humidity, the mixing layer height and temperature. Therefore the estimates of ground-level particulate matter concentrations simply from AOD usually have large uncertainties.

In this work we seek to improve the estimates of ground level $PM_{2.5}$ by including variables related to atmospheric transport. Upwind AOD were used as predictors of in-situ $PM_{2.5}$. In the following sections we will discuss the characteristics and properties of $PM_{2.5}$, the remote sensing and use of AOD to estimate concentrations of these pollutants, and the role of atmospheric transport and why upwind AOD is a valuable predictor.

1.1 Particulate material

Particulate matter is a mixture of both solid and liquid particles suspended in air (aerosols) and is usually classified in two distinct modes according to their aerodynamic diameter: the *fine* mode (<2.5 μm) and the *coarse* mode (2.5 - 10.0 μm). Fine and coarse particles differ in origin and physicochemical properties (Kaufman et al., 2002). Fine particulate matter ($\text{PM}_{2.5}$), also known as respirable particulate material, is usually related to anthropogenic pollution (Tandon et al., 2008). $\text{PM}_{2.5}$ is composed of varying amounts of sulfate, ammonium and nitrate ions, elemental carbon, organic carbon compounds, water, and small amount of soil dust, and trace species (Pb, Cd, V, Ni, Cu, Zn, Mn, Fe, etc); their lifetime ranges from days to weeks. On the other hand coarse particles (PM_{10}) are mainly formed by mechanical forces such as crushing, grinding, and abrasion of materials of geological origin. Pollen and spores are also included in the coarse particle range. Generally, coarse particles consist of aluminosilicate and other oxides of crystal elements, and major sources include dust from roads, industry, agriculture, construction and demolition; their lifetime being from minutes to hours (Roelofs et al., 2001; Tasić et al., 2006; Roelofs, 2012).

The results of a number of studies have shown that emission of aerosols of natural origin exceeds emission from anthropogenic sources on a global scale by a factor of 4 to 5 (Hobbs, 2000). However according to some estimations, emission from anthropogenic sources will, up to 2040, reach the level of the natural emission mainly as a consequence of fossil fuel combustion with a higher increase rate in China and India (European Environment Agency, 2003).

National air quality forecasts for major U.S. metropolitan areas have been provided to the public through a partnership between the U.S. Environmental Protection Agency and state and local agencies since 1997. These forecasts were done primarily for ground-level

ozone, but in October 2003 forecasts of particulate matter were started. Forecasters rely on a variety of information sources and tools to conduct these forecasts, including empirical and statistical models and, increasingly, numerical forecast guidance (e.g. McHenry et al. 2004).

Routinely made pollutant measurements at surface monitoring stations across the country can provide a direct way to obtain current information on air quality to generate an effective forecast. However, large regions of the United States are devoid of surface monitors, and coastal regions are often influenced by polluted air approaching over water. Pollution layers and plumes may also be transported aloft over long distances, undetected by surface monitors, and then descend to influence air at the ground. Observations made from satellites can help address the limited spatial coverage and irregular distribution of air pollution monitoring stations by augmenting the surface network. Satellite remote sensing has been widely used to detect aerosol since the late 1970s (King et al., 1999).

1.2 Remote sensing of particulate material

One of the instruments currently used in the remote sensing of particulate material is the Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS is a sensor onboard the National Aeronautics and Space Administration (NASA) Earth Observing System Terra and Aqua satellites. Both satellites are polar-orbiting, with Terra on a descending orbit (southward) over the equator about 10:30 local sun time, and Aqua on an ascending orbit (northward) over the equator about 13:30 local sun time. MODIS observes the entire globe on a nearly daily basis, and repeats orbits every 16 days (Salomonson et al., 1989; Barnes et al., 1998).

MODIS continuously scans the earth from 705 km above the surface with a field of view 110° across track. Each MODIS scan, approximately perpendicular to the direction of motion of the satellite ground track, is 2330 km (across track) by 10 km (along track). Due to spherical geometry, the size of each pixel increases from 1 km at nadir to nearly 2 km at the scan edges. Each scan is covered by 1354 pixels with a 1-km resolution at nadir. The MODIS orbit is separated into 5-minute sections called 'granules'. Each granule is about 2030 km (203 scans of 10 km) along the orbital path. Each granule is 1354 (across path) by 2030 (along path) pixels in the 1 km resolution (Remer et al., 2005).

MODIS continuously measures reflected and emitted radiance from the atmosphere in 36 channels from the near ultraviolet to the thermal infrared part of the electromagnetic spectrum ($0.405\mu\text{m}$ - $14.385\mu\text{m}$) at three different spatial resolutions: 250 m, 500 m and 1 km. These calibrated radiance measurements are processed to generate an aerosol optical depth (AOD) product. Aerosol optical depth is a measure of the column (surface to top of atmosphere) integrated light extinction (absorption plus scattering). There is a direct correlation between the aerosol load in the atmosphere and AOD: the higher the aerosol concentration is in the atmosphere, the higher the aerosol optical depth will be.

Radiance measurements are generated during daytime conditions only, so the maximum time between observations that cover the same location, for a single satellite sensor, is about 24hr over most of the US but increases to about 48hr for points south of $\sim 30^\circ\text{N}$ (Hutchison et al., 2008). Thus there is at least one measurement daily per satellite (Terra and Aqua) for locations at latitudes larger than 30°N though sometimes two measurements may be recorded at different times in a given 24hr period. MODIS thus provides a convenient method for aerosol remote sensing since it has near-daily global coverage.

Of the 36 channels available, the MODIS-AOD algorithm uses seven channels (see Table 1.1) over ocean (466nm – 2119nm) and two channels over land (466nm and 646nm) in conjunction with pre-computed look-up tables to generate the AOD product (Kaufman et al., 1997; Tanré et al., 1997; Remer et al., 2005; Levy et al., 2009). Channels 8 - 36 have other uses such as determination of atmospheric water vapor, ozone, oceanic biogeochemistry and surface temperature, among others (Gao and Kaufman, 1995; Guenther 1995; Levy et al., 2009; modis.gsfc.nasa.gov/about/specifications.php).

MODIS-AOD is generated by two independent algorithms, one for deriving aerosols over land (Kaufman et al., 1997) and the second for aerosols over ocean (Tanré et al., 1997). Both algorithms were conceived and developed before the Terra launch in 1999 (Aqua launched in 2002) (Kaufman et al., 2002).

Table 1.1 MODIS channels and primary uses for aerosol determinations (Guenther 1996; Levy et al., 2009).

Primary Use	Channel	Bandwidth (nm)	Weighted Central Wavelength (nm)	Resolution (m)
Land/Cloud/Aerosol Boundaries	1	620 - 670	646	250
	2	841 - 876	855	250
Land/Cloud/Aerosol Properties	3	459 - 479	466	500
	4	545 - 565	553	500
	5	1230 - 1250	1243	500
	6	1628 - 1652	1632	500
	7	2105 - 2155	2119	500

The MODIS-AOD algorithms use the spectral ‘reflectance’, a function of the measured spectral radiance, the solar zenith angle, and the solar irradiance (radiative flux) in a given wavelength band or channel (see Table 1.1). The retrieval algorithms apply screening to a 10x10km analysis area. The screening process eliminates certain pixels that don’t have

the right reflectance characteristics to avoid contamination by clouds, inland water, snow, ice, and solar glare over the ocean (Hutchison et al., 2005). Once these reflectance pixels have been screened, the values are averaged and the AOD retrieval is performed by using empirical relations and lookup values. These empirical correlations and lookup values take into account the brightness of the pixels, count of unscreened (good) pixels used in the calculation, the channel used, geography, vegetation, the observation angle of the instrument, and mode (coarse or fine particles). The resulting MODIS-AOD product is dimensionless and has a horizontal spatial resolution of 10 km, which yields a granule of 135×203 pixels.

MODIS-AOD has been comprehensively validated (Chu et al., 2002, 2003; Remer et al. 2005; Levy et al., 2005, 2009) by observations through the Aerosol Robotic Network (AERONET), a network of ground-based sun photometers which measure atmospheric aerosol properties (Holben et al., 1998). Nearly a ten-year record of MODIS aerosol data is now available on a near global basis. These data are being used in several applications including climate and air quality (Christopher et al., 2010a).

Since its inception in 1997 (Kaufman et al., 1997) the MODIS retrieval algorithm has undergone various improvements. Remer (2005) and Liang et al. (2006) have improved the AOD retrieval algorithms from MODIS by including different filters that correct reflectances according to surface type and vegetation. The original MODIS aerosol retrieval algorithm is based on the “Dark Target” approach and as a result it provides less reliable or no retrieval data over desert and other highly reflecting land surfaces. Wang et al. (2007) evaluated the MODIS-AOD retrieval algorithm over different ecosystems. The evaluation results showed different performances in the MODIS-AOD retrieval between different ecosystems and geographic locations (rural or urban areas) when validated using lidar. They found that the MODIS-AOD retrieval algorithm performs well over ‘dark’ surfaces, but

poorly over ‘bright’ surfaces. To improve the quality of bright surface retrievals, the aerosol team has implemented the “Deep Blue” retrieval algorithm which compliments the existing Dark Target method. In 2008 all the preexisting products were reprocessed with this upgraded algorithm (<http://ladsweb.nascom.nasa.gov/data/search.html>).

MODIS aerosol products have been used to analyze the temporal and spatial variability of aerosols over a wide area of interest ranging from radiation budget, cloud microphysics (Castleman 1974; Li et al., 1998; Li et al., 2004; Bellouin et al., 2005, 2008; Slater et al., 2004), analysis of air quality (Lyapustin et al., 2004; Gupta et al., 2006; Schäfer et al., 2008; Chang et al., 2010), epidemiological studies (Liu et al., 2007; Meng et al., 2012; Liu, 2013), and climate forcing of anthropogenic aerosols (Charlson et al., 1992).

1.3 Estimating PM_{2.5} using remotely sensed aerosol optical depth

MODIS-AOD has been used extensively to infer ground-level aerosol concentrations using simple linear regression models between AOD and *in situ* PM_{2.5} concentrations (Engl-Cox et al.; 2004; Hutchison et al., 2005; Liu et al., 2007; Schäfer et al., 2008; Li et al., 2009; Li et al., 2005a; Gupta and Christopher, 2008; Christopher et al., 2010b). With single linear regressions between AOD and ground-level PM_{2.5} goodness of fit ranges between $R^2 = 0.1-0.7$ (Hutchison et al., 2005; Gupta et al., 2006; Hutchison et al., 2008; Hu et al., 2009; Li et al., 2009)¹.

The concentrations of both coarse (PM₁₀) and respirable (PM_{2.5}) aerosol fractions and the aerosol extinction coefficient are related to meteorological factors such as relative

¹ In studies where high correlations were obtained ($R^2 > 0.5$) for single linear regressions of AOD and MODIS-PM_{2.5} (e.g. Hutchison et al., 2005), monthly averages PM_{2.5} and AOD were used. This approach may be unsuitable for use in issuing air quality forecasts (Hutchison et al., 2005; Li et al., 2009).

humidity, mixing layer height, temperature, wind speed and direction (Li et al., 2005b; Koelemeijer et al., 2006; Tandon et al., 2010). Gupta et al. (2006) determined that the $PM_{2.5}$ -AOD relationship strongly depends on ambient relative humidity, fractional cloud cover and height of the mixing layer. The highest correlation between MODIS AOD and $PM_{2.5}$ mass is found under clear sky conditions with less than 40–50% relative humidity and when atmospheric mixing height ranges from 100 to 200 m. These meteorological factors have been used as predictors in empirical models in an attempt to improve ground-level $PM_{2.5}$ estimates from remotely sensed AOD (Liu et al., 2007, 2009; Schäfer et al., 2008; Guo et al., 2009; Wang et al., 2010; Tian et al., 2010; Boyouk et al., 2010).

Multivariate linear regressions that include MODIS-AOD, relative humidity, and ground-based measurements of surface temperature and surface relative humidity have been used to improve the prediction of ground level $PM_{2.5}$. Liu et al. (2007) improved the estimates of daily ground-level $PM_{2.5}$ concentrations using empirical multivariate linear regression models that used wind speed, surface temperature, mixing layer height and MODIS-AOD as predictors. They also used two categorical variables: wind direction indicator, and seasonal indicator. Tian et al. (2010) developed similar multivariate models were able to provide better predictions for agricultural and rural areas than urban areas. Tandon et al. (2010) found that the incorporation of ground-level temperature and relative humidity is significant in improving the model predictability.

Besides multivariable linear regressions, other techniques such as nonlinear regressions and artificial neural networks have been used to relate the MODIS-retrieved AOD and meteorological variables with surface aerosol concentrations. Gupta et al. (2009) used neural networks to assess the correlation between ground-level $PM_{2.5}$, MODIS-AOD and meteorological measurements. They did not find significant improvement of ground-level $PM_{2.5}$ estimates with respect to the estimates generated with multivariable regressions.

Cobourn (2010) used nonlinear regression models to estimate ground-level PM_{2.5} concentrations. Meteorological variables were used as predictors along with a Lagrangian variable: PM_{2.5} concentration 24-hr upwind before the arrival to the target. The use of 24-hr upwind concentrations as a model input improved the ground-level PM_{2.5} forecasts. The models with the Lagrangian parameter performed significantly better than the basic nonlinear regression models with meteorological inputs.

The correlation between remotely sensed AOD and ground-level PM_{2.5} has deficiencies because while the aerosol concentration is measured only at surface level, MODIS determines AOD for the whole integrated atmospheric column. Thus the correlation between AOD and PM_{2.5} is influenced by the vertical distribution of aerosols. Engle-Cox et al. (2006) used lidar apportionment to determine the fraction of aerosol optical depth that is within the planetary boundary layer. This apportioned AOD yields better agreement with surface PM_{2.5} than does the total column amount. Hutchison et al. (2008) incorporated aerosol vertical profile inferred from the vertical component of a real-time trajectory-based forecast model and achieved improved MODIS-AOD and PM_{2.5} correlations.

Boyouk et al. (2010) found a large variability in the mixing layer height diurnal cycle and in the aerosol extinction vertical distribution using lidar analyses. They determined that the lidar backscattering coefficient at the lowest possible level is well correlated to the ground-level PM_{2.5}. Wang et al. (2010) found that the correlation between AOD and the aerosol extinction coefficient at surface level is improved as a result of the vertical correction and by the relative humidity correction for PM_{2.5}.

Another factor that affects the ground-level PM_{2.5} and MODIS-AOD correlations is seasonality. Wang et al. (2003) found agreement between monthly mean PM_{2.5} and MODIS-AOD with the largest correlations during the summer months. They also found that PM_{2.5}

has a distinct diurnal signature with maximum concentration values in the early morning (6:00 - 8:00AM) due to increased traffic flow and restricted mixing depths during these hours. Schaap et al. (2008) found a strong seasonal signature in the estimation of $PM_{2.5}$ inferred by MODIS-AOD. They found that this pollutant tends to be overestimated during the summer.

MODIS-AOD and ground-level $PM_{2.5}$ linear regressions are also sensitive to the temporal scale used in the regression. Li et al. (2009) found that empirical models with temporal scale of 30 days display the best performance in estimating ground-level $PM_{2.5}$. Beyond this time frame, $PM_{2.5}$ estimates lose predictive value. Because they can vary from different seasons and even different days, correlation coefficients between AOD and ground-level $PM_{2.5}$ cannot be predetermined in real-time daily air quality estimations and need to be calculated for each day's run. Guo et al. (2009) found that MODIS-AOD had a better positive correlation with hourly average $PM_{2.5}$ concentration compared to daily average $PM_{2.5}$ due to diurnal variation in particulate matter measurements.

Previous authors have attempted to improve the estimates of $PM_{2.5}$ by coupling MODIS-AOD with other instruments. Liu et al. (2007) compared the ability of MODIS and the Multiangle Imaging SpectroRadiometer (MISR) to predict ground-level $PM_{2.5}$. Their results suggest that air quality forecasting can be improved by combining the predicting accuracy of MISR and the better spatial coverage of MODIS.

Hu et al. (2009) developed a method to retrieve the $PM_{2.5}$ concentration from MODIS/MISR satellite measurements. A global chemical transport model (GEOS-CHEM) was used to generate the information on aerosol composition, size distribution and vertical profiles relative to retrievals of $PM_{2.5}$ to reduce uncertainties in satellite retrievals of aerosol properties. They found that high values of $PM_{2.5}$ occur over most parts of Eastern Europe,

Eastern United States, and areas surrounding the Saharan region, associated with biomass burning, industrial pollution and dust storms.

1.4 The role of transport

As previously discussed, the lifetime of respirable particulate matter ranges from days to weeks (Roelofs et al., 2002; Tasić et al., 2006; Roelofs, 2012). Therefore, $PM_{2.5}$ that is being measured at a given monitoring station, though associated with local meteorological conditions, may not have been generated locally but at a distant upwind location (NARSTO, 2003). Cobourn (2010) found that the dependence of $PM_{2.5}$ on local meteorology is usually weaker than for other pollutants (e.g. O_3) given their extended average atmospheric lifetime. Thus, most of the airborne $PM_{2.5}$ consists of relatively aged particles rather than particles formed recently from processes influenced by local weather. Therefore atmospheric transport must provide important information on the estimation of ground level $PM_{2.5}$.

For several decades, air parcel trajectory models have been commonly used tools for estimating the atmospheric transport pathways of air parcels. A modeled trajectory is an estimate of the movement of a single air parcel through the atmosphere. Single trajectories provide guidance on potential downwind impact areas (for forward trajectories) and upwind source regions (for backward trajectories), and are thus often used in air quality studies. Reviews of various aspects of trajectory models, including computational details, model accuracy, and applications, have been provided by Danielsen (1961), Miller (1987) and Stohl (1998). Besides estimating single trajectories, current trajectory models have capabilities for quantifying the dispersion due to turbulent mixing and wind shear by coupling the advection (single trajectory) component with particle dispersion models (Draxler and Hess, 1997; Stohl

et al., 1995). These model features allow the computation of gas and particle concentration and deposition, as well as different types of trajectory ensembles.

Despite the availability of particle dispersion models, however, the calculation of single air parcel trajectories to estimate atmospheric transport pathways continues to be a widely used application of trajectory models. The use of single trajectories has been frequently reported in the recent literature in a diverse range of studies including volcanic plumes (Revuelta et al., 2012), carbon cycles (Xueref-Remy et al., 2011), cloudwater chemistry (Hutchings et al., 2009) and aerosol transport (Wu et al., 2012; Koçak et al., 2012; Batmunkh et al., 2011; Kharol et al., 2011; Orru et al., 2010; Jaffe and Reidmiller, 2009; Occhipinti et al., 2008; Kouyoumdjian and Saliba, 2006).

Trajectory models have been used to aid the improvement of ground-level $PM_{2.5}$ estimates coupled with MODIS-AOD. Al-Saadi et al. (2005) assessed MODIS-AOD as a forecasting tool in conjunction with lidar and sun photometers. Trajectory models were used to make air quality forecasts where MODIS detected high AOD. It was found that the actual relation between AOD and $PM_{2.5}$ varied due to changes aerosol composition as inferred by the source location determined with the air trajectory models. Hutchison et al. (2008) used backward-trajectory models to assess the aerosol vertical structure the lower tropospheric levels thus improving the $PM_{2.5}$ estimations. Cobourn (2010) used 24-hr upwind $PM_{2.5}$ as a predictor in empirical nonlinear regression models for determining $PM_{2.5}$

In this work we seek to improve estimates of ground-level $PM_{2.5}$ using upwind MODIS-AOD as predictors. In the following chapters we will discuss the objectives of this work, the data used, the selection of the correct trajectories, the improvement over the state-of-the-art and the statistical significance of the improvement of the empirical models obtained.

2. Objectives

The main objective of this work is to improve the estimates of $PM_{2.5}$ concentrations by using MODIS-retrieved AOD and Lagrangian² variables. As previously mentioned, the estimation of $PM_{2.5}$ through AOD can be improved by including meteorological variables in the regressions. Even though the use of Lagrangian variables as predictors in the estimation of $PM_{2.5}$ has been explored before, none of the studies have fully addressed the predictive potential of Lagrangian MODIS-AOD.

Aerosol concentrations *in situ* are dependent mainly on the following processes:

- (a) Transport: airborne $PM_{2.5}$ consists of relatively aged particles rather than particles formed recently, therefore, the transport of material and the meteorological properties in the location where the aerosol was generated and advected from is of much importance even though the source may be far away from the site of interest. With this in mind, upwind $PM_{2.5}$ concentrations would be expected to be correlated with *in situ* $PM_{2.5}$ concentration. Similarly, upwind AOD would be expected to be correlated with *in situ* $PM_{2.5}$ and *in situ* AOD.
- (b) Sources and sinks: the strength of pollution sources is directly related to downwind $PM_{2.5}$ and certain chemical and physical processes may act as sinks of aerosols.

² In the context of this work we will refer to variables whose upwind locations have been determined with single backward trajectory models as Lagrangian variables.

- (c) Diffusion: as the material is transported it may be diffused and its concentration reduced. Atmospheric diffusion is related to wind speed and atmospheric stability.

We speculate that by using transport-related variables that affect *in situ* PM_{2.5} concentrations the estimates of PM_{2.5} can be improved when integrating them into the empirical models along with remotely sensed AOD. We propose a methodology designed to improve the estimations of PM_{2.5} that exploits the atmospheric transport of PM_{2.5}.

Possible applications of this project include an improvement in the ability to provide estimates of PM_{2.5} when there is missing *in situ* data; to provide data for remote or inaccessible regions where the study of aerosols is required, such as in the corrosion research carried out for sites of archeological interest (Reyes et al., 2009); and to improve air quality forecasts.

When calculating the upwind locations of MODIS-AOD we had to select the arrival elevation for the backward trajectory models. This led us to analyze and quantify the sensitivity of trajectory models to starting elevation in order to make a decision on the arrival elevation to be used. This sensitivity is discussed in Chapter 4.

3. Data

In this chapter we will provide a description of the locations selected for the present work. Details and sources of the data used will also be discussed in this chapter: the trajectory model and its associated meteorological data, ground-level $\text{PM}_{2.5}$, MODIS-AOD, and other meteorological variables that were used as predictors, i.e., relative humidity.

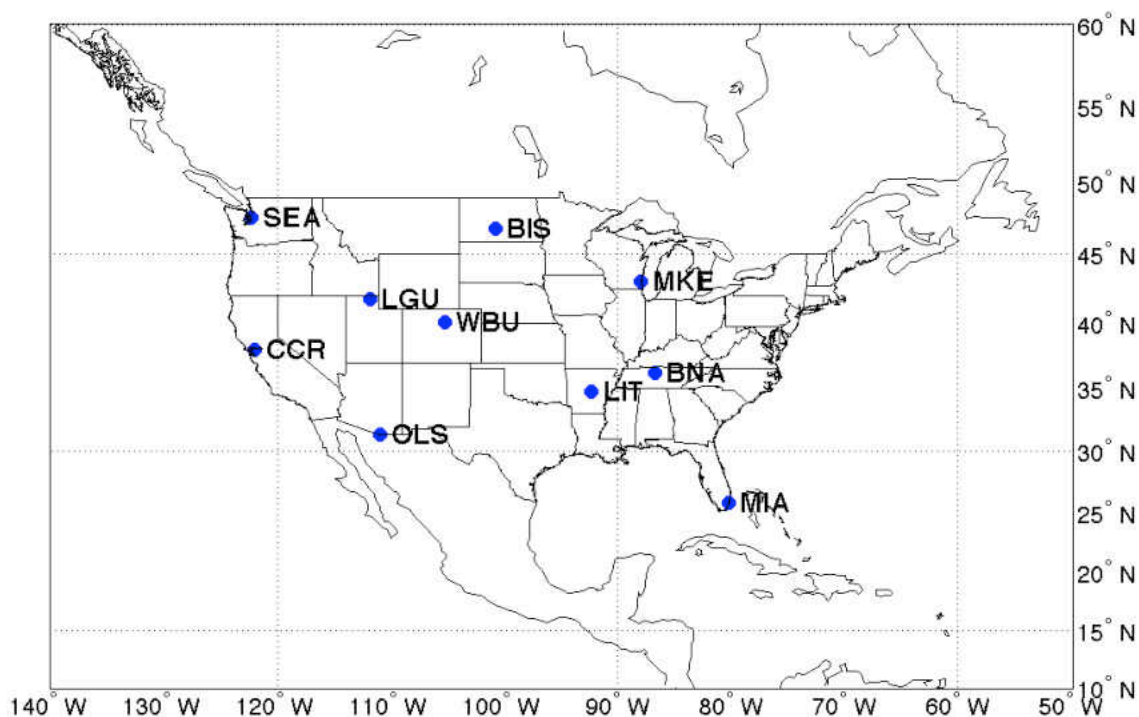


Figure 3.1 Study sites (details listed in Tables 3.1 and 3.2).

3.1 Study locations

Ten sites (Fig. 3.1) were selected for this work for which ground-level $\text{PM}_{2.5}$ measurements were obtained, backward trajectory models calculated and multivariate regression models

were constructed and evaluated. The sites were selected based on varied geography, climatology and population (Table 3.1).

The characteristics of the study sites, i.e., location, climate (Köppen classification), population, annual mean clear days, temperature, and precipitation are presented in Table 3.1. The data are climatological normals (daily and annual) 30-year average values computed from the data recorded during the period 1971-2000. Data in this table were obtained from the Normals Means and Extremes tables contained in the Local Climatological Data Annual Summary of the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA) (www.ncdc.noaa.gov/cdo-web/).

Table 3.1 Study sites and their geophysical characteristics.

Name	Climate	Population	Clear Days ¹⁰	Temp. (°C) ¹¹	Precip. (mm) ¹²
Nogales, AZ	BWh ¹	20,017	193	20.4	309
Logan, UT	BSk ²	49,549	125	11.1	419
N. Little Rock, AR	Cfa ³	60,140	119	17.1	1257
Bismarck, ND	Dfb ⁴	61,217	93	5.7	428
Boulder, CO	H ⁵	100,160	115	10.1	402
Concord, CA	Csb ⁶	122,224	160	14.6	511
Miami, FL	Af ⁷	433,136	74	24.8	1487
Milwaukee, WI	Dfa ⁸	605,013	90	8.6	884
Nashville, TN	Cfa	605,473	102	14.9	1222
Seattle, WA	Cfb ⁹	616,627	51	11.6	972

¹ BWh: Subtropical desert, low-latitude desert.

² BSk: Mid-latitude steppe, mid-latitude dry.

³ Cfa: Humid subtropical, mild with no dry season, hot summer.

⁴ Dfb: Humid continental, humid with severe winter, no dry season, warm summer.

⁵ H: Highland.

⁶ Csb: Mediterranean, mild with dry, warm summer,

⁷ Af: Tropical wet, no dry season.

⁸ Dfa: Humid continental, humid with severe winter, no dry season, hot summer.

⁹ Cfb: Marine west coast, mild with no dry season, warm summer.

¹⁰ Mean (rounded) number of clear days in a year (30-year period, 1971-2000).

¹¹ Normal daily mean (30-year period, 1971-2000).

¹² Normal annual mean (30-year period, 1971-2000).

3.2 Respirable particulate material

The PM_{2.5} data were obtained from the U.S. Environmental Protection Agency's Air Quality System (AQS) database (www.epa.gov/ttn/airs/airsaqs/). The monitoring data in AQS are the result of the various Clean Air Act requirements to provide a national database of ambient air pollution data: Criteria pollutant (SO₂, NO₂, O₃, CO, PM₁₀, PM_{2.5} and Pb) data, air toxic data, photochemical assessment data, and meteorological data. Individual observations, as well as summarized data are provided. The database contains values from 1957 through the present day.

In Table 3.2 we present the name of the ground level PM_{2.5} sites and three letter code according to the International Air Transport Association, latitude and longitude, sample collection frequency (SCF) of PM_{2.5} for the monitoring station, station elevation and elevation according to the trajectory algorithm³ and site ID according to the AQS.

Table 3.2 Study sites, codes and elevations.

Name	Code	Lat	Lon	PM _{2.5} SCF (days)	Station Elev. (m amsl)	CMDL Elev. (m amsl)	AQS Site ID
Nogales, AZ	OLS	31.34	-110.94	6	1176	937	04-023-0004
Logan, UT	LGU	41.73	-111.84	1	1380	1940	49-005-0004
N. Little Rock, AR	LIT	34.76	-92.28	1	80	147	05-119-0007
Bismarck, ND	BIS	46.83	-100.77	3	580	587	38-015-0003
Boulder, CO	WBU	40.02	-105.26	various	1619	2175	08-013-0012
Concord, CA	CCR	37.94	-122.03	3	27	401	06-013-0002
Miami, FL	MIA	25.79	-80.21	1	4	21	12-086-1016
Milwaukee, WI	MKE	43.02	-87.93	3	200	272	55-079-0010
Nashville, TN	BNA	36.18	-86.74	1	160	213	47-037-0023
Seattle, WA	SEA	47.57	-122.31	3	105	366	53-033-0080

³ The "CMDL elevation (m amsl)" column in Table 3.2 is discussed in detail in Chapter 4.

3.3 Air trajectory models

A trajectory model is an estimate of the movement of a single air parcel through the atmosphere. Atmospheric trajectory models use meteorological data and mathematical equations to simulate transport in the atmosphere. Generally, the position of particles or parcels of air is calculated based on meteorological data such as wind speed and direction, temperature, and pressure. Model results depend on the spatial and temporal resolution of the atmospheric data used, and on the complexity of the model itself. Simpler models may deal with only 2-dimensional transport by winds, while more complex models may include 3-dimensional chemical and thermodynamic processes such as aerosol formation, convection, and turbulent diffusion (Harris et al., 2005; Stohl et al., 1995; Kahl and Samson, 1986; Kuo et al., 1985).

The U.S. National Oceanic and Atmospheric Climate Monitoring and Diagnostics Laboratory (CMDL) model (Harris and Kahl, 1994) was used to calculate the trajectories in this study. Harris et al. (2005) determined that typical horizontal deviations between the CMDL model and other trajectory models such as HY-SPLIT (Draxler and Hess, 1997) are small thus the CMDL model is appropriate for the present study. The CMDL model is three-dimensional, with a fixed one-hour time step and a second-order Runge-Kutta integration technique for advection. The time resolution of the model input data is 6 hours (0000, 0600, 1200 and 1800 UTC). The trajectory model output is 240 hours with a resolution of 3 hours. Data interpolation is bilinear (horizontal) and linear (vertical and temporal). Additional model details are provided by Harris et al. (2005).

Meteorological data was supplied to the trajectory model through $2.5^{\circ} \times 2.5^{\circ}$ latitude/longitude global gridded fields of wind, temperature and geopotential height provided by the European Center for Medium-Range Weather Forecasts (Simmons et al.,

1995; Persson and Grazzini, 2005). This horizontal resolution is common for multi-day trajectory models (Harris et al., 2005) and is sufficient to capture the synoptic-scale meteorological features which control long-range atmospheric transport. Moreover, this resolution is frequently used in air quality and atmospheric transport studies employing trajectory models (e.g. Prijith et al., 2012; Rozwadowska et al., 2010; Ming et al., 2008; Meloni et al., 2007; Abdalmogith et al., 2006).

Trajectory model users have to choose a trajectory arrival elevation to be input to the model. This problem had to be addressed in this work. While the sensitivity of trajectory models to starting elevation has been discussed previously (Gebhart et al., 2005), it has never been quantified. As part of this work we analyzed and quantified this sensitivity. The sensitivity results are reported in Chapter 4.

Daily single backward trajectories were calculated for the sites described in Table 3.1 arriving at four different elevations: 500 m, 1000 m, 1500 m and 2000 m agl (above ground level) for the period 2000-2007. Only trajectories arriving at 1800 UTC were used for the present work. The arrival time of 1800 UTC is the most likely time for which daylight may be available for all sites throughout the year, given their geographic and latitudinal distribution (Fig. 3.1). This is relevant since the MODIS algorithm can only retrieve AOD during daylight. Furthermore, the upwind locations for Lagrangian AOD along the trajectories were selected under this criteria: 0, 3, 6, 18, 24 and 48 hours upwind of the study sites, all times for which daylight is generally available at the study sites.

3.4 Aerosol optical depth

The aerosol optical depth data were retrieved from the MODIS Level 1 and Atmosphere Archive and Distribution System (LAADS) of the NASA. 'Level 2 - Aerosol' products were

downloaded for both the Terra and Aqua satellites (<http://ladsweb.nascom.nasa.gov/data/search.html>) for the period 2003-2007. MODIS data retrieved before 2003 are scarce and incomplete. The MODIS ‘Optical_Depth_Land_And_Ocean’ product was used for this study. This product is filtered for higher quality assurance confidence (Levy et al., 2009). It contains aerosol optical depth determined using the 0.55 μm channel for both Ocean (best product) and Land (corrected product) with best quality data combined. ‘Level 2 - Aerosol’ processed data is obtained through several calculation algorithms using irradiance measured by the MODIS (described in Chapter 1); the resulting AOD pixels have a 10-km resolution. All of the MODIS data products are archived and distributed by the Goddard Distributed Active Archive Center (DAAC). Data products from Terra use the product ID MODxx, whereas those from Aqua use the product ID MYDxx (e.g. MYD04).

Daily average fields were produced to simplify the extraction of MODIS-AOD at the different upwind locations as determined by the trajectory model. Another purpose of the averaged AOD grid is to facilitate the reconstruction of missing data values for future work as certain reconstruction techniques require uniform spatiotemporal fields such as single value decomposition (Beckers and Rixen, 2003). The daily average AOD fields generated comprised the area from 10°N to 60°N and from 50°W to 140°W covering most of North and Central America (Fig. 3.1). The averaging process will be described in detail in Chapter 5.

3.5 Relative Humidity

As part of this work, surface relative humidity will be used as a predictor in the multivariate linear regressions models in an attempt to obtain further improvement. Relative humidity data were retrieved from the National Climatic Data Center’s Integrated Surface Database

(ISD) (cdoncdc.noaa.gov). ISD consists of global hourly and synoptic observations compiled from numerous sources. Relative humidity measurements are available one or more times per hour. While hourly relative humidity data were collected, only the relative humidity recorded closest to 1800 UTC was used to match the arrival time of the trajectories calculated.

4. Trajectory Model Sensitivity to Starting Elevation

Trajectory model users have to choose an arrival elevation to be input to the model. While the sensitivity of trajectory models to starting elevation has been discussed previously (Gebhart et al., 2005), it has never been quantified. We analyzed and quantified this sensitivity to make an informed choice of this important parameter.

Trajectory model calculations are sensitive to several factors, including the advection algorithm employed (Harris et al., 2005; Stohl and Seibert, 1998; Walmsley and Mailhot, 1983), spatial and temporal interpolation of meteorological data (Stohl et al., 1995; Kahl and Samson, 1986; Kuo et al., 1985), and the complexity of the meteorological scenario (Kahl, 1993, 1996; Merrill et al., 1985). The largest source of sensitivity in trajectory models is the choice of meteorological data base used in the calculations (Gebhart et al., 2005; Pickering et al., 1994; Kahl et al., 1989), accounting for 30-40% of the overall model sensitivity (Harris et al., 2005). Due to the extreme difficulty in providing ground truth for trajectory models (Riddle et al., 2006; Machta, 2002; Koffi et al., 1998), Harris et al. (2005) concluded that, “it would be difficult to prove that in all situations a particular data set was superior to another”. Gebhart et al. (2005) reached a similar conclusion.

In addition to choosing a meteorological data set and computational algorithms, trajectory model users must specify a starting elevation for the calculations. The starting elevation for backward trajectories would be the trajectory arrival altitude at the receptor, while for forward trajectories it would represent the starting altitude of the source. Specification of an appropriate starting elevation is sometimes arbitrary and may be problematic (Stohl, 1998).

The starting elevation for forward trajectories often represents the effective source height for the pollutant being studied, and during an accidental release this height may be unknown. For backward trajectories the problem is more pronounced. Backward trajectory models are often used to identify potential upwind source regions contributing to surface-based measurements of airborne concentrations. The starting elevation for these trajectory calculations is typically specified to be well above the surface, as surface winds do not necessarily represent the movement of contaminants (Samson, 1988). Differences between actual topography and the model's representation of topography create additional difficulties in specifying a starting elevation (Stohl, 1998)⁴.

How sensitive are trajectory model calculations to the starting elevation? It is well known that wind speed and direction vary with altitude, so trajectory models are expected to produce different results for different starting elevations. In cases where the starting elevation is ill-defined, knowledge of this sensitivity could be of great value in interpreting the model results. While a few previous studies have noted the importance of the starting elevation in trajectory model calculations, none have directly and systematically addressed the model sensitivity to this issue. Gebhart et al. (2005) studied the sensitivity of trajectory calculations to starting elevation and other factors including the choice of trajectory model, the meteorological data base used, and user input parameters and found that some trajectory models were more sensitive to the starting elevation than others. Baumann and Stohl (1997) found that isobaric trajectory model accuracy as verified by long-range balloon tracks was related to the sensitivity of the trajectory calculations to the starting elevation. Stunder (1996) found that forecast trajectory model accuracy was relatively insensitive to the starting elevation.

⁴ Compare the “Station Elevation (m amsl)” and “CMDL Elevation (m amsl)” columns in Table 3.2.

The objective of this Chapter is to quantify the sensitivity of trajectory calculations to the starting elevation. The results are intended to aid trajectory model users in their choice of starting elevation and in the interpretation of their model results. The results will help us choose the arrival elevations of the trajectories that will be used to find the location of Lagrangian MODIS-AOD.

4.1 Trajectory calculations

Ten sites across the continental United States were chosen for this study (Fig. 3.1, Tables 3.1 and 3.2). Three-dimensional backward trajectories were calculated for each site for the eight-year period of January 1, 2000, to December 31, 2007, at four different starting elevations, at 500 m, 1000 m, 1500 m, and 2000 m above ground level. We expect these starting elevations to be generally found within the atmospheric boundary layer (Yi et al., 2001). The duration of each trajectory is 48 hours. All trajectories arrive once daily at 1800 UTC. A total of 116,880 trajectories were calculated. The details of the meteorological data employed and the trajectory model employed are given in Section 3.2.

A trajectory model simulates atmospheric transport by estimating successive positions (trajectory endpoints) of a moving air parcel. To assess the sensitivity of trajectories to arrival altitude, we analyze the horizontal separation and directional difference between endpoints corresponding to trajectories with different starting elevations. These calculations are made for endpoints 3-hr, 12-hr, 24-hr, 36-hr and 48-hr upwind, for the 500, 1000, 1500 and 2000 m agl arrival altitudes.

Great circle trigonometry was used to calculate the distance between trajectory endpoints (Vincenty, 1975). The horizontal separation (HS) between trajectory endpoints can be calculated as:

$$HS_{ab}^t = r_e \arctan \left(\frac{\sqrt{(\cos \phi_a^t \sin \Delta \lambda_{ab}^t)^2 + (\cos \phi_a^t \sin \phi_b^t - \sin \phi_a^t \cos \phi_b^t \cos \Delta \lambda_{ab}^t)^2}}{\sin \phi_a^t \sin \phi_b^t + \cos \phi_a^t \cos \phi_b^t \cos \Delta \lambda_{ab}^t} \right), \quad (4.1)$$

where t is the upwind time (3, 12, 24, 36 or 48 hours), a and b represent different starting elevations, ϕ_a^t and ϕ_b^t are the latitudes of the trajectory endpoints being compared, $\Delta \lambda_{ab}^t = \lambda_b^t - \lambda_a^t$ is the difference of their longitudes, and $r_e = 6371$ km is the average radius of the Earth. With four arrival altitudes, there are six possible combinations of trajectory pairs (ab). Also using great circle trigonometry we can calculate the length of individual trajectories (TL) from the starting point (receptor) to the endpoint corresponding to a particular time t :

$$TL_t = r_e \sum_{i=0}^N \arctan \left(\frac{\sqrt{(\cos \phi_i \sin \Delta \lambda_{i, i+1})^2 + (\cos \phi_i \sin \phi_{i+1} - \sin \phi_i \cos \phi_{i+1} \cos \Delta \lambda_{i, i+1})^2}}{\sin \phi_i \sin \phi_{i+1} + \cos \phi_i \cos \phi_{i+1} \cos \Delta \lambda_{i, i+1}} \right), \quad (4.2)$$

where i identifies the trajectory endpoint ($i = 0$ indicates the starting point) and $N = t/3hr$ is the number of trajectory endpoints corresponding to upwind time t .

The procedure for determining the directional difference between trajectory endpoints is somewhat more involved. In order to calculate the arrival direction of the trajectory with starting elevation a , we calculate the approximate Δx_a and Δy_a components of the distance from the upwind location (S) to the receptor (R) given their longitudes (λ_a^S and λ_a^R) and latitudes (ϕ_a^S and ϕ_a^R):

$$\Delta x_a = \frac{2\pi r_e}{360^\circ} \cdot \Delta \lambda_a^{SR} \cdot \cos \left(\frac{\phi_a^S + \phi_a^R}{2} \right), \quad (4.3)$$

$$\Delta y_a = \frac{2\pi r_e}{360^\circ} \Delta\phi_a^{SR}, \quad (4.4)$$

$$\begin{aligned} \Delta\phi_a^{SR} &= (\phi_a^S - \phi_a^R) \\ \Delta\lambda_a^{SR} &= (\lambda_a^S - \lambda_a^R). \end{aligned} \quad (4.5)$$

Using the Δx_a and Δy_a components, we then calculate an acute positive reference angle α_a between the parallel (latitude circle) of the receptor and a straight line drawn between the upwind location and the receptor:

$$\alpha_a = \arctan \left| \frac{\Delta y_a}{\Delta x_a} \right|. \quad (4.6)$$

The reference angle α_a is then adjusted to conform to standard meteorological convention for wind direction, yielding the adjusted angle θ_a :

$$\theta_a = \begin{cases} 90^\circ - \alpha_a & \text{if } \Delta\phi_a^{SR} > 0 \text{ and } \Delta\lambda_a^{SR} > 0 \\ 90^\circ + \alpha_a & \text{if } \Delta\phi_a^{SR} < 0 \text{ and } \Delta\lambda_a^{SR} > 0 \\ 270^\circ + \alpha_a & \text{if } \Delta\phi_a^{SR} > 0 \text{ and } \Delta\lambda_a^{SR} < 0 \\ 270^\circ - \alpha_a & \text{if } \Delta\phi_a^{SR} < 0 \text{ and } \Delta\lambda_a^{SR} < 0 \end{cases}. \quad (4.7)$$

Equations (4.3) - (4.7) are repeated for upwind travel times of 3-hr, 12-hr, 24-hr, 36-hr and 48-hr and for all starting elevations. We then determine the directional difference $\Delta\theta$ between pairs of endpoints:

$$\Delta\theta_{ab}^t = \theta_b^t - \theta_a^t, \quad (4.8)$$

where a and b are different starting elevations ($b > a$). Finally, we correct the directional difference, if necessary, to account for cases in which the directional difference crosses the

0° line (e.g. a trajectory arriving at 1000 m agl from 45° (northeast) compared to a trajectory arriving at 500 m agl from 315° (northwest)):

$$DD_{ab}^t = \begin{cases} \Delta\theta_{ab}^t & \text{if } -180^\circ < \Delta\theta_{ab}^t < 180^\circ \\ \Delta\theta_{ab}^t - 360^\circ & \text{if } \Delta\theta_{ab}^t \geq 180^\circ \\ \Delta\theta_{ab}^t + 360^\circ & \text{if } \Delta\theta_{ab}^t \leq -180^\circ \end{cases} . \quad (4.9)$$

The calculation procedure [Eq. (4.3) - (4.9)] yields positive (negative) values of directional difference (DD) for a clockwise (counterclockwise) rotation from lower to higher starting elevations. HS and DD analysis were performed for the ten sites shown in Fig. 3.1 and described in Table 3.2.

4.2 Horizontal separation and directional difference

To better understand the behavior of the sensitivity measures we begin our analysis using a sample day (Fig. 4.1). The sample trajectories depict the pathways of air arriving at Milwaukee, Wisconsin at 1800 UTC on January 5th, 2000 at the four starting elevations chosen for the study: 500 m, 1000m, 1500 m, and 2000 m agl.

The 48-hr trajectories that arrive at the higher elevations show transport from south-central Canada. The lower altitude trajectories begin just north of Minnesota and North Dakota. The air parcels move southeast and turn northeast before arrival at Milwaukee. The trajectories, particularly those at the two highest starting elevations, exhibit an overall descending motion.

The sample trajectories shown in Fig. 4.1 are clearly sensitive to their starting elevations. The trajectory with a 2000 m starting elevation depicts a 48-hr endpoint in southern Alberta, 2290 km northwest of the receptor, traveling at an average velocity of

14.6 m/s. The 500 m trajectory, on the other hand, has a 48-hr endpoint near International Falls, 788 km north-northwest of Milwaukee, traveling at roughly half the speed (7.7 m/s).

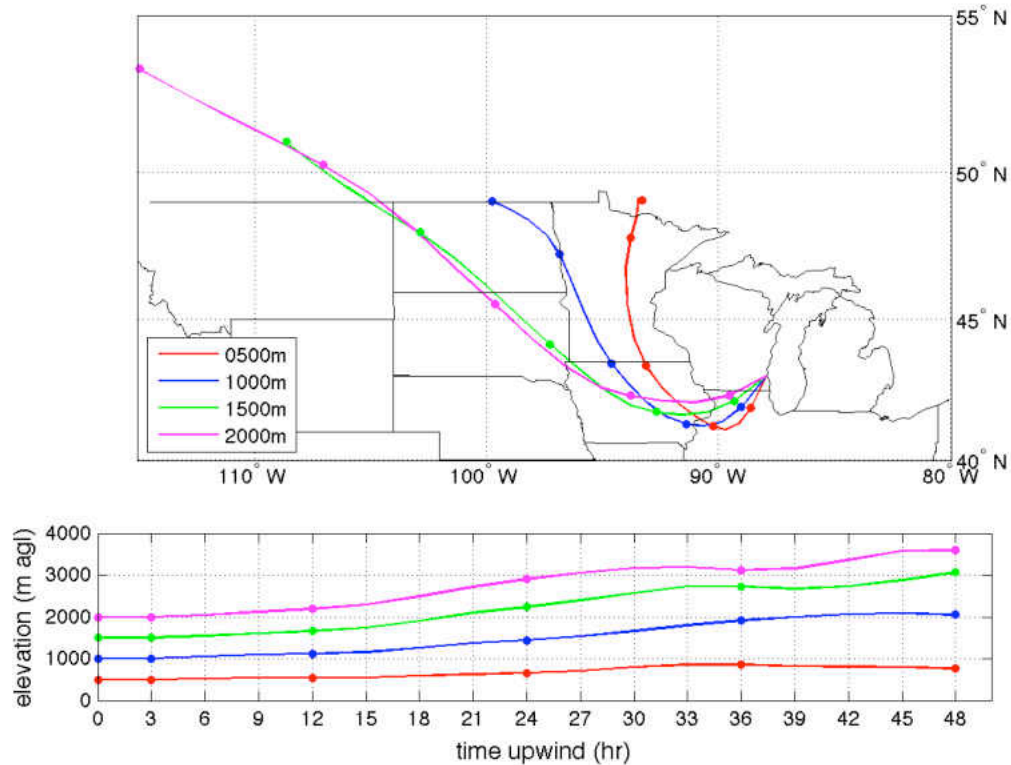


Figure 4.1 (top) Sample 48-hr back-trajectories for Milwaukee, Wisconsin arriving at 1800 UTC on January 5th, 2000. The filled circles along each trajectory are drawn at 3, 12, 24, 36 and 48-hr upwind. (bottom) Trajectory altitude as a function of travel time.

This sensitivity may be quantified further by examining the horizontal separation (HS) and directional difference (DD) of the sample trajectories. The HS of trajectories arriving at 1000 m, 1500 m and 2000 m agl is compared to the trajectory arriving at 500 m agl (Fig. 4.2). The largest separations occur for the 500 m - 2000 m agl comparison. At 24-hr upwind there is a horizontal separation of 570 km between endpoints and at 48-hr upwind this separation is 1577 km, a nearly three-fold increment. This latter separation (Fig. 4.1) is the distance between the 48-hr endpoints in British Columbia (2000 m agl trajectory) and southern Ontario (500m agl trajectory). Similar behavior is exhibited by the other

comparisons (i.e. 500 m - 1000 m agl and 500 m - 1500 m agl): as travel time increases from 24-hr to 48-hr the HS approximately triples.

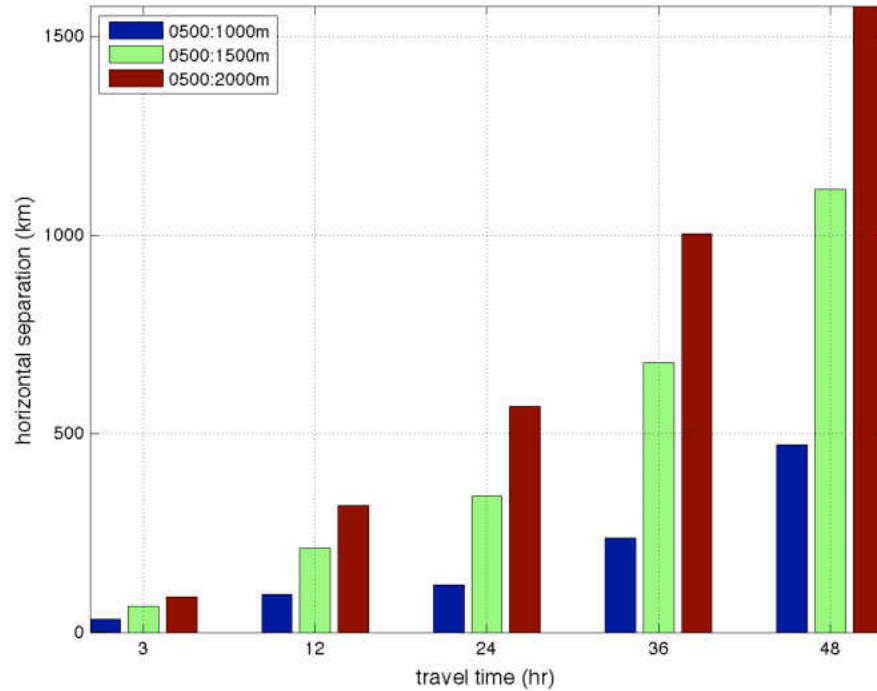


Figure 4.2 Horizontal separation analysis for sample back-trajectories arriving at Milwaukee, Wisconsin at 1800 UTC on January 5th, 2000.

The directional differences (DD) between the sample trajectories are shown in Fig. 4.3. The DD is largest for trajectories with the greatest difference in starting altitudes. The DD between trajectories with 500 m and 2000 m starting elevations is around 35 degrees. This is roughly consistent with the expected turning of the wind that occurs at increasing altitudes within the boundary layer. At further upwind segments, however, the DD decreases and becomes negative.

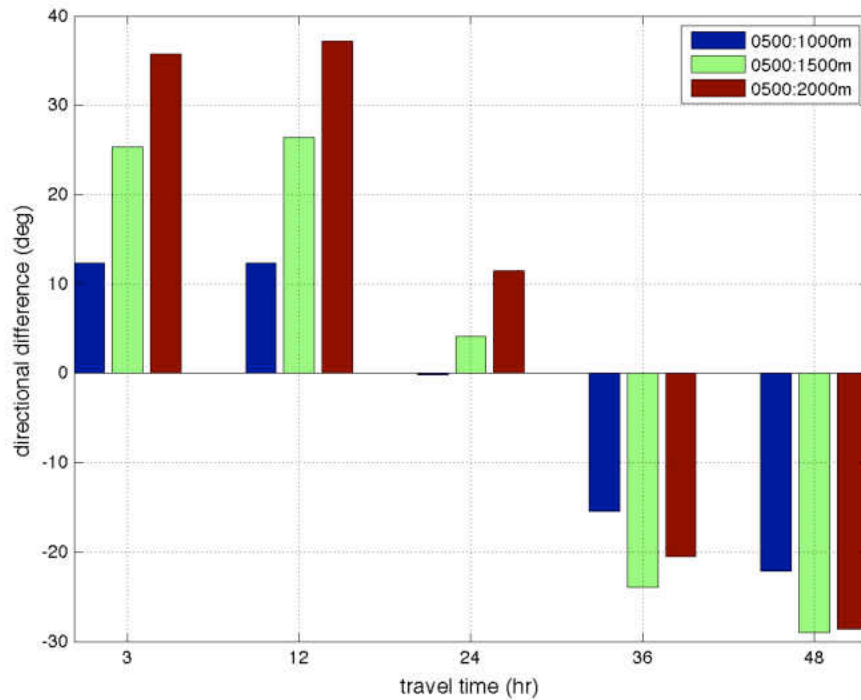


Figure 4.3 Directional difference analysis for sample back-trajectories arriving at Milwaukee, Wisconsin at 1800 UTC on January 5th, 2000.

When trajectories at different altitudes cross, as occurs in the sample trajectories between 3-hr and 24-hr upwind (Fig. 4.1), the DD changes sign. This is a result of curvature caused by trajectories sampling winds along different horizontal and vertical pathways. For curving trajectories, directional differences therefore represent sensitivities to both vertical and horizontal positions. The DD results for 3-hr upwind (Fig. 4.3) are unaffected by curvature, as the output resolution of the CMDL trajectory model is 3 hours. As the focus of this analysis is on trajectory model sensitivity to initial *vertical* position, we will restrict the DD analysis to the earliest (3-hr) trajectory segment.

An eight-year analysis is now presented for trajectories arriving at one particular site, Milwaukee, Wisconsin. Distributions of trajectory lengths as a function of upwind time (Fig. 4.4) show that trajectories arriving at 2000 m agl are 30% longer on the average than

trajectories arriving at 500 m agl. At 24-hr upwind, median trajectory length increased from 659 km to 865 km for 500 m agl and 2000 m agl starting elevations, respectively. At 48-hr upwind the median lengths increased from 1305 km to 1697 km.

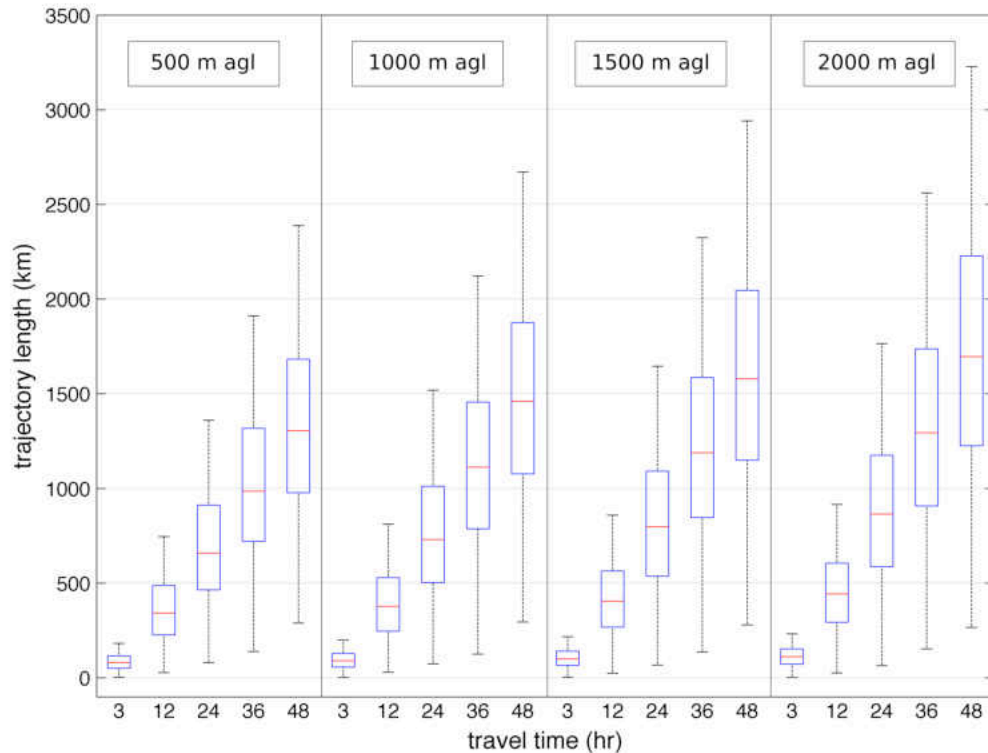


Figure 4.4 Length of trajectories arriving at Milwaukee, Wisconsin during January 1st, 2000 - December 31st, 2007. Results are shown for trajectories with starting elevations of 500 m, 1000 m, 1500 m, and 2000 m agl. The box-plots show the median (horizontal line within the box), the 25th and 75th percentiles (vertical extent of box) and the 5th and 95th percentiles (vertical extent of dashed lines).

The eight-year horizontal separation (HS) analysis for Milwaukee is shown in Fig. 4.5. As in the trajectory length case, HS increases with increasing travel time. For the 500 m - 2000 m agl starting elevation comparison, the median 24-hr HS is 422 km, nearly doubling at 48-hr upwind to 815 km. The median HS is smaller between the endpoints of the trajectories arriving at 500 m and 1000 m agl, however it also approximately doubles between 24-hr (165 km) and 48-hr (316 km) upwind. The 500 m - 1500 m agl starting elevation

comparison shows a similar near-doubling in median HS from 24-hr (301 km) to 48-hr (577 km). The HS analysis (Fig. 4.4) is more useful to trajectory model users than the trajectory length analysis (Fig. 4.5) since typically the interest is in estimating the locations of upwind sources or downwind receptors, rather than in estimating the length of the transport pathway from source to receptor.

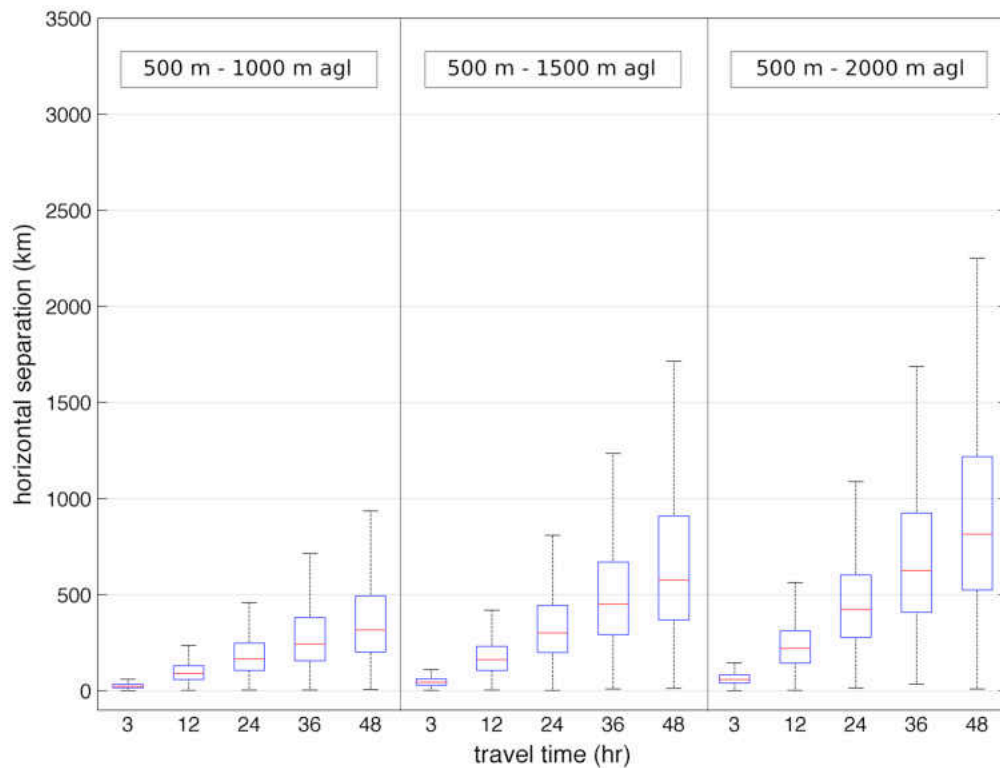


Figure 4.5 Horizontal separation analysis for trajectories arriving at Milwaukee, Wisconsin during January 1st, 2000 - December 31st, 2007. Results are shown for pairs of trajectories with starting elevations of 500 m, 1000 m, 1500 m, and 2000 m agl. The box-plots show the median (horizontal line within the box), the 25th and 75th percentiles (vertical extent of box) and the 5th and 95th percentiles (vertical extent of dashed lines).

We now examine the sensitivity of trajectory direction to starting elevation by analyzing the directional difference (DD) of 3-hr endpoints for eight years of trajectories arriving at Milwaukee (Fig. 4.6). The largest differences are found for the 500 m - 2000 m agl

comparison with a median DD of 15° and an interquartile range of 36° . Smaller differences occur for comparisons that do not include the near-surface (500 m) altitude.

The comparison of 1500 m - 2000 m agl reveals a 3-hr median DD of 0.7° , with an interquartile range of 11° . As noted earlier the positive median values of DD, especially for the 500 m - 2000 m comparison, is consistent with the expected turning of the wind at increasing elevations within the boundary layer (e.g. Arya, 2001).

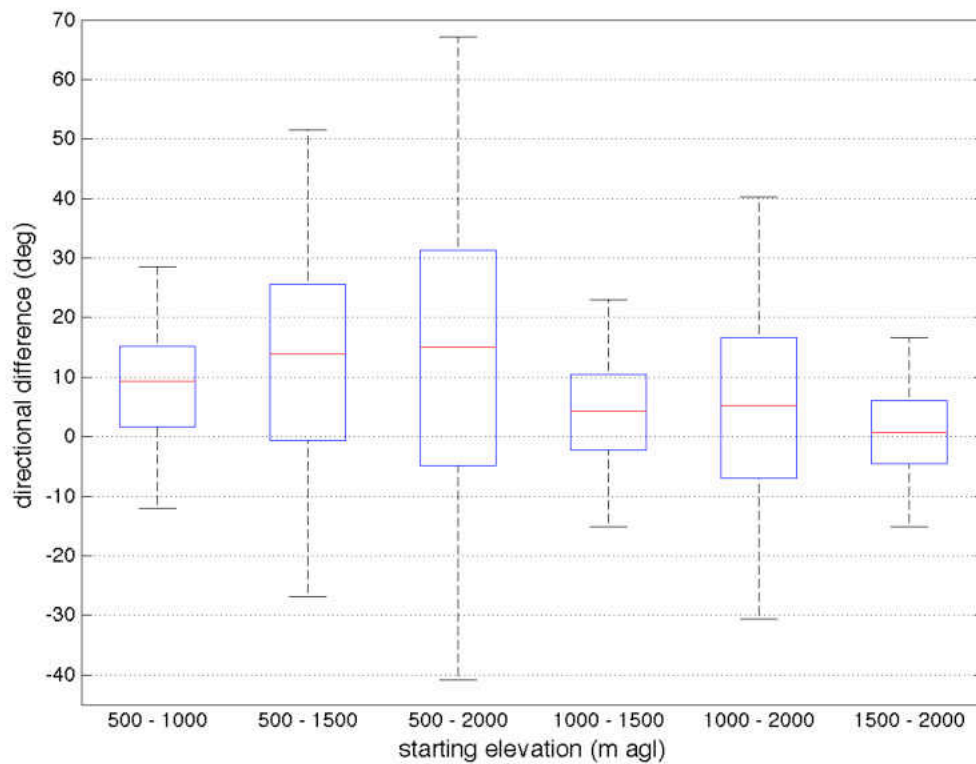


Figure 4.6 Directional difference analysis of 3-hr endpoints for trajectories arriving at Milwaukee, Wisconsin during January 1st, 2000 - December 31st, 2007. Results are shown for pairs of trajectories with starting elevations of 500 m, 1000 m, 1500 m, and 2000 m agl. The box-plots show the median (horizontal line within the box), the 25th and 75th percentiles (vertical extent of box) and the 5th and 95th percentiles (vertical extent of dashed lines).

The 3-hr DD distribution for the 500 m - 2000 m comparison is further explored in Fig. 4.7. A significant portion of the 3-hr DD distribution for Milwaukee is close to zero.

Similar behavior of DD distributions was observed for other sites (not shown). The clustering of the DD distribution about zero, together with the complications noted earlier due to trajectory curvature, suggest that the DD is not a particularly useful measure of trajectory model sensitivity to starting elevation.

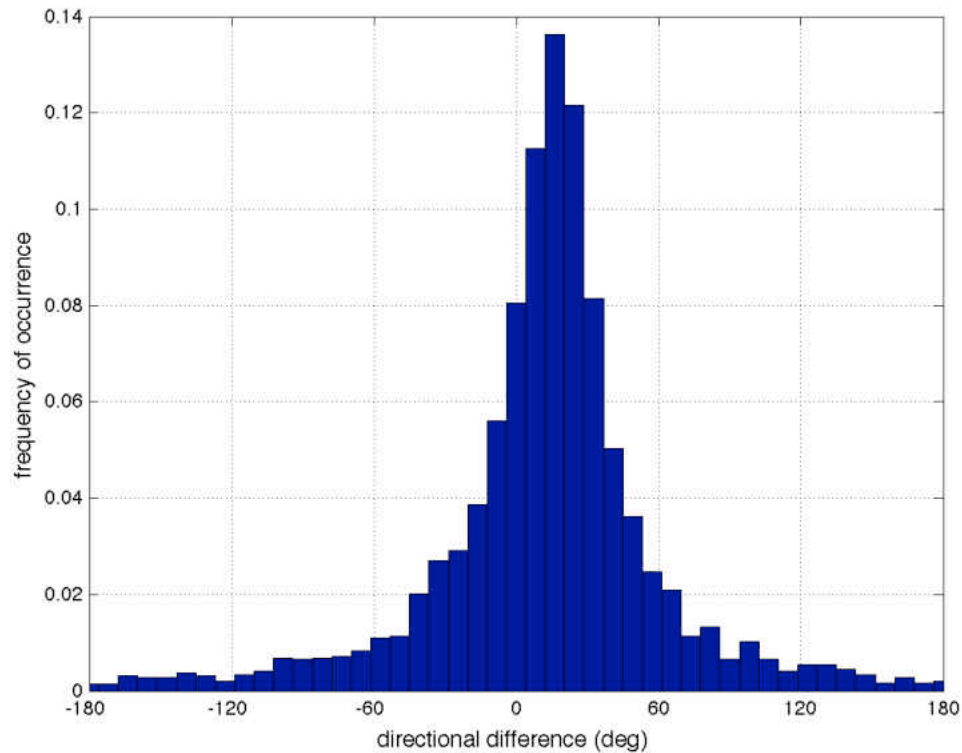


Figure 4.7 Directional difference analysis for trajectories arriving at Milwaukee, Wisconsin during January 1st, 2000 - December 31st, 2007. The histogram shows directional differences of 3-hr endpoints for trajectories calculated with 500 m and 2000 m agl starting elevations.

The HS results for all 10 sites, analogous to the results shown for Milwaukee in Fig's 4.4 - 4.6, are summarized in Table 4.1. Annual and seasonal analysis are included for individual sites and also for all sites combined, for the eight-year period of January 1st, 2000 through December 31st, 2007.

Table 4.1 Mean horizontal separation (km) of 48-hr trajectories with different starting elevations, January 1st, 2000 to December 12th, 2007. Standard deviation in parenthesis.

500 m agl - 1000 m agl						
Site	ID	Winter	Spring	Summer	Fall	Annual
Miami, Florida	MIA	403 (274)	304 (181)	129 (91)	228 (164)	265 (214)
Nogales, Arizona	OLS	507 (380)	328 (258)	175 (97)	301 (295)	327 (300)
North Little Rock, Arkansas	LIT	490 (319)	389 (324)	214 (150)	337 (233)	357 (284)
Nashville, Tennessee	BNA	530 (374)	391 (318)	229 (160)	376 (287)	381 (314)
Concord, California	CCR	388 (323)	310 (273)	216 (164)	266 (263)	295 (269)
Boulder, Colorado	WBU	356 (323)	245 (195)	136 (103)	250 (221)	246 (237)
Logan, Utah	LGU	452 (404)	253 (255)	121 (106)	288 (296)	277 (308)
Milwaukee, Wisconsin	MKE	450 (286)	413 (309)	304 (224)	395 (304)	390 (288)
Bismark, North Dakota	BIS	433 (292)	395 (269)	341 (218)	382 (283)	388 (269)
Seattle, Washington	SEA	469 (403)	304 (281)	176 (161)	376 (365)	330 (332)
All Sites	ALL	448 (344)	333 (276)	204 (169)	320 (282)	326 (288)

500 m agl - 1500 m agl						
Site	ID	Winter	Spring	Summer	Fall	Annual
Miami, Florida	MIA	695 (395)	553 (322)	237 (148)	414 (283)	474 (345)
Nogales, Arizona	OLS	944 (634)	659 (522)	338 (186)	573 (511)	627 (537)
North Little Rock, Arkansas	LIT	874 (480)	709 (506)	371 (233)	604 (388)	639 (454)
Nashville, Tennessee	BNA	921 (568)	687 (488)	384 (236)	657 (431)	661 (486)
Concord, California	CCR	747 (531)	636 (461)	479 (315)	568 (476)	607 (462)
Boulder, Colorado	WBU	716 (520)	550 (411)	319 (212)	551 (452)	533 (438)
Logan, Utah	LGU	970 (674)	592 (501)	291 (236)	668 (569)	628 (572)
Milwaukee, Wisconsin	MKE	807 (458)	727 (491)	530 (333)	683 (463)	686 (451)
Bismark, North Dakota	BIS	728 (437)	678 (403)	590 (314)	661 (411)	664 (396)
Seattle, Washington	SEA	993 (617)	726 (530)	459 (340)	831 (574)	751 (560)
All Sites	ALL	839 (549)	652 (472)	400 (283)	621 (474)	627 (481)

500 m agl - 2000 m agl						
Site	ID	Winter	Spring	Summer	Fall	Annual
Miami, Florida	MIA	921 (477)	766 (440)	335 (194)	552 (369)	642 (444)
Nogales, Arizona	OLS	1298 (822)	938 (702)	484 (264)	811 (674)	881 (712)
North Little Rock, Arkansas	LIT	1210 (598)	974 (607)	522 (323)	830 (529)	883 (582)
Nashville, Tennessee	BNA	1243 (658)	939 (605)	534 (317)	884 (558)	898 (604)
Concord, California	CCR	1072 (679)	930 (607)	718 (438)	852 (650)	892 (612)
Boulder, Colorado	WBU	1033 (650)	844 (573)	488 (319)	830 (630)	797 (592)
Logan, Utah	LGU	1374 (824)	913 (692)	475 (380)	1024 (776)	944 (760)
Milwaukee, Wisconsin	MKE	1097 (572)	964 (580)	719 (408)	924 (556)	925 (550)
Bismark, North Dakota	BIS	977 (525)	934 (497)	822 (403)	910 (511)	910 (489)
Seattle, Washington	SEA	1340 (699)	1039 (656)	743 (475)	1223 (687)	1082 (673)
All Sites	ALL	1156 (676)	924 (605)	584 (389)	883 (624)	886 (618)

Looking first at the annual results for all sites combined, we find that the mean 48-hr HS is strongly and consistently dependent on the difference in starting elevations. The smallest HS corresponds to a 500 m difference in starting elevation (the 500 m – 1000 m comparison) with an annual mean value of 326 km. When the difference in starting elevation doubles to 1000 m (the 500 m – 1500 m comparison), the all-site annual mean HS nearly doubles to 627 km. With a further increase in starting elevation difference (the 500 m – 2000 m comparison), the HS increases to 886 km.

Turning to the seasonal variation in HS, we find the smallest values in the summer months and the largest values during winter. For the all-site 500 m – 1000 m comparison, the mean HS is 204 km during summer and 448 km in winter. These values increase to 400 km (summer) and 839 km (winter) for the 500 m – 1500 m comparison, and to 584 km (summer) and 1156 km (winter) for the 500 m – 2000 m comparison.

The seasonal behavior in HS can be explained by seasonal variations in boundary layer thickness (Holzworth, 1964). As boundary layers are more shallow during the winter months, trajectories arriving at 2000 m are more likely to sample free tropospheric winds and thus be longer than those arriving at lower altitudes.

The HS results tend to be fairly consistent among the different sites for which trajectories were calculated. The inter-site variation in annual results ranges from 37% - 41% for the three starting elevation comparisons presented. Inter-site differences are likely due to geographical variations in boundary layer dynamics and synoptic-scale winds. No consistent relationship was observed between HS and receptor latitude, longitude or elevation.

The sensitivity of trajectory models to starting elevation is summarized in Fig. 4.8. Annual, all-site HS results are shown for both 24-hr and 48-hr upwind times. The results shown in Fig. 4.8 include the 1000 m – 1500 m, 1000 m – 2000 m, and 1500 m – 2000 m comparisons not shown in Table 4.1.

The dependence of HS on starting elevation is clearly evident and approximately linear (Fig. 4.8). Our results, based on an analysis of eight years of trajectories at 10 locations, indicate that a 500 m change in the starting elevation specified by the trajectory model user changes the 24-hr endpoint location by about 160 km. For 1000 m and 1500 m changes in starting elevation, this sensitivity in 24-hr endpoints increases to 300 km and 425 km, respectively. For 48-hr trajectory calculations, changing the starting elevation by 500 m, 1000 m and 1500 m changes the endpoint locations by 340 km, 630 km and 886 km, respectively.

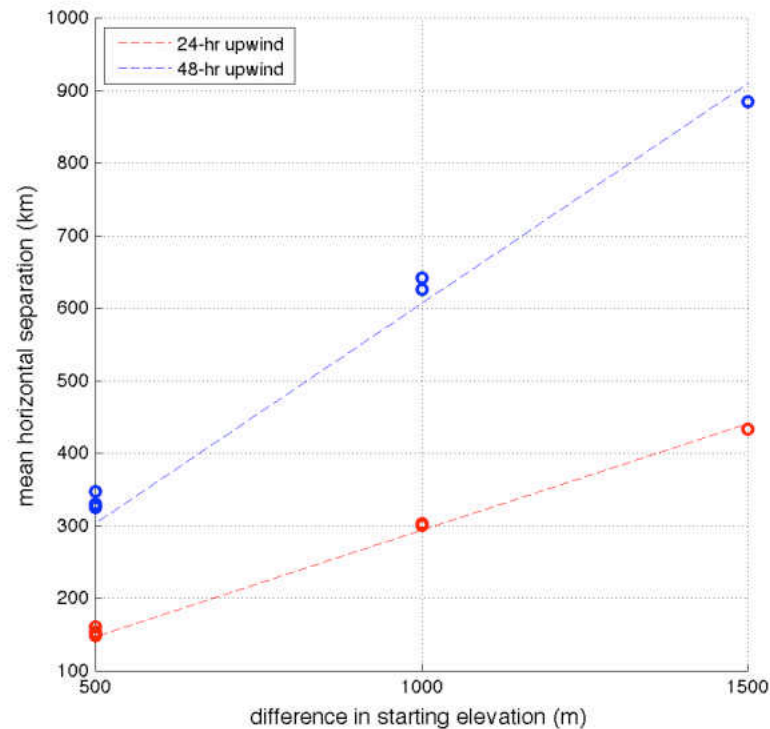


Figure 4.8 Mean horizontal separation as a function of the difference in starting elevation separation for all sites combined during January 1st, 2000 - December 31st, 2007.

The relationship between trajectory model sensitivity (HS) and difference in starting elevation (ΔSE) shown in Fig. 4.8 can be approximated by the regression equations:

$$HS_{24hr} = 0.294(\Delta SE) \quad (4.10)$$

and

$$HS_{48hr} = 0.607(\Delta SE), \quad (4.11)$$

where ΔSE has units of m, and HS_{24hr} and HS_{48hr} , the horizontal separation of 24-hr and 48-hr endpoints respectively, have units of km. Eqs. (4.10) and (4.11) were determined using the annual HS results for all sites combined. As discussed earlier, variations in model sensitivity (HS) exist among different sites and different seasons. Relationships analogous to (11) for 48-hr separations can be readily determined for specific sites and/or seasons by preparing figures similar to Fig. 4.8 using the relevant data from Table 4.1, and determining the empirical equations that describe them.

4.3 Conclusions

In this Chapter we explored the sensitivity of trajectory models to starting elevation. The analysis was based on daily 48-hr back-trajectories calculated for ten sites for the eight-year period of January 1st, 2000 - December 31st, 2007. Trajectories were calculated at four different starting elevations: 500 m, 1000 m, 1500 m and 2000 m agl. We developed two sensitivity indicators: horizontal endpoint separation (HS) and directional difference (DD). Seasonal and annual HS and DD analyses were performed for individual sites and all sites integrated.

Trajectory model calculations were found to be strongly sensitive to starting elevation. With a 500 m difference in starting elevation, the 48-hr all-site annual mean HS

was 326 km. When the difference in starting elevation was doubled to 1000 m, the all-site annual mean HS nearly doubled to 627 km. With a further increase in starting elevation difference (1500 m), the HS increased to 886 km. A seasonal dependence of this sensitivity to starting elevation was found, with the smallest mean HS occurring during the summer months, the largest during winter and intermediate values during the fall and spring.

A linear relationship was observed between trajectory model sensitivity (HS) and difference in starting elevation, and equations were presented to approximate this relationship. While our analysis was performed using backward trajectories, the results presented here are also directly applicable to forward trajectories.

Our sensitivity analysis was performed using a single trajectory model and a single meteorological data set. It is likely that the sensitivity results would be somewhat different if the analysis had been performed using different models and/or data sets. We expect, however, that the same general conclusions would hold: longer trajectories associated with higher elevations, increasing endpoint separations with increasing starting elevation differences, and a wintertime maximum in endpoint separations. The effect of model and data set choice on trajectory model sensitivity to starting elevation may be an area of interest for further research. Another suggestion for further research would be to assess the effect of starting elevation on the vertical component of trajectory model calculations; this aspect of model sensitivity was not addressed here.

Our results suggest that trajectory model users should be cautious when choosing the starting elevation for a trajectory calculation. Whenever possible it is advisable to test multiple starting elevations and/or to provide a strong argument justifying the specified starting elevation.

Al-Saadi et al. (2005) found that the actual relation between AOD and $PM_{2.5}$ varied due to changes in aerosol composition as inferred by the source location determined with

the air trajectory models. Since trajectory models are strongly sensitive to starting elevation, four arrival elevations (500 m, 1000 m, 1500 m and 2000 m agl) will be used in the present work as we would expect to find different sources of $PM_{2.5}$ with each arrival elevation for a given day and site.

5. AOD Spatial Correlation

In this work Lagrangian MODIS-AOD will be used as a predictor to estimate ground-level $PM_{2.5}$. Therefore, AOD has to be found for each selected upwind location, as determined by the calculated backward trajectories. There are 383,460 upwind locations that correspond to daily trajectories arriving at ten sites at four different arrival elevations during a five-year period. AOD needs to be retrieved for each of these upwind locations. This poses a problem since ‘Level 2 - Aerosol’ granules (described in Sections 1.2 and 3.4) are irregularly distributed in space and time.

To simplify the task of searching for the nearest individual granule in space and time, and then for the closest pixels for each upwind location, we decided to construct uniform daily averaged AOD fields covering the entire study area. The spatially uniform data fields will also be useful in future work derived from this study where we will seek to reconstruct incomplete MODIS-AOD fields. Certain techniques for the reconstruction of incomplete data sets require uniform spatiotemporal fields (Beckers et al., 2003).

A sample ‘Optical_Depth_Land_And_Ocean’ granule retrieved by Aqua on July 31st, 2007 at 1850 UTC, in both visible channel and the AOD product, is shown in Fig. 5.1 (a) and Fig. 5.1(b) respectively. The cloudy regions on the visible field match the areas of missing data on the AOD field. Also, note that the resolution of the AOD product (10 km) is coarse as compared to the sharper visible image (250 m resolution).

Even though the ‘Optical_Depth_Land_And_Ocean’ product granules are irregularly distributed in space, all of them have the same size and resolution: 135×203 pixels with a 10 km resolution. In Fig. 5.2 we show unprocessed MODIS-AOD from ‘Optical_Depth_Land_And_Ocean’ granules, retrieved by both Terra and Aqua satellites, for

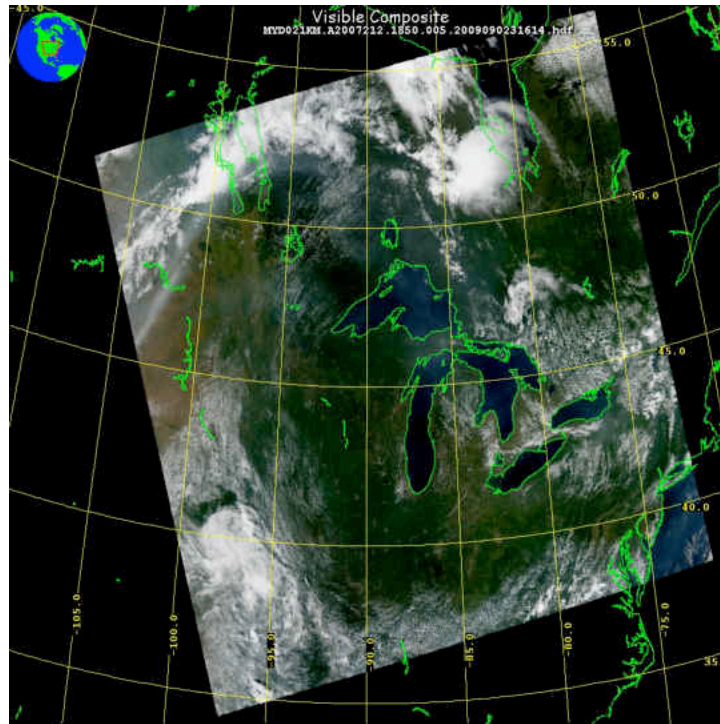


Figure 5.1(a) Sample MODIS granule retrieved by Aqua on July 31, 2007 at 1850 UTC in the visible band (images generated by MODIS LAADS Granule Browser).

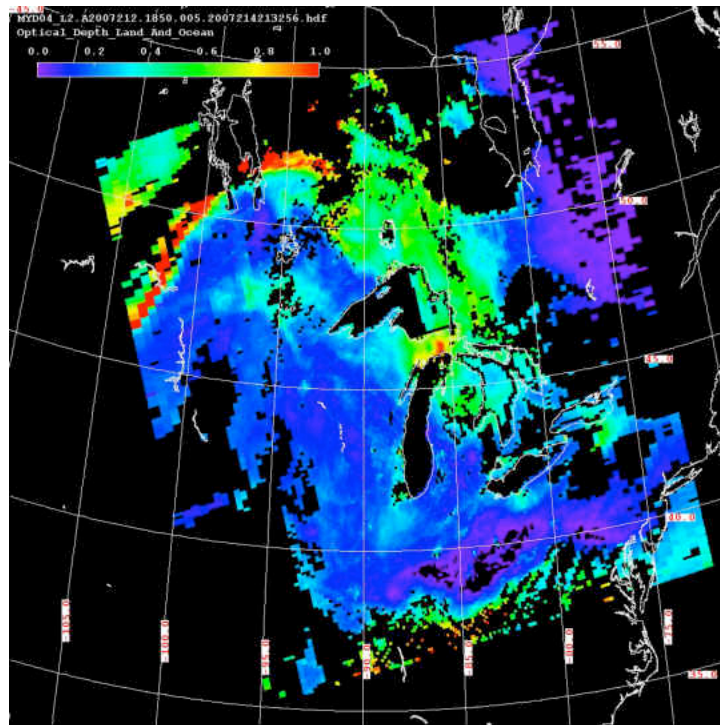


Figure 5.1(b) Sample MODIS granule retrieved by Aqua on July 31, 2007 at 1850 UTC in Optical Depth Land and Ocean (images generated by MODIS LAADS Granule Browser).

the same day shown in Fig. 5.1 (July 31st, 2007) for the area between 10°N to 60°N and 50°W to 140°W. This field contains 44 ‘Optical_Depth_Land_And_Ocean’ unprocessed granules.

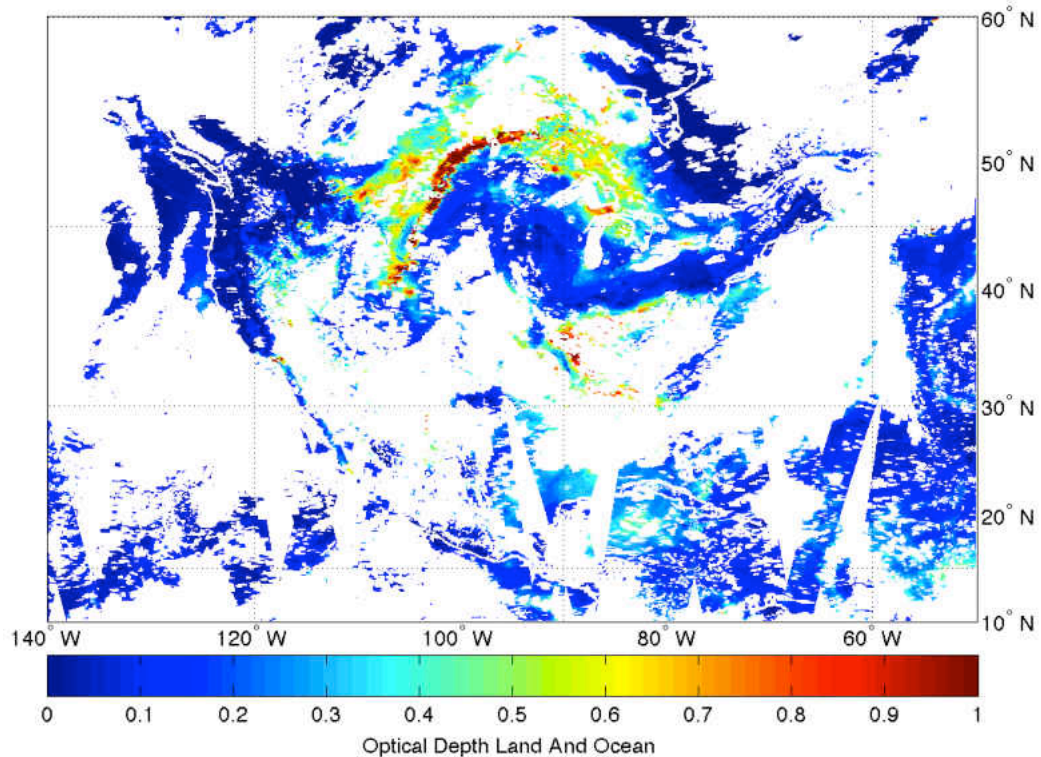


Figure 5.2 MODIS-AOD granules for July 31, 2007 retrieved by Terra and Aqua overlapped on the 10°N to 60°N and 50°W to 140°W region (unprocessed AOD).

The final averaged AOD field will have a $1^\circ \times 1^\circ$ latitude-longitude grid-point spacing which was arbitrarily selected. Even though it is arbitrary, this spatial resolution is sharper than the resolution of the meteorological data used to calculate the trajectories ($2.5^\circ \times 2.5^\circ$). Averaging the granule data over the study area shown in Fig. 5.2 will yield a final product of 51×91 pixels with a 1° latitude-longitude resolution.

The actual (real world) size of each $1^\circ \times 1^\circ$ grid box physically shrinks when moving from the equator toward the pole due to latitudinal convergence. At 10°N each $1^\circ \times 1^\circ$ grid cell is $12,176 \text{ km}^2$ in size whereas at the 60°N each $1^\circ \times 1^\circ$ grid cell is only 6182 km^2 .

The temporal resolution of the averaged field is set to one day. Though also arbitrary, this temporal resolution is a reasonable choice given the extended average atmospheric lifetime of $\text{PM}_{2.5}$. The value of each node in the field will be obtained by a distance weighted average of the unprocessed AOD pixels within a radius of influence. The 1° latitude-longitude averaged product resolution is the same as the resolution of the averaged 'Level 3 - Atmosphere' product generated by the NASA Goddard Space Flight Center. The main difference between our average product and the product generated by the NASA Goddard Space Flight Center is that our product considers a distance weighted averaged as opposed to unweighted averages of pixels within a grid cell (King et al., 2003; Hubanks et al., 2008).

5.1 AOD pixel correlation

The correlation of MODIS-AOD pixels with respect to distance has to be determined in order to choose the radius of influence that will be used in the distance weighted averaging. Each value of AOD in the unprocessed 135×203 pixels granule is accompanied by latitude and longitude information.

Using great circle trigonometry (Equation 4.1) the distance from each pixel with respect to the other 27,404 pixels in the granule was calculated using their correspondent latitudes and longitudes. This yields 375,503,310 pairs of AOD pixels for each granule. Not all of the combinations are available since many of the values are missing due to cloud cover. Once the distances between pixel combinations were calculated, the pairs were sorted

according to distance and separated into 10-km increment bins. The spatial autocorrelation between binned pixels was calculated as

$$r(d) = \frac{\sum_{k=1}^n \left[\left(AOD_k^a - \overline{AOD^a} \right) \left(AOD_k^b - \overline{AOD^b} \right) \right]}{\sqrt{\sum_{k=1}^n \left(AOD_k^a - \overline{AOD^a} \right)^2 \sum_{k=1}^n \left(AOD_k^b - \overline{AOD^b} \right)^2}}, \quad (5.1)$$

where AOD^a and AOD^b are paired pixels within a distance bin with means $\overline{AOD^a}$ and $\overline{AOD^b}$ respectively, n is the number of AOD pairs within a bin and d is the midpoint distance of said bin.

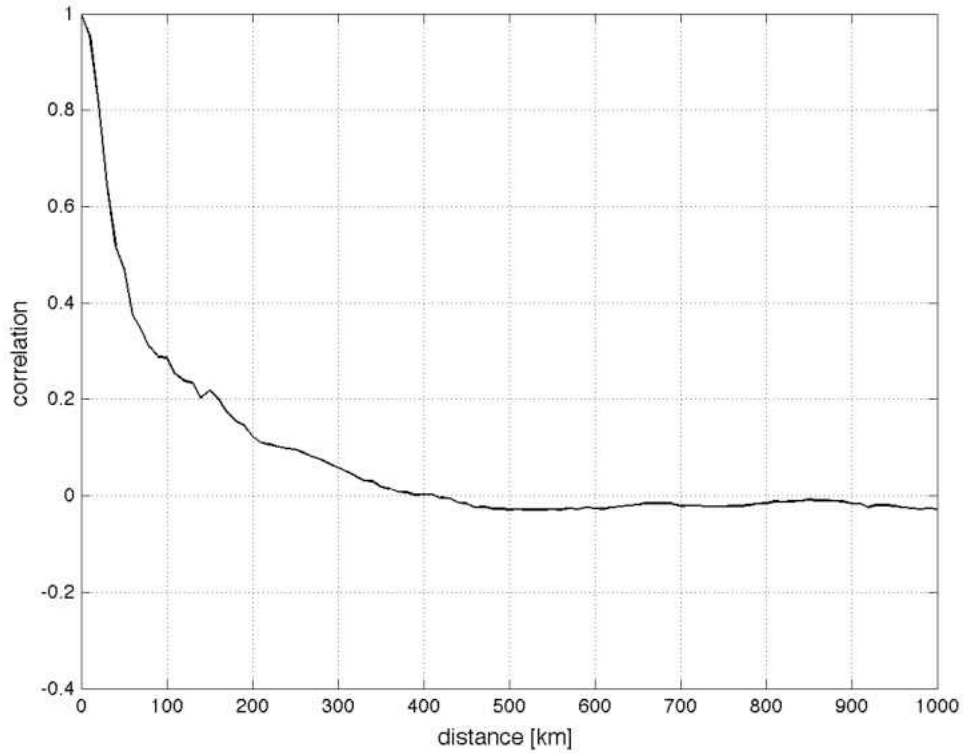


Figure 5.3 Correlation of MODIS-AOD pixels (10km distance bin) for a sample granule for the Aqua - MODIS granule for July 31st, 2007 at 1850 UTC in Figure 5.1(b).

In Fig. 5.3 we show the spatial autocorrelation of MODIS-AOD pixels in a 1000 km range for the sample granule in Fig. 5.1(b) (AOD retrieved by Aqua on July 31, 2007 at 1850 UTC). The AOD correlation decreases with distance as expected; for this particular example, the correlation drops below 0.5 beyond a 30 km separation between pixels. This decrement of correlation between AOD pixels with distance is caused by large and small-scale structures such as those occurring in Fig. 5.2. For example, notice the structure of high AOD in Fig. 5.2 that extends from Manitoba, Canada to North Dakota, USA. Taking any AOD point in this structure (e.g. 50°N/100°W) the pixels immediately around it are similar (there is high correlation) but as we move farther away from this point the AOD values become increasingly different.

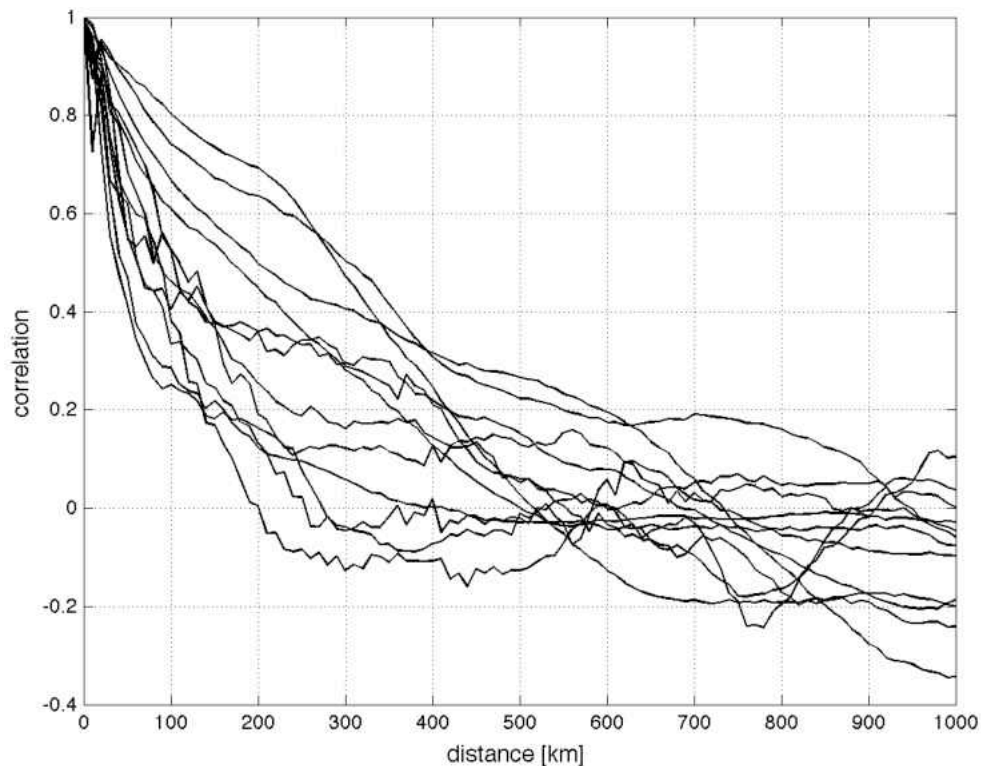


Figure 5.4 Correlation of MODIS-AOD pixels (10km distance bin) for twelve random granules.

The behavior of AOD pixel correlation with respect to distance varies from granule to granule. In Fig. 5.4 we show the correlation of MODIS-AOD pixels with respect to distance for twelve granules that were randomly selected from the available data (the spatial autocorrelation analysis shown in Fig. 5.3 is also included).

The correlation decays with increasing distance faster for some granules than for others, and in some cases the curves are smoother than others. This erratic behavior of some of the curves is likely due to the inhomogeneity of the data fields (presence of small-scale AOD structures) and to cloud cover which produces gaps between pixels.

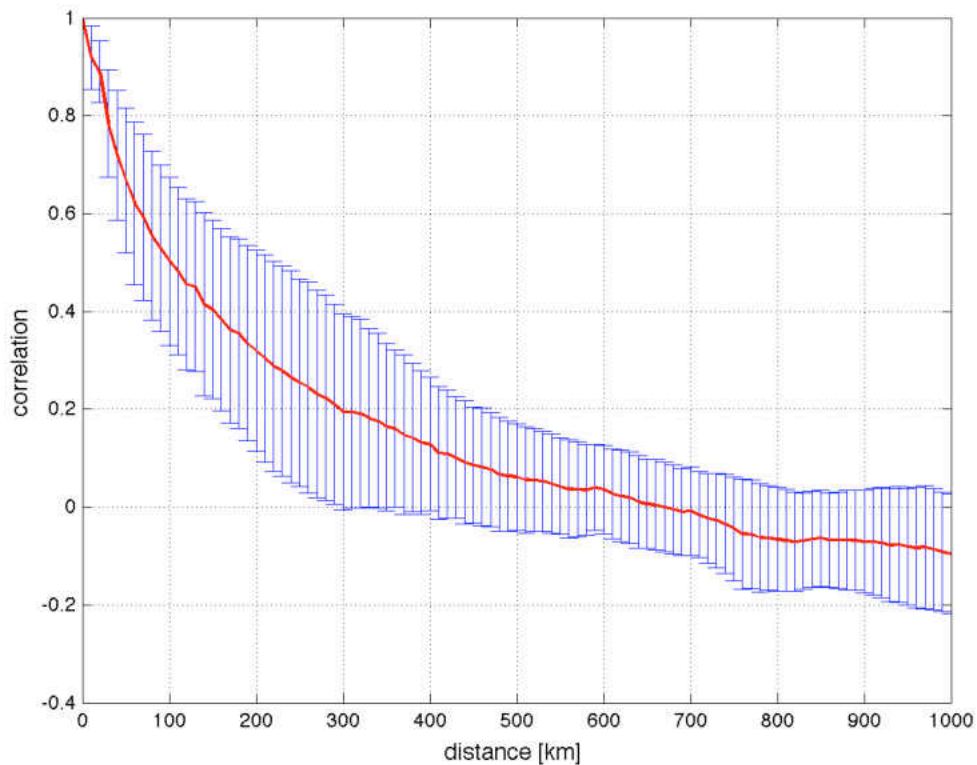


Figure 5.5 Mean correlation of MODIS-AOD with respect to distance (1 standard deviation bars) from 100 randomly selected granules.

To gain insight on the general AOD pixel correlation behavior with respect to distance, 100 randomly selected granules were processed using the same methodology

described above. In Fig. 5.5 we show the mean and standard deviation of 100 correlation analyses analogous to those shown in Fig. 5.4. The mean exhibited trend is that the correlation between AOD pixels drops below 0.5 beyond a 100 km separation. Thus we decided to set the influence radius for the distance weighted averaging procedure to 100km.

5.2 Distance weighted average

Pixels that are closer to a given grid node, as shown by the correlation analysis, are expected to be more representative of the actual AOD value of the node than pixels that are further away from it. With this in mind we decided to use a distance weighted average as opposed to an arithmetic mean. Distance weighted averages have been previously used in interpolation techniques of atmospheric data (Kahl and Samson, 1988). The general idea behind a distance weighted average is that data closest to center of averaging has a larger influence on the final value of the mean. The weighted average AOD for a pixel within the daily gridded AOD product is calculated as

$$\overline{AOD} = \frac{\sum_{k=1}^n W_k AOD_k}{\sum_{k=1}^n W_k}, \quad (5.2)$$

where AOD_k is the AOD pixel within the averaging radius, W_k is the weighting factor and n is the amount of pixels within the averaging radius. The weighting factor is defined as

$$W_k = \frac{1}{r^2}, \quad (5.3)$$

where r is the distance between the AOD_k pixel and the grid node within the averaging radius, calculated using Equation 4.1. This weighting factor (Equation 5.3) decreases rapidly

with increasing distance. Therefore the pixels closest to the node have a more significant contribution to the mean value than pixels further away. By doing this, the decreasing spatial autocorrelation of AOD pixels with distance is taken into account. Notice that if the weighting factors were $W_k = 1$, Equation 5.2 would be the arithmetic mean.

A sample daily average AOD field is shown in Fig. 5.6. This field corresponds to July 31, 2007, the data of the unprocessed field shown in Fig. 5.2. A total of 74738 MODIS-AOD granules were processed to generate 1826 daily average fields for the period 2003-2007. These spatially uniform AOD fields facilitate finding Lagrangian AOD using the the upwind locations calculated with backwards trajectories described in previous chapters.

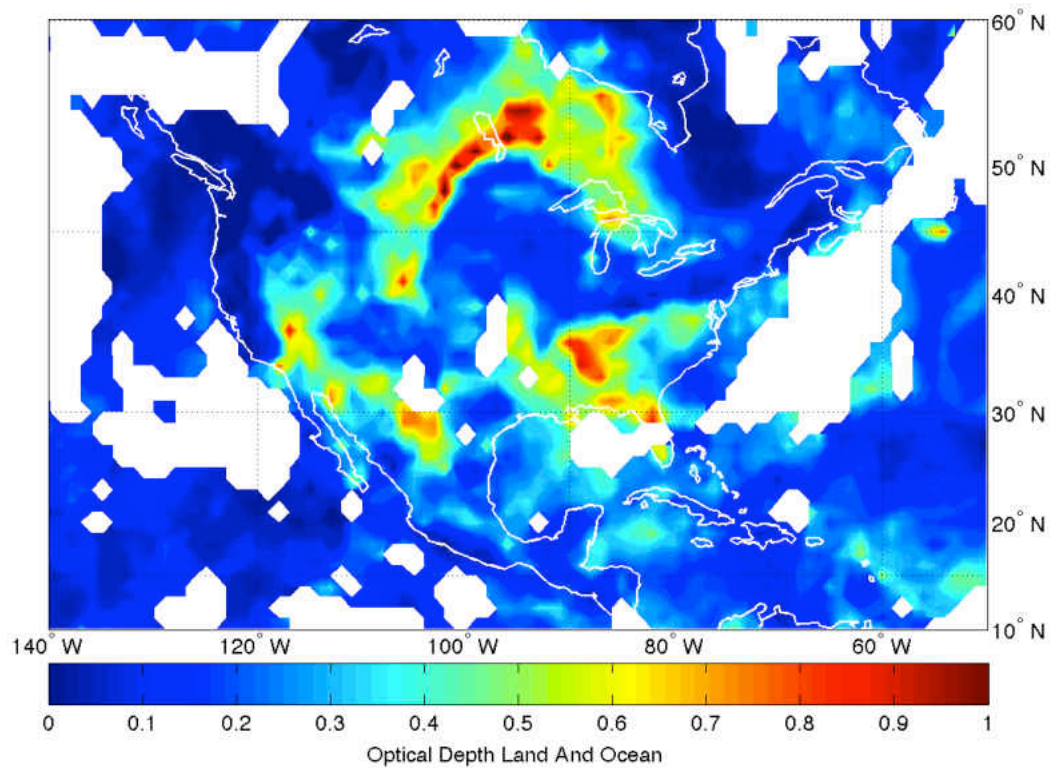


Figure 5.6 Sample AOD fields for July 31st, 2007 (distance weighted average).

It is important to note that values of AOD appeared where there were large gaps in the unprocessed field (e.g. in 55°N/98°W and 30°N/115°W). This is an artifact of the averaging procedure; if there are pixels within the influence radius, they are still taken into account in the averaging, even if there are few of them. This artificial spatial expansion of sparse Level 2 is also observed in 'Level 3 - Atmosphere' MODIS data. For example, a sizable region might have only a few scattered Level 2 retrievals, but Level 3 might show solid coverage of the averaged $1^\circ \times 1^\circ$ aerosol product. Even though this was not the main purpose of generating average grids, this is advantageous since previously missing data can be generated by distance weighted interpolation.

Also the high AOD structures from the unprocessed data (Fig. 5.1(b)) are retained after the distance weighted average has been applied to the unprocessed field (Fig's 5.2 and 5.6). This indicates that the averaging procedure was able to capture the original features of the AOD field.

The latitude and longitude information of the receptors and upwind locations obtained from the calculated air trajectories was used to estimate the distance to the nodes of daily weighted average AOD pixels using Equation 4.1. Once these distances were estimated, the AOD value for the trajectory endpoint is calculated using the distance weighted interpolation described by Equations 5.2 and 5.3 using an influence radius of 150 km. A data matrix was prepared that includes daily averaged ground-level PM_{2.5}, *in situ* distance weighted average AOD and distance weighted average AOD at 0, 3, 6, 18, 24, and 48 hours upwind. This data matrix, used in the multivariate linear regressions, is described in Chapter 6.

6. Estimation of PM_{2.5} Using Lagrangian MODIS-AOD

Aerosol optical depth (AOD) retrieved by the Moderate Resolution Imaging Spectroradiometer (MODIS) has been used as a predictor in empirical models to estimate ground-level concentrations of fine suspended particulate material (PM_{2.5}). These empirical models obtained by single linear regressions of observed PM_{2.5} and AOD have deficiencies. This is because while the aerosol concentration is measured at surface level, MODIS determines AOD for the whole integrated atmospheric column. Other researchers have improved the estimates of PM_{2.5} by including meteorological variables that affect the concentration and radiative properties of this pollutant (e.g. mixing layer height, relative humidity, temperature, wind speed and direction) in the empirical models (see Chapter 1). In this work we take a different approach to this problem: we seek to improve the estimates of ground-level PM_{2.5} by exploiting the fact that most PM_{2.5} consist of aged particles that have been advected to the site where they are being measured. We speculate that by using Lagrangian MODIS-AOD as predictors in multivariate linear regressions, the estimates of the concentrations of this pollutant can be improved.

The upwind locations of the Lagrangian variables were estimated using backward air trajectories arriving at four different elevations (see Chapter 4). We have assessed that a 500 m separation between trajectory arrival elevations can yield a horizontal separation of over 300 km between trajectory endpoints after 48 hours of transport. Since there is no “correct” arrival elevation for our problem, given that MODIS measures AOD for the integrated atmospheric column, using more than one starting elevation for the trajectories will help us account for different potential pollutant sources.

To obtain upwind values of AOD we created a daily averaged AOD product which was generated by distance-weighted averages of unprocessed MODIS-AOD data (see Chapter 5). In order to define a radius of influence for the distance-weighted average, we analyzed the spatial autocorrelation of the unprocessed AOD data. We assessed that on average the correlation of AOD pixels drops to 0.5 when separated by a distance of 100 km. Lagrangian AOD data were obtained through distance weighted interpolation of daily average MODIS-AOD using the latitude-longitude information of the upwind locations as determined by the backward trajectories and of the daily average AOD pixels.

The data set used in the linear regression models generated in this Chapter consisted of daily observations of ground-level $PM_{2.5}$ for the five-year period 2003-2007 at ten different sites. Each daily $PM_{2.5}$ observation is accompanied by *in situ* AOD and 5 upwind AOD values estimated using backward trajectories arriving at four different elevations: 500 m, 1000 m, 1500 m and 2000 m agl. However, AOD values (*in situ* and upwind) are frequently missing due to cloud cover. Only observations of $PM_{2.5}$ with a complete set of predictors are considered in the regressions.

It has been demonstrated that linear regressions of MODIS-AOD and ground-level $PM_{2.5}$ are sensitive to the temporal scale used in the regression (Li et al., 2009). In this work, regressions were performed for each site both over a one-year period and over seasons (i.e. five winters, five summers, etc.) We used different periods (years and seasons) because it was not clear which type of data aggregate would be more appropriate for testing our hypothesis that $PM_{2.5}$ can be improved by including upwind AOD in the regressions. In the present work we use the “meteorological” definition of seasons: winter (December - February); spring (March - May); summer (June - August) and fall (September - November).

In this chapter we will discuss the performance of the single ($PM_{2.5}$ and *in situ* AOD) and multivariate ($PM_{2.5}$ and *in situ* and upwind AOD) empirical models obtained through

linear regressions, the statistics used to test their validity, and the improvement of the estimates of ground-level $PM_{2.5}$ over the state-of-the-art.

6.1 Linear regressions of $PM_{2.5}$ and Lagrangian MODIS-AOD

We start our analysis by showing a sample regression of concentrations of ground-level $PM_{2.5}$ ⁵ and Lagrangian MODIS-AOD for one site. This sample corresponds to Milwaukee, Wisconsin with the location of Lagrangian AOD determined by trajectories arriving at 500 m agl for the combined winters of the period 2003-2007. The data set for this sample regression (Table 6.1) contains observed $PM_{2.5}$, AOD at 0, 3, 6, 18, 24, 48 hours upwind and estimates of $PM_{2.5}$ using the single and multivariate regression models. Even though a meteorological season consists of about ninety days (which totals close to 450 observations for a 5-year period), the sample size is greatly reduced ($N = 14$) due to cloud cover and missing $PM_{2.5}$ data.

The estimated $PM_{2.5}$ values in Table 6.1, Columns 9 and 10, were calculated with their respective regression coefficients (Table 6.2) according to the models:

single:

$$[PM_{2.5}]_{\text{predicted}}^{\text{single}} = \beta_1 \cdot AOD_0 + \beta_2, \quad (6.1)$$

multivariate:

$$[PM_{2.5}]_{\text{predicted}}^{\text{multivariate}} = \beta_1 \cdot AOD_0 + \beta_2 \cdot AOD_3 + \beta_3 \cdot AOD_6 + \beta_4 \cdot AOD_{18} + \beta_5 \cdot AOD_{24} + \beta_6 \cdot AOD_{48} + \beta_7. \quad (6.2)$$

⁵ Throughout the rest of the chapter we will refer to $PM_{2.5}$ measured at ground level simply as $PM_{2.5}$.

The AOD_i variables are the MODIS-AOD values at i hours upwind and β_j are the regression coefficients obtained using the least squares method for the data in Table 6.1 (Columns 2 through 8). The values of the coefficients obtained for Equation 6.1 are not the same as those for Equation 6.2 as seen in Table 6.2. The inclusion of additional explanatory variables will systematically alter the values of the previously existing regression coefficients β_j .

Table 6.1 Data set for Milwaukee, Wisconsin with the location of the Lagrangian AOD determined by the trajectories arriving at 500 m agl for the combined winters of 2003-2007. Included are predicted $PM_{2.5}$ using the single and multivariate regressions. AOD subscripts refer to the number of hours upwind.

n	Obs. $PM_{2.5}$ ($\mu g m^{-3}$)	AOD ₀	AOD ₃	AOD ₆	AOD ₁₈	AOD ₂₄	AOD ₄₈	Est. $PM_{2.5}$ ($\mu g m^{-3}$) Single	Est. $PM_{2.5}$ ($\mu g m^{-3}$) Multiple
1	23.40	0.039	0.291	0.299	0.032	-0.020	0.110	16.04	23.44
2	21.60	0.064	0.072	0.079	0.042	0.066	-0.034	16.56	18.71
3	25.30	0.359	0.324	0.225	-0.011	-0.022	-0.045	22.86	23.97
4	16.70	0.208	0.164	0.139	0.110	0.039	0.088	19.63	18.58
5	16.70	0.138	0.145	0.157	0.139	0.152	0.191	18.14	16.02
6	14.50	0.079	0.050	0.029	0.039	-0.011	0.012	16.88	15.48
7	15.90	0.083	0.143	0.185	0.083	0.275	0.188	16.98	17.03
8	15.30	0.097	0.048	0.075	0.048	0.069	0.325	17.27	14.25
9	12.00	0.081	0.110	0.110	0.020	0.019	0.110	16.93	17.79
10	16.60	0.084	0.102	0.092	0.078	0.086	0.190	16.98	12.48
11	11.60	0.062	0.059	0.035	0.034	0.086	0.058	16.51	12.10
12	21.30	0.226	0.200	0.178	0.074	0.046	0.028	20.01	22.19
13	11.00	0.040	0.044	0.037	0.128	0.171	-0.022	16.05	11.81
14	25.20	0.049	0.105	0.136	0.037	0.006	-0.012	16.24	23.24

Table 6.2 Regression coefficients for single and multivariate linear regressions for Milwaukee, Wisconsin with the location of the Lagrangian AOD determined by the trajectories arriving at 500 m agl for the combined winters of 2003-2007.

Model	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Single	21.33	15.20					
Multivariate	30.27	-101.80	131.30	-16.40	-18.85	-20.34	15.06

A single variable linear regression between $PM_{2.5}$ and MODIS-AOD yields a goodness of fit $R^2 = 0.1562$, whereas the goodness of fit for the multivariate model increases to $R^2 = 0.7644$ (Fig. 6.1). Clearly, in this case there is considerable improvement when including Lagrangian AOD in the regressions. However, we need to assess if this improvement over the single variable model is statistically significant.

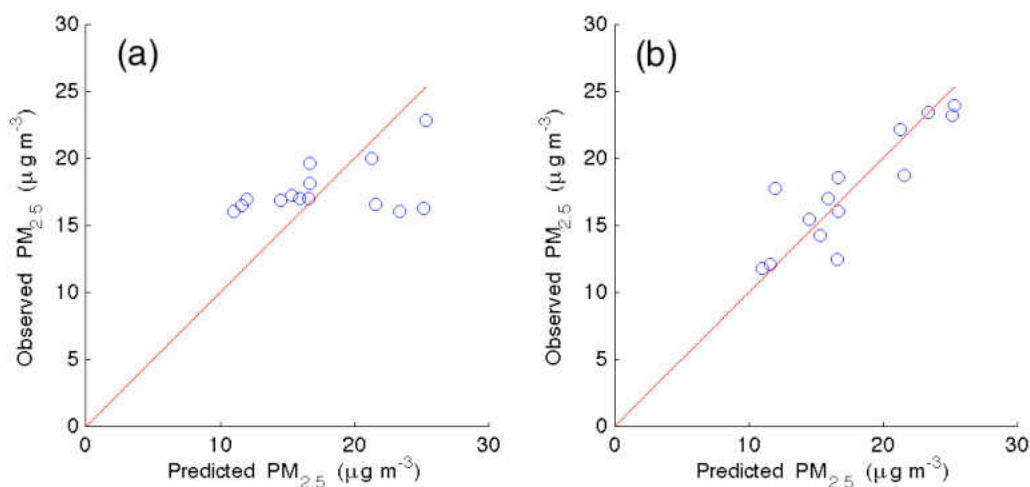


Figure 6.1 Scatter plots of observed vs. predicted $PM_{2.5}$ mass concentration for single (a) and multivariate (b) linear regression for Milwaukee, Wisconsin with the location of the Lagrangian AOD determined by the trajectories arriving at 500 m agl for the combined winters of 2003-2007. Red line indicates perfect fit between observed and predicted $PM_{2.5}$.

6.2 Statistical significance of the regression models

Just as the analysis of variance (ANOVA) can be used to test the null hypothesis that $\mu_1 = \mu_2 = \dots = \mu_p$ when looking at the means (μ) of several groups, it can also be used to assess the statistical significance of a regression. In the case of group means, the null hypothesis states that knowing the group membership provides no additional information. In the case of a regression, the corresponding null hypothesis would be that knowing the predictor X

provides no additional information about the response Y . If we were to guess the same Y value for every X , that would mean that the regression line had no slope ($\beta_1 = 0$ in Equation 6.1). Therefore, the null hypothesis for the ANOVA for a single predictor regression is $H_0: \beta_1 = 0$ and the alternate hypothesis is $H_A: \beta_1 \neq 0$.

The Extra Sum of Squares Principle allows us to compare two models for the same response where the full model (multivariate model) contains all the predictors in the reduced model (single variable model) and more. For example, the reduced model might contain m predictors while the full model contains p predictors, where p is greater than m and all the m predictors in the reduced model are among the p predictors of the full model, that is,

$$\text{reduced:} \quad Y = \beta_0 + \beta_1 \cdot X_1 + \cdots + \beta_m \cdot X_m + \varepsilon \quad (6.3)$$

$$\text{full:} \quad Y = \beta_0 + \beta_1 \cdot X_1 + \cdots + \beta_m \cdot X_m + \cdots + \beta_p \cdot X_p + \varepsilon. \quad (6.4)$$

The extra sum of squares principle allows us to determine whether there is statistically significant predictive capability in the set of additional variables. The specific null hypothesis test is

$$H_0 : \beta_{m+1} = \cdots = \beta_p = 0. \quad (6.5)$$

The method works by estimating the reduction in the residual sum of squares (or equivalently the increase in regression sum of squares) when the set of additional variables is added to the model. This change is divided by the difference in the number of degrees of freedom for the additional variables to produce a mean square. This mean square is compared to the residual mean square from the full model and the following F^* statistic can be generated:

$$F^* = \frac{\left(\frac{SSE(R) - SSE(F)}{df_R - df_F} \right)}{\left(\frac{SSE(F)}{df_F} \right)}, \quad (6.6)$$

where $SSE(R)$ and $SSE(F)$ are the residual sums of squares for the reduced and full models respectively and df_R and df_F are their respective degrees of freedom. Both $SSE(R)$ and $SSE(F)$ provide independent estimates of the sum of squares of the error ($SSE(R)$) associated with a regression model. If the addition of one or more variables does not substantially decrease the SSE , then $SSE(R)$ and $SSE(F)$ are essentially equal and the above ratio will be small, near zero. Alternatively, if the addition of one or more variables substantially decrease the SSE , then that ratio will be large (Kleinbaum et al., 1998). In the case of a single independent variable $SSE(R) = SST$, $SSE(F) = SSE$, $df_R = n-1$ and $df_F = n-2$, where SST is the total sum of squares and n is the number of observations.

The *p-value* is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. A high probability indicates the result is random, i.e., could be have been determined purely by chance. Thus, it is desirable to obtain *p-values* less than 0.05.

In the case of the sample regressions for Milwaukee, Wisconsin (Table 6.1) with the location of the Lagrangian AOD determined by the trajectories arriving at 500 m agl for the winters of 2003-2007, the extra sum of squares ANOVA yields a *p-value* of 0.0619. This means that even though the multivariate regression explains almost five times as much of the variance than the single model, the results are not statistically significant, i.e., the additional variables (Table 6.2) do not have significant predictive value at a 95% confidence level. The *p-value* is directly related to the variability within a given data set and indirectly related to the degrees of freedom, however. The fact that the sample multivariate regression failed the

statistical significance test may not be due to the explanatory capabilities of the additional independent variables but rather to the amount of degrees of freedom. This will be addressed in further detail later in this Chapter.

We now show results for single and multivariate regressions for each of the study sites for yearly (Table 6.3) and seasonal (Table 6.4) data aggregates. These regressions include Lagrangian variables with upwind locations determined by trajectories arriving at each of the four arrival elevations. Goodness of fit (R^2) for both single (Equation 6.1) and multivariate (Equation 6.2) linear regressions are shown in these tables, along with the test statistic F^* for the extra sum of squares (Equation 6.6), the F value at a 5% significance [$F^{\alpha}(p | dF_{numerator}, dF_{denominator})$] and the p -value estimated for an F-distribution considering the degrees of freedom for Equation 6.6. The null hypothesis (Equation 6.5) that the additional regression coefficients are equal to zero can be rejected at a 95% confidence level when $F^* > F$ or when $p\text{-value} < 0.05$.

To better understand the performance of the results presented in Tables 6.3 and 6.4 we show the distribution of R^2 for combined yearly (Table 6.3) and seasonal (Table 6.4) single and multivariate regressions using box plots (Fig. 6.2). We also include the distribution of R^2 for the regressions of a “control” data aggregate for each regression which has the same sample size as each multivariate data aggregate. These control runs have the same *in situ* AOD value as the single and multivariate regressions. However, the upwind values have been replaced by AOD values randomly selected from the entire population of daily-averaged AOD fields (as described in Chapter 5) for the period 2003-2007. These pseudo-random data aggregates will help us assess the predictive value of the trajectories in determining upwind AOD.

Table 6.3 Goodness of fit and statistical significance test for yearly regressions for the period 2003-2007. (Bold *p-values* indicate values less than 0.05.)

Sites	Elev. (m agl)	Year	R ²		F*	F ($\alpha = 0.05$)	p-value (< 0.05)	N
			Single	Multiple				
Nogales, Arizona (OLS)	500	2003	0.1341	0.2339	0.44	2.81	0.8124	24
		2004	0.1352	0.2637	0.80	2.64	0.5590	30
		2005	0.0261	0.1395	0.84	2.51	0.5292	39
		2006	0.1604	0.1808	0.18	2.47	0.9665	44
		2007	0.0279	0.1241	0.72	2.50	0.6100	40
	1000	2003	0.0826	0.4318	1.72	2.96	0.1946	21
		2004	0.1252	0.1991	0.46	2.60	0.8012	32
		2005	0.0271	0.0456	0.12	2.52	0.9868	38
		2006	0.1108	0.2470	1.19	2.50	0.3333	40
		2007	0.0337	0.1034	0.48	2.52	0.7868	38
	1500	2003	0.1005	0.5527	2.22	3.20	0.1251	18
		2004	0.1252	0.1583	0.20	2.60	0.9611	32
		2005	0.0062	0.0566	0.33	2.52	0.8905	38
		2006	0.1143	0.2786	1.32	2.55	0.2831	36
		2007	0.0416	0.2353	1.47	2.55	0.2302	36
	2000	2003	0.1005	0.3708	0.95	3.20	0.4898	18
		2004	0.1252	0.2198	0.61	2.60	0.6961	32
		2005	0.0025	0.1023	0.64	2.55	0.6677	36
		2006	0.1423	0.3705	2.10	2.55	0.0936	36
		2007	0.0480	0.2429	1.49	2.55	0.2227	36
Boulder, Colorado (WBU)	500	2003	0.1348	0.3176	2.68	2.40	0.0320	57
		2004	0.0029	0.0915	1.39	2.34	0.2403	78
		2005	0.0092	0.0545	0.68	2.34	0.6392	78
		2006	0.1846	0.2275	0.82	2.34	0.5383	81
		2007	0.1508	0.2407	1.80	2.33	0.1231	83
	1000	2003	0.0995	0.1992	1.22	2.40	0.3137	56
		2004	0.0029	0.1124	1.63	2.35	0.1644	73
		2005	0.0076	0.2450	4.65	2.34	0.0010	81
		2006	0.1775	0.2190	0.80	2.34	0.5542	82
		2007	0.1337	0.2294	1.84	2.34	0.1159	81
	1500	2003	0.1329	0.2709	1.93	2.40	0.1053	58
		2004	0.0070	0.0685	0.87	2.35	0.5056	73
		2005	0.0106	0.0570	0.69	2.35	0.6338	77
		2006	0.1982	0.2313	0.59	2.35	0.7043	76
		2007	0.0768	0.1746	1.75	2.34	0.1333	81
	2000	2003	0.1530	0.2058	0.69	2.39	0.6321	59
		2004	0.0013	0.0423	0.51	2.37	0.7646	67
		2005	0.0366	0.0602	0.34	2.35	0.8856	75
		2006	0.3017	0.3680	1.41	2.35	0.2336	74
		2007	0.0698	0.1970	2.31	2.34	0.0524	80

Table 6.3, continued.

Concord, California (CCR)	500	2003	0.0296	0.1236	2.66	2.29	0.0255	131
		2004	0.0092	0.0536	1.26	2.28	0.2868	141
		2005	0.0141	0.1026	2.88	2.28	0.0165	153
		2006	9.2E-06	0.0132	0.40	2.28	0.8515	155
		2007	0.0180	0.0587	1.49	2.27	0.1968	179
	1000	2003	0.0266	0.0411	0.36	2.29	0.8725	127
		2004	0.0112	0.0801	1.81	2.29	0.1155	128
		2005	0.0124	0.0562	1.27	2.28	0.2794	144
		2006	0.0038	0.0862	2.36	2.28	0.0433	138
		2007	0.0165	0.0365	0.65	2.27	0.6595	164
	1500	2003	0.0109	0.0337	0.46	2.31	0.8029	105
		2004	0.0117	0.0778	1.43	2.31	0.2193	107
		2005	0.0095	0.0613	1.28	2.29	0.2768	123
		2006	0.0088	0.0622	1.24	2.30	0.2952	116
		2007	0.0138	0.0356	0.58	2.29	0.7159	135
	2000	2003	0.0067	0.1019	1.70	2.33	0.1449	87
		2004	0.0029	0.0387	0.66	2.32	0.6579	95
		2005	0.0004	0.0310	0.64	2.30	0.6671	109
		2006	0.0269	0.0979	1.37	2.32	0.2432	94
		2007	0.0146	0.0413	0.59	2.30	0.7071	113
Miami, Florida (MIA)	500	2003	0.0348	0.0569	1.09	2.25	0.3661	240
		2004	0.0298	0.1142	5.67	2.24	0.0001	305
		2005	0.1058	0.1451	2.36	2.25	0.0404	264
		2006	0.1850	0.1886	0.27	2.24	0.9285	307
		2007	0.4212	0.4585	3.96	2.25	0.0017	294
	1000	2003	0.0344	0.0586	1.20	2.25	0.3101	240
		2004	0.0297	0.1243	6.48	2.24	9.7E-06	307
		2005	0.0932	0.1349	2.51	2.25	0.0303	268
		2006	0.1880	0.1935	0.42	2.24	0.8371	310
		2007	0.4223	0.4691	5.07	2.25	0.0002	295
	1500	2003	0.0375	0.0708	1.74	2.25	0.1259	250
		2004	0.0292	0.0868	3.76	2.24	0.0026	305
		2005	0.1020	0.1330	1.78	2.25	0.1165	257
		2006	0.1676	0.1788	0.82	2.24	0.5331	309
		2007	0.4300	0.4592	3.02	2.25	0.0114	287
	2000	2003	0.0333	0.0566	1.16	2.25	0.3275	243
		2004	0.0328	0.1019	4.52	2.24	0.0006	301
		2005	0.0736	0.0963	1.27	2.25	0.2767	260
		2006	0.1734	0.1844	0.84	2.24	0.5210	317
		2007	0.4302	0.4696	4.18	2.25	0.0011	288

Table 6.3, continued.

North Little Rock, Arkansas (LIT)	500	2003	0.2401	0.2862	2.13	2.27	0.0641	172
		2004	0.1055	0.1381	1.83	2.25	0.1071	249
		2005	0.2790	0.3717	7.14	2.25	3.0E-06	249
		2006	0.1394	0.2174	4.82	2.25	0.0003	249
		2007	0.0788	0.1078	1.58	2.25	0.1672	250
	1000	2003	0.2143	0.2601	2.13	2.27	0.0641	179
		2004	0.1158	0.1389	1.29	2.25	0.2704	247
		2005	0.2773	0.3722	7.29	2.25	2E-06	248
		2006	0.1442	0.2021	3.54	2.25	0.0041	251
		2007	0.0850	0.0965	0.63	2.25	0.6807	253
	1500	2003	0.2259	0.2603	1.65	2.27	0.1502	184
		2004	0.1190	0.1500	1.69	2.25	0.1367	239
		2005	0.2942	0.3834	7.00	2.25	4.0E-06	249
		2006	0.1423	0.1792	2.30	2.25	0.0458	263
		2007	0.1055	0.1289	1.36	2.25	0.2412	260
	2000	2003	0.2140	0.2456	1.44	2.27	0.2115	179
		2004	0.1389	0.1548	0.85	2.25	0.5149	232
		2005	0.3005	0.3666	4.95	2.25	0.0003	244
		2006	0.1409	0.1848	2.73	2.25	0.0203	260
		2007	0.0981	0.1280	1.73	2.25	0.1280	259
Milwaukee, Wisconsin (MKE)	500	2003	0.0133	0.2518	2.68	2.44	0.0345	49
		2004	0.2243	0.3190	1.47	2.39	0.2136	60
		2005	0.3377	0.4761	2.64	2.40	0.0340	57
		2006	0.0557	0.1814	1.72	2.38	0.1449	63
		2007	0.2344	0.3751	2.57	2.38	0.0365	64
	1000	2003	0.0260	0.2873	3.01	2.44	0.0211	48
		2004	0.2616	0.3891	2.05	2.40	0.0886	56
		2005	0.2773	0.3968	1.94	2.40	0.1045	56
		2006	0.0283	0.1617	2.00	2.36	0.0901	70
		2007	0.1736	0.2564	1.31	2.37	0.2708	66
	1500	2003	0.0202	0.3017	3.31	2.44	0.0134	48
		2004	0.2865	0.3441	1.02	2.37	0.4156	65
		2005	0.2053	0.4035	3.46	2.39	0.0090	59
		2006	0.0295	0.1102	1.11	2.37	0.3663	68
		2007	0.1656	0.3015	2.37	2.37	0.0493	68
	2000	2003	0.0254	0.2198	2.04	2.44	0.0925	48
		2004	0.2975	0.3485	0.84	2.39	0.5246	61
		2005	0.3105	0.4249	2.07	2.39	0.0842	59
		2006	0.0095	0.0965	1.12	2.37	0.3616	65
		2007	0.2197	0.3948	3.47	2.37	0.0080	67

Table 6.3, continued.

Bismark, North Dakota (BIS)	500	2003	0.2411	0.2760	0.39	2.45	0.8557	47
		2004	0.0476	0.2156	2.44	2.38	0.0450	64
		2005	0.1286	0.2240	1.28	2.39	0.2876	59
		2006	0.0510	0.1997	1.97	2.39	0.0983	60
		2007	0.0369	0.2017	2.35	2.38	0.0520	64
	1000	2003	0.2479	0.2683	0.21	2.46	0.9553	45
		2004	0.0587	0.1523	1.13	2.40	0.3583	58
		2005	0.0737	0.2492	2.43	2.39	0.0469	59
		2006	0.1062	0.4593	6.79	2.39	0.0001	59
		2007	0.0382	0.2500	2.88	2.40	0.0229	58
	1500	2003	0.2268	0.3001	0.77	2.47	0.5746	44
		2004	0.0919	0.1036	0.12	2.41	0.9868	54
		2005	0.0749	0.1503	0.92	2.39	0.4740	59
		2006	0.0530	0.2035	2.04	2.39	0.0875	61
		2007	0.0452	0.3362	4.65	2.39	0.0014	60
	2000	2003	0.2541	0.3251	0.67	2.51	0.6464	39
		2004	0.0631	0.0979	0.35	2.42	0.8815	52
		2005	0.0774	0.1328	0.69	2.39	0.6333	61
		2006	0.0439	0.1200	0.87	2.40	0.5109	57
		2007	0.0338	0.1485	1.35	2.40	0.2601	57
Seattle, Washington (SEA)	500	2003	0.0283	0.0621	1.18	2.27	0.3236	170
		2004	0.0281	0.0813	2.44	2.26	0.0354	218
		2005	0.0398	0.0820	2.06	2.25	0.0714	231
		2006	0.0301	0.1722	1.92	2.38	0.1051	63
		2007	0.0043	0.0508	1.91	2.26	0.0946	202
	1000	2003	0.0238	0.0638	1.34	2.27	0.2502	164
		2004	0.0355	0.1042	3.01	2.26	0.0123	203
		2005	0.0486	0.1324	4.15	2.26	0.0013	222
		2006	0.0279	0.1074	0.93	2.39	0.4715	59
		2007	0.0005	0.0793	3.16	2.26	0.0092	192
	1500	2003	0.0350	0.0924	1.62	2.29	0.1593	135
		2004	0.0374	0.0808	1.57	2.27	0.1699	174
		2005	0.0512	0.1015	2.07	2.26	0.0711	192
		2006	0.0001	0.1962	1.81	2.47	0.1357	44
		2007	0.0001	0.1560	5.84	2.27	0.0001	165
	2000	2003	0.0537	0.1166	1.41	2.31	0.2276	106
		2004	0.0462	0.1803	5.14	2.27	0.0002	164
		2005	0.0550	0.1105	1.80	2.28	0.1171	151
		2006	0.0001	0.1228	0.90	2.51	0.4959	39
		2007	0.0052	0.1161	3.31	2.28	0.0075	139

Table 6.3, continued.

Nashville, Tennessee (BNA)	500	2003	0.2409	0.3209	2.83	2.29	0.0190	127
		2004	0.2128	0.2207	0.44	2.26	0.8216	224
		2005	0.2813	0.3573	4.80	2.26	0.0004	210
		2006	0.1538	0.3749	17.05	2.25	2.0E-14	248
		2007	0.4442	0.4876	4.16	2.25	0.0012	252
	1000	2003	0.2278	0.3040	2.43	2.30	0.0394	118
		2004	0.1873	0.2106	1.29	2.26	0.2702	225
		2005	0.3179	0.3463	1.78	2.26	0.1185	212
		2006	0.3043	0.3522	3.41	2.25	0.0054	238
		2007	0.4546	0.4993	4.29	2.25	0.0009	247
	1500	2003	0.2571	0.3028	1.52	2.29	0.1889	123
		2004	0.2283	0.2358	0.41	2.26	0.8407	216
		2005	0.3224	0.3537	1.97	2.26	0.0853	210
		2006	0.3416	0.3581	1.20	2.25	0.3115	241
		2007	0.4560	0.5138	5.96	2.25	3.1E-05	258
	2000	2003	0.2820	0.3065	0.84	2.29	0.5243	126
		2004	0.2143	0.2260	0.62	2.26	0.6833	212
		2005	0.3458	0.3682	1.44	2.26	0.2106	210
		2006	0.3148	0.3578	3.14	2.25	0.0091	242
		2007	0.4508	0.5037	5.20	2.25	0.0001	251
Logan, Utah (LGU)	500	2003	0.0080	0.0435	1.30	2.27	0.2662	182
		2004	0.0546	0.0958	2.19	2.25	0.0564	247
		2005	0.0589	0.0769	0.66	2.27	0.6533	177
		2006	0.0257	0.0580	0.96	2.28	0.4444	147
		2007	0.0638	0.0781	0.69	2.25	0.6305	230
	1000	2003	0.0075	0.0361	1.00	2.27	0.4209	175
		2004	0.0551	0.1161	3.16	2.25	0.0089	236
		2005	0.0607	0.1021	1.56	2.27	0.1740	176
		2006	0.0319	0.1018	2.12	2.28	0.0668	143
		2007	0.0615	0.0631	0.08	2.26	0.9956	227
	1500	2003	0.0075	0.0422	1.17	2.27	0.3286	168
		2004	0.0537	0.0919	1.88	2.25	0.0986	231
		2005	0.0905	0.1603	2.65	2.27	0.0250	166
		2006	0.0306	0.1489	3.73	2.28	0.0034	141
		2007	0.0706	0.0743	0.17	2.26	0.9747	212
	2000	2003	0.0086	0.0614	1.73	2.27	0.1300	161
		2004	0.0473	0.0844	1.70	2.26	0.1350	217
		2005	0.0974	0.1893	3.40	2.27	0.0061	157
		2006	0.0361	0.1080	1.98	2.29	0.0855	130
		2007	0.0809	0.0873	0.26	2.26	0.9345	192

Table 6.4 Goodness of fit and statistical significance test for seasonal regressions for the period 2003-2007. (Bold *p-values* indicate values less than 0.05.)

Sites	Elev. (m agl)	Season	R ²		F*	F ($\alpha = 0.05$)	p-value (< 0.05)	N
			Single	Multi				
Nogales, Arizona (OLS)	500	Spring	0.0066	0.1766	1.32	2.51	0.2801	39
		Summer	0.0051	0.0554	0.41	2.46	0.8421	45
		Fall	0.0020	0.0892	0.82	2.43	0.5401	50
		Winter	0.1253	0.1456	0.17	2.48	0.9716	43
	1000	Spring	0.0042	0.3142	2.80	2.52	0.0335	38
		Summer	0.0098	0.0458	0.29	2.46	0.9175	45
		Fall	0.0027	0.0661	0.57	2.44	0.7217	49
		Winter	0.1045	0.2005	0.72	2.53	0.6133	37
	1500	Spring	0.0116	0.4308	4.27	2.55	0.0049	36
		Summer	0.0026	0.1277	1.06	2.47	0.3970	44
		Fall	0.0004	0.0383	0.30	2.46	0.9097	45
		Winter	0.1235	0.2520	0.96	2.56	0.4572	35
	2000	Spring	0.0136	0.2805	2.23	2.53	0.0776	37
		Summer	0.0019	0.1122	0.94	2.46	0.4638	45
		Fall	0.0024	0.0419	0.31	2.46	0.9016	45
		Winter	0.1344	0.1899	0.33	2.62	0.8906	31
Boulder, Colorado (WBU)	500	Spring	0.0677	0.0922	0.55	2.30	0.7417	108
		Summer	0.0506	0.1432	2.53	2.29	0.0327	124
		Fall	0.1101	0.1528	1.06	2.30	0.3872	112
		Winter	0.1547	0.2856	0.95	2.59	0.4643	33
	1000	Spring	0.0616	0.1156	1.25	2.30	0.2933	109
		Summer	0.0539	0.1611	2.99	2.29	0.0141	124
		Fall	0.1129	0.1482	0.81	2.31	0.5433	105
		Winter	0.1121	0.4288	3.10	2.56	0.0236	35
	1500	Spring	0.0754	0.1364	1.38	2.31	0.2368	105
		Summer	0.0517	0.1454	2.61	2.29	0.0281	126
		Fall	0.0983	0.1709	1.77	2.30	0.1256	108
		Winter	0.1920	0.3424	0.87	2.74	0.5199	26
	2000	Spring	0.1236	0.1631	0.89	2.31	0.4932	101
		Summer	0.0532	0.1207	1.83	2.29	0.1125	126
		Fall	0.1032	0.1792	1.80	2.31	0.1207	104
		Winter	0.0629	0.3572	1.56	2.81	0.2253	24
Concord, California (CCR)	500	Spring	0.0086	0.0869	2.45	2.28	0.0363	150
		Summer	0.0484	0.1298	1.33	2.34	0.2621	78
		Fall	0.0106	0.0304	1.01	2.25	0.4123	254
		Winter	0.0226	0.0385	0.89	2.25	0.4889	277

Table 6.4, continued.

	1000	Spring	0.0128	0.0877	2.10	2.29	0.0693	135
		Summer	0.0380	0.0783	0.63	2.34	0.6782	79
		Fall	0.0094	0.0225	0.63	2.25	0.6790	240
		Winter	0.0189	0.0460	1.36	2.25	0.2394	247
	1500	Spring	0.0104	0.1666	4.08	2.30	0.0019	116
		Summer	0.0286	0.0570	0.37	2.36	0.8646	69
		Fall	0.0116	0.0334	0.85	2.26	0.5150	196
		Winter	0.0194	0.0440	1.02	2.26	0.4072	205
	2000	Spring	0.0153	0.1557	2.83	2.32	0.0208	92
		Summer	0.0379	0.0503	0.14	2.38	0.9812	62
		Fall	0.0046	0.0448	1.34	2.27	0.2509	166
		Winter	0.0617	0.0758	0.52	2.27	0.7607	178
Miami, Florida (MIA)	500	Spring	0.3010	0.3641	7.84	2.24	4.7E-07	402
		Summer	0.1143	0.2605	14.08	2.24	1.4E-12	363
		Fall	0.0404	0.0693	1.92	2.24	0.0913	315
		Winter	0.0061	0.0416	2.39	2.24	0.0375	330
	1000	Spring	0.2953	0.3514	6.79	2.24	4.3E-06	400
		Summer	0.1166	0.2416	11.81	2.24	1.4E-10	365
		Fall	0.0425	0.0872	3.05	2.24	0.0104	319
		Winter	0.0050	0.0600	3.85	2.24	0.0021	336
	1500	Spring	0.2899	0.3287	4.53	2.24	0.0005	399
		Summer	0.1114	0.2304	11.01	2.24	7.2E-10	363
		Fall	0.0425	0.0752	2.22	2.24	0.0526	320
		Winter	0.0047	0.0525	3.22	2.24	0.0075	326
	2000	Spring	0.2927	0.3182	2.88	2.24	0.0144	393
		Summer	0.1184	0.2301	10.42	2.24	2.4E-09	366
		Fall	0.0298	0.0630	2.22	2.24	0.0526	320
		Winter	0.0077	0.0512	2.96	2.24	0.0125	330
North Little Rock, Arkansas (LIT)	500	Spring	0.1174	0.1431	1.86	2.24	0.1013	317
		Summer	0.1975	0.2802	8.45	2.24	1.4E-07	375
		Fall	0.2027	0.2503	3.77	2.24	0.0025	304
		Winter	0.0205	0.0597	1.39	2.27	0.2323	173
	1000	Spring	0.1168	0.1428	1.96	2.24	0.0844	331
		Summer	0.1919	0.2692	7.71	2.24	6.7E-07	371
		Fall	0.2089	0.2570	3.81	2.24	0.0023	301
		Winter	0.0231	0.0431	0.70	2.27	0.6220	175
	1500	Spring	0.1117	0.1426	2.35	2.24	0.0411	333
		Summer	0.2008	0.2705	7.05	2.24	2.6E-06	376
		Fall	0.2292	0.2664	3.16	2.24	0.0084	319
		Winter	0.0402	0.0682	0.96	2.27	0.4434	167

Table 6.4, continued.

	2000	Spring	0.1197	0.1650	3.49	2.24	0.0043	329
		Summer	0.2086	0.2712	6.36	2.24	1.1E-05	377
		Fall	0.2157	0.2459	2.45	2.24	0.0340	312
		Winter	0.0315	0.0741	1.37	2.27	0.2377	156
Milwaukee, Wisconsin (MKE)	500	Spring	0.2374	0.3037	1.18	2.36	0.3289	69
		Summer	0.1400	0.2512	3.44	2.29	0.0061	123
		Fall	0.2676	0.5202	8.42	2.33	2.0E-06	87
		Winter	0.1562	0.7644	3.61	3.97	0.0619	14
	1000	Spring	0.2147	0.2389	0.39	2.36	0.8513	69
		Summer	0.1409	0.2350	2.76	2.30	0.0219	119
		Fall	0.2454	0.4970	8.60	2.32	1.2E-06	93
		Winter	0.0792	0.2236	0.30	3.69	0.9011	15
	1500	Spring	0.2342	0.2547	0.36	2.36	0.8755	72
		Summer	0.1451	0.2233	2.30	2.29	0.0497	121
		Fall	0.2502	0.5090	9.49	2.32	2.7E-07	97
		Winter	0.0615	0.4931	1.87	3.20	0.1791	18
	2000	Spring	0.2118	0.2662	0.98	2.35	0.4374	73
		Summer	0.1487	0.2503	3.01	2.30	0.0139	118
		Fall	0.2443	0.4952	8.45	2.32	1.6E-06	92
		Winter	0.0299	0.7183	4.89	3.33	0.0160	17
Bismark, North Dakota (BIS)	500	Spring	0.0255	0.1291	1.79	2.34	0.1261	82
		Summer	0.0734	0.2478	5.38	2.29	0.0002	123
		Fall	0.2195	0.3275	2.31	2.34	0.0524	79
		Winter	0.3056	0.6507	0.59	9.01	0.7160	10
	1000	Spring	0.0130	0.1385	2.16	2.34	0.0682	81
		Summer	0.0818	0.2194	3.99	2.29	0.0023	120
		Fall	0.2036	0.2767	1.21	2.37	0.3136	67
		Winter	0.6007	0.6903	0.23	6.26	0.9300	11
	1500	Spring	0.0154	0.1163	1.64	2.34	0.1592	79
		Summer	0.0884	0.1827	2.63	2.29	0.0273	121
		Fall	0.2084	0.3057	1.77	2.36	0.1328	70
		Winter	0.1373	0.9448	2.92	230.16	0.4159	9
	2000	Spring	0.0227	0.0404	0.25	2.35	0.9361	76
		Summer	0.0797	0.1690	2.36	2.30	0.0444	117
		Fall	0.2043	0.3739	3.25	2.37	0.0116	67
		Winter	0.4992	1.0000	-	-	-	9
Seattle, Washington (SEA)	500	Spring	0.0017	0.0222	0.98	2.25	0.4318	240
		Summer	0.0133	0.0500	2.06	2.25	0.0705	274
		Fall	0.0024	0.0476	2.43	2.25	0.0357	263
		Winter	0.0298	0.0812	1.12	2.31	0.3548	107

Table 6.4, continued.

	1000	Spring	0.0014	0.0313	1.29	2.26	0.2678	217	
		Summer	0.0146	0.0575	2.41	2.25	0.0372	271	
		Fall	0.0038	0.0422	1.91	2.25	0.0937	245	
		Winter	0.0222	0.1804	3.86	2.31	0.0031	107	
	1500	Spring	0.0012	0.0214	0.72	2.27	0.6103	181	
		Summer	0.0142	0.0518	1.94	2.25	0.0891	251	
		Fall	0.0072	0.0545	1.88	2.26	0.1000	195	
		Winter	0.0024	0.2349	4.62	2.33	0.0010	83	
	2000	Spring	0.0002	0.0108	0.34	2.27	0.8880	166	
		Summer	0.0123	0.0404	1.21	2.26	0.3076	213	
		Fall	0.0092	0.0505	1.29	2.28	0.2718	155	
		Winter	0.0167	0.2570	3.75	2.37	0.0052	65	
Nashville, Tennessee (BNA)	500	Spring	0.0945	0.2278	8.97	2.25	7.0E-08	267	
		Summer	0.3228	0.3732	5.37	2.24	0.0001	341	
		Fall	0.3858	0.4151	2.98	2.24	0.0122	305	
		Winter	1.0E-06	0.0919	2.85	2.28	0.0174	148	
		1000	Spring	0.1906	0.2228	2.13	2.25	0.0625	264
			Summer	0.3225	0.3789	6.08	2.24	2.1E-05	342
			Fall	0.3618	0.3944	3.04	2.25	0.0109	289
			Winter	0.0001	0.0544	1.59	2.28	0.1677	145
		1500	Spring	0.2406	0.2427	0.15	2.25	0.9801	272
			Summer	0.3227	0.3756	5.69	2.24	4.7E-05	343
			Fall	0.3786	0.4079	2.82	2.25	0.0167	292
			Winter	0.0012	0.0092	0.22	2.28	0.9544	141
		2000	Spring	0.2537	0.2568	0.22	2.25	0.9548	268
			Summer	0.3116	0.3633	5.40	2.24	0.0001	340
			Fall	0.3770	0.4097	3.14	2.25	0.0089	291
			Winter	0.0045	0.0448	1.14	2.28	0.3432	142
Logan, Utah (LGU)	500	Spring	0.0049	0.0098	0.26	2.25	0.9362	269	
		Summer	0.3340	0.3641	2.99	2.24	0.0119	323	
		Fall	0.0092	0.0196	0.60	2.25	0.6994	291	
		Winter	0.0382	0.0637	0.51	2.31	0.7705	100	
		1000	Spring	0.0055	0.0172	0.60	2.25	0.6991	259
			Summer	0.3383	0.3650	2.65	2.24	0.0229	323
			Fall	0.0102	0.0352	1.38	2.25	0.2329	273
			Winter	0.0170	0.0582	0.83	2.31	0.5317	102
		1500	Spring	0.0047	0.0299	1.28	2.25	0.2736	253
			Summer	0.3217	0.3479	2.52	2.24	0.0293	321
			Fall	0.0152	0.0504	1.86	2.25	0.1027	257
			Winter	0.0489	0.1204	1.30	2.33	0.2717	87
		2000	Spring	0.0039	0.0462	1.93	2.26	0.0899	225
			Summer	0.3078	0.3412	3.10	2.24	0.0096	313
			Fall	0.0219	0.0490	1.36	2.25	0.2420	245
			Winter	0.0099	0.0821	1.05	2.35	0.3936	74

On each box, the central mark is the median of the R^2 distribution, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. The whisker length corresponds to approximately $\pm 2.7\sigma$ and 99.3% coverage if the data were normally distributed.

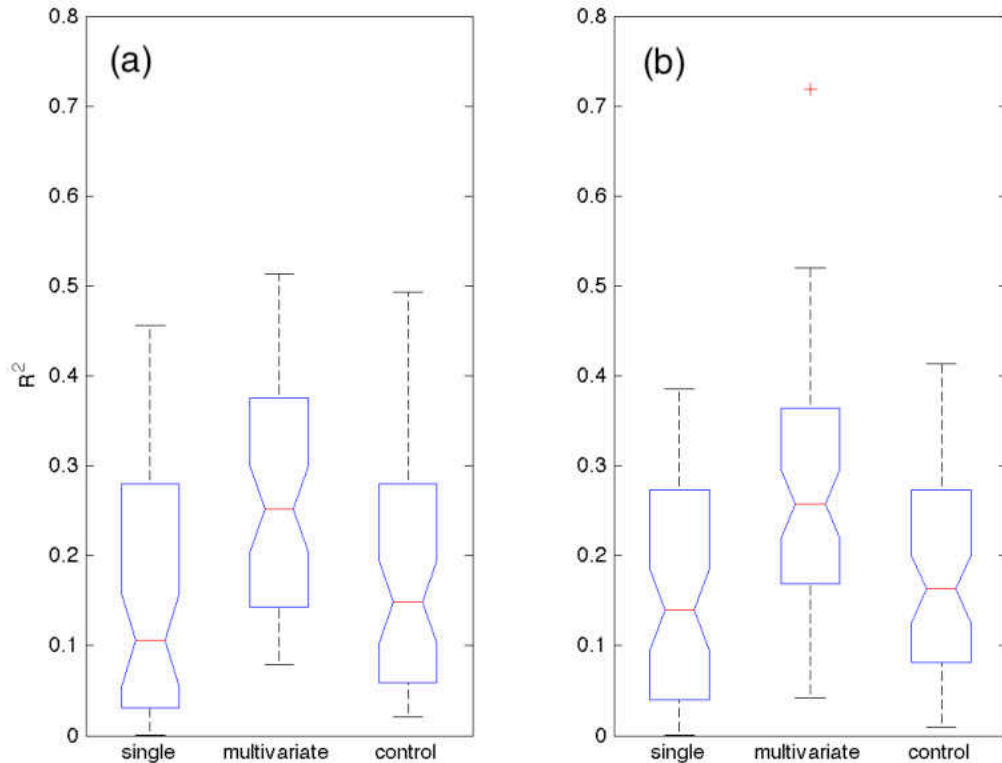


Figure 6.2 Comparison of the distribution of R^2 for yearly (a) and seasonal (b) data aggregates for each site for the period 2003-2007. Single, multivariate and control regressions are shown. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.

Two medians are significantly different at the 5% significance level if their intervals do not overlap. Interval endpoints are the extremes of the notches on the box plots (Fig. 6.2). The extremes correspond to

$$q_2 \pm 1.57(q_3 - q_1) / \sqrt{n}, \quad (6.7)$$

where q_2 is the median (50th percentile), q_1 and q_3 are the 25th and 75th percentiles, respectively, and n is the number of observations (McGill et al., 1978).

In Fig. 6.2 it can be observed that the multivariate regressions exhibit improvement over the single variable regressions when considering the distribution of R^2 solely, as their respective medians are different at a 95% confidence level. Thus, the multivariate model performs better than the single regression model in both seasonal and yearly linear regressions. Nevertheless, of the 200 yearly multivariate regressions (Fig. 6.2a) only 29% exhibit statistically significant improvement over the single variable regressions when applying the extra sum of squares principle. Furthermore, of the 160 seasonal regressions (Fig. 6.2b), 41% exhibit improvement that is statistically significant. This suggests a seasonal variability in the performance of the multivariate regression models.

It can also be observed in Fig. 6.2 that the control regressions exhibit a better performance than the single variable regressions. This is due to the fact that in the control runs we replaced the upwind predictors with randomly selected values of AOD, however, the $PM_{2.5}$ and *in situ* AOD pairs are the same as in the single and multivariate regressions. Adding additional predictors to a regression equation improves the performance somewhat, even if those additional predictors are uncorrelated to the response (Kleinbaum et al., 1998). Nevertheless, the multivariate model (in both the yearly and seasonal data aggregates) has a better performance than the single and the control models. The median R^2 for the multivariate model (Fig. 6.2a and Fig. 6.2b) is significantly higher at a 95% confidence level than the single and control median R^2 . This indicates that Lagrangian AOD truly has predictive capabilities and that the increment in R^2 is not merely an artifact of an increased amount of explanatory variables.

Having demonstrated that the overall performance of the multivariate regressions is better than that exhibited by the single regressions we now seek to further understand the

performance of yearly and seasonal regressions. We begin by grouping the yearly linear regressions for each site and trajectory arrival elevation into years (Fig. 6.3).

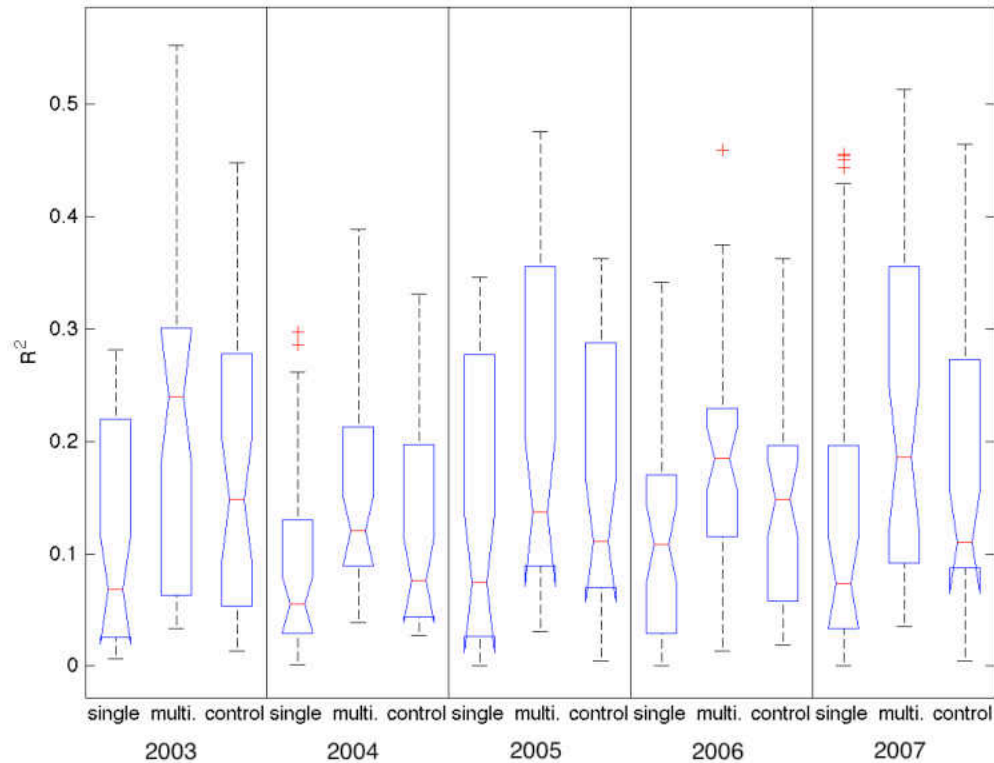


Figure 6.3 Goodness of fit distribution of yearly linear regressions for individual sites for the period 2003-2007. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.

It can be observed that the median R^2 of the multivariate regressions is significantly different from their respective single variable regressions at a 95% confidence level for all years except 2005. The multivariate regressions exhibit the highest median values of R^2 on are 2003 and 2007. The median R^2 for the control models is lower when compared to the multivariate models, however, none of the medians are different at a 95% confidence level when comparing the control and the multivariate regressions. A significant difference cannot be discerned between regressions in different years.

On the other hand, when the single, multivariate and control linear regressions are grouped into seasons (Fig. 6.4) a significant difference can be observed. The median R^2 for the summer multivariate regressions is higher than the other seasons and is also the only one different from the control at a 95% confidence level. These results are consistent with the findings of Wang et al. (2003) and Schaap et al. (2008), who found a strong seasonal signature in the estimation of $PM_{2.5}$ inferred by MODIS-AOD. Furthermore, Wang et al. (2003) found the largest correlations between monthly mean $PM_{2.5}$ and MODIS-AOD during the summer months.

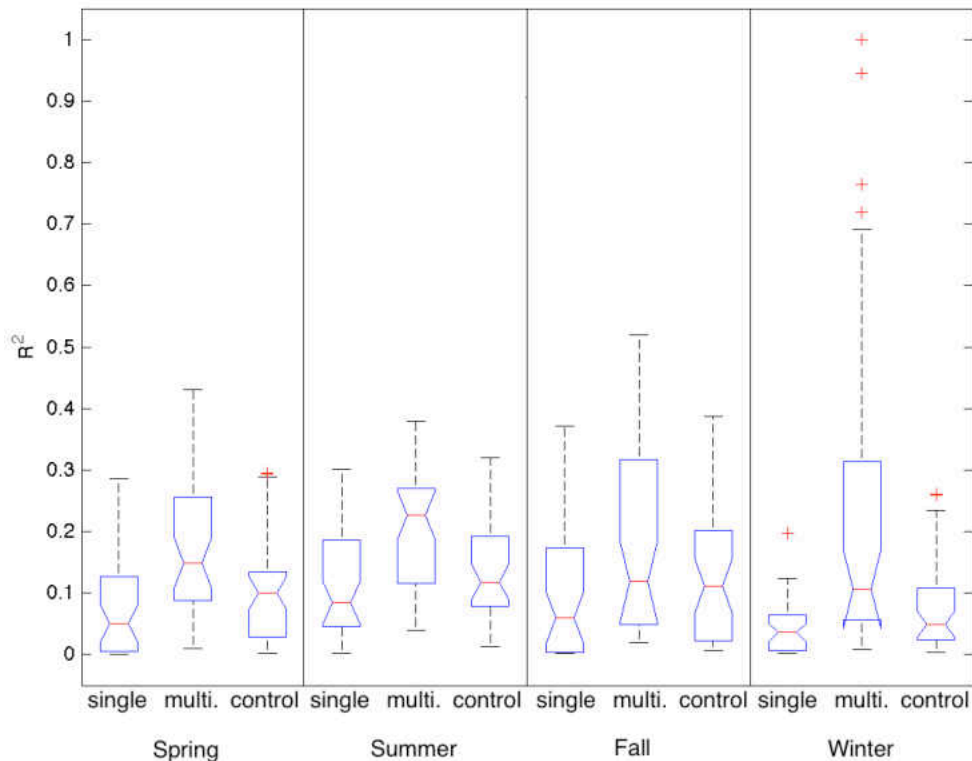


Figure 6.4 Goodness of fit distribution of seasonal linear regressions for individual sites for the period 2003-2007. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.

Winter regressions exhibit a broader distribution of R^2 than any other season. This can be attributed to a reduced sample size in the winter, which also explains the increased

outlier values of R^2 in the winter. The multivariate model performs better than the single variable model considering that their medians are significantly different at a 95% confidence level regardless of grouping (seasonal or yearly).

We now present the distributions of R^2 for multivariate linear regressions grouped into arrival elevations of the trajectories used to determine the upwind locations of Lagrangian AOD (Fig. 6.5). It can be observed that the R^2 distributions are similar regardless of arrival elevation. The median R^2 of the multivariate regressions is different (at a 95% confidence level) than the median R^2 of the single variable regressions for all trajectory arrival elevations.

As has been discussed, a modeled trajectory is an estimate of the movement of a single air parcel through the atmosphere. It has been demonstrated that transport has a direct effect on the optical properties of the atmosphere (Rozwadowska et al., 2010). We speculate that air parcels keep some of their optical properties as they are transported and therefore the AOD of parcels upwind is correlated to the AOD of air parcels downwind. With this in mind, any air parcel advected into a site of interest, regardless of arrival elevation, contributes to the optical properties of the atmospheric column over the site of interest for which AOD is being measured.

Other researchers have investigated how the atmospheric concentration profile is related to the estimates of $PM_{2.5}$ using AOD (Hutchison et al. 2008; Hu et al., 2009). In future work we'll investigate the vertical component of the trajectories and how it is related to downwind estimates of $PM_{2.5}$ using AOD.

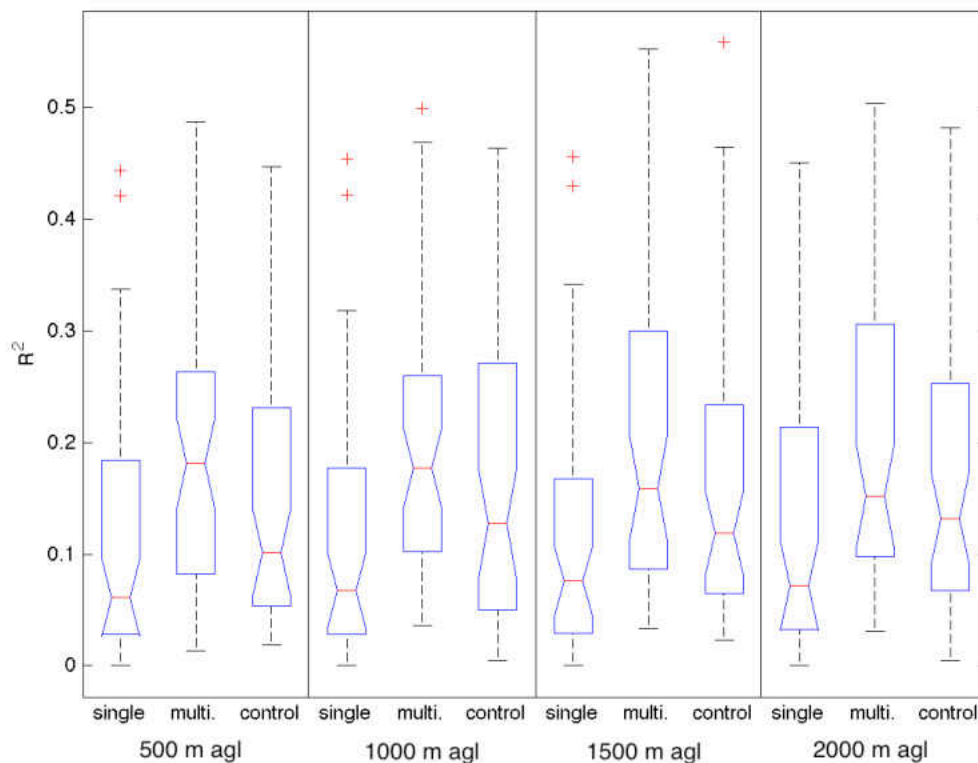


Figure 6.5 Goodness of fit distribution for yearly linear regressions grouped into arrival elevations for the period 2003-2007. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.

6.3 Multicollinearity

At this point we have demonstrated that Lagrangian AOD has value as a predictor of ground-level $PM_{2.5}$ concentration. Furthermore, we have demonstrated that Lagrangian AOD explains some of the variance of ground-level $PM_{2.5}$ regardless of trajectory arrival elevation. However, we have yet to address the variability of the goodness of fit of the regression models with respect to seasons, the sensitivity of the statistical tests used to validate the models to the degrees of freedom, and the low percentage of cases that exhibit statistically significant improvement over the single variable model.

It is relevant to understand how the explanatory variables are related to each other. We analyze the correlation of the upwind AOD with respect to the *in situ* AOD for the entire data set (Fig. 6.6). The correlation between *in situ* AOD and AOD at 3 and 6 hours upwind is high ($r > 0.8$). If we assume a moderate value of wind speed of 30 km hr^{-1} at the trajectory arrival elevations (Thuillier and Lappe, 1964; Wilson and Flesch, 2004; Buligon et al., 2010) then air parcels will have traveled roughly 90, 180, 540, 720 and 1440 km after 3, 6, 18, 24 and 48 hours respectively. Since the radius of influence in the determination of the upwind AOD was set to $r = 150 \text{ km}$, we expect at least the first two upwind predictors to be similar in value to the *in situ* AOD.

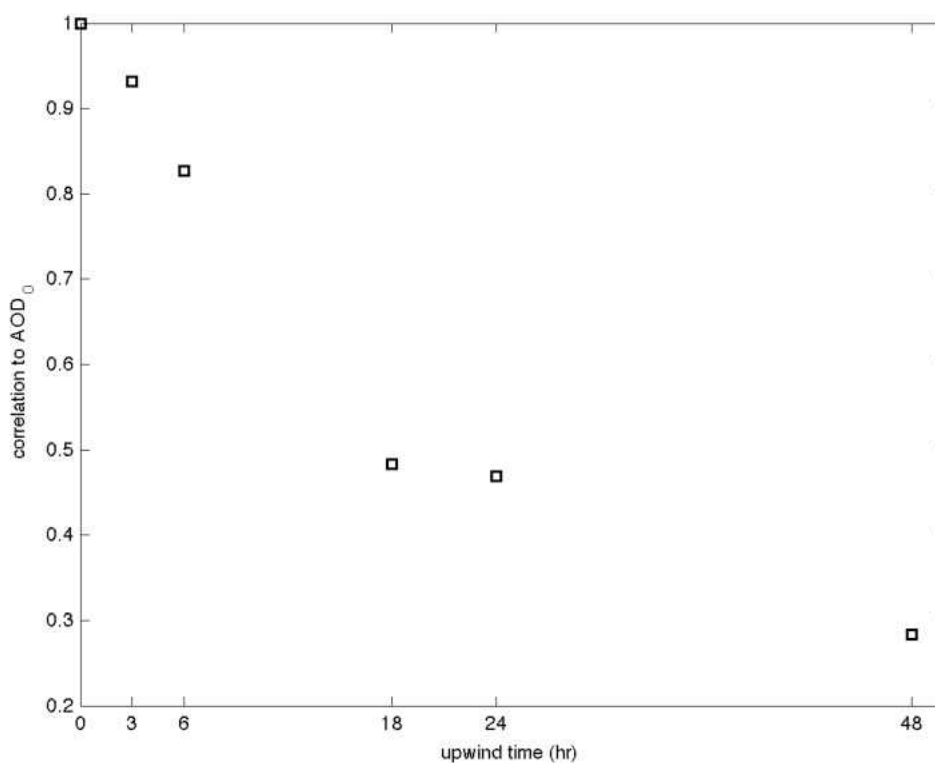


Figure 6.6 Correlation of upwind AOD to *in situ* AOD (AOD_0) as a function of upwind time.

The results shown in Fig. 6.6 have important implications in linear regression theory. Multicollinearity is the statistical phenomenon in which two or more predictor variables in a

multiple regression model are highly correlated. It does not affect the properties of the ordinary least squares estimators: these estimators remain unbiased and efficient. However, when multicollinearity is present in the data, the least squares are imprecisely estimated. If the goal is simply to estimate a response variable from a set of explanatory variables, then multicollinearity is not a problem. The estimates will still be accurate and the overall R^2 quantifies how well the model predicts the response values. However, if the goal is to understand how individual predictors affect the variability of the response, then multicollinearity is a problem (Farrar and Glauber, 1967; Silvey, 1969; Hawking and Pendleton, 1983).

To further explore the multicollinearity condition we calculate the correlation of Lagrangian AOD at the upwind times of interest for the complete data set (Table 6.2). It can be observed that AOD values spaced within 6 hours of transport (i.e. 0 and 3, 0 and 6, 3 and 6, 18 and 24 hours) are highly correlated ($r > 0.8$). Generally multicollinearity increases the standard errors.

Table 6.2 Correlation matrix of Lagrangian AOD predictors. AOD subscripts refer to the number of hours upwind.

	AOD ₀	AOD ₃	AOD ₆	AOD ₁₈	AOD ₂₄	AOD ₄₈
AOD ₀	1	0.9316	0.8271	0.4836	0.4695	0.2840
AOD ₃	0.9316	1	0.9175	0.4877	0.4734	0.2856
AOD ₆	0.8271	0.9175	1	0.4764	0.4691	0.2875
AOD ₁₈	0.4836	0.4877	0.4764	1	0.8445	0.3531
AOD ₂₄	0.4695	0.4734	0.4691	0.8445	1	0.3763
AOD ₄₈	0.2840	0.2856	0.2875	0.3531	0.3763	1

Large standard errors can be caused by things besides multicollinearity, however (Farrar and Glauber, 1967; Silvey, 1969; Hawking and Pendleton, 1983). Thus, coefficients

will have to be larger in order to be statistically significant when multicollinearity is present. In future work will seek to quantify and nullify any adverse effects that multicollinearity may have in our regression models.

6.4 Sample size and seasonal variability

So far we have seen that multivariate models explain more of the variance in the observed $PM_{2.5}$ than the single variable models. We have also shown that the improvement over the single variable models when including Lagrangian AOD as predictors is statistically significant. Nevertheless, only a small percentage of the regressions show improvement over the single variable model that is statistically significant. We have to assess the possibility that some of the cases may have failed the extra sum of squares ANOVA test due to the reduced sample size.

As has been discussed, the *p-value* is directly related to the variability within a given data set and indirectly related to the number of observations. In order to analyze the sensitivity of the *p-value* to the degrees of freedom of the sample we introduce a set of additional variables: binary categorical variables. These new binary categorical variables CAT_{spring} , CAT_{summer} and CAT_{fall} are a function of the season in which the response (i.e. $PM_{2.5}$) was observed. For example, if $PM_{2.5}$ was measured on a spring day, according to the meteorological definition of a season, then $CAT_{spring} = 1$ and $CAT_{summer} = CAT_{fall} = 0$. This new set of variables will help us address the problem of reduced sample size, as well as the seasonal variability of the regression models.

There is no binary categorical variable for winter. During winter, therefore, $CAT_{spring} = CAT_{summer} = CAT_{fall} = 0$. This choice was arbitrary and we would not expect different

results if the missing categorical variable was set for any other season. The new extended model with seasonal categorical variables is

$$\begin{aligned} [PM_{2.5}]_{\text{predicted}}^{\text{multivariate+seasons}} = & \beta_1 \cdot AOD_0 + \beta_2 \cdot AOD_3 + \beta_3 \cdot AOD_6 + \beta_4 \cdot AOD_{18} + \beta_5 \cdot AOD_{24} \cdot \\ & + \beta_6 \cdot AOD_{48} + \beta_7 CAT_{\text{spring}} + \beta_8 CAT_{\text{summer}} + \beta_9 CAT_{\text{fall}} + \beta_{10} \end{aligned} \quad (6.8)$$

Equation 6.8 allows us to increase the degrees of freedom of the numerator in the F* statistic (Equation 6.6) and also to account for the variability of the regressions model performance across seasons. The benefit of including the seasonal categorical variables is shown in Fig. 6.7. The results are for yearly regressions for individual sites for the period 2003-2007 and for each arrival elevation ($N = 200$). Only cases that exhibit statistically significant improvement over the single regression model at a 95% confidence level were considered in the distributions.

When the ANOVA was performed for the $PM_{2.5}$ - Lagrangian AOD regressions using the extra sum of squares principle we found that only 29% of the 200 cases exhibited statistically significant improvement over the single model at a 95% confidence level (Fig. 6.2a and Fig. 6.7b). However, when we include the seasonal categorical variables (Fig. 6.7a), 71% of the 200 cases exhibited statistically significant improvement over the single variable model. Furthermore, the median R^2 for the multivariate+seasons model is statistically higher than the median R^2 of the multivariate model at a 95% confidence level. This indicates an overall improvement from the original multivariate model (Equation 6.2) to the new extended model (Equation 6.8).

The improvement shown by the addition of seasonal categorical variables indicate that seasons explain part of the variance of ground-level $PM_{2.5}$. We now compare the

influence of seasonality (as shown by the seasonal categorical variables) with that of transport (as shown by the Lagrangian AOD variables).

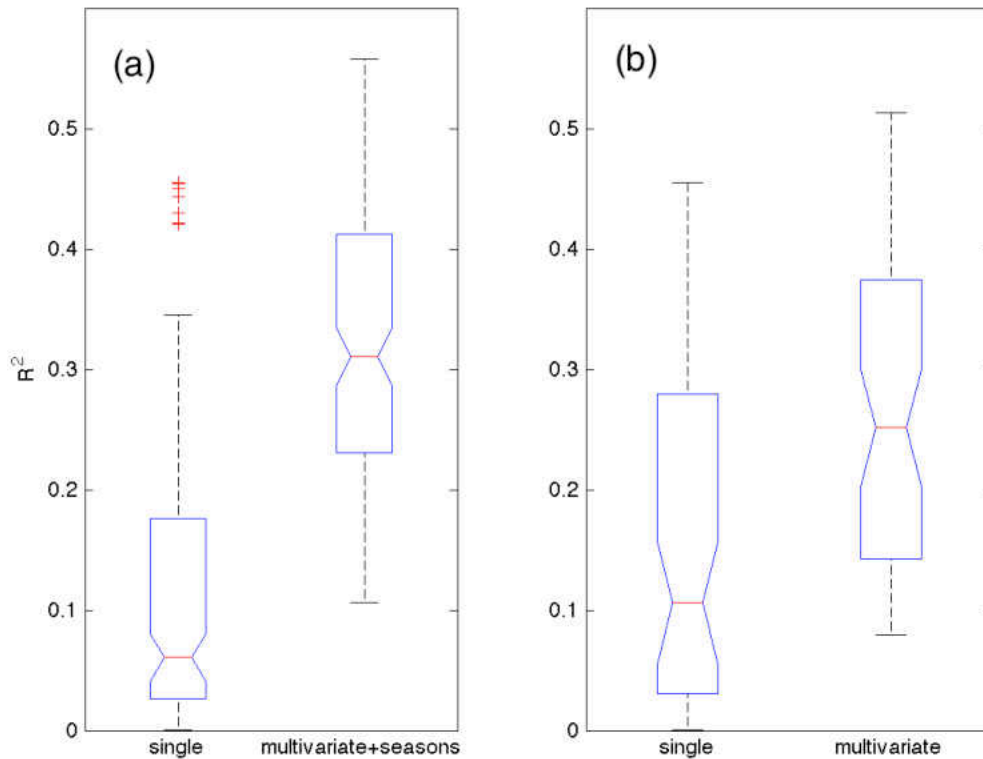


Figure 6.7 Improvement over single regressions of PM_{2.5} - Lagrangian AOD - Seasonal Categorical Variables (multivariate +seasons) (a) and yearly multivariate regressions of PM_{2.5} - Lagrangian AOD (multivariate) (b) for individual study sites at each of the four trajectory arrival elevations for the period 2003-2007. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.

To this end, we performed yearly linear regressions in which we excluded upwind AOD and only used *in situ* AOD and the seasonal categorical variables. Of these regressions, 61% exhibited statistically significant improvement over the single variable model (Equation 6.1). Labeled as “single+seasons” in Fig. 6.8, this model, which can be described as

$$\left[PM_{2.5} \right]_{\text{predicted}}^{\text{single+seasons}} = \beta_1 \cdot AOD_0 + \beta_2 CAT_{\text{spring}} + \beta_3 CAT_{\text{summer}} + \beta_4 CAT_{\text{fall}} + \beta_5, \quad (6.9)$$

has a higher median R^2 than the original multivariate model (Equation 6.2) that only considers *in situ* and upwind AOD as predictors.

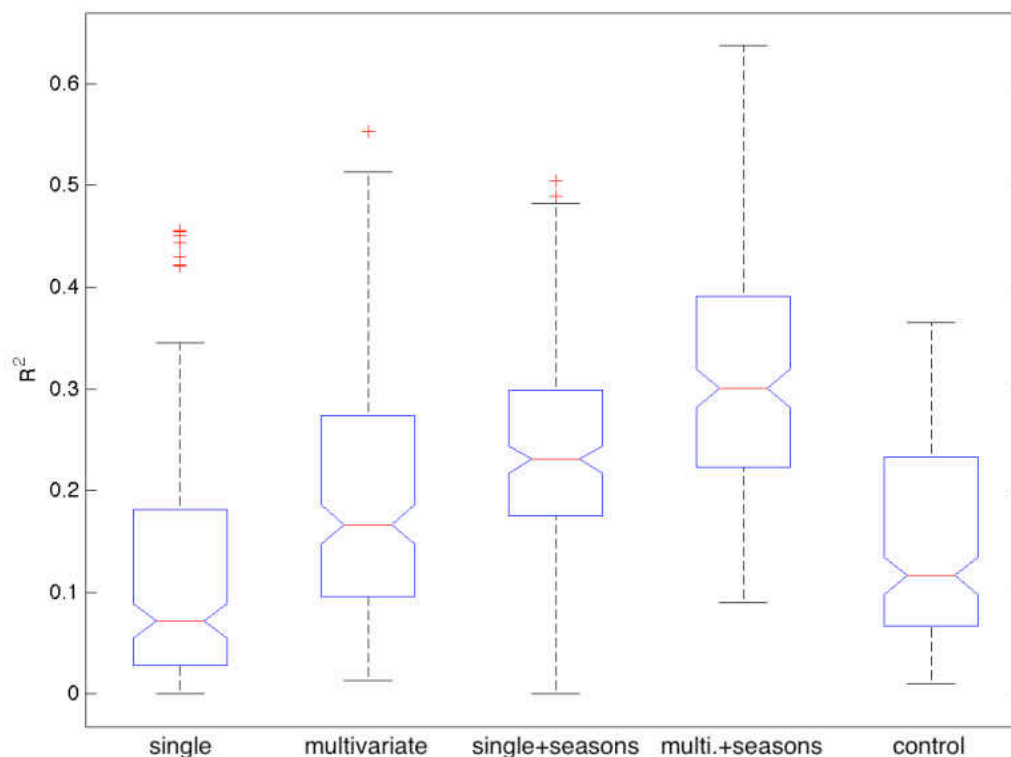


Figure 6.8 Goodness of fit distribution for four types of yearly linear regressions for the period 2003-2007 (individual sites and elevations). Multivariate regression with control sample AOD predictors is shown as a control. All cases considered regardless of statistical significance of improvement over the single model.

Thus, the seasonal effects have a strong signature that cannot be neglected. The multivariate model (Equation 6.2) has a more extended interquartile range and fourth quartile when compared to the “single+seasons” model (Equation 6.8), however. It is difficult to determine whether a model outperforms the other solely by comparing their medians. Nevertheless, transport (as indicated by Lagrangian AOD), contributes important information to the variability of ground-level aerosol concentrations. The complete model,

i.e., “multivariate+seasons”, exhibits the highest values of R^2 and also the largest percentage of statistically significant cases (70.5%) at a 95% confidence level (Fig. 6.8). It can be observed in Fig. 6.8 that the control has a median R^2 that is significantly higher (at a 95% confidence level) than the median R^2 for the single model. As previously mention, the additional predictors improve the performance to some extent, even if those additional predictors are uncorrelated to the response. Nevertheless, the median R^2 for the multivariate models (“multivariate”, “single+seasons” and “multivariate+seasons” as seen in Fig. 6.8) is significantly higher (at a 95% confidence level) than the control median R^2 .

Now that the contribution of both transport and seasons to the variance of $PM_{2.5}$ has been investigated we seek to understand the effect that a reduced amount of predictors would have on the goodness of fit of the regressions. In Fig. 6.9 we explore the effect of different amounts of upwind AOD predictors in the regressions. The box plots show the distribution of R^2 for yearly linear regressions for each site and arrival elevation using the seasonal categorical variables. The first box is equivalent to the regressions of Equation 6.9, “single+seasons” in Fig. 6.8. The sixth box is equivalent to the “multi.+seasons” regressions in Fig. 6.8. It can be observed that even though the median R^2 of some models are significantly different (at a 95% confidence level) the overall performance of the regressions is very similar. This result suggests that the “multi.+seasons” model (Equation 6.8) can be used even when some of the variables are missing. This is discussed further in Section 6.5.

When the number of predictors is increased (Fig. 6.9) the amount of cases that exhibit statistically significant improvement over the single regression model without seasonal effects also increases. This is a result of having increased the amount of degrees of freedom of the full model. The percentage of statistically significant cases increases from 64% to 71%.

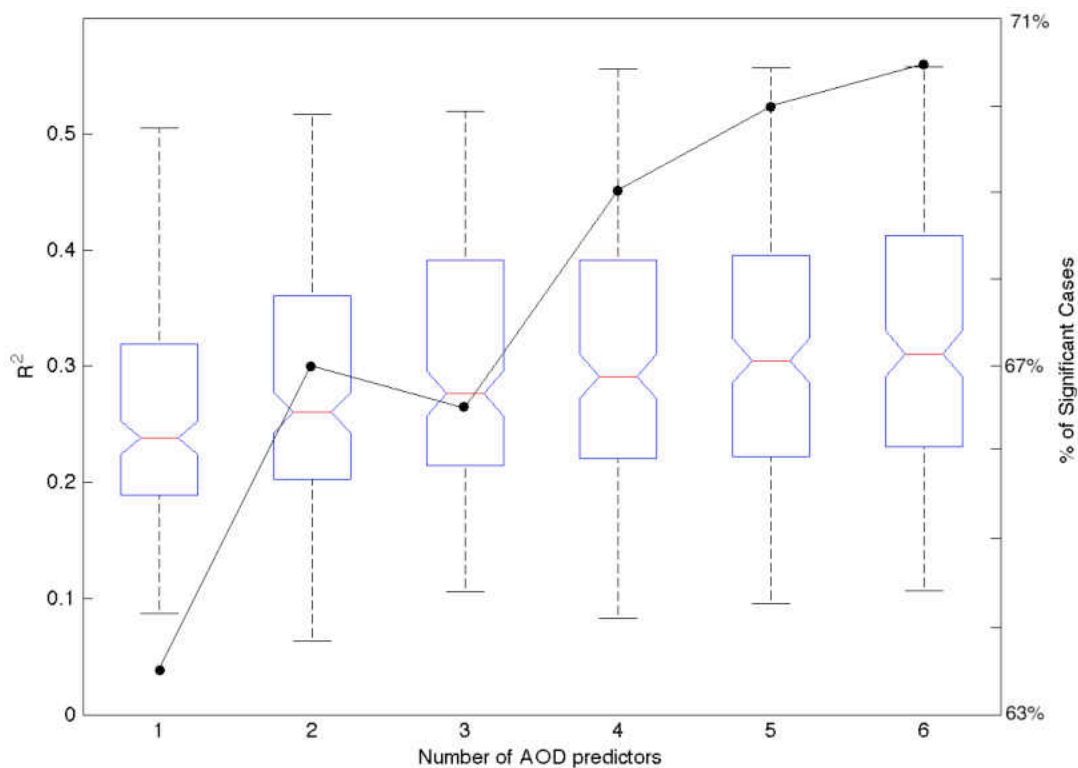


Figure 6.9 R^2 distribution of the yearly $PM_{2.5}$ - *Lagrangian* AOD - Seasonal Categorical Variable linear regression for the period 2003-2007 as a function of number of AOD predictors. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions. Percentage of statistically significant cases with respect to single variable regression without seasonal variables also shown (solid black line, axis on the right).

So far we have analyzed the distribution of R^2 for various models for yearly and seasonal data aggregates for each site and arrival elevation of the trajectories used to determine the upwind location of Lagrangian AOD. However, we have yet to analyze the performance of the models at the different sites. When the R^2 results of the “multivariate +seasons” model (Equation 6.8 and Fig. 6.8 box “multi.+seasons”) are grouped by sites (Fig. 6.10) we can get an idea of the performance of the regression models at individual sites.

The percentage of statistically significant cases with respect to the single variable regression model is also shown (solid black line, axis on the right). The sites are arranged from left to right in increasing order of percentage of statistically significant cases. The

distribution of R^2 and also the amount of statistically significant cases varies greatly from site to site. Even though Nogales, Arizona (OLS) exhibits the highest R^2 values ($R^2 > 0.6$) it also exhibits the lowest percentage of statistically significant improvement over the single variable model (15%). On the other hand, Milwaukee, Wisconsin (MKE), which exhibits the highest median R^2 , and Nashville, Tennessee (BNA), appear to have a good balance between high R^2 and amount of statistically significant cases (70% and 90% respectively).

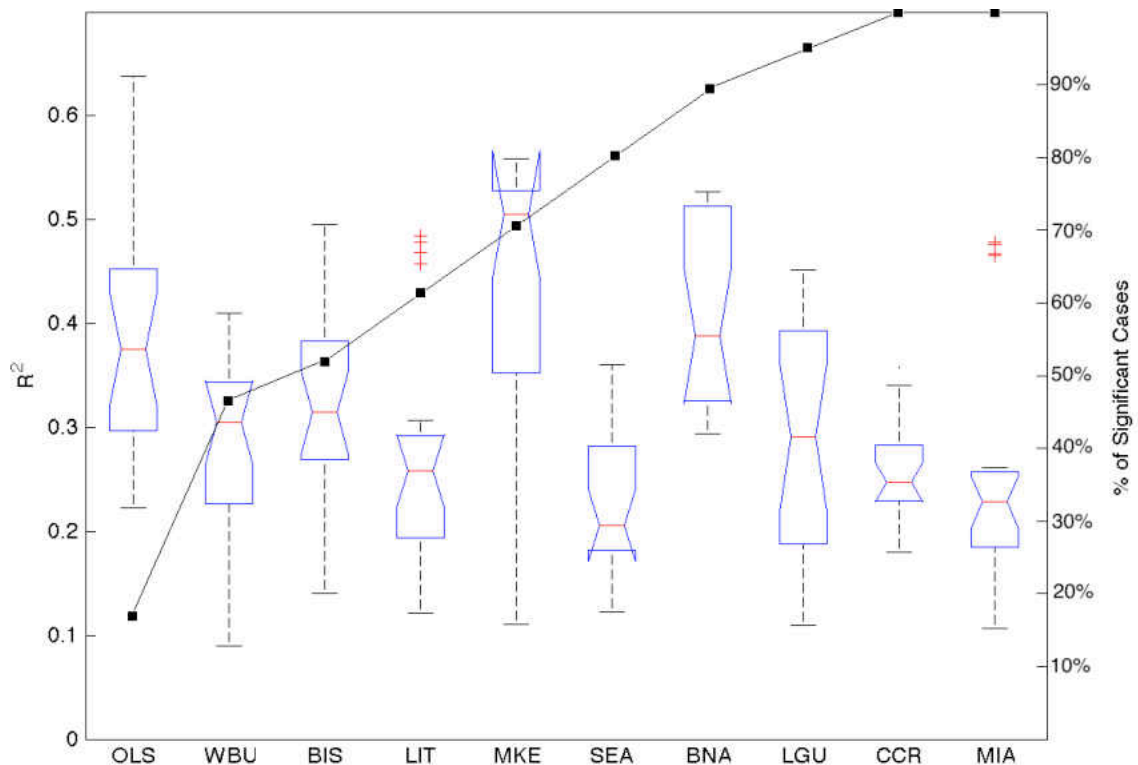


Figure 6.10 R^2 distribution of the yearly $PM_{2.5}$ - Lagrangian AOD - Seasonal Categorical Variable linear regression for the period 2003-2007 grouped into individual sites. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions. Percentage of statistically significant cases with respect to single variable regression without seasonal variables also shown (solid black line, axis on the right).

To better understand the behavior of the goodness of fit for the different sites we compare the median goodness of fit [$m(R^2)$] and the percentage of significant cases (%SF) for each site to their respective climatological properties in Table 6.6. In the last two rows of

Table 6.6 we show the correlation of the different site properties to percentage of statistically significant cases and to median R^2 .

The correlation of median R^2 to the properties of the sites is low ($|r| < 0.5$). The same can be said for the correlation between percentage of significant cases and site properties. However, there is negative correlation between median R^2 and %SF ($r = -0.61$). This negative correlation can be explained. As we only consider statistically significant cases, we neglect the values of certain regressions when we calculate the median R^2 , thus increasing the value of the median. The more cases are excluded, the higher the median R^2 will tend to be. This supports our previous observation of Milwaukee, Wisconsin (MKE) and Nashville, Tennessee (BNA) having a good balance between high R^2 and percentage of statistically significant cases.

Table 6.6 Comparison of the percentage of statistically significant “multivariate+seasons” regressions with respect to single model and median R^2 to the properties of the study sites.

Name	Climate	Pop.	Clear Days	Temp. (°C)	Precip. (mm)	%SF	m(R^2)
Nogales, AZ	BWh	20017	193	20.4	309	15	0.5186
Boulder, CO	H	100160	115	10.1	402	45	0.3377
Bismarck, ND	Dfb	61217	93	5.7	428	50	0.3834
N. Little Rock, AR	Cfa	60140	119	17.1	1257	60	0.2612
Milwaukee, WI	Dfa	605013	90	8.6	884	70	0.5149
Seattle, WA	Cfb	616627	51	11.6	972	80	0.2239
Nashville, TN	Cfa	605473	102	14.9	1222	90	0.3925
Logan, UT	BSk	49549	125	11.1	419	95	0.3031
Concord, CA	Csb	122224	160	14.6	511	100	0.2473
Miami, FL	Af	433136	74	24.8	1487	100	0.2282
r (SF%)	-	0.44	-0.43	0.10	0.48	1	-0.61
r (m(R^2))	-	0.01	0.37	-0.23	-0.37	-0.61	1

To complete our analysis of multivariate models, we develop one last model that includes relative humidity, a meteorological variable that has been used by previous authors

in the prediction of $PM_{2.5}$ using AOD. This model, labeled as “multivariate+seasons+rh” (Fig 6.11), has an overall performance that is better than the “multivariate+seasons model (Equation 6.8). However, the percentage of significant cases that exhibit statistically significant improvement over the single regression model is 63% as opposed to the percentage exhibited by the “multivariate+seasons” model 71%.

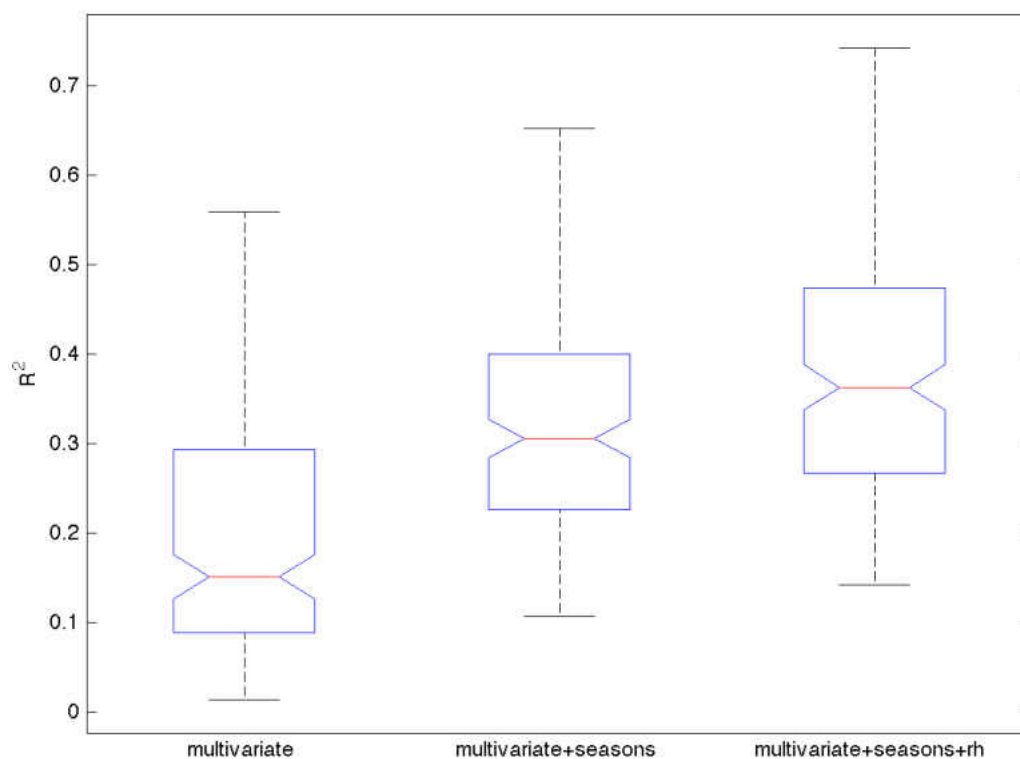


Figure 6.11 Goodness of fit distribution for three types of yearly linear regressions for the period 2003-2007 (individual sites and elevations). Multivariate, multivariate+seasons and a full model including relative humidity. Only cases that exhibit statistically significant improvement over the single regression model were considered in the distributions.

Relative humidity affects the optical properties of $PM_{2.5}$. Some of these particles can be either hygroscopic or hydrophobic. They may also react and change their composition in the presence of atmospheric moisture, thus having an effect on AOD. The size of water-soluble or any hydrophilic aerosol particle increases with growing relative humidity. This

changes the mass density of the particles towards the density of water even if the aerosol number density is constant. Due to the chemical difference in the aerosol types, their hygroscopic properties differ as well (Day and Malm, 2001).

Cobourn (2010) explored the relationship between $PM_{2.5}$ and meteorological variables. It was found that there is a strong relationship between $PM_{2.5}$ and daily maximum temperature. The interaction between $PM_{2.5}$ vs. relative humidity was also explored but was not statistically significant. Nevertheless, the contribution of relative humidity to the explanation of the variance of $PM_{2.5}$ was statistically significant in their multiple non-linear regression models, and added about 2% to the fraction of variation explained by the regression (R^2). As shown in Fig. 6.11, our results confirm those of Cobourn (2010) in that relative humidity explains some of the variance of $PM_{2.5}$.

6.5 Estimates of $PM_{2.5}$ in cases of missing *in situ* AOD

We have shown in previous sections that the “multivariate+seasons” model may be used even when some of the variables are missing. In this section we’ll test cases in which *in situ* AOD is missing. This test is relevant given that AOD values are occasionally missing due to cloud cover in locations where estimates of $PM_{2.5}$ are needed. We’ll estimate $PM_{2.5}$ for two sample days in Milwaukee, Wisconsin (Table 6.7). On the first test day, August 7, 2007, all of the predictors (*in situ* AOD and Lagrangian AOD) are available. This complete test day will be used as a control. The predicted values for this day (columns 12 and 13 of Table 6.7) were calculated by entering the data in Table 6.7 (columns 3 through 11) for this given day into Equation 6.8 using the regression coefficients (β_j) shown in Table 6.8. The regression coefficients in Table 6.8 were obtained by a yearly single linear regression of $PM_{2.5}$ and *in situ* AOD, and a yearly multivariate linear regression of $PM_{2.5}$ and *in situ* and Lagrangian AOD

with locations determined by trajectories arriving at 500 m agl for Milwaukee, Wisconsin for 2007. Only the days in 2007 that had all of the predictors available (i.e. *in situ* and Lagrangian AOD) were used in the estimation of the regression coefficients, however. The seasonal categorical variables were also included in the multivariate linear regression.

The second test day, August 19, 2007, is missing AOD *in situ* and at 3 and 6 hours upwind. In order to estimate ground-level PM_{2.5} concentrations for this day, the missing values were replaced with zeros and then entered into Equation 6.8 along with the rest of the available predictors. The same coefficients (Table 6.8) obtained from the yearly regression of the Milwaukee, Wisconsin data aggregate for 2007 with upwind locations of Lagrangian AOD estimated with trajectories arriving at 500 m agl are used in Equation 6.8 to obtain an estimate of PM_{2.5}.

Table 6.7 Sample test data for Milwaukee, Wisconsin on August 7 and 19, 2007 and August 7, 2007. The data corresponds to Lagrangian AOD determined by trajectories arriving at 500 m agl.

Day	PM _{2.5} ($\mu\text{g m}^{-3}$)	AOD hours upwind						Categorical			Estimated PM _{2.5} ($\mu\text{g m}^{-3}$)		error	
		0	3	6	18	24	48	spring	summer	fall	Single	Multi.	Single	Multi.
Aug 7	7.60	0.122	0.119	0.093	0.198	0.068	0.091	0	1	0	12.61	8.68	5.01	1.08
Aug 7*	7.60	-	0.119	0.093	0.198	0.068	0.091	0	1	0	11.10	6.7	3.50	-0.90
Aug 19	10.70	-	-	-	0.742	0.711	0.171	0	1	0	11.10	10.04	0.40	-0.66

**In situ* AOD was synthetically removed.

Table 6.8 Regression coefficients for single and multivariate+seasons linear models for Milwaukee, Wisconsin with Lagrangian AOD determined by trajectories arriving at 500 m agl for 2007.

Model	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
Single	12.39	11.10								
Multivariate	16.18	-4.76	17.87	5.29	2.58	-1.30	-14.43	-11.00	-5.40	15.50

The error of the $PM_{2.5}$ estimate obtained for the incomplete test day, August 19, 2007, (-.66) is lower in value than the error of the $PM_{2.5}$ estimate for the control day (1.08) where *in situ* AOD was available. Furthermore, we show a case in which *in situ* AOD is synthetically removed (August 7, 2007). When we compare the control day (August 7, 2007) with its synthetically incomplete counterpart in Table 6.7, we observe that the errors have actually decreased. The errors decreased from 5.01 to 3.50 and from 1.08 to -0.90 for the single and multivariate model respectively. We do not suggest that an estimate of $PM_{2.5}$ that has been obtained from an incomplete set of predictors is better than estimates from a complete predictor set, but rather that the errors are comparable in magnitude, and that the model can be used even in instances of missing *in situ* AOD.

These examples demonstrate that upwind AOD has predictive capabilities, even in cases when *in situ* AOD may be missing due to cloud cover. Nevertheless, the reconstruction of incomplete AOD fields is relevant to the production of more robust regression models. With this in mind, when an estimate of $PM_{2.5}$ is necessary but *in situ* AOD is missing due to cloud cover, we should be able to use the multivariate models obtained for the year corresponding to the day for which an estimate is needed, regardless of the missing *in situ* AOD value.

6.6 Summary

Previous authors have improved the estimates of $PM_{2.5}$ through MODIS-AOD by including meteorological variables relevant to the physicochemical properties of this pollutant. In this work we improved estimates of $PM_{2.5}$ over the single variable models by using Lagrangian AOD as predictors in multivariate empirical models. The performance of these multivariate empirical models, and their statistical validity was also assessed.

Even when the R^2 for the multivariate regressions was consistently higher compared to the single regressions, not all of the results were statistically better when tested using the extra sum of squares principle. This is due both to the variability of the predictors and to the degrees of freedom of the data set. Furthermore, we found that the goodness of fit of the models is different according to seasons: summer regressions exhibited the highest median R^2 . The seasonal variability in model performance was addressed by including a set of categorical seasonal variables. These variables increased the degrees of freedom of the data set and also helped us account for the seasonal variability. The models that included the seasonal variables performed better than the models that didn't.

Multicollinearity was found to exist between AOD predictors spaced within 6 hours of transport, i.e., 0 and 3, 0 and 6, 3 and 6, 18 and 24 hours. Given that the goal of this project was to improve the estimates of $PM_{2.5}$ from a set of explanatory variables (i.e. Lagrangian AOD), multicollinearity was not a problem. In future work we'll seek to understand how the various predictors impact the response.

It was found that the goodness of fit of the regressions varies greatly from site to site. The site properties did not support any explanation of the variability of the percentage of statistically significant cases nor of the variability median R^2 . An area of interest for future work is to understand why the performance of the regression models varies from site to site.

The multivariate regression models that were generated have the capability of estimating $PM_{2.5}$ even when *in situ* AOD retrievals are missing. This has relevant implications for $PM_{2.5}$ estimation in regions of cloud cover where MODIS-AOD is unavailable or in sites below the 30°N latitude where MODIS scans are scarce.

7. Summary, Conclusions and Future Work

Accurate estimates of ground level $PM_{2.5}$ concentrations are relevant in the analysis of air quality, epidemiology, cloud microphysics and climate forcing of anthropogenic aerosols. Exposure to suspended particulate matter with aerodynamic diameters less than $2.5 \mu m$ ($PM_{2.5}$) has negative effects on human health. Suspended particulate matter also plays an important role in the climate system by altering the Earth's energy budget. Any changes in the atmospheric aerosol load can result in global climate changes.

The limited spatial coverage of ground-based monitoring stations restricts the study of $PM_{2.5}$ and its effect on human health and the environment. Satellite remote sensing has been used to aid the prediction of ground-level $PM_{2.5}$ concentrations. The Moderate Resolution Imaging Spectroradiometer (MODIS) instruments provide a convenient method for aerosol remote sensing.

Aerosol optical depth (AOD) retrieved by MODIS has been used as a predictor in empirical models to estimate ground-level concentrations of $PM_{2.5}$. These estimates usually have large uncertainties and errors, however. Previous authors have improved the estimates of $PM_{2.5}$ by including meteorological variables relevant to the properties of this pollutant along with AOD. In this work we generated empirical models that included *in situ* AOD and AOD at various upwind locations. The main result of this work is the determination that estimates of $PM_{2.5}$ can be improved by using Lagrangian MODIS-AOD as predictors.

The upwind locations of the Lagrangian AOD were estimated using backward air trajectories arriving at four different elevations. Trajectory model users have to input a starting elevation into the trajectory models to specify the effective pollutant source height (forward trajectory) or an arrival elevation (backward trajectory). The starting elevation of

backward trajectory calculations are typically specified to be well above the surface, as surface winds do not necessarily represent the movement of contaminants. Given that MODIS measures AOD for the complete integrated atmospheric column, there is not a unique arrival elevation that can be specified, thus making the problem of choosing a starting elevation more pronounced. As a result, more than one starting elevation was used in the calculation of backward trajectories.

We explored the sensitivity of trajectory models to starting elevation using daily trajectories arriving at four different starting elevations (500, 1000, 1500 and 2000 m agl) at ten different sites for an eight year period. Trajectory model calculations were found to be strongly sensitive to starting elevation. With a 500 m difference in starting elevation, the 48-hr all-site annual mean horizontal separation of trajectory endpoints was 326 km. When the difference in starting elevation was doubled to 1000 m, the all-site annual mean horizontal separation of trajectory endpoints nearly doubled to 627 km. With a further increase in starting elevation difference (1500 m), the horizontal separation increased to 886 km. A seasonal dependence of this sensitivity to starting elevation was also found, with the smallest mean horizontal separation occurring during the summer months, the largest during winter and intermediate values during the fall and spring.

The effect of model and data set choice on trajectory model sensitivity to starting elevation may be an area of interest for further research. Another suggestion for further research would be to assess the effect of starting elevation on the vertical component of trajectory model calculations; this aspect of model sensitivity was not addressed here.

The trajectory information was used to determine AOD values at different locations. Since the spatiotemporal distribution of unprocessed AOD granules is irregular, we decided to generate uniform daily distance-weighted averaged AOD fields. As part of the production of daily average AOD fields we analyzed the spatial autocorrelation of unprocessed AOD

pixels. A general trend was found in which the correlation between AOD pixels drops below 0.5 beyond a 100 km separation. Thus we decided to set the influence radius for the distance weighted averaging procedure to 100 km.

Our daily average product was able to capture the features and structures of the original unprocessed AOD field. Furthermore, values of AOD were generated in locations where large gaps in the unprocessed field were previously present. This artificial spatial expansion of sparse Level 2 data is also observed in ‘Level 3 - Atmosphere’ MODIS data; which was not used in this work. Even though the reconstruction of missing data was not the main purpose of generating average grids it is advantageous since this can help us increase our sample size in a robust manner. An aspect for future research is to analyze the performance of our distance weighted AOD product vs. the performance of the NASA ‘Level 3 - Atmosphere’ averaged products.

Another area for further research is the reconstruction of missing MODIS-AOD data. In satellite data analysis, such as the determination of $PM_{2.5}$ concentrations from MODIS-retrieved AOD, missing data is a common problem that affects the quality of the empirical models. Data may be missing due to cloud cover, errors in the retrieval algorithms or even malfunctions of the instruments. Often these data gaps extend over large areas due synoptic-scale meteorological features.

Beckers et al. (2003) used an iterative method for the calculation of the empirical orthogonal functions (EOFs) of incomplete spatiotemporal oceanographic data using standard singular value decomposition to fill the missing data. They reconstructed the data fields by selecting the optimal amount of EOFs until convergence. The procedure was tested with a synthetic incomplete data field and an actual field with missing data yielding satisfactory results. One of the advantages of this technique is that, unlike interpolation, EOF reconstruction does not need “future” points to determine missing data in a time

series. We speculate that these techniques can be applied to incomplete AOD data fields to fill the missing data in a robust manner. These reconstructed AOD fields will provide valuable information in anthropogenic aerosol epidemiological studies and climate forcing studies. However, simpler methods, such as nearest neighbor or bilinear interpolation, may prove also useful and their potential in the reconstruction of missing MODIS-AOD data is an area of interest for future work.

Multivariate regression models that included AOD *in situ* and upwind as predictors of PM_{2.5} were generated by linear regressions using the least squares method. The statistical significance of the improvement of the multivariate models over the single models was tested using the extra sum of squares principle.

We found that in many cases, even when the R² was high for the multivariate models, the results were not statistically better than the single regression models. The empirical models were sensitive to seasonal variations. We addressed this by including a set of categorical seasonal variables. The models with these variables performed better than the models that didn't account for seasonal variability.

Multicollinearity was found to exist between AOD predictors spaced within 6 hours of transport, i.e., 0 and 3, 0 and 6, 3 and 6, 18 and 24 hours. The goal of this project was to improve the estimates of a response variable (PM_{2.5}) from a set of explanatory variables (Lagrangian AOD), therefore multicollinearity is not a relevant problem. In future work we will seek to understand how the different explanatory variables are related to the response.

We found that the goodness of fit of the regressions varies greatly from site to site. We did not find any correlation between the site properties to the percentage of statistically significant cases nor to median R². Another area for future work is to fully understand why there is such a variability of the regression models with location.

It was found that the multivariate regression models have the capability of estimating $PM_{2.5}$ even when *in situ* AOD retrievals are missing. This finding has important implications for $PM_{2.5}$ estimations in regions of cloud cover where *in situ* MODIS-AOD is unavailable or in sites between latitudes 30°N-30°S where MODIS scans are scarce. Nevertheless, the reconstruction of missing AOD data, another aspect for future work, is relevant in the production of more robust and significant regressions.

As has been discussed, *in situ* meteorological variables such as relative humidity, temperature, wind speed and mixing layer height are correlated to $PM_{2.5}$ concentrations. We speculate that just as we were able to use Lagrangian AOD as a predictor, other Lagrangian variables, such as upwind $PM_{2.5}$, relative humidity and mixing layer height can improve the models further. We also speculate that upwind MODIS-AOD may be correlated with *in situ* $PM_{2.5}$ and that *synthetic* $PM_{2.5}$ concentrations estimated with Gaussian plume or Gaussian puff models may also be correlated with *in situ* $PM_{2.5}$.

Bibliography

- Abdalmogith SS, Harrison RM, Derwent RD. Particulate sulphate and nitrate in Southern England and Northern Ireland during 2002/3 and its formation in a photochemical trajectory model. *Sci Tot Environ* 2006; 368: 769 – 80.
- Al-Saadi JA, Szykman J, Pierce R, Kittaka C, Neil D, Chu D, Remer LA, Gumley L, Prins E, Weinstock L. Improving national air quality forecasts with satellite aerosol observations. *Bull Amer Meteor Soc* 2005; 86: 1249 – 61.
- Arya SP. *Introduction to Micrometeorology*. 2nd ed. San Diego CA: Academic Press; 2001. International Geophysics Series Volume 79 p. 67 – 84.
- Barnes WL, Pagano TS, Salomonson V. Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS AM-1. *IEEE Trans Geosci Remote Sens* 1998; 36: 1088 – 100.
- Batmunkh T, Kim YJ, Lee KY, Cayetano MG, Jung JS, Kim SY, Kim KC, Lee SJ, Kim JS, Chang LS, An JY. Time-Resolved Measurements of PM_{2.5} Carbonaceous Aerosols at Gosan. Korea. *J Air & Waste Manage Assoc* 2011; 61: 1174 – 82.
- Baumann K, Stohl A. Validation of a long-range trajectory model using gas balloon tracks from the Gordon Bennett Cup 95. *J Appl Meteor* 1997; 36(6): 711 – 20.
- Beckers J, Rixen M. EOF Calculations and Data Filling from Incomplete Oceanographic Datasets. *J Atmos Oceanic Technol* 2003; 20: 1839 – 55.
- Bellouin N, Boucher O, Haywood J, Reddy MS. Global estimate of aerosol direct radiative forcing from satellite measurements. *Nature* 2005; 438: 1138 – 41.
- Bellouin N, Jones A, Haywood J, Christopher SA. Updated estimate of aerosol direct radiative forcing from satellite observations and comparison against the Hadley Centre climate model. *J Geophys Res* 2008; 113: D10205.
- Boyouk N, Léon J, Delbarre H, Podvin T, Deroo C. Impact of the mixing boundary layer on the relationship between PM_{2.5} and aerosol optical thickness. *Atmos Environ* 2010; 4(2): 271 – 77.
- Buligon L, Degrazia GA, Acevedo OC, Szinvelski CRP, Goulart AGO. A semi-analytical solution for the mean wind profile in the Atmospheric Boundary Layer: the convective case. *Atmos Chem Phys* 2010; 10: 2227 – 36.
- Castleman Jr. AW. Nucleation processes and aerosol chemistry. *Space Sci Rev* 1974; 15(5): 547 – 89.

- Chang S, Chou CC, Chan C, Lee C. Temporal characteristics from continuous measurements of PM_{2.5} and speciation at the Taipei Aerosol Supersite from 2002 to 2008. *Atmos Environ* 2010; 44(8): 1088 – 96.
- Charlson RJ, Schwartz SE, Hales JM, Cess D, Coakley JA, Hansen JE. Climate forcing of anthropogenic aerosols. *Science* 1992; 255: 423 – 30.
- Christopher SA, Gupta P. Satellite remote sensing of particulate matter air quality: The cloud-cover problem. *J Air & Waste Manage Assoc* 2010; 60(5): 596 – 602.
- Christopher SA, Jones TA. Satellite and surface-based remote sensing of Saharan dust aerosols. *Remote Sens Environ* 2010; 114(5): 1002 – 7.
- Chu DA, Kaufman YJ, Ichoku C, Remer LA, Tanré D, B.N. Holben BN. Validation of MODIS aerosol optical depth retrieval over land. *Geophys Res Lett* 2002; 29(12): 8007.
- Chu DA, Kaufman YJ, Zibordi G, Chern JD, Mao J, Li C, Holben BN. Global monitoring of air pollution over land from the Earth Observing System—Terra Moderate Resolution Imaging Spectroradiometer (MODIS). *J Geophys Res* 2003; 108(4): 1 – 18.
- Cobourn WG. An enhanced PM_{2.5} air quality forecast model based on nonlinear regression and back-trajectory concentrations. *Atmos Environ* 2012; 44(25): 3015 – 23.
- Danielsen EF. Trajectories: Isobaric, isentropic and actual. *J Meteor* 1961; 18: 479 – 86.
- Day DE, Malm WC. Aerosol light scattering measurements as a function of relative humidity: a comparison between measurements made at three different sites. *Atmos Environ* 2001; 35: 5169 – 76.
- Draxler RR, Hess GD. Description of the HYSPLIT4 modeling system. NOAA Technical Memorandum ERL ARL-224 1997. Air Resources Laboratory.
- EIIP Volume 9: CHAPTER 1: Getting Started: Emission Inventory Methods for PM_{2.5}, July 1999.
- Engle-Cox JA, Holloman CH, Coutant BW. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmos Environ* 2004; 38: 2495 – 509.
- Engle-Cox JA, Hoff RM, Rogers R, Dimmick F, Rush AC, Szykman JJ, Al-Saadi J, Chu DA, Zell ER. Integrating lidar and satellite optical depth with ambient monitoring for 3-dimensional particulate characterization. *Atmos Environ* 2006; 40(40): 8056 – 67.
- European Environment Agency. Air quality in Europe. State and Trends 1990-1999. Topic Report 4/2002, Copenhagen 2003.
- Farrar DE, Glauber RR. Multicollinearity in Regression Analysis: The Problem Revisited. *REStat* 1967; 49(1): 92 – 107.

- Gao B, Kaufman YJ. Selection of the 1.375 μm MODIS Channel for Remote Sensing of Cirrus Clouds and Stratospheric Aerosols from Space. *J Atmos Sci* 1995; 52(23): 4231 – 37.
- Gebhart KA, Schichtel BA, Barna MG. Directional biases in back trajectories caused by model and input data. *J Air & Waste Manag Assoc* 2005; 55(11): 1649 – 62.
- Guenther B, Barnes W, Knight E, Barker J, Harnden J, Weber R, Roberto M, Godden G, Montgomery H, Abel P. MODIS Calibration: A Brief Review of the Strategy for the At-Launch Calibration Approach. *J Atmos Oceanic Technol* 1996; 13: 274 – 85.
- Guo JP, Zhang XY, Che HZ, Gong SL, An X, Cao CX, Guang J, Zhang H, Wang YQ, Zhang XC, Xue M, Li XW. Correlation between PM concentrations and aerosol optical depth in eastern China. *Atmos Environ* 2009; 43(37): 5876 – 86.
- Gupta P, Christopher SA, Wang J, Gehrig R, Lee Y, Kumar N. Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmos Environ* 2006; 40(30): 5880 – 92.
- Gupta P, Christopher SA. An evaluation of Terra-MODIS sampling for monthly and annual particulate matter air quality assessment over the Southeastern United States. *Atmos Environ* 2008; 42(26): 6465 – 71.
- Gupta P, Christopher SA. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *J Geophys Res* 2009; 114(20205): 1 – 7.
- Hansen J, Sato M, Ruedy R. Radiative forcing and climate response. *J Geophys Res* 1997; 102, 6831 – 64.
- Harris JM, Kahl JD. Analysis of 10-day isentropic flow patterns for Barrow, Alaska: 1985-1992. *J Geophys Res* 1994; 9(12): 25845 – 55.
- Harris JM, Draxler RR, Oltmans SJ. Trajectory model sensitivity to differences in input data and vertical transport method. *J Geophys Res* 2005; 110: D14109.
- Hawking RR, Pendleton OJ. The regression dilemma. *Commun Stat-Theo Meth* 1983; 12: 497 – 527.
- Hobbs PV. *Introduction to Atmospheric Chemistry*. Cambridge University Press; 2000.
- Holben BN, Eck TF, Slutsker I, Tanré D, Buis JP, Setzer A, Vermote E, Reagan JA, Kaufman YJ, T. Nakajima, Lavenu F, Jankowiak I, Smirnov A. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sens Environ* 1998; 66(1): 1 – 16.
- Holzworth GC. Estimates of mean maximum mixing depth in the contiguous United States. *Mon Wea Rev* 1964; 92: 235 – 42.

- Hu R, Sokhi R, Fisher B. New algorithms and their application for satellite remote sensing of surface PM_{2.5} and aerosol absorption. *J Aerosol Sci* 2009; 40: 394 – 402.
- Hubanks PA, King MD, Platnick S, Pincus R. MODIS Algorithm Theoretical Basis Document No. ATBD-MOD-30 for Level-3 Global Gridded Atmosphere Products (08_D3, 08_E3, 08_M3). Collection 005 Version 1.1; 4 December 2008. ATBD-MOD-30 [modis-atmos.gsfc.nasa.gov/MOD08_M3/atbd.html].
- Hutchings JW, Robinson MS, McIlwraith H, Kingston JT, Herckes P. The Chemistry of Intercepted Clouds in Northern Arizona during the North American Monsoon Season. *Water Air Soil Pollut* 2009; 199: 191 – 202.
- Hutchison KD, Smith S, Faruqui SJ. Correlating MODIS aerosol optical thickness data with ground-based PM_{2.5} observations across Texas for use in a real-time air quality prediction system. *Atmos Environ* 2005; 39(37): 7190 – 203.
- Hutchison KD, Faruqui SJ, Smith S. Improving correlations between MODIS aerosol optical thickness and ground-based PM_{2.5} observations through 3D spatial analyses. *Atmos Environ* 2008; 42: 530 – 43.
- Jaffe DA, Reidmiller DR. Now you see it, now you don't: Impact of temporary closures of a coal-fired power plant on air quality in the Columbia River Gorge National Scenic Area. *Atmos Chem Phys* 2009; 9: 7997 – 8005.
- Kahl JD, Samson PJ. Uncertainty in Estimating Boundary-Layer Transport during Highly Convective Conditions. *J Appl Meteor* 1988; 27: 1024 – 35.
- Kahl JD, Harris JM, Herbert GA. Intercomparison of three long-range trajectory models applied to Arctic haze. *Tellus* 1989; 41B: 524 – 36.
- Kahl JD. A cautionary note on the use of air trajectories in interpreting atmospheric chemistry measurements. *Atmos Environ* 1993; 27A: 3037 – 38.
- Kahl JDW. On the prediction of trajectory model error. *Atmos Environ* 1996; 30(17): 2945 – 57.
- Kappos AD, Bruckmann P, Eikmann T, Englert N, Heinrich U, Höpfe P, Koch E, Krause GH, Kreyling WG, Rauchfuss K, Rombout P, Schulz-Klemp V, Thiel WR, Wichmann HE. Health effects of particles in ambient air. *Int J Hyg Envir Heal* 2004; 207(4): 399 – 407.
- Kaufman YJ, Tanré D, Remer LA, Vermote EF, Chu A, Holben BN. Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *J Geophys Res* 1997; 102(17): 17 – 51.
- Kaufman YJ, Boucher O. A satellite view of aerosols in the climate system. *Nature* 2002; 419: 215 – 23.

- Kharol SK, Badarinath KVS, Kaskaoutis DG, Sharma AR, Gharai B. Influence of continental advection on aerosol characteristics over Bay of Bengal (BoB) in winter: results from W- ICARB cruise experiment. *Ann Geophys* 2011; 29: 1423 – 38.
- King MD, Kaufman YJ, Tanré D, Nakajima T. Remote sensing of tropospheric aerosols from space: past, present, and future. *Bull Amer Meteor Soc* 1999; 80: 2229 – 59.
- King MD, Menzel WP, YJ Kaufman YJ. Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS. *IEEE Trans Geosci Remote Sens* 2003; 41(2): 442 – 58.
- Kleinbaum DG, Kpper LL, Muller KE, Nizam A. Applied regression analysis and other multivariate methods. 2nd ed. Pacific Grove, CA: Duxbury; 1998.
- Koçak M, Theodosi C, Zarmpas P, Séguret MJM, Herut B, Kallos G, Mihalopoulos N, Kubilay N, Nimmo M. Influence of mineral dust transport on the chemical composition and physical properties of the Eastern Mediterranean aerosol. *Atmos Environ* 2012; 57: 266 – 77.
- Koелеmeijer RB, Homan CD, Matthijsen J. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmos Environ* 2006; 40: 5304 – 15.
- Koffi EN, Nodop K, Benech B. Comparison of constant volume balloons, model trajectories and tracer transport during ETEX. *Atmos Env* 1998; 32: 4139 – 49.
- Kouyoumdjian H, Saliba NA. Mass concentration and ion composition of coarse and fine particles in an urban area in Beirut: effect of calcium carbonate on the absorption of nitric and sulfuric acids and the depletion of chloride. *Atmos Chem Phys* 2006; 6: 1865 – 77.
- Kuo YH, Skumanich M, Haagenson PL, Chang JS. The accuracy of trajectory models as revealed by the observing system simulation experiments. *Mon Wea Rev* 1985; 113(11): 1852 – 67.
- Levy RC, Remer LA, Martins JV, Kaufman YJ, Plana-Fattori A, Redemann J, Wenny B. Evaluation of the MODIS aerosol retrievals over ocean and land during CLAMS. *Journal of Atmospheric Sciences* 2005; 62: 974 – 92.
- Levy RC, Remer LA, Tanré D, Mattoo S, Kaufman YJ. Algorithm For Remote Sensing of Tropospheric Aerosol Over Dark Targets from MODIS: Collections 005 and 051: Revision 2; Feb 2009.
- Li CC, Lau AK, Mao JT. Retrieval, validation and application of the 1- km aerosol optical depth from MODIS measurement over Hong Kong. *IEEE Trans Geosci Remote Sens* 2005; 43: 2650 – 58.

- Li CC, Mao JT, Lau AK, Yuan Z, Wang M, Liu X. Application of MODIS aerosol product in the study of air pollution in Beijing. *Science in China – Series, D Earth Science* 2005; 35: 177 – 86.
- Li F, Vogelmann A, Ramanathan V. Saharan dust aerosol radiative forcing measured from space. *J Climate* 2004; 17: 2558 – 71.
- Li H, Faruque F, Williams W, Al-Hamdan M, Luvall J, Crosson W, Rickman D, Limaye A. 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* 2009; 43(28): 4303 – 10.
- Li Z. Influence of absorbing aerosols on the inference of solar surface radiation budget and cloud absorption. *J Climate* 1998; 11: 5 – 17.
- Liang S, Zhong B, Fang H. Improved estimation of aerosol optical depth from MODIS imagery over land surfaces. *Remote Sens Environ* 2006; 104(4): 416 – 25.
- Liu Y, Franklin M, Kahn R, Koutrakis P. Using aerosol optical thickness to predict ground-level PM_{2.5} concentrations in the St. Louis area: A comparison between MISR and MODIS. *Remote Sens Environ* 2007; 107(1-2): 33 – 44.
- Liu Y, Paciorek CJ, Koutrakis P. Estimating Regional Spatial and Temporal Variability of PM_{2.5} Concentrations Using Satellite Data, Meteorology, and Land Use Information. *Environ Health Perspect* 2009; 17(6): 886 – 92.
- Liu Y. New Directions: Satellite driven PM_{2.5} exposure models to support targeted particle pollution health effects research. *Atmos Environ* 2013; 68: 52 – 3.
- Lyapustin A, Williams D, Markham B, Irons J, Holben BN, Wang Y. A Method for Unbiased High-Resolution Aerosol Retrieval from Landsat. *J Atmos Sci* 2004; 61(11): 1233 – 44.
- Machta L. Meteorological benefits from atmospheric nuclear tests. *Health Phys* 2002; 82(5): 635 – 43.
- McGill R, Tukey JW, Larsen WA. Variations of Boxplots. *Am Stat* 1978; 32(1): 12 – 16.
- McHenry JN, Ryan WF, Seaman NL, Coats Jr. CJ, Padykiewicz J, Arunachalam S, Vukovich JM. A real-time Eulerian photochemical model forecast system. *Bull Amer Meteor Soc* 2004; 85: 525 – 48.
- Meloni D, di Sarra A, Biavati G, DeLuisi JJ, Monteleone F, Pace G, Piacentino S, Sferlazzo DM. Seasonal behavior of Saharan dust events at the Mediterranean island of Lampedusa in the period 1999–2005. *Atmos Environ* 2007; 41: 3041 – 56.
- Meng Q, Williams R, Pinto JP. Determinants of the associations between ambient concentrations and personal exposures to ambient PM_{2.5}, NO₂, and O₃ during DEARS. *Atmos Environ* 2012; 63: 109 – 16.

- Merrill JT, Bleck R, Avila L. Modeling atmospheric transport to the Marshall Islands. *J Geophys Res* 1985; 90(12): 927 – 36.
- Miller JM. The use of back air trajectories in interpreting atmospheric chemistry data: a review and bibliography. NOAA Technical Memorandum ERL ARL-155, 1987.
- Ming J, Cachier H, Xiao C, Qin D, Kang S, Hou S, Xu J. Black carbon record based on a shallow Himalayan ice core and its climatic implications. *Atmos Chem Phys* 2008; 8: 1343 – 52.
- North American Research Strategy for Tropospheric Ozone (NARSTO) 2003: Particulate matter science for policy makers: A NARSTO assessment. [cgenv.com/Narsto/].
- National Aeronautics and Space Administration (NASA): LAADS Server [ladsweb.nascom.nasa.gov].
- National Aeronautics and Space Administration (NASA): MODIS Web - MODIS Specifications. [modis.gsfc.nasa.gov/about/specifications.php].
- Occhipinti C, Aneja VP, Showers W, Niyogi D. Back-Trajectory Analysis and Source-Receptor Relationships: Particulate Matter and Nitrogen Isotopic Composition in Rainwater. *J Air & Waste Manage Assoc* 2008; 58: 1215 – 22.
- Orru H, Kimmel V, Kikas Ü, Soon A, Künzli N, Schins RPF, Borm PJA, Forsberg B. Elemental composition and oxidative properties of PM_{2.5} in Estonia in relation to origin of air masses - results from the ECRHS II in Tartu. *Sci Tot Environ* 2010; 408: 1515 – 22.
- Persson A, Grazzini F. User Guide to ECMWF forecast products. Meteorological Bulletin M3.2, European Centre for Medium Range Weather Forecasts, Reading, UK, 2005; 162.
- Pickering KE, Thompspon AM, McNamara DP, Schoeberl MR. An intercomparison of isentropic trajectories over the South Atlantic. *Mon Weath Rev* 1994; 122, 864 – 79.
- Pope CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J Am Med Assn* 2002; 287(9): 1132 – 41.
- Prijith SS, Aloysius M, Mohan M, Beegum N, Moorthy KK 2012. Role of circulation parameters in long range aerosol transport: Evidence from Winter-ICARB. *J Atmos Solar-Terr Phys* 2012; 77: 144 – 51.
- Remer LA, Kaufman YJ, Tanré D, Mattoo S, Chu DA, Martins JV, Li RR, Ichoku C, Levy RC, Kleidman RG, Eck TF, Vermote E, Holben BN. The MODIS aerosol algorithm, products, and validation. *J Atmos Sci* 2005; 62(4): 947 – 73.
- Revuelta MA, Sastre M, Fernández AJ, Martín L, García R, Gómez-Moreno FJ, Artúñano B, Pujadas M, Molero F. Characterization of the Eyjafjallajökull volcanic plume over the

- Iberian Peninsula by lidar remote sensing and ground-level data collection. *Atmos Environ* 2012; 48: 46 – 55.
- Reyes J, Torres F, Miss M, Corvo F, Bravo H, Bartolo-Pérez P, Azamar-Barrios JA. Effect of the environment on the degradation of Colonial and prehispanic buildings in Campeche, Mexico. *Environmental Degradation of Infrastructure and Cultural Heritage in Coastal Tropical Climate*; 2009. p. 115 – 142.
- Riddle EE, Voss PB, Stohl A, Holcomb D, Maczka D, Washburn K, Talbot RW. Trajectory model validation using newly developed altitude-controlled balloons during the International Consortium for Atmospheric Research on Transport and Transformations 2004 campaign. *J Geophys Res* 2006; 111(D23): S57.
- Roelofs GJ, Kasibhatla P, Barrie L, Bergmann D, Bridgeman C, Chin M, Christensen J, Easter R, Feichter J, Jeuken A, Kjellstrom E, Kock D, Land C, Lohmann U, Rasch P. Analysis of regional budgets of sulphur species modeled for the COSAM exercise. *Tellus* 2001; 53: 673 – 94.
- Roelofs GJ. Aerosol lifetime and climate change. *Atmos Chem Phys Discuss* 2012; 12: 16493 – 514.
- Rozwadowska A, Zielinski T, Petelski T, Sobolewski P. Cluster analysis of the impact of air back-trajectories on aerosol optical properties at Hornsund, Spitsbergen. *Atmos Chem Phys* 2010; 10: 877 – 93.
- Salomonson VV, Barnes WL, Maymon PW, Montgomery HE, Ostrow H. MODIS: advanced facility instrument for studies of the earth as a system. *IEEE Trans Geosci Remote Sens* 1989; 2: 145 – 53.
- Samson PJ. A simple methodology for the determination of back trajectories - Comment. *J Air Pollut Control Assoc* 1988; 38(10): 1290 – 91.
- Schäfer K, Harbusch A, Emeis S, Koepke P, Wiegner M. Correlation of aerosol mass near the ground with aerosol optical depth during two seasons in Munich. *Atmos Environ* 2008; 42(18): 4036 – 46.
- Schaap M, Timmermans R, Koelemeijer R, Leeuw G, Bultjes P. Evaluation of MODIS aerosol optical thickness over Europe using sun photometer observations. *Atmos Environ* 2008; 42: 2187 – 97.
- Silvey SD. Multicollinearity and Imprecise. *J Roy Statist Soc Ser B* 1969; 31(3): 539 – 52.
- Simmons AJR, Mureau R, Petroligis T. Error growth and estimates of predictability from the ECMWF forecasting system. *QJR Meteorol Soc* 1995; 121: 1739 – 71.
- Slater JF, Dibb JE, Campbell JW, Moore TS. Physical and chemical properties of surface and column aerosols at a rural New England site during MODIS overpass. *Remote Sens Environ* 2004; 92: 173 – 80.

- Stohl A, Wotawa G, Seibert P, Kromp-Kolb H. Interpolation errors in wind fields as a function of spatial and temporal resolution and their impact on different types of kinematic trajectories. *J Appl Meteor* 1995; 34(10): 2149 – 65.
- Stohl A. Computation, accuracy and applications of trajectories – A review and bibliography. *Atmos Environ* 1998; 32(6): 947 – 66.
- Stohl A, Seibert P. Accuracy of trajectories as determined from the conservation of meteorological tracers. *QJR Meteorol Soc* 1998; 124(549): 1465 – 84.
- Stunder B. An assessment of the quality of forecast trajectories. *J Appl Meteor* 1996; 35(8): 1319 – 31.
- Tandon A, Yadav S, Attri AK. City-wide sweeping a source for respirable particulate matter. *Atmos Environ* 2008; 42: 1064 – 69.
- Tandon A, Yadav S, Attri AK. Coupling between meteorological factors and ambient aerosol load. *Atmos Environ* 2010; 44(9): 1237 – 43.
- Tanré D, Kaufman YJ, Herman M, Mattoo S. Remote sensing of aerosol properties over oceans from EOS-MODIS. *J Geophys Res* 1997; 102: 16971 – 88.
- Tasić M, Rajšić S, Novaković V, Mijić Z. Atmospheric aerosols and their influence on air quality in urban areas. *Facta Univ Ser Phys Chem Techn* 2006; 4(1): 83 – 91.
- Thuillier RH, Lappe UO. Wind and temperature profile characteristics from observations on a 1400 ft tower. *J Appl Meteor*, 1964; 3: 299 – 306.
- Tian J, Chen D. A semi-empirical model for predicting hourly ground-level fine particulate matter (PM_{2.5}) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sens Environ* 2010; 114(2): 221 – 9.
- Twomey SA, Piegras M, Wolfe TL. An assessment of the impact of pollution on the global cloud albedo. *Tellus* 1984; 36B: 356 – 66.
- United States Environmental Protection Agency (EPA). Air Quality System Data Mart: AirExplorer. [epa.gov/airexplorer/].
- Vincenty T. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review* 1975; 23(176): 88 – 93
- Walmsley JL, Mailhot J. On the numerical accuracy of trajectory models for long-range transport of atmospheric pollutants. *Atmos-Ocean* 1983; 21(1): 14 – 39.
- Wang J, Christopher SA. Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies. *Geophys Res Lett* 2003; 30: 1 – 4.

- Wang L, Xin J, Wang Y, Li Z, Liu G, Li J. Evaluation of the MODIS aerosol optical depth retrieval over different ecosystems in China during EAST-AIRE. *Atmos Environ* 2007; 41(33): 7138 – 49.
- Wang L, Wang Y, Xin J, Li Z, Wang X. Assessment and comparison of three years of Terra and Aqua MODIS Aerosol Optical Depth Retrieval (C005) in Chinese terrestrial regions. *Atmos Res* 2010; 97(1-2): 229 – 40.
- Wilson JD, Flesch TK. An idealized mean wind profile for the atmospheric boundary layer. *Bound-Layer Meteor* 2004; 110: 281 – 99.
- Wu Y, Cordero L, Gross B, Moshary F, Ahmed S. Smoke plume optical properties and transport observed by a multi-wavelength lidar, sunphotometer and satellite. *Atmos Env* 2012; 63: 32 – 42.
- Xueref-Remy I, Messenger C, Filippi D, Pastel M, Nedelec P, Ramonet M, Paris JD, Ciais P. Variability and budget of CO₂ in Europe: analysis of the CAATER airborne campaigns – Part 1: Observed variability. *Atmos Chem Phys* 2011; 11: 5655 – 72.
- Yi C, Davis KJ, Berger BW, Bakwin PS. Long-term observations of the dynamics of the continental planetary boundary layer. *J Atmos Sci* 2001; 58, 1288 – 99.

CURRICULUM VITAE

Rolando Olivas Saunders
Place of birth: Mexico City, Mexico, 1988

EDUCATION

2013 PhD Atmospheric Sciences (graduation date: May, 2013)
University of Wisconsin - Milwaukee

Thesis: “Improved Estimation of PM_{2.5} Using Lagrangian Satellite-Measured Aerosol Optical Depth”

2007 MS Sustainable Processes
Universidad Autónoma de Nuevo León

Thesis: “Modeling of the Cr³⁺-biosorption process by residual *Saccharomyces cerevisiae* in packed columns”

2004 BS Chemical Engineering
Universidad Autónoma de Nuevo León

REFEREED JOURNAL PUBLICATIONS

Saunders RO, Scotty E, Kahl JD. The Sensitivity of Single Air Parcel Trajectory Calculations to Starting Elevation. *Sci Tot Environ* 2013; In review.

Kravtsov S, **Saunders RO**. A closer look at data independence: Comment on “Lies, damned lies, and statistics (in geology)”. *Eos Trans AGU* 2011; 92(8): 65.

Rodriguez-Cantú ME, Miranda-Guardiola RC, Sosa-Blanco CA, **Saunders RO**. Effects of the Operating Conditions on the Biosorption of Pb²⁺, Cd²⁺ y Cr³⁺ in Solution by Residual *Saccharomyces cerevisiae*. *Inf Tecnol* 2008; 19(6): 47 - 55.

BOOK CHAPTERS

Kahl JDW, **Saunders RO**. *Meteorological Analysis of Acid Deposition in Advances and Perspectives in Acid Deposition in Mexico*. Universidad Autonoma del Carmen. Cerón R, Griselda J, Guerra J, eds, Ciudad del Carmen, Mexico, 2012.

AWARDS

2010, 2011 Dhirenda Sikdar Memorial Scholarship in atmospheric science for excellent academic record and research accomplishments. University of Wisconsin – Milwaukee.

- 2008 National Council of Science and Technology, Mexico (CONACyT) graduate scholarship to pursue Doctoral studies.
- 2007 Master's Degree Thesis Project awarded 1st Place in the Mexican Academy for Research and Teaching in Chemical Engineering (AMIDIQ).

CONTINUED PROFESSIONAL DEVELOPMENT

- 2011 "Write Winning Grants"/"Write Successful CAREER Award Proposals to NSF"
University of Wisconsin - Milwaukee. Graduate School (12hrs).
- 2006 "Sorption and Biosorption of Heavy Metals in Solution"
Universidad Autónoma de San Luis Potosi. Bohumil B. Volesky (20hrs).
- 2006 "Meteorological Aspects of Air Pollution"
Universidad Autónoma de Nuevo León. Jonathan Kahl (50hrs).
- 2006 "BET and gas absorption methods for surface determination"
Universidad Autónoma de Nuevo León. Fernando Rojas González (25hrs).
- 2005 "Graphic methods for determination of chemical species"
Universidad Autónoma de Nuevo León. Isabel Gómez del Río (20hrs).