Study of Satellite Retrieved Aerosol Optical Depth Spatial Resolution Effect on Particulate Matter Concentration Prediction

Johan Strandgren
2014

Master of Science in Engineering Technology
Space Engineering

Luleå University of Technology
Department of Computer Science, Electrical and Space Engineering

LULEÅ UNIVERSITY OF TECHNOLOGY

Study of Satellite Retrieved Aerosol Optical Depth Spatial
Resolution Effect on Particulate Matter Concentration Prediction

Author:

JOHAN STRANDGREN

Examiner:

Dr. JANA MENDROK

Supervisors:

Dr. LINLU MEI

Dr. MARCO VOUNTAS

A thesis submitted in partial fulfillment for the degree of
Master of Space Technology

Division of Space technology

Department of Computer Science, Electrical and Space Engineering

September 29, 2014

Abstract

In this thesis, the effect of the spatial resolution of the aerosol optical depth (AOD) on the linear relationship between satellite retrieved AOD and ground level concentrations of fine particulate matter ($PM_{2.5}$) is investigated. 946 $PM_{2.5}$ monitoring stations across the US are used together with AOD at three different spatial resolutions (1 km, 3 km and 10 km) to investigate the linear relationship at different spatial scales (urban scale, meso-scale and continental scale).

The AOD data used for the study originates from the Moderate-Resolution Imaging Spectroradiometer (MODIS) and is derived by the MODIS standard algorithm and the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm. The analysis is done separately for the two sets of AOD data.

The results supports patterns in the linear relationship presented by previous researchers and provides new conclusions related to the spatial resolution effect as well as the significance of the size and characteristics of the study region. The main conclusions are: (1) the linear correlation between AOD and $PM_{2.5}$ increases with higher spatial resolution at all spatial scales, (2) the correlation decreases with enlarged study region in the eastern part of the US, but increases with enlarged study region in the western part and (3) the AOD- $PM_{2.5}$ correlation is higher and more stable over the eastern part of the US due to dissimilarities in surface characteristics, meteorological conditions and the aerosol fine mode fraction.

Acknowledgments

First of all I would like to thank Prof. John P. Burrows for giving me the opportunity to write my master thesis at the Institute of Environmental Physics at the University of Bremen. I also wish to express my gratitude to Dr. Marco Vountas for the opportunity to conduct my work within the CASPAR research group and for his help throughout my work. I would also like to express my great appreciation to Dr. Linlu Mei for his constant support, patient guidance and suggestions related to my work. I would also like to thank the additional members of the CASPAR group and staff at the Institute of Environmental Physics who have contributed to my work.

I would like to express my appreciation to Dr. Johnny Ejemalm, Mrs. Maria Winnebäck, Dr. Anette Ladstätter-Weissenmayer, Mr. Lars Jeschke, Mrs. Anja Gatzka and the international offices at Luleå University of Technology and the University of Bremen for all help and assistance related to my exchange year at the University of Bremen. I would also like to thank the Erasmus programme for the financial support during my stay in Bremen.

My special thanks are extended to my girlfriend for her support and our time together in Bremen as well as to my family for their constant support.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Previous Work	1
1.3	Purpose and Goal with the Thesis	3
1.4	Used Software	4
1.5	Outline of the Thesis	5
2	Theory and Background Information	6
2.1	Aerosols and Particulate Matter	6
2.1.1	Size Categorization	6
2.1.2	Sources	7
2.1.3	Lifetime and Sinks	7
2.2	Effects of Aerosols/Particulate Matter	8
2.3	Basic Related Radiative Transfer	10
2.3.1	Retrieval of AOD From Space	12
2.4	Typical AOD Values and the AOD Distribution	14
2.5	Use AOD as a Proxy for PM _{2.5} Concentration Prediction	14
3	Material and Study Region	18
3.1	Data Sets	18
3.1.1	Satellite Aerosol data	18
3.1.2	Ground Level PM _{2.5} Data	25
3.1.3	Modeled Meteorological Data	25
3.2	Regions and Time of Study	25
4	Method	28
4.1	Data Structure and Processing	28
4.1.1	Structure and Processing of MAIAC L3 Aerosol Data	28
4.1.2	Structure and Processing of MODIS L2 Aerosol Data	28
4.1.3	Structure and Processing of Ground Level PM _{2.5} Data	29
4.1.4	Structure and Processing of Modeled Meteorological Data	29
4.2	Matching Satellite AOD and Ground Level PM _{2.5} Data	29
4.3	Removal of Outliers	30
4.4	Corrections of Satellite AOD Data	31
4.5	Evaluation of the Linear Correlation Between AOD and PM _{2.5}	31
5	Results	34
5.1	MAIAC AOD Data	34
5.1.1	Study I - Region of Continental Spatial Scale	34
5.1.2	Study II - Regions of Varying Spatial Scale	37
5.2	MYD04 AOD Data	43
5.2.1	Study I - Region of Continental Spatial Scale	43
5.2.2	Study II - Regions of Varying Spatial Scale	44
5.3	Comparison of MAIAC AOD and MYD04 AOD Data	50
5.4	Results Using RH and BLH Corrections	51

6	Conclusions	52
6.1	Possible Improvements for Further Work	52
Appendix A	Scatter Plots Excluded From Chapter 5	58
A.1	MAIAC Aerosol Data	59
A.1.1	Study II - Regions of Varying Scale	59
A.2	MYD04 Aerosol Data	65
A.2.1	Study II - Regions of Varying Scale	65
Appendix B	Matlab Scripts	71
B.1	Grid AOD Data	72
B.2	Read PM _{2.5} Data	77
B.3	Match Gridded AOD Data and PM _{2.5} Data	79
B.4	Analyze Data for Study I	83
B.5	Analyze Data for Study II	88

List of Figures

1.1	Illustration of the theoretical advantage of AOD at a high spatial resolution for $PM_{2.5}$ prediction. Left: 9 measured $PM_{2.5}$ concentrations at ground level. Center: An ideal AOD retrieval from a satellite over the same region using 1 km spatial resolution. Right: The AOD retrieval over the same region using a coarser (3 km) resolution.	4
2.1	Size comparison between two aerosols with diameters 2.5 and 10 μm , a human hair and a sand grain (credit: Environmental Protection Agency).	7
2.2	Illustration of an aerosol acting as a CCN to form a cloud droplet from condensing water vapor.	8
2.3	Illustration of how the cloud properties changes with increased aerosol (CCN) concentration for the same amount of water.	9
2.4	Illustration of the differences between propagation path and vertical (normal) path.	11
2.5	Simple illustration of a passive remote sensing setup (based on Figure 2.1 from Kokhanovsky and de Leeuw (2009)).	12
2.6	Contributing factors to the TOA reflectance observed by the satellite.	13
2.7	The AOD over US averaged for January, April, July and October in 2008	15
2.8	AOD (left) and $PM_{2.5}$ concentrations (right) in the US averaged for January, April, July and October in 2008.	15
2.9	Impact of FMF and coarse particles on the AOD- $PM_{2.5}$ correlation and regression slope. Left: low FMF with mostly fine particles; yields good AOD- $PM_{2.5}$ correlation. Center: higher FMF with some coarse particles leading to too high AOD values compared to the measured $PM_{2.5}$ concentrations, yields lower correlation and gentler slope. Right: even higher FMF with more coarse particles; yields even lower correlation and even gentler slope.	16
3.1	The scanning principle of an across track scanner.	19
3.2	The bow tie effect present for the MODIS instrument (credit: Deutsches Zentrum für Luft- und Raumfahrt).	20
3.3	Simplified flowchart for the DT-land algorithm.	22
3.4	Illustration of the time series of AOD data used in the MAIAC algorithm. The larger parallelograms represent three blocks of AOD data holding 25×25 AOD pixels/grid cells. The smaller parallelograms represent the first AOD pixel (1,1) and an arbitrary AOD pixel (i,j). Figure inspiration from Lyapustin and Wang (2008).	23
3.5	Simplified flowchart for the MAIAC algorithm.	24
3.6	AOD on January 1 st 2008 from MYD04 (left) and MAIAC (right).	25
3.7	The EPA $PM_{2.5}$ monitoring stations and the different study regions.	26
4.1	Visualization of the areas where AOD- $PM_{2.5}$ match-ups are considered to be outliers.	30
5.1	Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in the US in January 2008 at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.	34
5.2	Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in the US in April 2008 at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.	35
5.3	Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in the US in July 2008 at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.	36

5.4	Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in the US in October 2008 at 10 km (left), 3 km (center) and 1 km (right) spatial resolution. . .	36
5.5	Relative humidity at surface level (1000 hPa) in January (left) and July (right) 2008	37
5.6	Boundary layer height in January (left) and July (right) 2008	37
5.7	AOD over Indianapolis (left) and Atlanta (right) at urban scale (visualization of the AOD spatial variability). The red stars show the respective city centers and the black dots the $PM_{2.5}$ monitoring stations included in the urban scale analysis.	39
5.8	The aerosol fine mode fraction over the US for the four months in 2008. The white stars shows the location of the six cities.	40
5.9	Fine mode fraction over Salt Lake City (left) and Atlanta (right) at urban scale. The red stars show the respective city centers and the black dots the $PM_{2.5}$ monitoring stations included in the urban scale analysis.	42
5.10	Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in the US in January 2008 at 10 km (left) and 3 km (right) spatial resolution.	44
5.11	Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in the US in April 2008 at 10 km (left) and 3 km (right) spatial resolution.	44
5.12	Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in the US in July 2008 at 10 km (left) and 3 km (right) spatial resolution.	45
5.13	Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in the US in October 2008 at 10 km (left) and 3 km (right) spatial resolution.	45
5.14	Scatter plot with MYD04 AOD at 3 km resolution and $PM_{2.5}$. Visualization of the impact of an outlier on the linear correlation.	47
5.15	AOD over Los Angeles at 3 km resolution (visualization of the high spatial AOD variability). The red star shows the city center and the black dots the $PM_{2.5}$ monitoring stations included in the urban scale analysis.	49
5.16	Fine mode fraction over Los Angeles at 10 km (left) and 3 km (right) spatial resolution. The red stars show the respective city centers and the black dots the $PM_{2.5}$ monitoring stations included in the urban scale analysis.	50

List of Tables

2.1	Aerosol size categorization	6
2.2	The air quality condition for different PM _{2.5} concentrations levels	10
2.3	AOD levels and the corresponding atmospheric conditions (credit: Levy et al. (2014))	14
3.1	Aqua's orbital characteristics	18
3.2	The bands of MODIS used for aerosol retrieval	19
5.1	Statistical data related to the linear relationship between MAIAC AODs and PM _{2.5} concentrations for the whole contiguous US	35
5.2	Statistical data related to the linear relationship between MAIAC AODs and PM _{2.5} concentrations in the eastern part of the US 2008 for different spatial scales	38
5.3	Statistical data related to the linear relationship between MAIAC AODs and PM _{2.5} concentrations in the western part of the US 2008 for different spatial scales . . .	41
5.4	Statistical data related to the linear relationship between MYD04 AODs and PM _{2.5} concentrations for the whole US	43
5.5	Statistical data related to the linear relationship between MYD04 AODs and PM _{2.5} concentrations in the eastern part of the US 2008 for different spatial scales	46
5.6	Statistical data related to the linear relationship between MAIAC AODs and PM _{2.5} concentrations in the western part of the US 2008 for different spatial scales . . .	48

List of Acronyms

AOD	Aerosol Optical Depth
BLH	Boundary Layer Height
BRF	Bi-directional Reflectance Factor
C005	Collection 5
C006	Collection 6
DT	Dark Target
ECMWF	European Centre for Medium-Range Weather Forecasts
EEA	European Environmental Agency
EPA	Environmental Protection Agency
FEM	Federal Equivalent Method
FRM	Federal Reference Method
GMT	Generic Mapping tools
L0	Level 0
L1A	Level 1A
L1B	Level 1B
L2	Level 2
L3	Level 3
LUT	Look-Up Table
MAIAC	Multi-Angle Implementation of Atmospheric Correction
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	MODerate-resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
PM	Particulate Matter
QAC	Quality Assurance Confidence
ROD	Rayleigh Optical Depth
RH	Relative Humidity
RMSE	Root Mean Square Error
TOA	Top Of Atmosphere
WHO	World Health Organization

List of Symbols

ρ	reflectance	[unitless]
T	transmittance	[unitless]
S	spherical albedo	[unitless]
θ_0	solar zenith angle	[degrees]
θ_s	satellite view zenith angle	[degrees]
φ	relative azimuth angle	[degrees]
λ	wavelength	[μm]
L	radiance	[$\text{W m}^{-2} \text{sr}^{-1}$]
E	irradiance	[W m^{-2}]
α^S	scattering coefficient	[m^{-1}]
α^a	absorption coefficient	[m^{-1}]
ϵ	extinction coefficient	[m^{-1}]

Chapter 1 Introduction

1.1 Motivation

Particulate matter (PM) is microscopic particles suspended in the air. Despite their small size an excess of PM in the ambient air can have many negative side effects, for example on the climate and the human health. The PM can be divided into different size categories and in this thesis the focus will be on those particles with a diameter less than $2.5\text{ }\mu\text{m}$ known as fine particles and referred to as $\text{PM}_{2.5}$. Understanding the composition and distribution as well as the sources and sinks of particulate matter is an essential part of present research related to atmospheric physics, climate physics and the human health ([McMurry et al., 2004](#)).

Currently the $\text{PM}_{2.5}$ concentrations at ground level can be accurately measured in-situ using instruments located at the Earth's surface. Those instruments can only perform point measurements which leads to poor spatial coverage with very limited information especially in unpopulated regions. Using satellite retrieved aerosol optical depth (AOD) for $\text{PM}_{2.5}$ prediction is challenging but still an important field of research since a satellite has the ability to provide both global and regional information about the AOD.

Increasing the spatial resolution of the AOD is a balance, on one hand a higher resolution can represent the spatial variability of the $\text{PM}_{2.5}$ concentrations better, but on the other hand a high resolution is more vulnerable to errors that are well mitigated using a coarser resolution. To understand this balance is important in order to know to which extent it is worth to increase the spatial resolution i.e. when the errors due to the increase in resolution are small but the resolution still high enough to observe the spatial variability of $\text{PM}_{2.5}$. At present, this balance is not well understood and this thesis aims to improve this understanding. The correlation cannot be expected to improve by just increasing the spatial resolution and it is important to understand the impact of the study region. A continental scale study may represent different and varying surface characteristics, meteorological and aerosol conditions compared to an urban scale study where the characteristics and conditions can be expected to be more homogeneous. To understand how the correlation varies with size and geographical position of the study region is also a key part of this research in order to improve $\text{PM}_{2.5}$ modeling using satellite AOD data.

1.2 Previous Work

The AOD refers to the scattering and absorption of solar radiation in the atmosphere and the $\text{PM}_{2.5}$ concentration refers to the mass concentrations of particles in the ambient air. So both quantities are related to the amount of particles in the atmosphere which suggests that the two quantities should be related. In this section some results from previous work related to this relationship will be presented. Without previous knowledge about this field of research it might be helpful to read Section 2.5 first, where the idea of using AOD for $\text{PM}_{2.5}$ prediction is explained.

Several results related to the relationship between satellite retrieved AOD and ground level $\text{PM}_{2.5}$ concentrations have been published for different time periods, regions and satellite sensors. In the majority of the publications a linear relationship between the two quantities is evaluated. Even though the relationship is not perfectly linear it is assumed to be a sufficiently good proxy in relation to its simplicity. Other relationships that have been used are mix-effect

regression where extra meteorological parameters have been included in the relationship and physical-based methods where the PM was assumed to be a function of the AOD and the aerosol fine mode fraction (FMF). In this section the previous work related to the linear relationship between AOD and PM_{2.5} will be discussed since this is the approach used in this thesis. For the same reason only previous results of the correlation obtained for the US are presented in this section.

Engel-Cox et al. (2004) studied the linear correlation between AOD from MODIS and the daily PM_{2.5} concentrations in the contiguous US for the summer months in 2002. They presented a correlation coefficient equal to 0.43 with generally better correlation in the eastern part of the US. The correlation coefficients for some individual PM_{2.5} monitoring stations were presented as well with generally higher correlation for the individual stations compared to the whole US. This pattern did not apply for the stations located in the western part of the US where the correlation was very low with values of the correlation coefficient between -0.221 and 0.13).

Toth et al. (2014) studied the same AOD-PM_{2.5} correlation in the contiguous US using MODIS AOD data as well, but for a time period of two years (2008-2009). In this study a significantly lower correlation was obtained (0.19) compared to the results by Engel-Cox et al. (2004). The US was also divided in four parts based on the time zones and the correlation was also analyzed by season (Dec-May and Jun-Nov). Clear indications were presented that the correlation was higher in the eastern part of the US compared to the western part as well as during summer and fall and compared to winter and spring. The same analysis was done using AOD from the Multi-angle Imaging SpectroRadiometer (MISR) as well, with similar results as for the MODIS AOD data. It was also shown that the aerosols in general were located closer to the surface level in the eastern part compared to the western part which is one possible explanation for the better correlation in the eastern part of the US.

Zhang et al. (2009) did a similar study using MODIS for the time period 2005-2006, but with hourly PM_{2.5} data and with the contiguous US divided in ten regions. Regarding the seasonal and geographical differences in the correlation the same results as before were obtained with higher correlation for the regions in the eastern part of the US as well as for the months during summer and fall. By studying different seasons as well as regions Zhang et al. (2009) could also show that the seasonal pattern changed slightly between the two coasts of the US. In the eastern part the correlation was generally lowest during winter and second lowest during spring, while for the western part of the US the correlation was lowest during spring and not as bad during winter time.

Wang and Christopher (2003) studied the AOD-PM_{2.5} correlation using MODIS AOD and daily PM_{2.5} for a smaller study region using 7 PM_{2.5} monitoring stations in Alabama for 2002 and presented a correlation coefficient 0.70. Engel-Cox et al. (2006) studied the correlation for only a few PM_{2.5} monitoring stations as well, but in Baltimore, Maryland, for the summer of 2004. In this study a correlation coefficient equal to 0.57 was obtained between MODIS AOD and the daily PM_{2.5} concentrations.

Chudnovsky et al. (2013a,b) did recently investigate the AOD's spatial resolution effect on the linear AOD-PM_{2.5} relationship for a limited region in the US. The hypothesis was that AOD at higher spatial resolution would have better capabilities to distinguish the spatial variability of the air quality and hence improve the AOD-PM_{2.5} correlation. In the first study (2013a) Massachusetts and Connecticut with extra focus on Boston was analyzed for one year (2003) and in the second study (2013b) a multiyear analysis (2002-2008) was done for New England (divided into three sub-regions). The first results presents clear indications that the correlation between AOD and PM_{2.5} improves with higher spatial resolution for the selected study regions when AOD data from the Multi-Angle Implementation of Atmospheric Correction for MODIS (MAIAC) was analyzed at four spatial resolutions. However the second results gave no clear indication that a higher resolution yields higher AOD-PM_{2.5} correlation and similar correlation coefficients were obtained for both the fine resolution (MAIAC 1 km) and the coarse resolution (MODIS standard product 10 km) for all three sub-regions within New England. The fact that AOD from two different algorithms were used in the second study could be one reason why the results did not

indicate a higher correlation using a higher resolution.

From the previous research within this field it is clear that the linear correlation between AOD and $PM_{2.5}$ varies for different studies. For example did [Engel-Cox et al. \(2004\)](#) obtain a considerably higher correlation than [Toth et al. \(2014\)](#) even though the same region was studied. This can probably be explained by the fact that [Toth et al. \(2014\)](#) studied two complete years and included all seasons while [Engel-Cox et al. \(2004\)](#) only studied one summer. Since clear evidence have been presented from several publications that the correlation is higher during the summer compared to winter and spring the differences between the two studies are not as surprising. For the two studies by [Wang and Christopher \(2003\)](#) and [Engel-Cox et al. \(2006\)](#) the same argumentation does however not hold since a higher correlation was obtained by [Wang and Christopher \(2003\)](#) even though a whole year was studied compared to the summer study by [Engel-Cox et al. \(2006\)](#). This indicates that temporal as well as regional variations most likely plays a role in the AOD- $PM_{2.5}$ correlation as well, where the latter has been proven by e.g. [Toth et al. \(2014\)](#) and [Zhang et al. \(2009\)](#) as shown above.

For more details and results please see the respective publications or the critical review done by [Hoff and Christopher \(2009\)](#), where the research related to the relationship between AOD and $PM_{2.5}$ before 2008 is summarized and reviewed.

1.3 Purpose and Goal with the Thesis

This thesis aims to further investigate the linear relationship between satellite retrieved AOD and ground level $PM_{2.5}$ concentrations with emphasis on the effect of the AODs spatial resolution. The topic is poorly understood with limited publications like [Chudnovsky et al. \(2013a,b\)](#). In this thesis AOD from both MAIAC and the MODIS standard product will be used, from now one referred to as MAIAC AOD and MYD04 AOD respectively. MYD04 provides AOD at 10 km and 3 km resolution and MAIAC provides AOD at 1 km which in this thesis has been aggregated to the additional spatial resolutions 10 km and 3 km in order to be able to (1) evaluate the spatial resolution effect separately for the MYD04 AOD and the MAIAC AOD and (2) compare the correlations obtained from the two data sets.

Since a ground based $PM_{2.5}$ monitor only provides point measurements it is reasonable to believe that using AOD with a high spatial resolution would increase the capability to represent and predict the $PM_{2.5}$ conditions at ground level given that the retrieval conditions and AOD data quality is equal for the different resolutions. The point measurement by a $PM_{2.5}$ monitoring instrument is only representative for the air near the instrument, so if AOD at a very coarse resolution is used there is no chance that the AOD can represent the conditions near the individual instrument if the spatial variability of the $PM_{2.5}$ concentrations is high. Figure 1.1 clarifies this theoretical hypothesis. If nine $PM_{2.5}$ monitoring stations are positioned in a 3×3 km region it is clear that AOD at 1 km resolution would predict the $PM_{2.5}$ concentrations better than AOD at 3 km resolution which would not be able to distinguish any variability and would only provide a mean value for the whole scene, given that the observation conditions and AOD data quality is the same for the two spatial resolutions. This phenomena was clearly visualized at urban scale by [Chudnovsky et al. \(2013a\)](#) where four $PM_{2.5}$ monitoring stations were located inside a 10×10 km area. For this scene the $PM_{2.5}$ representation was significantly better for AOD at 1 km compared to AOD at 10 km since the AOD at this high resolution had the ability to observe the variability within the 10×10 km region and between the four stations.

Even though a high resolution seems more superior theoretically the risk for introducing errors increases if the spatial resolution is increased. Error sources like cloud screening from nearby pixels, increase of the pixel footprint area with larger view angle etc. are well mitigated using a coarse resolution, but may reduce the AOD quality using a high resolution given that the observation conditions (atmospheric condition, surface characteristics etc.) are the same for both spatial resolutions. If the extra error introduced at a high resolution is too high, the AOD quality may become low enough to overcome the theoretical advantage of the higher spatial resolution,

Figure 1.1: Illustration of the theoretical advantage of AOD at a high spatial resolution for $PM_{2.5}$ prediction. Left: 9 measured $PM_{2.5}$ concentrations at ground level. Center: An ideal AOD retrieval from a satellite over the same region using 1 km spatial resolution. Right: The AOD retrieval over the same region using a coarser (3 km) resolution.

leading to worse correlation with the $PM_{2.5}$ compared to a coarser spatial resolution. The main goal with this thesis is to investigate this balance by analyzing the linear AOD- $PM_{2.5}$ correlation using two AOD data sets at three different spatial resolutions in total.

Additionally the effect of the study region will be investigated. An extensive study with several regions with varying spatial scales has not been published and different publications have to be compared in order to evaluate the impact of the study region on the AOD- $PM_{2.5}$ correlation. The correlation is believed to change with varying area of the study region due to increased spatial variability of several factors. If the study region is small, for example one single city, the conditions such as surface properties, boundary layer height, relative humidity etc. should be relatively homogeneous, which will result in a correlation representative for the conditions within that particular study region. When the study area is gradually increased the conditions will get less homogeneous and more variability will be observed. This will probably lead to changes in the correlation since various conditions is observed for the larger study region. By comparing results from previous research there are indications that smaller study regions generally provides higher AOD- $PM_{2.5}$ correlation compared to larger study regions like the whole US. If this is a solid conclusion will be investigated in this thesis by studying the correlation and the spatial resolution effect for three spatial scales across the US.

The analysis is divided in two parts, in *Study I - Region of Continental Spatial Scale* the spatial resolution effect is analyzed for the whole contiguous US for the four seasons separately and in *Study II - Regions of Varying Spatial Scale* the spatial resolution effect is analyzed for several study regions with varying spatial scales, but with all seasons combined.

1.4 Used Software

Three software packages have been used throughout the work with this thesis

- **MATLAB** has been used to handle and analyze all satellite aerosol data, $PM_{2.5}$ data and meteorological data used for this study and to visualize the linear relationships.
- **Generic Mapping Tools (GMT)** has been used to grid the AOD data and to visualize geolocated data and information.
- **Inkscape** has been used to create some of the images/illustrations showed in this thesis.

All MATLAB scripts are available in Appendix B, but their faultless functioning cannot be guaranteed on other computers than the authors. If not stated otherwise the figures in this thesis have been made by the author.

1.5 Outline of the Thesis

In chapter 1 the motivation, previous research related to this topic and the purpose and goal of the thesis is explained and the major reasons why this is a relevant field of research are explained.

Chapter 2 aims to provide the reader with sufficient background information in order to understand the topic of this thesis and follow the subsequent chapters. However some pre-existing background knowledge related to atmospheric and climate science is necessary. Aerosols and particulate matter will be introduced, with information about their sources, sinks and lifetimes. The effects of aerosols and particulate matter will be explained, related to the climate as well as the human health. A brief introduction to radiative transfer will be given and the retrieval of AOD from space explained. Finally the use of AOD as a proxy for $PM_{2.5}$ prediction will be explained.

In Chapter 3 the AOD, $PM_{2.5}$ and meteorological data sets used for the study will be introduced. For this part the emphasis will be on the satellite retrieved data. Information will be given about the MODIS instrument in particular and the satellite on which it is flying in general. Additionally the aerosol retrieval algorithms used to derive the AOD data used in this thesis will be explained. Finally the regions and time period that are analyzed in this thesis will be introduced and motivated.

In Chapter 4 the method and work related to this thesis will be described. The steps between the raw data and the final results will be explained. The reader will for example understand how the different data sets have been processed, how the AOD and $PM_{2.5}$ data have been matched and how the linear relationships between the two have been determined.

In Chapter 5 the results of the thesis are presented and discussed. Followed by the conclusions of the study and suggestions of possible improvements for further work related to this topic presented in Chapter 6.

Chapter 2 Theory and Background Information

This second chapter aims to give the background information relevant for the topic of this thesis. As a first step some basic knowledge of aerosols and particulate matter will be provided in order to introduce readers that are not familiar with with field of research. Details will initially be avoided as far as possible, but with the readers progress more details will be provided for full preparation for the upcoming chapters of this thesis.

2.1 Aerosols and Particulate Matter

Aerosols and PM are two terms that refers to liquid and solid particles suspended in the atmosphere except cloud droplets. The term PM refers to the suspended particle alone while the term aerosol includes the suspending gas as well. In practice the term aerosol usually refers to the suspended matter as well ([Burrows et al., 2011](#)), so throughout this thesis both terms will be used referring to the same type of particles in the atmosphere.

In order to be classified as an aerosol the particle has to be small enough not to sediment down from the atmosphere immediately. This can only be achieved by extremely small particles that are affected by such a weak gravitational pull from the Earth that it can be balanced by winds and turbulent flow, forcing the aerosol to follow the movement of the air. An exact size range for aerosols is not set, but particles with a diameter from a few nanometers (nm) to hundreds of micrometers (μm) are generally considered to be aerosols ([Wallace and Hobbs, 2006](#)).

2.1.1 Size Categorization

Particulate matter can be divided into three size categories; the ultrafine mode, the accumulation mode and the coarse mode. Aerosols with a diameter less than $0.1 \mu\text{m}$ belongs to the ultrafine mode, particles larger than $0.1 \mu\text{m}$ but smaller than $2.5 \mu\text{m}$ belongs to the accumulation mode and particles larger than $2.5 \mu\text{m}$ are considered to be coarse. Particles smaller than $2.5 \mu\text{m}$ i.e. both particles in the ultrafine and accumulation mode can also be combined to form a fourth category called the fine mode ([McMurry et al., 2004](#)). Table 2.1 clarifies this size categorization concept.

Table 2.1: Aerosol size categorization

Fine Mode $\leq 2.5 \mu\text{m}$		Coarse Mode $2.5\text{-}10 \mu\text{m}$
Ultrafine Mode $< 0.1 \mu\text{m}$	Accumulation Mode $0.1\text{-}2.5 \mu\text{m}$	

When monitoring the air quality, three different quantities based on the aerosol size are used: $\text{PM}_{2.5}$, PM_{10} and $\text{PM}_{2.5\text{-}10}$. $\text{PM}_{2.5}$ measures the total mass concentration of dry particles that are smaller than $2.5 \mu\text{m}$ i.e. the total mass concentration of fine particles. Similarly, $\text{PM}_{2.5\text{-}10}$ measures the total mass of all dry particles smaller than $10 \mu\text{m}$ but larger than $2.5 \mu\text{m}$, i.e. the

total mass of coarse particles. PM_{10} measures the total dry mass of both coarse and fine particles i.e. the total mass of all particles smaller than $10\text{ }\mu\text{m}$. To get a better size perspective and to understand how tiny those particles are a size comparison is given in Figure 2.1.

Figure 2.1: Size comparison between two aerosols with diameters 2.5 and $10\text{ }\mu\text{m}$, a human hair and a sand grain (credit: Environmental Protection Agency).

2.1.2 Sources

Aerosols in the atmosphere originate either from natural or anthropogenic sources. **Natural** sources release aerosols into the atmosphere independent on human activity and are responsible for $\sim 90\%$ of the total aerosol mass in the atmosphere. Examples of aerosols that originates from natural sources are pollen, smoke from forest fires, ash from volcanic eruptions, desert dust from sand storms and sea salt from wind driven sea spray. **Anthropogenic** sources involves human activity and are responsible for the remaining $\sim 10\%$ of the atmospheric aerosol mass. Examples of anthropogenic sources are emissions from fossil fuel combustion, biomass burning and agricultural activities (NASA, 2014b).

Aerosols are divided into two categories based on their formation, primary and secondary aerosols. **Primary** aerosols are particles that directly enters the atmosphere as aerosols, without additional processes. Examples of primary aerosols are pollen, desert dust and sea salt particles. **Secondary** aerosols are particles that are formed in the atmosphere by oxidation and/or reactions of so called precursor gases. Two examples are the precursor gases NH_3 and SO_2 which reacts in the atmosphere creating ammonia and sulfate compounds which condense to the liquid state and thus forming liquid aerosols. The formation of secondary aerosols depends on the physical and chemical state of the atmosphere like the concentration of precursor gases, free radicals and/or other reactive species. A higher amount of reactive species in the atmosphere increases the overall reactivity and hence the possibility for secondary aerosol formation (EEA, 2013).

2.1.3 Lifetime and Sinks

The two major removal processes of aerosols are wet and dry deposition. **Wet** deposition is the process when an aerosol is brought to the surface by interaction with water, for example when aerosols are captured by falling raindrops. **Dry** deposition is the process when an aerosol is brought to the surface without any interaction with water, for example when a particle grows to a mass so large that the gravitational pull overcomes the atmospheric motion, forcing the aerosol to sediment down to the surface. Wet deposition is the main removal process and accounts for around 80-90% of the aerosol mass removed from the atmosphere (Wallace and Hobbs, 2006).

Two additional removal processes are coagulation of aerosols and the process when an aerosol acts as a cloud condensation nuclei (CCN). Coagulation is the process when two aerosol particles

collide and merge. This means that unlike from the deposition processes, the total aerosol mass stays constant, however the total number of aerosols is reduced which for example can lead to improved visibility in a heavily polluted region. In order to form cloud droplets the water vapor in the atmosphere have to condensate on a non-gaseous surface, namely a CCN. When an aerosol acts as a CCN, water vapor condenses on its surface forming a cloud droplet. Similarly as for the coagulation process an aerosol acting as a CCN is not removed from the atmosphere, but as it is embedded in a cloud droplet it can no longer be referred to as an aerosol. Figure 2.2 illustrates this process, please note that in order to get a better visualization the scale is far from correct, in reality the diameter of a CCN is around $0.2\ \mu\text{m}$, a cloud droplet around $20\ \mu\text{m}$ and water vapor at molecule size level ($\sim 0.2\ \text{nm}$).

Figure 2.2: Illustration of an aerosol acting as a CCN to form a cloud droplet from condensing water vapor.

The lifetime of an aerosol depends on its size. Small particles in the ultrafine mode will grow fast by coagulation to the accumulation mode. This together with the deposition processes results in a lifetime of minutes to hours. The coarse mode particles have a similar lifetime. This is due to the rapid sedimentation (dry deposition) of such comparably large and heavy particles. Aerosols in the accumulation mode do generally stay in atmosphere for days to months. This is because the particles are large enough to have a very slow or negligible growth rate by coagulation and small enough not to be rapidly removed by dry deposition (McMurry et al., 2004). With such a long atmospheric lifetime the aerosols can be transported far distances and depending on the atmospheric state the aerosol transport can be from intercontinental to global scale. For example can desert dust from the Sahara desert be transported to North America and in 2010 the eruption of the Eyjafjallajökull volcano on Iceland shut down the air traffic in large parts of Europe as a result of the aerosols released by the volcano.

2.2 Effects of Aerosols/Particulate Matter

Aerosols can affect climate both directly and indirectly. A **direct effect** by aerosols is when the particles interacts with the solar radiation traveling through the atmosphere or changes the surface reflectance of the Earth. If an aerosol reflects the incoming solar radiation it means that less radiation reaches the surface leading to a cooling effect. If the radiation is absorbed the aerosol layer will be heated and on a longer time scale where atmospheric circulation and mixing is included the warming effect will extend to the surface level as well (IPCC, 2013). The scattering/absorption properties of an aerosol are determined by the physical and chemical properties of the aerosol. For example are sulfates and salt particles highly reflective and reflects nearly all incident radiation whereas soot (black carbon) is highly absorbing (NASA, 2014b).

In 1991 after the volcanic eruption of Mount Pinatubo a huge amount of SO_2 was ejected into the troposphere and stratosphere where it reacted and formed small sulfuric acid droplets ($< 1.0\ \mu\text{m}$). Those droplets/aerosols located in the stratosphere above all precipitating clouds were not subjected to wet deposition which allowed the sulfate aerosols to remain in the stratosphere for several years. This high amount of reflective aerosols present in the atmosphere for such a long time period resulted in an observed net cooling of 0.5°C globally and 0.7°C in the northern hemisphere (Dutton and Christ, 1992).

Another direct climate effects induced by aerosols is when high amounts of aerosols deposits on a surface and creates a surface layer that changes the reflectance characteristics leading to a change in the amount of reflected solar radiation. One example, which is a problem in the arctic region, is when soot is transported northward and forms the so called Arctic Haze (Mitchell, 1957). When those soot particles deposits on the sea ice the surface color changes from very bright to dark leading to more absorption of solar radiation leading to heating and melting of the sea ice (Clarke and Noone, 1985).

An **indirect effect** by aerosols on the climate is when the aerosols act as CCNs and contributes to the cloud droplet formation as illustrated in Figure 2.2. More aerosols in the air means more CCNs which lead to more and smaller cloud droplets compared to a case with low level of pollution and few aerosols. A cloud that contains more and smaller droplets is more reflective and will appear as brighter than a cloud with few and larger droplets. This means that a more polluted region tends to be locally cooled due to increased cloud reflectance (Twomey, 1977). Figure 2.3 illustrates how more aerosols acting as CCNs can lead to more and smaller cloud droplets which in turn leads to higher cloud reflectivity for the same amount of water vapor.

Figure 2.3: Illustration of how the cloud properties changes with increased aerosol (CCN) concentration for the same amount of water.

The overall impact by aerosols on the Earth's climate and energy budget has been shown to be a cooling effect, both for the direct and indirect effects (IPCC, 2013).

Several epidemiological studies have shown that both short and long term exposure to particulate matter increases the risk for morbidity and mortality. The problem is global and both developing and developed countries suffers from PM induced health problems (Pope et al., 2002; WHO, 2013, 2006b). Even though both PM_{10} and $PM_{2.5}$ can induce health problems $PM_{2.5}$ is considered to be more harmful because of its small size. At inhalation the small particles can penetrate deep into the respiratory system as well as into the bloodstream, where it can interact with the body tissue both physically and chemically. The most well-known and generally accepted effects from PM exposure is disorders in the respiratory, cardiovascular and immune systems which can lead to premature death, cardiovascular diseases as well as respiratory disabilities such as coughing, asthma, painful breathing, chronic bronchitis and cancer (EEA, 2013; EPA, 2014).

(Lim, 2012) showed that in 2010, 3.1 million deaths worldwide and 3.1% of the total disability-

adjusted life years¹ could be attributed to the exposure to ambient PM_{2.5}. In the European Union the reduced lifetime due to PM_{2.5} exposure was estimated to 7.7 months in 2010 (3.4 in Sweden and 7.5 in Germany) (WHO, 2006a).

There is no known concentration threshold where PM_{2.5} no longer is harmful and even very low concentrations of PM_{2.5} may induce health problems, especially for the most vulnerable target groups (WHO, 2006b). To keep the concentrations of particulate matter in control and minimize major health problems it has become more common that national or regional PM standards are set. In US the 24-hour maximum concentrations of PM_{2.5} is 35 µg/m³. At a particular PM_{2.5} monitoring station the concentration is allowed to occasionally be higher than 35 µg/m³, but averaged over three years the daily mean concentrations must not exceed the limit (EPA, 2013). Additionally the world health organization (WHO) has set their own standard for the maximum PM_{2.5} daily mean concentration to 25 µg/m³. This is not a level that must be met by any countries or regions, but a recommended limit from the WHO.

Table 2.2 shows the air quality levels for different levels of PM_{2.5} concentrations (EPA, 2012). The categorizations in the right column are known as the air quality index levels used to communicate the air quality conditions in the US.

Table 2.2: The air quality condition for different PM_{2.5} concentrations levels

Daily mean concentration [µg/m ³]	Air quality level
0.0 - 12.0	Good
12.1 - 35.4	Moderate
35.5 - 55.4	Unhealthy for Sensitive Groups
55.5 - 150.4	Unhealthy
150.5 - 250.4	Very Unhealthy
250.5 - 500	Hazardous

2.3 Basic Related Radiative Transfer

Atmospheric radiative transfer describes the propagation of electromagnetic radiation through the atmosphere. When solar radiation propagates through the atmosphere it will most likely interact with the atmospheric species by scattering and absorption which leads to attenuation of the radiation. **Scattering** is the process when photons of the incoming solar radiation are extracted and scattered into one or several directions and thus changing the propagation path of the radiation beam. The angular distribution of the scattered solar radiation depends on the wavelength of the incident radiation, the size of the species and its refractive index. Scattering is divided into two main types; Rayleigh scattering and Mie scattering. Rayleigh scattering describes the scattering by tiny particles like atoms and molecules and Mie scattering describes the scattering by larger particles whose size is in the size range of the wavelength of the incident radiation. **Absorption** does usually happen in conjunction with scattering and is the phenomena when photons are absorbed rather than redirected and converted into other forms of energy (Mishchenko et al., 2002). The absorption is dominated by the atmospheric gases and wavelengths where the gas absorption is weak are used for aerosol retrieval in order to minimize the light absorption effect.

The attenuation of the solar radiation with a spectral radiance, L_λ , due to absorption and scattering along a path ds is given by

$$\frac{dL_\lambda^a}{ds} = -\alpha_\lambda^a L_\lambda \quad (2.1a)$$

¹a measure of how much lifetime that is lost due to illness, disability or premature death

$$\frac{dL_\lambda^S}{ds} = -\alpha_\lambda^S L_\lambda, \quad (2.1b)$$

where α_λ^a is the total absorption coefficient and α_λ^S is the total scattering coefficient at wavelength λ . The total attenuation of the radiation induced by scattering and absorption by aerosols, cloud particles and atmospheric gases within the path of the radiation is referred to as the **extinction**, where the extinction coefficient, ϵ_λ , is given by the sum of the absorption and scattering coefficients. Using the extinction coefficient, Equations (2.1a) and (2.1b) can be rewritten to

$$\frac{dL_\lambda}{ds} = -\alpha_\lambda^a L_\lambda - \alpha_\lambda^S L_\lambda = -\epsilon_\lambda L_\lambda. \quad (2.2)$$

The **optical depth** of the atmosphere, $\tau_\lambda^{\text{atm}}$, is a measure of the atmospheric transparency at wavelength λ . It can be calculated by integrating the extinction coefficient over the propagation path, from the surface to the top of the atmosphere (TOA)

$$\tau_\lambda^{\text{atm}} = \int_0^{s^{\text{TOA}}} \epsilon_\lambda(s) ds. \quad (2.3)$$

Usually the optical depth refers to the *normal optical depth*, which is the extinction coefficient integrated vertically, with zero solar zenith angle ($\theta_0 = 0$), and not over the propagation path. The normal optical depth (from now on referred to as the optical depth) of an atmospheric column is given by

$$\tau_\lambda^{\text{atm}} = \int_0^{z^{\text{TOA}}} \epsilon_\lambda(z) dz, \quad (2.4)$$

where 0 indicates surface level, z^{TOA} the top of the atmosphere altitude at zenith and $dz = \cos(\theta_0) ds$ (Schlundt, 2013). Figure 2.4 clarifies the geometric difference between Equations (2.3) and (2.4).

Figure 2.4: Illustration of the differences between propagation path and vertical (normal) path.

The atmospheric optical depth of the atmosphere is a measure of the extinction of the solar radiation caused by all atmospheric species and Equation (2.4) can accordingly be divided into several parts each represent one contributing factor as shown in (2.5). Please note that Equations (2.5), (2.6) and (2.7) are simplified versions of the optical depths and in reality the contribution from the different parts is not simply linear. For a fully correct interpretation an additional term should be added representing the coupled extinction from the four parts.

$$\tau_{\lambda}^{\text{atm}} = \int_0^{z^{\text{TOA}}} \epsilon_{\lambda}^{\text{aerosols}}(z) dz + \int_0^{z^{\text{TOA}}} \epsilon_{\lambda}^{\text{clouds}}(z) dz + \int_0^{z^{\text{TOA}}} \epsilon_{\lambda}^{\text{gases}}(z) dz + \int_0^{z^{\text{TOA}}} \epsilon_{\lambda}^{\text{molecules}}(z) dz, \quad (2.5)$$

which can be rewritten to

$$\tau_{\lambda}^{\text{atm}} = \tau_{\lambda}^{\text{aerosols}} + \tau_{\lambda}^{\text{clouds}} + \tau_{\lambda}^{\text{gases}} + \tau_{\lambda}^{\text{molecules}}. \quad (2.6)$$

Hence the aerosol optical depth (AOD) is defined as the extinction-contribution by aerosols to the atmospheric optical depth and given by

$$\tau_{\lambda}^{\text{aerosols}} = \int_0^{z^{\text{TOA}}} \epsilon_{\lambda}^{\text{aerosols}}(z) dz. \quad (2.7)$$

So in order to retrieve the AOD the extinction contribution from clouds, gases and molecules have to be derived and subtracted from the atmospheric optical depth (Levy et al., 2014).

2.3.1 Retrieval of AOD From Space

A satellite sensor that looks down towards the surface of the Earth will observe radiance originating from the sun backscattered by clouds, atmospheric aerosols, atmospheric molecules and the surface. Assuming that the Earth's surface can be treated as a Lambertian surface, the upwelling spectral radiance, L_{λ}^{TOA} , from below (backscattered solar radiation) and the incoming solar spectral irradiance, E_{λ}^0 , at TOA (corrected for the solar zenith angle) can be used to calculate the TOA reflectance i.e. the fraction of incoming solar radiation that is backscattered and observed by the satellite

$$\rho_{\lambda}^{\text{TOA}}(\theta_0, \theta_s, \varphi, \tau_{\lambda}) = \frac{\pi L_{\lambda}^{\text{TOA}}}{E_{\lambda}^{\text{TOA}} \cos(\theta_0)}. \quad (2.8)$$

The angles θ_0, θ_s and φ are the solar zenith angle, satellite view zenith angle and the relative azimuth angles respectively. Figure 2.5 illustrates a simple passive remote sensing setup where the meaning of the three angles is clarified.

Figure 2.5: Simple illustration of a passive remote sensing setup (based on Figure 2.1 from Kokhanovsky and de Leeuw (2009)).

The observed TOA reflectance consist of three major parts; (1) the backscattered solar radiation within the atmosphere (by aerosols, clouds and molecules), without surface interaction referred to as the *atmospheric path reflectance*, (2) direct and diffuse solar radiation reflected by the surface

and directly transmitted to the sensor and (3) direct and diffuse solar radiation reflected by the surface and diffusely transmitted to the sensor. A clarification of the three parts of the observed TOA reflectance can be seen in Figure 2.6.

Figure 2.6: Contributing factors to the TOA reflectance observed by the satellite.

Combining the three contributing reflectances the following relationship for the observed TOA reflectance at wavelength λ can be derived

$$\rho_{\lambda}^{\text{TOA}}(\theta_0, \theta_s, \varphi, \tau_{\lambda}) = \rho_{\lambda}^a(\theta_0, \theta_s, \varphi) + \frac{T_{\lambda}(\theta_0)T_{\lambda}(\theta_s)\rho_{\lambda}^s}{1 - S_{\lambda}\rho_{\lambda}^s}, \quad (2.9)$$

where

$\rho_{\lambda}^{\text{TOA}}$ is the reflectance observed by the satellite at TOA

ρ_{λ}^a is the reflectance from scattering by aerosols, molecules and clouds in the atmosphere

ρ_{λ}^s is the surface reflectance

$T_{\lambda}(\theta_0)$ is the downward transmission (direct and diffuse) from TOA towards the surface

$T_{\lambda}(\theta_s)$ is the upward transmission (direct and diffuse) from the surface towards TOA

S_{λ} is the spherical albedo.

The first term on the right-hand side of Eq. 2.9 is the *atmospheric path reflectance* i.e. the backscattered solar radiation that has not interacted with the Earth's surface and the second term is the backscattered solar radiation that has interacted with the surface. For more detailed information about the exact formulas of the different reflectances and the derivation of Equation (2.9) please see Kokhanovsky and de Leeuw (2009) and Vermote et al. (2006).

The keen part in the aerosol retrieval is that depending on whether the solar radiation is backscattered by aerosols, molecules, clouds or the surface the spectral signal will be different and by using different wavelengths, view angles and polarization the signal originating from aerosols can be estimated. However the spectral signals are not completely independent which puts high demands on the aerosol retrieval algorithms (Levy et al., 2013). The backscattered signal by clouds is very strong which makes it impossible to distinguish the aerosol signal and hence estimate the AOD for a cloudy scene, which limits the AOD retrieval to cloud free regions.

To retrieve aerosol data for a cloud free scene the molecular scattering, known as the Rayleigh scattering and the absorption by atmospheric gases (water vapor, CO_2 , CH_4 etc.) have to be removed from the first term in Equation (2.9), additionally the contribution from the surface has to be removed (the second term) (Levy et al., 2013). If the surface would be very dark, meaning that it has a low reflectivity the second term becomes very low and can be assumed to have zero contribution to the observed reflectance at TOA ($\rho_{\lambda}^{\text{TOA}}$) meaning that the only contribution is from

the atmosphere. However this is usually not the case and normally the second term cannot be excluded. However, all quantities on the right-hand side of Equation (2.9), except for the surface reflectance, can be calculated using radiative transfer if the observation geometry is known (see Figure 2.5) and the atmospheric conditions (such as AOD and aerosol type(s) etc.) are known or can be estimated. So if the surface reflectance can be estimated and the *atmospheric path reflectance* corrected for Rayleigh scattering and absorption by gases the AOD can be retrieved inversely by calculating the TOA reflectance theoretically using atmospheric/aerosol properties such that the error between observation and theoretical calculation is minimized.

Since radiative transfer calculations are very complex and time consuming a common feature for aerosol retrieval and remote sensing in general is to use a so called look-up table (LUT). A LUT contains a set of pre-calculated TOA reflectances for a wide range of observing geometries, wavelengths and atmospheric/aerosol conditions. Using this approach the pre-calculated TOA reflectances can be fitted to the observed TOA reflectance with reduced computational time, since the time consuming radiative transfer calculations have already been done when the LUT was created.

2.4 Typical AOD Values and the AOD Distribution

AOD values range from zero and upwards. A higher AOD means higher attenuation and hence more aerosols in the atmosphere. Table 2.3 shows what conditions the respective AOD values represents.

Table 2.3: AOD levels and the corresponding atmospheric conditions (credit: Levy et al. (2014))

AOD	Corresponding Condition
0	Clear sky, free from aerosols
0.02	Very clean, usually over isolated areas
0.2	Fairly clean air
0.6	Polluted air
1.5	Heavy smoke/dust event (e.g. volcanic eruption)
> 3	The solar disk is obscured

The AOD over US is in general relatively low, with the maximum values in larger cities, industrial regions and dusty regions. Figure 2.7 shows the AOD at 10 km spatial resolution over the US averaged for January, April, July and October in 2008. Please note that there are some values larger than 0.5, which are represented by the red color in the figure.

2.5 Use AOD as a Proxy for PM_{2.5} Concentration Prediction

The aerosol optical depth corresponds to the attenuation of the solar radiation due to aerosol interaction throughout the whole atmospheric column from ground level up to the satellite sensor while measurements of the PM_{2.5} concentration at ground level provides point measurements, both spatially and vertically, and is only representative for the air surrounding the monitor.

Despite the difference between AOD and PM_{2.5} and the respective retrieval methods, the two quantities relate to the amount of particles in the air and should be related to each other. Figure 2.8 shows the AOD at 10 km resolution (same as Figure 2.7) together with the daily mean PM_{2.5} concentrations in the US for the same four months in 2008. For this comparison the PM_{2.5} data have been gridded to 50×50 km. Even though the spatial coverage of the PM_{2.5} is comparably poor, similar patterns can be found. For example both graphs shows that the cleanest air is in the central US and that the highest air pollution is at the south-west coast, where a streak of both high AODs and PM_{2.5} concentrations are retrieved. Additionally both plots can distinguish

Figure 2.7: The AOD over US averaged for January, April, July and October in 2008

higher air pollution in the area around New York, Philadelphia and Washington D.C. So just by this comparison it is clear that there is a correlation between the two quantities, even though it is not perfect or exists for each particular area. From Figure 2.8 the advantage using satellite retrieved AOD for $PM_{2.5}$ prediction becomes clear with regards to the differences in spatial coverage.

Figure 2.8: AOD (left) and $PM_{2.5}$ concentrations (right) in the US averaged for January, April, July and October in 2008.

Several conditions influence the AOD data quality and its correlation with ground level $PM_{2.5}$ concentrations. To retrieve a good correlation between AOD and $PM_{2.5}$ and a linear relationship that can describe the surface $PM_{2.5}$ levels the conditions have to be favorable. In the following paragraphs the impact of some important factors will be described and the ideal retrieval conditions for an accurate AOD- $PM_{2.5}$ correlation will be explained based on those factors. Understanding this section is essential in order to follow the discussion of the obtained results in Chapter 5.

Despite the similarities between AOD and $PM_{2.5}$ there are of course differences due to the different retrieval methods that affect the linear correlation between AOD and $PM_{2.5}$. Two of the major differences are (1) the fact that AOD is a measure of the attenuation throughout the whole atmospheric column while a $PM_{2.5}$ monitor only provides a point measurement and (2) the fact that for the AOD includes the attenuation from both wet and dry particles while a $PM_{2.5}$ monitor only measures the mass concentration of dry particles (Hoff and Christopher, 2009). The boundary layer refers to the near surface atmospheric layer and the boundary layer height (BLH) is

the height of this layer. Under high BLH conditions there is a stronger exchange of particles between the boundary layer and free troposphere (Venzac et al., 2009) which may produce a more homogenous vertical aerosol distribution, making the surface measurements of $PM_{2.5}$ more representative to the total column AOD (Monache et al., 2004).

The relative humidity (RH) is a measure of the amount of water vapor in the atmosphere and defined as the ratio between the partial water vapor pressure and the saturated water vapor pressure (Seinfeld and Pandis, 2006). So the RH indicates the amount of wet particles in the atmosphere since the hygroscopic growth increases with increased RH (especially for $RH > 90\%$) (Gupta et al., 2006). Increased hygroscopic growth means more wet particles and thus higher extinction caught by the satellite. A $PM_{2.5}$ instrument, on the other hand, which does not include wet particles are independent of the RH. So for conditions with low RH the AOD is less affected by wet particles and hence more representative for the $PM_{2.5}$ concentration. With the goal to reduce the dissimilarities arising due to the differences between AOD and $PM_{2.5}$ corrections of the AOD for the RH and BLH conditions were implemented. This will be further explained in Section 4.4.

The temporal variability of the AOD does also play an important role. Since a satellite is constantly moving it has a poor temporal resolution compared to the $PM_{2.5}$ monitoring stations that measures the mass concentrations continuously and calculates the daily mean concentration. So if the temporal AOD variability is high the AOD retrieved at the time of the satellite overpass may not be representative for the daily mean value. Favorably is if the AOD/ $PM_{2.5}$ concentration is as constant as possible during the diurnal cycle, such that the satellite AOD is a good representation of the daily mean concentration.

Another factor is the impact of coarse particles. A $PM_{2.5}$ instrument is designed to filter out particles larger than $2.5 \mu m$ and does not include the coarser particles in the mass concentration. The coarse particles are included in the AOD on the other hand, which can lead to dissimilarities between the AOD and $PM_{2.5}$ if the amount of coarse particles in the atmosphere is high. The contribution by fine particles to the retrieved AOD is given by the fine mode fraction (FMF) which varies between 0 and 1, where a high value indicates a high contribution from fine particles and thus a small amount of coarse particles in the atmosphere, and vice versa for a low value. If the FMF is low the AOD will be too high in relation to the measured $PM_{2.5}$ concentration compared to if the FMF is high. This results in lower correlation (even negative) and can be identified by a gentler regression slope (which indicates overestimated AODs with respect to the measured $PM_{2.5}$ concentrations). Figure 2.9 shows the impact of the FMF on the linear AOD- $PM_{2.5}$ correlation and the corresponding regression slope.

Figure 2.9: Impact of FMF and coarse particles on the AOD- $PM_{2.5}$ correlation and regression slope. Left: low FMF with mostly fine particles; yields good AOD- $PM_{2.5}$ correlation. Center: higher FMF with some coarse particles leading to too high AOD values compared to the measured $PM_{2.5}$ concentrations, yields lower correlation and gentler slope. Right: even higher FMF with more coarse particles; yields even lower correlation and even gentler slope.

A high AOD data quality is also essential in order to obtain a good correlation. The quality is based on many factors, but one key factor is the surface characteristics. If the surface is too bright the aerosol signal gets weaker with regards to the total signal which reduces the data quality and hence the quality of the retrieval AOD-PM_{2.5} relationship.

So for an optimal correlation between AOD and PM_{2.5} the surface should be dark, the FMF and BLH high and the RH and temporal variability of the AOD low.

Chapter 3 Material and Study Region

3.1 Data Sets

In order to achieve information about the AOD, the $PM_{2.5}$ concentrations and the meteorological conditions for the regions studied in this thesis, several different data sets have been used, provided by four organizations and research teams. AOD data with varying spatial resolution as well as $PM_{2.5}$ data are necessary in order to compare AOD and $PM_{2.5}$ and derive the linear relation between the two quantities. Meteorological data have been used for the relative humidity and boundary layer height corrections of the AOD data.

3.1.1 Satellite Aerosol data

Two different AOD data sets have been used in this study, one from the MODIS standard product (MYD04) and one from the MAIAC algorithm. The AOD from both data sets originates from the MODIS sensor flying onboard the satellite Aqua.

Aqua was launched in 2002 as one of the Earth observing satellites and is part of the A-train, which is the name of the constellation of six Earth observation satellites flying in formation. Aqua is flying in a low Earth orbit with an altitude of 705 km. The inclination of the orbit is 98 degrees which makes it to a near polar orbit with global coverage. Additionally the orbit of Aqua is sun-synchronous, which means that the satellite always passes over the same latitude at the same local mean solar time. Due to this feature the illumination angle from the sun is nearly the same for each individual location every time the satellite passes over it (Berlin, 2007), which is advantageous for remote sensing where the solar radiation is used for the data retrieval. In Aqua's case, it will always cross the equator at the ascending node at 1.30 p.m. local time (NASA, 2014a). Aqua passes over the US roughly between 4 p.m. UTC and 9 p.m. UTC every day and four orbits are required to cover the whole contiguous US from east to west. A summary of Aqua's orbital parameters can be seen in Table 3.1.

Table 3.1: Aqua's orbital characteristics

Orbit type	Syn-synchronous low Earth orbit
Equatorial crossing	1.30 p.m., ascending node
Altitude	705 km
Inclination	98°

MODIS is an across track scanner (also called whisk broom scanner), which is a type of scanner used for passive satellite remote sensing. The scanner consist of a rotating mirror that reflects light from below to a receiving detector. The mirror rotates in such a way that the scan direction is across the satellites flight direction. A simplified sketch of the scanning principle of an across track scanner can be seen in Figure 3.1. For each revolution of the mirror a stripe extending 10 km (at nadir) along the satellite flight direction and 2330 km across the satellite track is scanned. With such a large swath width, the MODIS instrument can deliver a near global coverage in less than two days (NASA, 2014a).

Figure 3.1: The scanning principle of an across track scanner.

A unique feature of MODIS is the large spectral coverage. MODIS covers a spectral range from $0.41\ \mu\text{m}$ to $14.39\ \mu\text{m}$ divided into 36 bands. Another feature is that it also operates at different spatial resolutions. Band 1 and 2 have a spatial resolution of 250 m, while the bands 3-7 and 8-36 have a spatial resolution of 500 m and 1000 m respectively. Table 3.2 shows the characteristics of the five bands that are used for the aerosol retrieval over land (NASA, 2014a).

Table 3.2: The bands of MODIS used for aerosol retrieval

Band	Bandwidth [μm]	Spatial resolution [m]
1	0.620-0.670	250
2	0.841-0.876	250
3	0.459-0.479	500
5	1.230-1.250	500
7	2.105-2.155	500

One side effect with an across track scanner is that the pixels varies in size depending on the scan angle. The spatial resolutions mentioned above are at nadir and towards the swath edges the spatial resolution increases, both in length and width. A pixel that is $1 \times 1\ \text{km}$ at nadir grows gradually towards the swath edge where it is $4.8 \times 2\ \text{km}$ (Levy et al., 2013). This effect is extra apparent for the MODIS sensor which scans several stripes simultaneously. At nadir the combined swath length (along track) is 10 km, which means 10, 20 and 40 scan stripes with 1000 m, 500 m and 250 m spatial resolution respectively. This leads to a bow tie like scan pattern, where regions toward the swath edges will be scanned twice by two subsequent scans. From the shape of the scan pattern this is called the *bow tie effect*. Figure 3.2 illustrates MODIS' bowtie effect where three subsequent scans are visualized. It can be seen that at the swath edges the overlap between two subsequent scans is quite significant and with decreasing scan angle (closer to nadir) the effect decreases. The bow tie effect is present for scan angles larger than $\sim 15^\circ$ (NASA, 2014a; Liu et al., 2008).

3.1.1.1 MODIS Raw Data Processing

The raw data that is registered by the detector is transmitted to the ground where it is processed. Processed data from MODIS are delivered at different levels; Level 0 (L0) is the raw data from the sensor, Level 1A (L1A) is similar to L0, but here the data have been organized into so called

Figure 3.2: The bow tie effect present for the MODIS instrument (credit: Deutsches Zentrum für Luft- und Raumfahrt).

granules, which are five minute scan segments. For the next step, Level 1B (L1B), the data have been calibrated and geolocated. The L1B data is the input data for the aerosol retrieval algorithms used in this study.

As mentioned before two AOD data sets are used for this study. MYD04 provides Level 2 AOD data at 10 and 3 km spatial resolution and MAIAC provides gridded Level 3 (L3) AOD data at 1 km spatial resolution. In this section the basic steps in the aerosol retrieval processes for the two sets of AOD data are explained.

MYD04 Aerosol Data - Collection 6

The main algorithm used for aerosol retrieval from the MODIS sensor was first developed over 20 years ago i.e. before the launch of the first MODIS sensor on-board the satellite Terra. The algorithm is called *Dark Target* (DT) and is divided into two parts, for the retrieval over land and ocean surfaces respectively. The DT-land algorithm was originally formulated by [Kaufman et al. \(1997\)](#) and the DT-ocean algorithm by [Tanre et al. \(1997\)](#). Since then the algorithms have been further developed and improved by the MODIS aerosol science team and released in different versions, referred to as Collections ([Levy et al., 2007, 2013](#)). In 2013 the 6th Collection (C006) of the MODIS aerosol data was released. Except for general improvements and refinements of the algorithm the main difference to Collection 5 (C005) is that the aerosol data are provided at an additional higher spatial resolution (3 km) and the additional *Deep Blue* algorithm ([Hsu et al., 2004, 2006, 2013](#)) has been implemented and merged with the DT algorithms for the aerosol data at the original 10 km spatial resolution. Since all PM_{2.5} monitoring stations are located on land no data derived using the DT-ocean algorithm have been used in this study. And since the *Deep Blue* algorithm has only been implemented at 10 km spatial resolution and not at 3 km resolution the aerosol data derived using the DT-land algorithm only has been used in this study. Therefore the DT-land algorithm only will be explained further in this section.

The aerosol data retrieval is nearly identical for 10 km and 3 km resolution, therefore the retrieval of aerosol data at 10 km spatial resolution will be explained in the next paragraphs and afterwards the differences for the retrieval of aerosol data at 3 km resolution will be given.

The first step in the DT-land algorithm is to use geolocation information to distinguish the surface type for the observed L1B scene consisting of 20×20 500 m pixels. If all pixels are over water the DT-ocean algorithm is used, but if at least 24 of the pixels are over land the DT-land algorithm is used. Any water pixel from deep/shallow water, rivers etc. are filtered out in the

DT-land algorithm as a subsequent step. If there are less than 24 pixels over land or if a water scene also includes pixels over shallow water the scene is excluded and no retrieval is made (Levy et al., 2014).

The next steps are to filter out pixels that are too bright and pixels where the surface is too heterogeneous for the DT-land algorithm to handle, like pixels over snowy, icy and/or desert regions as well as cloud contaminated pixels. When all those clearly improper pixels have been filtered out the *Dark Target* pixels are identified using the relationship: $0.01 \leq \rho_{2.12}^{\text{TOA}} \leq 0.25$. From those *Dark Target* pixels, the 50% of the brightest and 20% of the darkest pixels are removed, this is done in order to remove possible cloud brightening, cloud shadowing and too non-uniform surface regions. After the pixel filtering each pixel is corrected for the radiative absorption by atmospheric gases. If there are at least 12 pixels left at this stage the average reflectance is calculated for those corrected pixels (Levy et al., 2013, 2014; Munchak et al., 2014).

The DT-land look-up table (LUT) is calculated for a black surface, which means that the TOA reflectances are calculated for zero surface reflectance. So in case of an observed reflecting surface, its reflectance contribution (ρ_{λ}^s) to the observed TOA reflectance (see Equation (2.9)) has to be estimated and corrected for before the pre-computed reflectances from the LUT can be fitted to the observed TOA reflectance (Kokhanovsky and de Leeuw, 2009; Levy et al., 2013).

For MODIS Collection 4 Kaufman et al. (1997) proposed that the aerosols play a negligible role at $2.12 \mu\text{m}$ and that the atmosphere can be assumed to be transparent at this wavelength. Using this assumption the surface reflectance in the visible channels could be estimated using two constant empirical relationships

$$\rho_{0.66}^s \approx 0.50 \rho_{2.12}^{\text{TOA}} \quad (3.1a)$$

$$\rho_{0.47}^s \approx 0.25 \rho_{2.12}^{\text{TOA}}. \quad (3.1b)$$

In C005 and C006 the atmosphere is no longer assumed to be transparent (since some coarse particles still interferes with the radiation at this wavelength) and the two constant empirical relationships have been improved such that the reflectance in the visible channels ($0.47 \mu\text{m}$ and $0.66 \mu\text{m}$) are functions of the reflectance at $2.12 \mu\text{m}$, the scattering angle and the normalized difference vegetation index ¹ (NDVI) (Levy et al., 2013, 2014)

As the surface reflectance can be estimated using this parameterization assuming that the surface is Lambertian, the *atmospheric path reflectance* (ρ_{λ}^a) can be retrieved by correcting the observed TOA reflectance for the surface reflectance. The next step is to select a LUT to use based on the aerosol type, which is a function of time and geographical location (Levy et al., 2007). The reflectances in the LUT are pre-computed from a huge set of combinations of sun/satellite/surface scattering geometries and a bunch of atmospheric/aerosol properties such as Rayleigh optical depths (RODs), AODs, aerosol types, FMF, spherical albedo etc. and can furthermore be fitted to the observed *atmospheric path reflectance* such that the squared error between observation and theoretical calculation is minimized. At this stage the aerosol properties, such as the AOD and FMF, used to calculate the reflectance with the smallest error compared to the observation for the analyzed scene has been inversely estimated. The retrieved parameters at $0.47 \mu\text{m}$ and $0.66 \mu\text{m}$ are furthermore interpolated in order to obtain the same parameters at $0.55 \mu\text{m}$ (Levy et al., 2013; Kokhanovsky and de Leeuw, 2009).

As a final step after the aerosol data have been determined using the LUT the data quality is determined. At predetermined steps in the algorithm the data quality is evaluated, and if there are non-ideal conditions the data quality is degraded. Factors that determines the quality are for example the number of pixels used for the retrieval, the difference between observed and calculated reflectances and the surface brightness. The final product is assigned one of the four quality assurance confidence (QAC) flags depending on the overall judgment (3 = high confidence, 2 = good confidence, 1 = marginal confidence, 0 = no confidence) (Levy et al., 2013, 2014).

¹Indicates which type of vegetation that is present for the analyzed surface

As mentioned above one new feature in C006 is the aerosol products at 3 km spatial resolution. The original idea with the MODIS data was to use it for climate applications, therefore a relatively coarse resolution (10 km) was chosen in order to reduce the error. But since the usage of MODIS data to predict surface PM concentrations was found a higher spatial resolution became desirable. The derivation is almost identical for both products, with the only differences in the first steps: for data at 3 km resolution only 6×6 500 m pixels are collected and only 6 valid pixels are required after filtering for further processing. This means that pixels discarded for a 10 km pixel might be used for a 3 km pixel (Levy et al., 2013; Munchak et al., 2014). The advantage with an increased spatial resolution is the higher level of detail for spatial variability, however there is of course a risk that the increased amount of errors with this higher resolution leads to worse data quality and AOD-PM_{2.5} correlation.

A simplified flowchart of DT-land algorithm can be seen in Figure 3.3, the numbers inside parentheses represents the 3 km aerosol data retrieval. More detailed information about MYD04 Collection 6 and the Dark Target algorithms can be found in the paper by (Levy et al., 2013, and references therein).

Figure 3.3: Simplified flowchart for the DT-land algorithm.

MAIAC Aerosol Data

The main difference for the MAIAC algorithm compared to the DT algorithm is that the retrieval is based on image based processing and for a time series of satellite data. The MAIAC algorithm takes the L1B data from MODIS as input and grids the data to 1×1 km. The gridded data is divided into so called blocks and tiles which consist of 25×25 and 600×600 grid cells respectively. Data over the same region but from different days/orbits are collected and put in a queue which form the time series of data from which the processing is made. The queue consist of data from 5-16 days depending on latitude. One key advantage that arise when using a time series of data is that for each grid cell the observation geometry changes for each observation day/orbit so by using a time series of data retrieved using different observation geometries the surface bidirectional reflectance factor (BRF) can be estimated. This requires one assumption, that the surface

BRF is constant throughout the time period that is processed. Another key advantage is the possibility to perform the cloud masking based on temporal variations of the surface appearance (Lyapustin et al., 2011; Lyapustin and Wang, 2008).

Similarly as for the DT algorithm the main goal is to calculate a theoretical TOA reflectance and fit it to the observed TOA reflectance in order to inversely derive the aerosol information. In contrast to the DT-algorithm, MAIAC does not treat the Earth's surface as a Lambertian surface so Equation (2.9) cannot be used to describe the TOA reflectance observed by the satellite. Instead the TOA reflectance at wavelength λ at one specific day (k) within the time series and in one grid cell (i,j) is calculated using the following formula

$$\rho_{i,j}^{TOA,k} \cong \rho^D(\lambda, \tau^k) + b_{i,j}(\lambda) \cdot Y_{i,j}(\lambda, \tau^k), \quad (3.2)$$

where $\rho^D(\lambda, \tau^k)$ is the *atmospheric path reflectance*, $b_{i,j}(\lambda)$ the so called *spectral regression coefficients* (SRCs) and $Y_{i,j}(\lambda, \tau^k)$ a parameter calculated for a given sun-satellite-surface geometry, AOD, wavelength and the surface BRF. Figure 3.4 illustrates the time series of the AOD blocks and the individual pixels/grid cells.

Figure 3.4: Illustration of the time series of AOD data used in the MAIAC algorithm. The larger parallelograms represent three blocks of AOD data holding 25×25 AOD pixels/grid cells. The smaller parallelograms represent the first AOD pixel (1,1) and an arbitrary AOD pixel (i,j). Figure inspiration from Lyapustin and Wang (2008).

To detect and exclude cloudy scenes a covariance approach is applied to the time series of data. The temporal variability of the surface for a block ($25 \times 25 \text{ km}^2$) can be assumed to be very low in absence of heavy snowfall, ablation, fires etc. However the temporal variability of clouds is known to be high. So if the same scene is observed at different times where one of the cases is a known clear sky observation and the spatial covariance between the two images is high it is reasonable to believe that the second image is a clear sky observation as well (low variability between the observation time). If the spatial covariance is low on the other hand it is likely that clouds are present in the image. The reference clear sky observations retrieved from cases where two images have a high spatial covariance. Every time such a case occurs the reference image is updated in order to adapt to the seasonal changes in the surface patterns. To make sure the high covariance is not due to two similar but cloudy pixels a surface type test is done where the reflectances, brightness temperatures and the normalized difference vegetation/snow/ice indices (NDVI, NDSI, NDII) are examined. To proceed from this cloud masking step the number of cloud free days/orbits have to be at least three (out of initial 5-16 days/orbits) and the water vapor column for a whole block and the BRFs for the individual grid cells have to be known.

After the cloud masking the goal is to derive one AOD and 25×25 SRCs ($b_{i,j}(\lambda)$) for each block k in the time series. First the SRC for all blocks and grid cells are calculated using Equation (3.2) for AOD=0 in order to determine the clearest day which is defined as the day with the lowest SRC on average. Furthermore the AOD difference between the block with the clearest scene (τ^{clear}) and the other days are calculated ($\Delta\tau^k$). By assuming that the surface BRF is constant for all days in the time series $\Delta\tau^k$ can be estimated by finding the AOD for which the overall difference between $b_{i,j}^{\text{clear}}$ and $b_{i,j}^k$ for each block is as smallest. Since the AOD for the clearest block was initially set to zero the AOD for each block yielding the smallest overall difference in

$b_{i,j}$ between the days is known as $\Delta\tau^k$. Knowing $\Delta\tau^k$ for all days the SRCs can be recalculated using Equation (3.2) by minimizing the root mean square error (RMSE) between the measured radiance and the theoretically calculated radiance for all days separately. Until now the clearest day has been assumed to have an AOD equal to zero, which most likely is not true. So the next step is to repeat the steps after the cloud mask with increasing AODs until a sufficiently small RMSE has been obtained between the measured and calculated reflectances. When the RMSE is sufficiently low the SRCs have been derived for all days and individual pixels but the AOD is only known per block i.e. at 25 km spatial resolution.

Using the following relationship

$$\rho_{0.47}^s(\theta_0, \theta_s, \varphi, \tau) = b_{i,j}(\lambda) \cdot \rho_{2.12}^{\text{TOA}}(\theta_0, \theta_s, \varphi, \tau) \quad (3.3)$$

the surface reflectance at 0.47 μm can be calculated for each day and pixel using the previously derived SRCs ($b_{i,j}$). Furthermore the aerosol type is chosen in a similar manner as in the DT algorithm and the AOD and FMF can be inversely derived by fitting the pre-calculated TOA reflectances (using Equation (3.2)) inside the LUT to the observed TOA reflectance.

A simplified flowchart of the MAIAC algorithm can be seen in Figure 3.5. More detailed information about the MAIAC algorithm can be found in the paper by Lyapustin et al. (2011) and in the algorithm theoretical basis document (Lyapustin and Wang, 2008).

Figure 3.5: Simplified flowchart for the MAIAC algorithm.

If the surface is too bright the MAIAC algorithm assigns the pixel a pre-defined AOD value of ~ 0.0619 (at 0.55 μm). This is not a valid AOD value, but used for the atmospheric correction part of the algorithm. Therefore a vertical pattern can be seen in the scatter plots between the MAIAC AOD and the $\text{PM}_{2.5}$ concentrations for this particular AOD in Chapter 5 and Appendix A. This is especially clear during winter time when there is less vegetation and the risk for ice and snow is higher and in the western part of the US where the surface in general is brighter.

Difference between MYD04 AOD and MAIAC AOD data

From the previous sections it is clear that the two AOD retrieval algorithms differ from each other. To better understand how those differences affect the obtained AOD, the AOD at 3 km

spatial resolution on January 1st 2008 can be seen in Figure 3.6. The left subfigure shows the MYD04 AOD data and the right subfigure the MAIAC AOD data. It is clear that MAIAC provides a better coverage compared to MYD04, but also that close to the cloudy regions MAIAC provides comparably high AODs which probably indicates cloud screening from nearby pixels. For this particular day major part of the northern half of the US as well as the southern parts of Texas and Florida was covered by clouds and a relatively clear border can be seen especially for the MAIAC AOD data which provides data for the majority of the cloud free regions. Additionally the above mentioned pre-defined AOD values of 0.0619 assigned to some too bright study scenes in the MAIAC algorithm can be seen in the right subfigure in the center of the US ranging from approximately 90°W to 110°W and 40°N to 45°N.

Figure 3.6: AOD on January 1st 2008 from MYD04 (left) and MAIAC (right).

3.1.2 Ground Level PM_{2.5} Data

PM_{2.5} data provided by the US Environmental Protection Agency (EPA) have been used for this study. EPA provides several data products related to air pollution and particulate matter concentrations, but for this study the daily mean PM_{2.5} concentration data have been used. This means that each instrument measures the PM_{2.5} concentration continuously and for each day the average value is calculated. For the time period studied in this thesis 946 EPA PM_{2.5} monitoring stations provided daily PM_{2.5} mass concentrations across the contiguous US. All PM_{2.5} monitoring stations in the US can be seen in Figure 3.7 as black dots further down in this chapter. EPA provides two parameters of the daily mean PM_{2.5} concentrations called *PM_{2.5} - Local Conditions* and *Acceptable PM_{2.5} AQI & Speciation Mass* respectively. Data belonging to the first parameter is derived from federal approved or equivalent methods (FRM/FEM) and is considered to have a higher reliability, therefore this parameter only is chosen for this study.

3.1.3 Modeled Meteorological Data

Two data sets are used for the meteorological correction of the AOD data, one for boundary layer height (BLH) and one for relative humidity (RH). The BLH data is forecast data while the RH data is analysis data and both data sets are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). Both the BLH and RH data are gridded to 0.25°×0.25° and provided at four times per day with six hours spacing. Additionally the RH is provided at 15 different pressure levels.

3.2 Regions and Time of Study

The choice of geographical regions to study in this thesis has been dependent on the data availability. The coverage of the MODIS data is near global since the satellite continuously travels

around the rotating Earth. The meteorological data that have been used for this study are modeled products and thus gives a global coverage as well. Large data sets of $PM_{2.5}$ data with large spatial coverage are publicly available for the US and Europe. The constrain lies in the AOD data; AOD from the MODIS standard product (MYD04) is available globally however the AOD data from the MAIAC algorithm could only be obtained over the contiguous US for this study. Therefore the main study region in this thesis is the contiguous US. The contiguous US excludes Alaska and all off-shore US states and territories.

The study is divided into two parts. In the first part, referred to as *Study I - Region of Continental Spatial Scale*, the impact of the AOD spatial resolution is analyzed for the whole contiguous US and in the second part, referred to as *Study II - Regions of Varying Spatial Scale*, the impact of the spatial scale of the study region is analyzed as well. The second study is done for six US cities located in the eastern and western part of the US and three corresponding spatial scales; *urban scale*, *meso-scale* and *continental scale*. The six cities are Washington D.C, Atlanta and Indianapolis in the eastern part and Salt Lake City, Phoenix and Los Angeles in the western part. The *urban scale* refers to the actual city and only $PM_{2.5}$ monitoring stations inside the respective cities are used for the analysis. The *meso-scale* refers to an $8^\circ \times 8^\circ$ area centered at the respective city centers and the *continental scale* refers to the whole US. The different study regions and the EPA $PM_{2.5}$ monitoring stations can be seen in Figure 3.7. The black dots represent each $PM_{2.5}$ monitoring station, the red stars show the location of the six cities and the blue rectangles shows the respective meso-scale study regions. The individual boundaries of the respective urban scale study regions are clarified in the sub-maps on each side of the US map (please note that this is not the meso-scale regions). For those sub-maps the red stars represent the city centers and the black dots the $PM_{2.5}$ monitoring stations included for the urban scale analysis for the respective cities. Additionally the continental scale represented by the contiguous US can be seen as the white non-ocean area.

Figure 3.7: The EPA $PM_{2.5}$ monitoring stations and the different study regions.

The choice of time period to study has also been constrained by the data availability. The modeled ECMWF data is available from 1979-present and sets no constraints with regards the choice of study period. Instead the constrain backwards in time is set by the satellite and $PM_{2.5}$ data. A Time period before the launch of Aqua in 2002 is of course not possible to study and with regards to the ground level $PM_{2.5}$ data the period of study should ideally be kept as late as possible, since the size of the EPA network expands for every year. The MYD04 aerosol data at 10 km and 3 km spatial resolution developed by the MODIS aerosol science team are available from launch to present, however the MAIAC 1 km data do only provide a full coverage over the US until 2008. Therefore the year 2008 is chosen for this study, since it is the latest year where all data products are available.

As shown in the first chapter clear seasonal variations have been observed in the AOD-PM_{2.5} relationship in the US. Therefore four months in 2008 have been studied, one representing each season; January, April, July and October. Only one month per season is chosen in order to save computational time. For the first study the four months are studied separately but for the second part the four months have been combined in order to maintain a large number of points for all combinations of study region and spatial resolution

Chapter 4 Method

This chapter will explain the steps and work that has been done throughout this study in order to obtain the results presented in Chapter 5. All work described in this chapter has been conducted by the author of this thesis alone.

4.1 Data Structure and Processing

Except downloading all data sets the data have also been processed in different ways since the different data are delivered in different formats and structures.

4.1.1 Structure and Processing of MAIAC L3 Aerosol Data

The AOD data provided by the MAIAC algorithm is gridded to 1 km and divided into so called tiles. Each tile has one hdf-file holding the latitude and longitude data together with several hdf-files holding the AOD data for different days/orbits. For each day of the four studied months the valid AOD data was extracted and paired with the geodetic information and the satellite overpass UTC time. The MAIAC AOD was provided at $0.47 \mu\text{m}$, but using a supplied conversion table and linear interpolation the AOD was converted to $0.55 \mu\text{m}$ in order to match the MYD04 data (personal communication with Alexei Lyapustin, 2014).

Even though the MAIAC AOD data is gridded Level 3 data the AOD and the corresponding overpass times are regridded in order to have the same grid structure for both the MAIAC AOD and the MYD04 AOD. To obtain more spatial resolutions to analyze, the MAIAC data have also been aggregated to 3 km and 10 km spatial resolution based on the assumption that AOD at a coarser spatial resolution can be represented by simple aggregation of the 1 km AOD data (Chudnovsky et al., 2013a). The gridding is done for each of the observed days and the overpass time is included in the gridding in case two scans from adjacent orbits overlap, in order to more accurately correct the AOD values using the meteorological data later on. The gridding method used for the AOD data in this study is simple averaging, which means that if more than one AOD pixel center is located within a grid cell the mean value is calculated.

After the gridding, the empty grid cells were excluded before the pre-processing of the MAIAC AOD data was finished. The Matlab code for the pre-processing of the MAIAC data can be seen in Appendix B.1 At this stage the data were ready to be matched with corresponding $\text{PM}_{2.5}$ concentrations and analyzed (see Section 4.2).

4.1.2 Structure and Processing of MODIS L2 Aerosol Data

The level 2 aerosol data provided by the MODIS aerosol science team is divided into so called granules, where each granule holds aerosol data for a five minute scan period along the satellite orbit. Each granule is embedded in an hdf-file which contains information on several aerosol parameters as well as the geodetic latitudes and longitudes and other information related to the satellite observation. The parameters used in this study (except for latitude and longitude) are

- **Optical_Depth_Land_And_Ocean**, which is the AOD at $0.55 \mu\text{m}$ over land and ocean retrieved using the *Dark Target* algorithms. Only the highest quality AOD (QAC flag = 3) has been included over land for this parameter.

- **Optical_Depth_Ratio.Small_Land**, which is the fraction of the AOD contributed to by fine mode particles, i.e. the fine mode fraction.
- **Land_sea_Flag**, which has been used to exclude pixels over ocean since all PM_{2.5} monitors are located on land.

Usually 8-11 granules of data are available per day over the US and because of the wide swath width of MODIS and the bow tie effect the chance for pixel overlap is quite high at the swath edges, especially for the higher resolution. So the first step was to find all pixels over land with valid AOD values together with the corresponding geodetic coordinates and satellite overpass UTC times for each day of the study period and grid the data. The gridding was performed due to two reasons: (1) average the values where several pixels overlap and (2) obtain uniform pixel sizes corresponding to the spatial resolution (originally the pixel size increases towards the swath edge).

After gridding the empty grid cells were excluded before the pre-processing of the MYD04 AOD data was finished. The Matlab code for the pre-processing of the MYD04 data can be seen in Appendix B.1

4.1.3 Structure and Processing of Ground Level PM_{2.5} Data

From the EPA website the daily PM_{2.5} concentrations can be downloaded for all stations in the US year by year. The data is stored in a csv-file and for each PM_{2.5} monitoring station the geographical coordinates and the corresponding daily PM_{2.5} concentration are provided along with a bunch of other parameters. For each station all PM_{2.5} concentrations measured in January, April, July or October was extracted together with the station coordinates and the date of retrieval.

If more than one daily mean concentration was provided at the same site for the same day e.g. if more than one monitor is located at the same site, the mean concentration for that site and day was calculated, such that all stations provides only one PM_{2.5} concentration value per day. The Matlab code for the pre-processing of the ground level PM_{2.5} data can be seen in Appendix B.2

4.1.4 Structure and Processing of Modeled Meteorological Data

The modeled data from ECMWF was provided day by day as one netCDF file each for the relative humidity and boundary layer height data. Each netCDF file contains the latitude, longitude and the global RH/BLH data as well as the UTC times for the four sets of data provided for each day. Additionally data for the 15 pressure levels were provided for the RH data. The RH and BLH data were provided in a good structure and were not processed before the corrections were applied to the AOD.

4.2 Matching Satellite AOD and Ground Level PM_{2.5} Data

At this point both AOD data sets and all spatial resolutions had a similar structure and could more easily be matched to the PM_{2.5} data. For each PM_{2.5} monitoring station the coordinates were extracted and a square was created and centered at the PM_{2.5} station. The size of the square depends on the spatial resolution that was investigated, for 10 km spatial resolution the square was 10×10 km² etc. Furthermore a check was made whether any of the gridded AOD coordinate centers were located within this area for each day of the studied time period. Using this approach the PM_{2.5} concentrations could be matched with pixels of aerosol data covering the PM_{2.5} monitoring station, however there will of course be cases where no pixel will cover a PM_{2.5} monitoring station and for those cases the measured PM_{2.5} concentration was excluded. A pair of a PM_{2.5} concentration and overlaying AOD is in this thesis referred to as an AOD-PM_{2.5} match-up.

The reason for creating squares around the PM_{2.5} monitors which performs point measurements and not around the AOD pixel centers is because with this approach the same results are achieved, but the computational time is considerable lower. The Matlab code for the matching of AOD and PM_{2.5} data can be seen in Appendix B.3

4.3 Removal of Outliers

When a measured $\text{PM}_{2.5}$ concentrations is matched with the AOD retrieved over the same region there are three possible sources of outliers. If the scene is cloud contaminated the AOD will be too high to be representative for the measured $\text{PM}_{2.5}$ concentration compared to other possible AOD- $\text{PM}_{2.5}$ match-ups. The other possible sources are if the satellite passes over a $\text{PM}_{2.5}$ monitoring station when the pollution level is high or low compared to the daily mean provided by the instrument and used for this study. In such a case the AOD will be overestimated or underestimated compared to the daily mean AOD and the corresponding $\text{PM}_{2.5}$ concentration. The impact of those outliers on the results does of course depend on the number of outliers, but if just very few match-ups are found and analyzed just a couple of outliers can have a large impact and yield misleading and inaccurate results.

To minimize the impact from those match-ups two empirical removal criteria have been developed and used in the analysis. The principle is to remove pairs of AOD and a $\text{PM}_{2.5}$ where the $\text{PM}_{2.5}$ concentration is too low (indicates cloud contamination or overestimated AOD with regards to the daily mean concentration) or too high (indicates underestimated AOD with regards to the daily mean concentration) compared to the corresponding AOD. Values are removed from the analysis if

$$\text{PM}_{2.5} < 10 \cdot \text{AOD} \quad \text{or} \quad (4.1a)$$

$$\text{PM}_{2.5} > \max\{700 \cdot \text{AOD}, 1.5b\} \quad (4.1b)$$

where b is the intersection of the linear regression fit, which will be introduced in Section 4.5. A lower limit of $1.5b$ is used for the second criteria in order to avoid removal of points near the regression line, because if a point is located near the regression line it is a reasonable value even if the measured $\text{PM}_{2.5}$ concentrations is very high compared to the retrieved AOD. Figure 4.1 shows the the regions where AOD- $\text{PM}_{2.5}$ match-ups are considered to be outliers. The lower region corresponds to the first removal criteria (Equation (4.1a)) and the upper region to the second criteria (Equation (4.1b)). For this illustration a slope of $b = 8 \mu\text{m}^3$ has been used.

Figure 4.1: Visualization of the areas where AOD- $\text{PM}_{2.5}$ match-ups are considered to be outliers.

The two removal criteria have been developed to remove a similar and comparably low fraction of points compared to the total number of match-ups independent on spatial resolution and study region. On average around 2.0% of the total number of points have been removed for every combination of AOD data set, spatial resolution city with a slightly higher fraction for the MYD04 AOD data set compared to the MAIAC AOD data set. The Matlab code for the removal of outliers can be seen in Appendix B.4 and B.5 for *Study I* and *Study II* respectively

4.4 Corrections of Satellite AOD Data

To correct for the differences in the retrieval of AOD and PM_{2.5} boundary layer height and relative humidity corrections were implemented.

The boundary layer is the near surface atmospheric layer and the correction was a simple division of the retrieved AOD by the BLH

$$\text{AOD}_{\text{corr}} = \frac{\text{AOD}}{\text{BLH}} \quad [\text{km}^{-1}], \quad (4.2)$$

used to estimate the AOD close to the surface since PM_{2.5} instruments only performs point measurements and the AOD refers to the extinction within the whole atmospheric column. The correction relies on two assumptions; that all aerosols are located within the boundary layer and that the aerosols are homogeneously distributed within the layer. The corresponding BLH for each AOD-PM_{2.5} match-up was found by identifying which BLH grid cell that covered the PM_{2.5} monitoring station. Furthermore linear interpolation was used for the four sets of data (provided each day) to find the BLH at the time of the satellite overpass for each AOD-PM_{2.5} match-up.

Two RH corrections were implemented and tested based on two publications by Wang et al. (2010, 2014). The idea with the relative humidity correction was to remove the impact of wet particles from the satellite retrieved AOD since wet particles are not included in the PM_{2.5} mass concentration measurements. The estimation of the corresponding RH at surface level for each AOD-PM_{2.5} match-up followed the same principle as for the BLH. Only the RH at surface level (1000 hPa) was used in this study since the BLH correction aimed to estimate the surface level contribution to the AOD.

4.5 Evaluation of the Linear Correlation Between AOD and PM_{2.5}

First of all the two vectors of matched AOD data and PM_{2.5} concentrations were plotted against each other for a visual understanding of the linear correlation. The exact linear relation between the AOD and ground level PM_{2.5} concentrations was determined using linear regression, the Pearson correlation coefficient, R , and the statistical significance using the p -value, p .

The **linear regression** was derived using the least squares approach according to the formulation by Schlitzer (2006). The idea is to determine the coefficients a and b for the linear equation

$$\text{PM} = a \cdot \text{AOD} + b, \quad (4.3)$$

such that the sum of the squared error between measured PM_{2.5} concentrations and the estimated ones using the linear regression is as small as possible. First when this sum is minimized the optimal solution has been obtained. This is done using matrix formulation where Equation (4.3) for all match-ups is written as

$$\mathbf{d} = \mathbf{A}\mathbf{x}, \quad (4.4)$$

where $\mathbf{d}$ is the *response vector*, $\mathbf{A}$ the *design matrix* and $\mathbf{x}$ the *parameter vector*. The matrix and the two vectors are obtained by applying the two linear parameters a and b in Equation (4.3) to each matching pair of the AOD pixels and the corresponding PM_{2.5} concentrations:

$$\begin{aligned} \text{PM}_{2.5}^1 &= a \cdot \text{AOD}^1 + b \\ \text{PM}_{2.5}^2 &= a \cdot \text{AOD}^2 + b \\ \text{PM}_{2.5}^3 &= a \cdot \text{AOD}^3 + b & \rightarrow \\ &\vdots \\ \text{PM}_{2.5}^N &= a \cdot \text{AOD}^N + b \end{aligned}$$

$$\begin{bmatrix} \text{PM}_{2.5}^1 \\ \text{PM}_{2.5}^2 \\ \text{PM}_{2.5}^3 \\ \vdots \\ \text{PM}_{2.5}^N \end{bmatrix} = \begin{bmatrix} a \cdot \text{AOD}^1 + b \\ a \cdot \text{AOD}^2 + b \\ a \cdot \text{AOD}^3 + b \\ \vdots \\ a \cdot \text{AOD}^N + b \end{bmatrix} \rightarrow \begin{bmatrix} \text{PM}_{2.5}^1 \\ \text{PM}_{2.5}^2 \\ \text{PM}_{2.5}^3 \\ \vdots \\ \text{PM}_{2.5}^N \end{bmatrix} = \begin{bmatrix} \text{AOD}^1 & 1 \\ \text{AOD}^2 & 1 \\ \text{AOD}^3 & 1 \\ \vdots & \vdots \\ \text{AOD}^N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \rightarrow \mathbf{d} = \mathbf{Ax}.$$

where AOD^1 and $\text{PM}_{2.5}^1$ is one out of N AOD- $\text{PM}_{2.5}$ match-ups.

The error between the measured $\text{PM}_{2.5}$ concentrations and the estimated concentrations are known as the residuals and given by $e = \mathbf{Ax} - \mathbf{d}$. The sum of the squared errors (residuals) is in matrix formulation written as

$$F = (\mathbf{Ax} - \mathbf{d})^T (\mathbf{Ax} - \mathbf{d}) \quad (4.5)$$

and known as the *cost function*. The minimum of the cost function and hence the optimal solution is found by setting the gradient of the cost function equal to zero, which after some math yields the *normal equation*.

$$\vec{\nabla}_x(F) = 0 \rightarrow \mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{d}. \quad (4.6)$$

From the *normal equation* the unknown coefficients giving the optimal linear fit can be determined by solving Equation (4.6) for $\mathbf{x}$

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{d}). \quad (4.7)$$

The formulation by [Schlitzer \(2006\)](#) is used to calculate the **correlation coefficient** as well using the following formula

$$R = \frac{\sigma_{\text{AOD}, \text{PM}_{2.5}}^2}{\sigma_{\text{AOD}} \cdot \sigma_{\text{PM}_{2.5}}}, \quad (4.8)$$

where σ_{AOD} , $\sigma_{\text{PM}_{2.5}}$ and $\sigma_{\text{AOD}, \text{PM}_{2.5}}^2$ are the standard deviations of the AODs and $\text{PM}_{2.5}$ concentrations and the covariance between the two respectively. The three quantities are given by

$$\sigma_{\text{AOD}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\text{AOD}^i - \mu_{\text{AOD}})^2} \quad (4.9a)$$

$$\sigma_{\text{PM}_{2.5}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\text{PM}_{2.5}^i - \mu_{\text{PM}_{2.5}})^2} \quad (4.9b)$$

$$\sigma_{\text{AOD}, \text{PM}_{2.5}}^2 = \frac{1}{N-1} \sum_{i=1}^N (\text{AOD}^i - \mu_{\text{AOD}})(\text{PM}_{2.5}^i - \mu_{\text{PM}_{2.5}}), \quad (4.9c)$$

where μ represents the respective mean values. The correlation coefficient can vary between -1 and 1, where

- $R = 1$ means perfect positive linear correlation
- $R = -1$ means perfect negative linear correlation
- $R = 0$ means zero linear correlation.

The statistical significance of this study is evaluated using the p-value (p), which determines whether the null hypothesis can be rejected or not. The null hypothesis claims that there is no relationship between the $\text{PM}_{2.5}$ concentrations and AODs and that any pattern of relationship occurs by chance and/or sampling errors. To prove that this is not the case and that the two parameters in fact are related to each other, it has to be proven that the null hypothesis is false. The *p-value* itself is a measure of the probability of obtaining the observed AOD- $\text{PM}_{2.5}$ pattern if the null hypothesis is true. If the p-value is very low the data is not consistent with the null hypothesis and the null hypothesis can be rejected. For this study the results have been regarded as significant if $p < 0.001$.

The Matlab code for the evaluation of the linear AOD- $\text{PM}_{2.5}$ correlation can be seen in Appendix [B.4](#) and [B.5](#) for *Study I* and *Study II* respectively.

Chapter 5 Results

In the following chapter the results of this thesis will be presented. The results are separated in two major parts; in the first part the results obtained using the MAIAC AOD data are presented and in the second part the results obtained using the MYD04 AOD data are presented. Information given in the first part will be used for the analysis in the second part as well, so reading the two parts in the order presented here is recommended.

The BLH and RH corrections of the AOD did not turn out as expected and the correlation did in general decrease when the corrections were applied. Therefore the results presented in this part are derived **without** any corrections and instead the results obtained with the corrections will be discussed shortly in the last section of this chapter.

The results are presented using tables and scatter plots, where the tables contain information about the number of points (N), the linear correlation coefficient (R), linear regression slopes and intersects and the p -value (p). The scatter plots shows the data points for the AOD and $PM_{2.5}$ concentration match-ups together with the corresponding linear regression fits.

5.1 MAIAC AOD Data

The MAIAC team provides AOD at 1 km spatial resolution, which in this thesis has been aggregated to the two additional resolutions at 10 km and 3 km. In this subsection the linear relationship between the MAIAC AOD data and the $PM_{2.5}$ concentrations measured by the 946 EPA $PM_{2.5}$ monitoring stations are presented. First the results for the study of the whole US (the continental scale study) are presented followed by the results of the study of varying spatial scales.

5.1.1 Study I - Region of Continental Spatial Scale

Figure 5.1 shows the AODs and corresponding $PM_{2.5}$ concentrations in January 2008 together with the linear regressions at 10 km (left), 3 km (center) and 1 km (right) spatial resolution. The similar results for April, July and October are shown in Figure 5.2, 5.3 and 5.4 respectively. Table 5.1 summarizes the statistical information related to the retrieved linear relationships for the four months.

Figure 5.1: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in the US in January 2008 at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Table 5.1: Statistical data related to the linear relationship between MAIAC AODs and PM_{2.5} concentrations for the whole contiguous US

	Statistics	MAIAC 10 km	MAIAC 3 km	MAIAC 1 km
January	<i>N</i>	4756	3704	2999
	<i>R</i>	0.143	0.222	0.283
	Slope	18.21	32.60	45.27
	Intersect	9.21	7.87	6.85
	<i>p</i>	< 0.001	< 0.001	< 0.001
April	<i>N</i>	5296	4427	3683
	<i>R</i>	0.239	0.262	0.272
	Slope	13.57	16.75	19.15
	Intersect	8.00	7.76	7.61
	<i>p</i>	< 0.001	< 0.001	< 0.001
July	<i>N</i>	6886	5147	3751
	<i>R</i>	0.562	0.566	0.587
	Slope	29.66	31.21	33.64
	Intersect	7.40	7.55	7.30
	<i>p</i>	< 0.001	< 0.001	< 0.001
October	<i>N</i>	7824	6708	5859
	<i>R</i>	0.214	0.272	0.313
	Slope	14.59	22.28	31.10
	Intersect	7.81	7.34	6.89
	<i>p</i>	< 0.001	< 0.001	< 0.001

Figure 5.2: Linear relationship between MAIAC AODs and PM_{2.5} concentrations in the US in April 2008 at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

First of all it can be seen that the *p-value* is less than 0.001 for all combinations of different months and spatial resolutions which tells us that the results are statistically significant.

For all months the linear correlation between AOD and PM_{2.5} becomes better with an increased spatial resolution. This means that the theoretical hypothesis is supported which says that the AOD at 1 km spatial resolution should provide better correlation with PM_{2.5} concentrations than AOD at 3 km resolution which in turn should provide better correlation compared to the AOD at 10 km resolution. The increments of the correlation coefficient are in general not significant, but still it is clear that AOD at a higher spatial resolution better describes the PM_{2.5} concentration at ground level.

The slope of the linear fit does always increase with increased spatial resolution. This means that for the same PM_{2.5} measurement the corresponding AOD retrieved at 1 km should be lower than the AODs retrieved at 3 km and 10 km. A possible explanation for this pattern is that for a

Figure 5.3: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in the US in July 2008 at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure 5.4: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in the US in October 2008 at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

larger pixel of 10 km a small cloud can be harder to distinguish and successfully remove leading to overestimated AOD values and hence a gentler slope.

By studying the correlation coefficients for the different months a pattern of seasonal variability of the linear correlation between AOD and $PM_{2.5}$ can be observed. The highest correlation occurs in summer time with values of the correlation coefficient up to 0.587. In winter time the correlation reaches its minimum with values down to 0.143. In spring and autumn the correlations are quite similar and generally higher compared to wintertime correlation. Similar seasonal variations in the correlation between AOD and $PM_{2.5}$ concentrations have been presented by previous research for major parts of the US as shown in Chapter 1. Two possible underlying reasons for the seasonal variation and the considerably higher correlation in July could be the variability in AOD data quality and meteorological conditions.

With higher AOD quality a better correlation with $PM_{2.5}$ concentrations is expected. One major source of reduced AOD quality is the surface reflectivity. With a low surface reflectivity the aerosol signal is stronger leading to higher AOD quality (Lyapustin et al., 2011). During summertime large parts of the US is covered by green vegetation which reduces the surface reflectance compared to wintertime when the area covered by vegetation is smaller. So it is reasonable to think that the AOD's quality and hence its correlation with $PM_{2.5}$ concentrations should follow the seasonal pattern of surface vegetation. By studying the Figures 5.1 - 5.4 and Table 5.1 it can be seen that this is the case.

As described in Section 2.5 the AOD- $PM_{2.5}$ correlation should be higher if the BLH is high and the RH is low due to the differences between AOD and $PM_{2.5}$. Figure 5.5 and 5.6 shows the relative humidity and boundary layer height respectively for January (left) and July (right). It is clear that in July both the RH is lower and the BLH higher compared to January which supports the idea that low RH and high BLH favors the linear correlation between AOD and the $PM_{2.5}$ concentrations.

Figure 5.5: Relative humidity at surface level (1000 hPa) in January (left) and July (right) 2008

Figure 5.6: Boundary layer height in January (left) and July (right) 2008

5.1.2 Study II - Regions of Varying Spatial Scale

In this section the linear relationship between MAIAC AOD and $\text{PM}_{2.5}$ concentrations is investigated for different spatial resolutions as well as for regions of different spatial scales. The relationship is investigated for six US cities, each representing three spatial scales; urban scale, meso-scale and continental scale. Details about the different spatial scales of the study regions can be found in Section 3.2. In this study the four months are combined in order to maintain a sufficient number of points.

Previous research have presented significant differences in the AOD- $\text{PM}_{2.5}$ correlation between the eastern and western parts of the US (e.g. Zhang et al., 2009; Paciorek et al., 2008; Toth et al., 2014). For this reason the two sides of US have been studied separately and it is important to understand the differences in surface and aerosol properties between the two parts. The surface in the eastern part of the US is covered by green vegetation and fine particles is strongly dominating in the atmosphere. The western part of the US on the other hand, has in general a more desert like surface with comparably low amounts of vegetation and with higher amounts of coarse particles.

5.1.2.1 The Eastern Part of the US

For the study over the eastern part of the US, Washington D.C, Atlanta and Indianapolis have been analyzed (see Figure 3.7 for geographical locations). Table 5.2 shows the statistical information for the linear correlation between AOD and $\text{PM}_{2.5}$ for the three cities and the corresponding study regions for the three spatial resolutions. Please note that the continental scale is the same for all six cities and repeated only for easier comparison for the individual cities. The corresponding scatter plots are not included in the chapter but can be found in Appendix A for better visualization of the results.

Table 5.2: Statistical data related to the linear relationship between MAIAC AODs and PM_{2.5} concentrations in the eastern part of the US 2008 for different spatial scales

	Region Size	Statistics	MAIAC 10 km	MAIAC 3 km	MAIAC 1 km
Washington D.C	Urban scale	<i>N</i>	191	152	123
		<i>R</i>	0.634	0.661	0.665
		Slope	32.35	33.41	35.66
		Intersect	6.84	7.24	7.32
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Meso-scale	<i>N</i>	4994	3992	3256
		<i>R</i>	0.514	0.5317	0.5323
		Slope	27.51	30.75	33.83
		Intersect	7.83	7.64	7.49
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Continental scale	<i>N</i>	24749	19973	16291
		<i>R</i>	0.452	0.468	0.478
		Slope	25.72	28.81	31.61
		Intersect	7.29	7.14	7.01
		<i>p</i>	< 0.001	< 0.001	< 0.001
Atlanta	Urban scale	<i>N</i>	211	149	135
		<i>R</i>	0.627	0.703	0.759
		Slope	28.94	36.35	37.65
		Intersect	8.35	7.54	7.22
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Meso-scale	<i>N</i>	3244	2554	2080
		<i>R</i>	0.564	0.583	0.586
		Slope	26.01	28.76	31.56
		Intersect	7.75	7.70	7.62
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Continental scale	<i>N</i>	24749	19973	16291
		<i>R</i>	0.452	0.468	0.478
		Slope	25.72	28.81	31.61
		Intersect	7.29	7.14	7.01
		<i>p</i>	< 0.001	< 0.001	< 0.001
Indianapolis	Urban scale	<i>N</i>	220	141	119
		<i>R</i>	0.555	0.680	0.662
		Slope	30.28	46.10	47.53
		Intersect	7.66	5.97	5.80
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Meso-scale	<i>N</i>	3925	3238	2571
		<i>R</i>	0.478	0.502	0.508
		Slope	25.87	30.32	32.72
		Intersect	7.80	7.46	7.35
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Continental scale	<i>N</i>	24749	19973	16291
		<i>R</i>	0.452	0.468	0.478
		Slope	25.72	28.81	31.61
		Intersect	7.29	7.14	7.01
		<i>p</i>	< 0.001	< 0.001	< 0.001

By studying Table 5.2 from left to right it can be seen that similarly to the study of the whole US the slope increases with increased spatial resolution. Additionally it can be seen that with only one exception the linear correlation increases with increased spatial resolution, for all cities and spatial scales. This indicates that the theoretical hypothesis does not only hold for the continental scale as shown in the previous section but also for study regions of smaller spatial scales. One example is for Atlanta at urban scale where the correlation increases from 0.627 to 0.703 to 0.759 when the spatial resolution increases from 10 km to 3 km and 1 km respectively. The

one exception from this pattern occurs for Indianapolis at urban scale where the correlation decreases from 0.680 to 0.662 when the spatial resolution is increased from 3 km to 1 km. The reason for this exception could be related to the spatial variability of the AOD in Indianapolis. If the variability is very low, i.e. the opposite to what is visualized in Figure 1.1, a lower resolution could be sufficient to catch the $\text{PM}_{2.5}$ concentration pattern at ground level and a very high resolution could introduce enough error to overcome the theoretical advantage of the higher resolution leading to a lower correlation. Figure 5.7 shows the AOD at 1 km spatial resolution (from MAIAC) over Indianapolis (left) and Atlanta (right) at urban scale averaged over the respective days when matching AODs and $\text{PM}_{2.5}$ concentrations were found. In Atlanta the correlation increases significantly when the spatial resolution is increased from both 10 km to 3 km and from 3 km to 1 km, which indicates a comparably high spatial variability of the AOD which requires a high spatial resolution for good $\text{PM}_{2.5}$ prediction capability.

Figure 5.7: AOD over Indianapolis (left) and Atlanta (right) at urban scale (visualization of the AOD spatial variability). The red stars show the respective city centers and the black dots the $\text{PM}_{2.5}$ monitoring stations included in the urban scale analysis.

From the two sub-figures it is clear that within the area defined as the respective urban scales the spatial AOD variability is comparably lower in Indianapolis. Also for the nearby pixels of the respective $\text{PM}_{2.5}$ monitoring stations the variability is comparably lower in Indianapolis compared to Atlanta. The difference is not significant but it is in line with idea that the variability in Indianapolis is too low for the AOD at 1 km spatial resolution to further improve the linear correlation with the $\text{PM}_{2.5}$ concentrations, causing the exception from the clear pattern of better linear correlation with higher spatial resolution.

Two other clear patterns in the AOD- $\text{PM}_{2.5}$ correlation in the eastern part of the US are (1) the decrease of the linear regression slope and (2) the decrease in correlation with increased area of the study region. For all three cities and three spatial resolutions both the slope and correlation decreases when the study region is enlarged from urban scale to meso- and continental scale. Those patterns are reasonable and most likely related to the fine mode fraction (FMF) of the study scene. The FMF describes the contribution of fine particles to the retrieved AOD. As mentioned before the FMF varies between zero and one and a high value indicates a study scene where mainly fine particles contribute to the AOD while a low value indicates a study scene where mainly coarse particles contribute to the AOD. The FMF at 10 km spatial resolution (from MYD04) in the contiguous US can be seen in Figure 5.8 and the major pattern is that in general the FMF is much higher over the eastern part of the US where the surface is vegetated compared to the western part where the surface is more desert like. The problem with a low FMF and more coarse particles is that the $\text{PM}_{2.5}$ instruments are designed to measure the mass concentrations

of particles smaller than $2.5\ \mu\text{m}$ and do not take those coarse particles into account. So for a study scene with a low FMF the satellite will retrieve a too high AOD compared to the measured $\text{PM}_{2.5}$ concentration due to the extinction by coarse particles. Those too high AODs compared to the $\text{PM}_{2.5}$ concentrations will lead to decreased correlation as well as gentler regression slopes as illustrated in Figure 2.9.

When the eastern cities are studied at urban scale the FMF is high due to local emissions of fine aerosols, but as the area of the study region is increased regions with lower FMF and more coarse particles are included which reduces the linear correlation and flattens the regression slopes. This is especially clear for the increase to continental scale when the western part of US is included, where the FMF is considerably lower.

Figure 5.8: The aerosol fine mode fraction over the US for the four months in 2008. The white stars shows the location of the six cities.

5.1.2.2 The Western Part of US

For the study over the western part of US, Salt Lake City, Phoenix and Los Angeles have been analyzed (see Figure 3.7 for geographical locations). Table 5.3 shows the statistical information for the linear correlation between AOD and $\text{PM}_{2.5}$ for the three cities and the corresponding spatial scales for the three spatial resolutions. Please note that the continental scale is the same for all six cities and repeated only for easier comparison for the individual cities. The corresponding scatter plots are not included in the chapter but can be found in Appendix A.

Table 5.3: Statistical data related to the linear relationship between MAIAC AODs and PM_{2.5} concentrations in the western part of the US 2008 for different spatial scales

	Region Size	Statistics	MAIAC 10 km	MAIAC 3 km	MAIAC 1 km
Salt Lake City	Urban scale	<i>N</i>	215	176	144
		<i>R</i>	-0.165	-0.058	0.070
		Slope	-14.92	-3.78	3.66
		Intersect	12.75	10.74	9.76
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Meso-scale	<i>N</i>	875	696	571
		<i>R</i>	-0.030	0.111	0.153
		Slope	-2.47	7.40	10.24
		Intersect	10.01	8.46	8.26
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Continental scale	<i>N</i>	24749	19973	16291
		<i>R</i>	0.452	0.468	0.478
		Slope	25.72	28.81	31.61
		Intersect	7.29	7.14	7.01
		<i>p</i>	< 0.001	< 0.001	< 0.001
Phoenix	Urban scale	<i>N</i>	63	58	54
		<i>R</i>	-0.118	0.021	0.117
		Slope	-6.60	1.31	7.58
		Intersect	10.72	10.04	9.53
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Meso-scale	<i>N</i>	609	528	484
		<i>R</i>	0.125	0.141	0.188
		Slope	10.62	13.36	17.66
		Intersect	7.81	8.06	7.34
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Continental scale	<i>N</i>	24749	19973	16291
		<i>R</i>	0.452	0.468	0.478
		Slope	25.72	28.81	31.61
		Intersect	7.29	7.14	7.01
		<i>p</i>	< 0.001	< 0.001	< 0.001
Los Angeles	Urban scale	<i>N</i>	300	272	249
		<i>R</i>	0.019	0.049	0.053
		Slope	1.64	4.46	5.06
		Intersect	14.91	14.91	14.95
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Meso-scale	<i>N</i>	2232	2061	1871
		<i>R</i>	0.250	0.275	0.299
		Slope	22.06	24.93	28.92
		Intersect	10.23	10.02	9.82
		<i>p</i>	< 0.001	< 0.001	< 0.001
	Continental scale	<i>N</i>	24749	19973	16291
		<i>R</i>	0.452	0.468	0.478
		Slope	25.72	28.81	31.61
		Intersect	7.29	7.14	7.01
		<i>p</i>	< 0.001	< 0.001	< 0.001

Similarly to the study over the western part there are strong indications that a higher spatial resolution increases the correlation both at city, meso- and continental scale. For the eastern cities and their corresponding study regions this is always the case without any exceptions.

For Phoenix and Salt Lake City negative correlations are retrieved which means that with an increased AOD the PM_{2.5} concentration should decrease, which is not in line with the physical understanding of the AOD-PM_{2.5} correlation. The negative correlation is thought to be related to the FME. If the extinction by coarse particles is very high the corresponding AODs will be

overestimated for the measured $\text{PM}_{2.5}$ concentrations which can lead to negative correlations (as visualized in Figure 2.9). Figure 5.9 shows the FMF at 3 km spatial resolution (MYD04) over Salt Lake City (left) and Atlanta (right) at urban scale averaged over the respective days when matching AODs and $\text{PM}_{2.5}$ concentrations were found. Atlanta, where no negative correlations were found, is again used for comparison reasons.

Figure 5.9: Fine mode fraction over Salt Lake City (left) and Atlanta (right) at urban scale. The red stars show the respective city centers and the black dots the $\text{PM}_{2.5}$ monitoring stations included in the urban scale analysis.

From Figure 5.9 it is clear that the FMF is much lower over Salt Lake City compared to Atlanta, which supports the idea that the FMF could be related to the negative correlation in Salt Lake City at urban scale. It is reasonable to believe that this also can explain the negative correlation in Phoenix were the FMF at urban scale is very low as well.

The change in the linear AOD- $\text{PM}_{2.5}$ correlations and the regression slopes with increased area of the study region is completely reverse compared to the eastern part of the US and both the correlations and the slopes increases with enlarged study region for all three cities and spatial resolutions. Again this is probably related to the FMF. By studying Figure 5.8 and the position of the three western cities it can be seen that all of them are located in regions with very low FMF. At urban scale it seems like the coarse particles strongly dominates despite the local emission of fine particles (as seen in Figure 5.9 (left) for Salt Lake city and 5.16 for Los Angeles) resulting in the lowest correlation at urban scale, but as the study area is enlarged regions with higher FMF and better AOD- $\text{PM}_{2.5}$ correlation are included which increases the total correlations and the regression slopes.

It is clear that the two parts of the US provides different correlations, with much better $\text{PM}_{2.5}$ prediction capabilities in the eastern part of the US and generally gentler regression slopes in the western part. The last sections have pointed out the importance of the coarse particles and the FMF in this pattern, but there are other factors as well. For example the surface characteristics; in the eastern part the surface is covered by green vegetation, which has been explained to improve the AOD retrieval and hence the AOD- $\text{PM}_{2.5}$ correlation. For the western part on the other hand the surface is brighter which reduces the AOD quality and hence the AOD- $\text{PM}_{2.5}$ quality.

5.2 MYD04 AOD Data

The MODIS aerosol science team provides MYD04 AOD data at 10 km and 3 km spatial resolution. In the following subsections the linear relationship between the two MYD04 AOD data products and the $PM_{2.5}$ concentrations will be presented in a similar way as in the previous sections. It should be clarified that even though both the MAIAC AOD data and the MYD04 AOD data are based on data from MODIS they cannot be expected to provide the same results since the data originates from different algorithms.

5.2.1 Study I - Region of Continental Spatial Scale

Figure 5.10 shows the AODs and corresponding $PM_{2.5}$ concentrations in January 2008 together with the linear regressions at 10 km (left) and 3 km (right) spatial resolution. The similar results for April, July and October are shown in Figures 5.11, 5.12 and 5.13 respectively. Table 5.4 summarizes the statistical information related to the retrieved linear relationships for the four months.

Table 5.4: Statistical data related to the linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations for the whole US

	Statistics	MYD04 10 km	MYD04 3 km
January	<i>N</i>	290	402
	<i>R</i>	0.428	0.471
	Slope	57.04	44.00
	Intersect	8.79	7.25
	<i>p</i>	< 0.001	< 0.001
April	<i>N</i>	1228	849
	<i>R</i>	0.238	0.244
	Slope	9.69	7.79
	Intersect	9.80	9.60
	<i>p</i>	< 0.001	< 0.001
July	<i>N</i>	1359	853
	<i>R</i>	0.616	0.641
	Slope	24.22	23.27
	Intersect	7.93	6.41
	<i>p</i>	< 0.001	< 0.001
October	<i>N</i>	1808	1212
	<i>R</i>	0.379	0.399
	Slope	22.39	17.71
	Intersect	7.56	7.04
	<i>p</i>	< 0.001	< 0.001

Similarly as for the MAIAC AOD the increase in correlation with increased spatial resolution is consistent, with a higher correlation using the 3 km AOD compared to the 10 km AOD for all months. The pattern of the intersection is similar as well with rather low and stable intersections for the separate combinations of month and spatial resolution. Additionally it can be seen that the steepness of the linear regression slopes does generally decrease with increased resolution. As mentioned in the previous section a possible explanation for this pattern is that for a larger pixel of 10 km a small cloud can be harder to distinguish and successfully remove leading to overestimated AOD values and hence a gentler slope.

Figure 5.10: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in the US in January 2008 at 10 km (left) and 3 km (right) spatial resolution.

Figure 5.11: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in the US in April 2008 at 10 km (left) and 3 km (right) spatial resolution.

A seasonal pattern in the correlation can be observed using the MYD04 AOD data as well and similarly to the MAIAC AOD- $PM_{2.5}$ results the correlation is highest in July with correlations up to 0.641 using 3 km MYD04 AOD. For the MYD04 the worst correlation is obtained in April for both 10 km and 3 km spatial resolution with 0.238 as the lowest correlation. Zhang et al. (2009) presented a similar seasonal pattern in the AOD- $PM_{2.5}$ correlation pattern for the western part of the US, which might indicate that using the MYD04 a larger fraction of the total amount of AOD- $PM_{2.5}$ match-ups are retrieved over the western part of the US. It could also indicate that more factors than those investigated in this thesis have a large impact on the seasonal variation in the AOD- $PM_{2.5}$ correlation.

5.2.2 Study II - Regions of Varying Spatial Scale

In this section the linear relationship between MYD04 AOD and $PM_{2.5}$ concentrations is investigated for different spatial resolutions as well as for regions of different spatial scales. Like for the MAIAC AOD data the relationship is investigated for six US cities, each representing the spatial scales; urban scale, meso-scale and continental scale. The four months have been combined for this part as well to maximize the number of points; however the points are in some cases low and the corresponding results may not be as trustful as before where the number of points were comparably higher. As before the analysis has been divided into two parts based on the different characteristics between the eastern and western parts of US.

Figure 5.12: Linear relationship between MYD04 AODs and PM_{2.5} concentrations in the US in July 2008 at 10 km (left) and 3 km (right) spatial resolution.

Figure 5.13: Linear relationship between MYD04 AODs and PM_{2.5} concentrations in the US in October 2008 at 10 km (left) and 3 km (right) spatial resolution.

5.2.2.1 The Eastern Part of the US

Table 5.5 shows the statistical information related to the linear correlation between AOD and PM_{2.5} for Washington D.C, Atlanta and Indianapolis, the corresponding study regions and the two spatial resolutions. Please note that the continental scale is the same for all six cities and repeated only for easier comparison for the individual cities. The corresponding scatter plots are not included in the chapter but can be found in Appendix A for better visualization of the results.

Table 5.5: Statistical data related to the linear relationship between MYD04 AODs and PM_{2.5} concentrations in the eastern part of the US 2008 for different spatial scales

City	Region Size	Statistics	MYD04 10 km	MYD04 3 km
Washington D.C	Urban scale	<i>N</i>	27	16
		<i>R</i>	0.883	0.878
		Slope	41.52	20.78
		Intersect	2.54	6.01
		<i>p</i>	< 0.001	< 0.001
	Meso-scale	<i>N</i>	849	673
		<i>R</i>	0.558	0.505
		Slope	27.72	18.58
		Intersect	8.10	8.43
		<i>p</i>	< 0.001	< 0.001
	Continental scale	<i>N</i>	4698	3315
		<i>R</i>	0.546	0.565
		Slope	22.31	20.03
		Intersect	8.08	7.27
		<i>p</i>	< 0.001	< 0.001
Atlanta	Urban scale	<i>N</i>	37	16
		<i>R</i>	0.904	0.938
		Slope	39.31	35.98
		Intersect	7.47	7.75
		<i>p</i>	< 0.001	< 0.001
	Meso-scale	<i>N</i>	753	512
		<i>R</i>	0.695	0.729
		Slope	30.21	31.20
		Intersect	8.51	7.77
		<i>p</i>	< 0.001	< 0.001
	Continental scale	<i>N</i>	4698	3315
		<i>R</i>	0.546	0.565
		Slope	22.31	20.03
		Intersect	8.08	7.27
		<i>p</i>	< 0.001	< 0.001
Indianapolis	Urban scale	<i>N</i>	36	35
		<i>R</i>	0.599	0.624
		Slope	23.94	23.95
		Intersect	5.98	4.20
		<i>p</i>	< 0.001	< 0.001
	Meso-scale	<i>N</i>	869	596
		<i>R</i>	0.587	0.529
		Slope	28.33	19.68
		Intersect	8.09	8.07
		<i>p</i>	< 0.001	< 0.001
	Continental scale	<i>N</i>	4698	3315
		<i>R</i>	0.546	0.565
		Slope	22.31	20.03
		Intersect	8.08	7.27
		<i>p</i>	< 0.001	< 0.001

Generally the correlation between MYD04 AOD and the $PM_{2.5}$ concentrations is high with values of the correlation coefficient between 0.505 and 0.938. Regarding the impact of the spatial resolution on the correlation the results are not as consistent for the eastern part of the US as for the MAIAC AOD data. At both urban scale and meso-scale for Washington D.C the correlation decreases when the resolution is increased. This is also the case for Indianapolis at meso-scale. For Washington D.C at urban scale the decrease is very slight and with only 16 points at 3 km spatial resolution the results are not very trustful and very sensitive to uncertainty and outliers that may have slipped through the empirical removal criteria. Figure 5.14 shows the scatter plot for Washington D.C at urban scale with AOD at 3 km resolution. And it can be seen that there is one outlier that has not been removed successfully, where the AOD is comparably high with respect to the measured $PM_{2.5}$ concentration. Removal of this point leads to an increased correlation to 0.891, which means that without this point the hypothesis holds for Washington D.C at urban scale as well with a higher correlation using the higher spatial resolution.

Figure 5.14: Scatter plot with MYD04 AOD at 3 km resolution and $PM_{2.5}$. Visualization of the impact of an outlier on the linear correlation.

The results using MYD04 AOD data presents strong indications that the AOD- $PM_{2.5}$ correlation and the regression slope decreases with increased area of the study region. The only exceptions from this pattern are when the study region is enlarged from meso- to continental scale for Washington D.C and Indianapolis using 3 km AOD. It is probably not a coincidence that those are the exceptions mentioned above as well, where the correlation did not increase when the resolution was increased from 10 km to 3 km. It is also likely to believe that the correlation and slope does not decrease for the shift between meso- and urban scale for Washington D.C and Indianapolis using 3 km AOD due to the fact that the correlation does not increase as expected at meso-scale when the resolution is increased. The behavior of those exceptions, with an increase in both correlation and slope when the study region is enlarged from meso- to continental scale could indicate a comparably high FMF for the meso-scale for those study regions. However it was shown in the previous section that for the eastern cities the total FMF decreases gradually for each increase in size of the study region. It is believed that the underlying reason(s) for those two exceptions is either related to the increased amount of error that is introduced for the 3 km resolution and mentioned as a potential problem for this comparably high resolution (Levy et al., 2013) or other aerosol/atmospheric parameters not discussed in this thesis.

5.2.2.2 The Western Part of US

Table 5.6 shows the statistical information related to the linear correlation between MYD04 AOD and $PM_{2.5}$ for Salt Lake City, Phoenix and Los Angeles, their corresponding study regions and the two spatial resolutions of the MYD04 AOD. Please note that the continental scale is the same for all six cities and repeated only for easier comparison for the individual cities. The corresponding scatter plots are not included in the chapter but can be found in Appendix A.

Table 5.6: Statistical data related to the linear relationship between MAIAC AODs and PM_{2.5} concentrations in the western part of the US 2008 for different spatial scales

City	Region Size	Statistics	MYD04 10 km	MYD04 3 km
Salt Lake City	Urban scale	N	35	27
		R	-0.153	-0.118
		Slope	-2.19	-1.58
		Intersect	11.98	11.41
		p	< 0.001	< 0.001
	Meso-scale	N	128	125
		R	0.117	0.502
		Slope	2.88	9.45
		Intersect	9.52	5.94
		p	< 0.001	< 0.001
	Continental scale	N	4698	3315
		R	0.546	0.565
		Slope	22.31	20.03
		Intersect	8.08	7.27
		p	< 0.001	< 0.001
Phoenix	Urban scale	N	9	5
		R	-0.296	-0.111
		Slope	-2.35	-2.19
		Intersect	11.48	12.23
		p	< 0.001	< 0.001
	Meso-scale	N	60	35
		R	0.239	0.524
		Slope	4.22	9.25
		Intersect	7.80	5.98
		p	< 0.001	< 0.001
	Continental scale	N	4698	3315
		R	0.546	0.565
		Slope	22.31	20.03
		Intersect	8.08	7.27
		p	< 0.001	< 0.001
Los Angeles	Urban scale	N	47	46
		R	0.249	-0.143
		Slope	4.88	-4.05
		Intersect	13.27	17.45
		p	< 0.001	< 0.001
	Meso-scale	N	493	426
		R	0.279	0.322
		Slope	9.10	10.70
		Intersect	11.70	10.54
		p	< 0.001	< 0.001
	Continental scale	N	4698	3315
		R	0.546	0.565
		Slope	22.31	20.03
		Intersect	8.08	7.27
		p	< 0.001	< 0.001

For the western cities and the corresponding study regions the indications are again strong that a higher resolution increases the linear AOD-PM_{2.5} correlation. With only one exception for all cities and spatial scales the correlation coefficient increases when MYD04 AOD at 3 km resolution is used compared to the MYD04 AOD at 10 km. The only exception that occurs is for Los Angeles at urban scale where the correlation is significantly reduced when the resolution is increased to 3 km. Figure 5.15 shows the AOD over Los Angeles at 3 km spatial resolution (MYD04) at urban scale averaged over the respective days when matching AODs and PM_{2.5} concentrations were found. It can be seen that the spatial AOD variability is high which indicates the need of a high spatial resolution for a good PM_{2.5} representation. So the low correlation at urban scale cannot be explained due to low spatial AOD variability as for Indianapolis using the MAIAC AOD data. Another possible explanation could be if the 3 km AOD better retrieves the large amounts of coarse particles in Los Angeles compared to the 10 km AOD, which would result in a decrease in the correlation for the 3 km AOD. Figure 5.16 shows the FMF over Los Angeles at 10 km (left) and 3 km (right) spatial resolution (MYD04) at urban scale averaged over the respective days when matching AODs and PM_{2.5} concentrations were found and it is clear that this is not the case. Both at 10 km and 3 km spatial resolution the high FMF is observed around all PM_{2.5} monitoring stations. With regards to this, this exception is again believed to be related either to the lower quality of the 3 km MYD04 AOD data or other factors not taken into account in this thesis.

Figure 5.15: AOD over Los Angeles at 3 km resolution (visualization of the high spatial AOD variability). The red star shows the city center and the black dots the PM_{2.5} monitoring stations included in the urban scale analysis.

As for the MAIAC AOD data negative correlations and slopes are obtained at urban scale for the western cities, but to a greater extent using the MYD04 AOD data. The underlying reason is assumed to be related to the low FMF as explained and visualized before. The greater extent of negative values could be related to the fact that fewer points are obtained using the MYD04 data or the fact that AOD retrieval over bright surfaces is more accurate using MAIAC compared to MYD04 (Lyapustin et al., 2011).

Again a very clear pattern of increased AOD-PM_{2.5} correlation and regression slope with increased size of the study region for the western cities was obtained. Without any exceptions the correlation and the regression slope increases when the study region is enlarged from urban scale to meso and continental scale respectively, for both 10 km and 3 km AOD.

As before it is clear that the correlation varies between the eastern and western parts of the US with considerably higher correlations at both city and meso-scale for the eastern part of the US. However the 3 km MYD04 AOD can actually represent the PM_{2.5} concentrations quite well even

Figure 5.16: Fine mode fraction over Los Angeles at 10 km (left) and 3 km (right) spatial resolution. The red stars show the respective city centers and the black dots the $\text{PM}_{2.5}$ monitoring stations included in the urban scale analysis.

for the western part of the US at meso-scale.

5.3 Comparison of MAIAC AOD and MYD04 AOD Data

As previously explained it is not fair to directly compare the results using data from two different algorithms due to the differences in the retrieved AODs as visualized in Figure 3.6 and identical results cannot be expected for the two data products. However there are a lot of similar patterns that can be seen in the AOD- $\text{PM}_{2.5}$ correlation using both the MAIAC data and the MYD04.

For the continental scale analysis in *Study I* both AOD data sets provides a similar pattern for the AOD- $\text{PM}_{2.5}$ correlation with increased spatial resolution. For both data sets and all four months the correlation did always increase with higher spatial resolution. Regarding the seasonal variations the highest correlation was obtained in July for both data sets. However the month with lowest correlation was not consistent. Using the MAIAC data the correlation was lowest in January which is in line with the previous research related to the seasonal variations of the AOD- $\text{PM}_{2.5}$ correlation in the US. For the MYD04 AOD the lowest correlation was obtained in April instead. Zhang et al. (2009) showed a similar pattern for the western part of the US which might indicate that a larger fraction of the total number of match-ups are from the western part of the US for the MYD04 AOD compared to the MAIAC AOD.

For the study of varying spatial scales in *Study II* both AOD data products showed similar patterns with general increments in correlation with higher spatial resolution and smaller (larger) study region for the eastern (western) part of the US. The patterns were not as consistent for the MYD04 AOD data as for the MAIAC AOD data, which might indicate that the decrease in AOD quality is larger between 10 km and 3 km for the MYD04 AOD data compared to the corresponding resolution increase for the MAIAC data since the two MYD04 data products have been derived separately while the coarser MAIAC resolutions are derived by simple aggregation of the MAIAC AOD data at 1 km resolution. Another important factor is the number of points; for the MYD04 AOD data the number of match-ups are considerably lower compared to the MAIAC data and especially at urban scale the points can in some cases be too few in order to give reliable results.

Another notable pattern is that for all spatial scales and study regions the correlation is generally higher for the MYD04 AOD even compared to the MAIAC AOD at 1 km resolution.

5.4 Results Using RH and BLH Corrections

Despite a lot of work the corrections of the AOD did not turn out right, in the majority of cases the correlation was reduced when the corrections were applied which is the opposite to the expectations. The corrections were both tested separately and together, but with similar results. For this reason no results including the RH and/or BLH corrections are included in this thesis.

The reason why the boundary layer did not work properly is most likely due to the assumption that all atmospheric particles are homogeneously distributed within the boundary layer. For a good estimation of the AOD near the surface the vertical profile of the aerosol distribution would be needed. The RH corrections were originally derived for another region than those studied in this thesis, which most likely is one reason why it was not successful for this study. Additionally the data provided by ECMWF is modeled and cannot provide the exact BLH and RH conditions at the time for the $PM_{2.5}$ concentration measurements and AOD retrievals.

Chapter 6 Conclusions

The effect of the AOD spatial resolution have been analyzed for the linear relationship between AOD retrieved from MODIS and ground level $PM_{2.5}$ concentrations measured by 946 $PM_{2.5}$ monitoring stations across the contiguous US. AOD from the MODIS standard product (MYD04) and the MAIAC algorithm have been used to compare the correlation at 1 km (MAIAC), 3 km (MYD04 and MAIAC) and 10 km (MYD04 and MAIAC) spatial resolution. The study has been divided into two parts, in *Study I* the spatial resolution effect on the AOD- $PM_{2.5}$ correlation is analyzed for the whole contiguous US while in *Study II* the spatial resolution effect is analyzed for different spatial scales.

Without taking the spatial resolution effect into account the study confirms two major patterns from previous research related to the linear AOD- $PM_{2.5}$ correlation:

- the correlation is in general higher during summer and fall and lower during winter and spring
- the correlation is higher in the eastern part of the US compared to the western part of the US.

Regarding the spatial resolution effect very strong indications were obtained saying that the correlation does increase if the spatial resolution of the AOD is increased which is in line with the theoretical hypothesis stated and described in the first chapter. The continental scale analysis did also show that the correlation increased with higher spatial resolution for the four seasons separately. The correlation was shown to increase with higher spatial resolution not only for the continental scale study of the whole contiguous US but for the urban scale and meso-scale analysis as well. This means that for the spatial scales investigated in this thesis the increased amount of error for the higher resolutions is not large enough to overcome the advantage of better spatial variability representation using a higher resolution and the spatial resolution could probably be further increased for slightly better AOD- $PM_{2.5}$ correlation.

The correlation was also shown to vary consistently with the area of the studied region for all spatial resolutions of the AOD. The pattern of the correlation variation did however change differently depending on the geographical location of the study regions. For the eastern part of the US the correlation was shown to decrease with an enlarged study region, while for the western part of the US the correlation was shown to increase when a larger study region was analyzed, by studying the similar pattern of the linear regression slope this was concluded to be related to the differences in the fine mode fraction across the US, with generally higher amounts of coarse particles in the western part of US, which is unfavorable for $PM_{2.5}$ prediction using AOD.

From the results it can be seen that the AOD cannot always be used directly to predict the $PM_{2.5}$ concentrations, especially not at the smaller scales in the western part of the US. The optimal conditions for an accurate $PM_{2.5}$ prediction have been shown to be a high spatial resolution of the AOD and a small study region in the eastern part of the US with a high a boundary layer height, low relative humidity and low concentrations of coarse particles.

6.1 Possible Improvements for Further Work

For further work related to this research topic there are some major parts that could improve the general understanding as well as the derived AOD- $PM_{2.5}$ correlations. First of all the study could

be extended with regards to the study regions. In this thesis only the contiguous US has been analyzed, similar studies for other regions such as Europe and Asia would help to further improve the understanding of the spatial resolution effect on the AOD-PM_{2.5} correlation. To analyze the AOD-PM_{2.5} correlation using AOD data from more satellite sensors would also be valuable for the further understanding of the spatial resolution effect. Regarding the AOD-PM_{2.5} correlation it could probably have been improved if more time would have been available for this thesis. Five suggestions that could improve the results are:

1. Improve the BLH and RH humidity corrections. Theoretically the corrections should yield better correlation and if more time would be spent on this part the AOD would probably be able to represent the PM_{2.5} concentrations better.
2. Study the vertical distribution of the aerosols using LIDAR data. The AOD data itself do not provide any information where in the atmospheric column the aerosols are located which makes it hard to judge whether a retrieved AOD is representative or not for a PM_{2.5} concentrations that measures the concentration at surface level. The vertical profile would be an important part for an improved BLH correction as well.
3. The empirical selection criteria used to remove outliers could be improved. For this thesis the aim was to have rather simple selection criteria applicable for both the MAIAC and MYD04 AOD data and all spatial resolutions, study regions and spatial scales. To develop more criteria, for example one for each data set and study region would probably improve the removal of outliers and hence improve the results.
4. Investigate the fine mode fraction further. From the results it was clear that the fine mode fraction had a large impact on the correlation. Implementing a successful correction for removal of the coarse particle contribution to the retrieved AOD would surely improve the correlation considerably for the whole US in general and for the western part in particular.
5. Increase the temporal resolution of the PM_{2.5} data. In this study where daily mean concentrations of PM_{2.5} have been used the AOD data retrieved at the time of the satellite overpass have been assumed to be representative for the whole day (i.e. the same time period as the sampling time for the PM_{2.5} concentrations). To use hourly data would provide more representative AOD-PM_{2.5} match-ups with increased correlation as an expected result.

Bibliography

- Berlin, P. *Satellite Platform Design*. Department of Computer Science, Electrical and Space Engineering, 5 edition, 2007.
- Burrows, J. P., Platt, U., and Borrell, P., editors. *The Remote Sensing of Tropospheric Composition from Space*. Springer, 2011.
- Chudnovsky, A., Tang, C., Lyapustin, A., Wang, Y., Schwartz, J., and Koutrakis, P. A critical assessment of high resolution aerosol optical depth (AOD) retrievals for fine particulate matter (PM) predictions. *Atmospheric Chemistry and Physics*, 13:14581–14611, 2013b. doi:[10.5194/acpd-13-14581-2013](https://doi.org/10.5194/acpd-13-14581-2013).
- Chudnovsky, A. A., Kostinski, A., Lyapustin, A., and Koutrakis, P. Spatial scales of pollution from variable resolution satellite imaging. *Environmental Pollution*, 172:131–138, 2013a. doi:[10.1016/j.envpol.2012.08.016](https://doi.org/10.1016/j.envpol.2012.08.016).
- Clarke, A. D. and Noone, K. J. Soot in the arctic snowpack: A cause for perturbations in radiative transfer. *Atmospheric Environment*, 19(12):2045–2053, 1985. doi:[10.1016/0004-6981\(85\)90113-1](https://doi.org/10.1016/0004-6981(85)90113-1).
- Dutton, E. G. and Christ, J. R. Solar radiative forcing at selected locations and evidence for global lower tropospheric cooling following the eruptions of el chichon and pinatubo. *Geophysical Research Letter*, 19(23):2313–2316, 1992. doi:[10.1029/92GL02495](https://doi.org/10.1029/92GL02495).
- EEA. Air quality in Europe — 2013 report. Technical report, European Environment Agency, 2013.
- Engel-Cox, J. A., Holloman, C. H., Coutant, B. W., and Hoff, R. M. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment*, 38:2495–2509, 2004. doi:[10.1016/j.atmosenv.2004.01.039](https://doi.org/10.1016/j.atmosenv.2004.01.039).
- Engel-Cox, J. A., Hoff, R. M., Rogers, R., Dimmick, F., Rush, A. C., Szykman, J. J., Al-Saadi, J., Chu, D. A., and Zell, E. R. Integrating lidar and satellite optical depth with ambient monitoring for 3-dimensional particulate characterization. *Atmospheric Environment*, 40:8056–8067, 2006. doi:[10.1016/j.atmosenv.2006.02.039](https://doi.org/10.1016/j.atmosenv.2006.02.039).
- EPA. Revised air quality standards for particle pollution and updates to the air quality index (AQI), 2012.
- EPA. National ambient air quality standards for particulate matter; final rule, 2013. Federal Register.
- EPA, 2014. URL <http://www.epa.gov/>.
- Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y., , and Kumar, N. Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, 40:5880–5892, 2006. doi:[10.1016/j.atmosenv.2006.03.016](https://doi.org/10.1016/j.atmosenv.2006.03.016).
- Hoff, R. M. and Christopher, S. A. Remote sensing of particulate pollution from space: Have we reached the promised land? *Journal of the Air and Waste Management Association*, 59:645–675, 2009. doi:[10.3155/1047-3289.59.6.645](https://doi.org/10.3155/1047-3289.59.6.645).

- Hsu, N. C., Jeong, M.-J., Bettenhausen, C., Sayer, A. M., Hansell, R., Seftor, C. S., Huang, J., , and Tsay, S.-C. Enhanced deep blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research*, 118:9296–9315, 2013. doi:[10.1002/jgrd.50712](https://doi.org/10.1002/jgrd.50712).
- Hsu, N. C., Tsay, S.-C., King, M. D., and Herman, J. R. Aerosol properties over bright-reflecting source regions. *IEEE Transactions of Geoscience and Remote Sensing*, 42(3):557–569, 2004. doi:[10.1109/TGRS.2004.824067](https://doi.org/10.1109/TGRS.2004.824067).
- Hsu, N. C., Tsay, S.-C., King, M. D., and Herman, J. R. Deep blue retrievals of asian aerosol properties during ACE-Asia. *IEEE Transactions of Geoscience and Remote Sensing*, 44(11):3180–3195, 2006. doi:[10.1109/TGRS.2006.879540](https://doi.org/10.1109/TGRS.2006.879540).
- IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, 2013.
- Kaufman, Y. J., Tanr, D., Remer, L. A., Vermote, E. F., Chu, A., and Holben, B. N. Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *Journal of Geophysical Research*, 102:17,051–17,067, 1997.
- Kokhanovsky, A. A. and de Leeuw, G., editors. *Satellite Aerosol Remote Sensing Over Land*. Springer, 2009.
- Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C. The collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, 6:2989–3034, 2013. doi:[10.5194/amt-6-2989-2013](https://doi.org/10.5194/amt-6-2989-2013).
- Levy, R., Mattoo, S., Munchak, L., Kleidman, R., Patadia, F., and Gupta, P. MODIS atmosphere team webinar series #2: Overview of collection 6 dark-target aerosol product, 2014. URL <http://aerocenter.gsfc.nasa.gov/ext/registration/>.
- Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., , and and, Y. J. K. Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of moderate resolution imaging spectroradiometer spectral reflectance. *Journal of Geophysical Research*, 112, 2007. doi:[10.1029/2006JD007811](https://doi.org/10.1029/2006JD007811).
- Lim, S. S. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *The Lancet*, 380, 2012. doi:[10.1016/S0140-6736\(12\)61766-8](https://doi.org/10.1016/S0140-6736(12)61766-8).
- Liu, L., Wen, X., Dong, X., and Dai, Q. A new prompt algorithm for removing bowtie effect of MODIS L1b data. In *2008 Congress on Image and Signal Processing*, volume 5, pages 663–667. 2008. doi:[10.1109/CISP.2008.404](https://doi.org/10.1109/CISP.2008.404).
- Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R., and Reid, J. S. Multiangle implementation of atmospheric correction (MAIAC): 2. aerosol algorithm. *Journal of Geophysical Research*, 116, 2011. doi:[10.1029/2010JD014986](https://doi.org/10.1029/2010JD014986).
- Lyapustin, A. and Wang, Y. MAIAC: Multi-angle implementation of atmospheric correction for MODIS algorithm theoretical basis document. Technical report, Goddard Earth Science and Technology Center UMBC and NASA GSFC, 2008. URL <https://neptune.gsfc.nasa.gov/bsb/index.php?section=101>.
- McMurry, P., Shepherd, M., and Vickery, J., editors. *Particulate Matter Science for Policy Makers A NARSTO Assessment*. Cambridge University Press, 2004.
- Mishchenko, M. I., Travis, L. D., and Lacis, A. A. *Scattering, Absorption and Emission of Light by Small Particles*. Cambridge University Press, 2002.

- Mitchell, J. M. Visual range in the polar regions with particular reference to the alaskan arctic. *J. Atmos. Terr. Phys.*, pages 195–211, 1957.
- Monache, L. D., Perry, K. D., Cederwall, R. T., and Ogren, J. A. In situ aerosol profiles over the southern great plains cloud and radiation test bed site: 2. effects of mixing height on aerosol properties. *J. Geophys. Res.*, 109, 2004. doi:10.1029/2003JD004024.
- Munchak, L., Levy, R., Mattoo, S., and Remer, L. MODIS atmosphere team webinar series #5: Overview of the 3 km aerosol product in collection 6", 2014. URL <http://aerocenter.gsfc.nasa.gov/ext/registration/>.
- NASA, 2014a. URL <http://modis.gsfc.nasa.gov/data/>.
- NASA, 2014b. URL <http://earthobservatory.nasa.gov/>.
- Paciorek, C. J., Liu, Y., and Kondragunta, H. M.-M. S. Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level PM2.5. *Environmental Science and Technology*, 42:5800–5806, 2008. doi:10.1021/es703181j.
- Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., and Thurston, G. D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association*, 287(9), 2002.
- Schlitzer, R. Introduction to Inverse Methods and Mathematical Data Analysis - Lecture Script, 2006.
- Schlundt, C. Atmospheric Aerosol - Lecture Script, 2013.
- Seinfeld, J. H. and Pandis, S. N. *Atmospheric Chemistry and Physics, From Air Pollution to Climate Change*. John Wiley & Sons, Inc., 2 edition, 2006.
- Tanre, D., Kaufman, Y. J., Herman, M., and Mattoo, S. Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. *Journal of Geophysical Research*, 102: 16,971–16,988, 1997.
- Toth, T. D., Zhang, J., Campbell, J. R., Hyer, E. J., Reid, J. S., Shi, Y., and L. Westphal, D. Impact of data quality and surface-to-column representativeness on the PM2.5 / satellite AOD relationship for the contiguous United States. *Atmospheric Chemistry and Physics*, 14:6049–6062, 2014. doi:10.5194/acp-14-6049-2014.
- Twomey, S. The influence of pollution on the shortwave albedo of clouds. *J. Atmos. Sci.*, 34(7): 1149–1152, 1977. doi:10.1175/1520-0469(1977)034<1149:TIOPOT>2.0.CO;2.
- Venzac, H., Sellegri, K., Villani, P., Picard, D., and Laj, P. Seasonal variation of aerosol size distributions in the free troposphere and residual layer at the puy de dôme station, france. *Atmospheric Chemistry Physics*, 9:1465–1478, 2009. doi:10.5194/acp-9-1465-2009.
- Vermote, E., Tanré, D., Deuzé, J. L., Herman, M., Morcrette, J. J., and Kotchenova, S. Y. *Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6SV)*, 2006. 6S User Guide Version 3.
- Wallace, J. M. and Hobbs, P. V. *Atmospheric Science an Introductory Survey*. Elsevier, 2 edition, 2006.
- Wang, J. and Christopher, S. A. Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies. *Geophysical Research Letters*, 2003. doi:10.1029/2003GL018174.
- Wang, Z., Chen, L., Tao, J., Zhang, Y., and Su, L. Satellite-based estimation of regional particulate matter (PM) in beijing using vertical-and-RH correcting method. *Remote Sensing of Environment*, 114:50–63, 2010. doi:10.1016/j.rse.2009.08.009.

- Wang, Z., Chen, L., Tao, J., Liu, Y., Hu, X., and Tao, M. An empirical method of RH correction for satellite estimation of ground-level pm concentrations. *Atmospheric Environment*, 95:71–81, 2014. doi:[10.1016/j.atmosenv.2014.05.030](https://doi.org/10.1016/j.atmosenv.2014.05.030).
- WHO. Health risks of particulate matter from long-range transboundary air pollution. Technical report, World health organization, 2006a. WHO Regional Office for Europe.
- WHO. WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide - global update 2005. Summary of risk assessment. Technical report, World health organization, 2006b.
- WHO. Review of evidence on health aspects of air pollution – REVIHAAP project. Technical report, World health organization, 2013. WHO Regional Office for Europe.
- Zhang, H., Hoff, R. M., and Engel-Cox, J. A. The relation between moderate resolution imaging spectroradiometer (modis) aerosol optical depth and pm_{2.5} over the united states: A geographical comparison by u.s. environmental protection agency regions. *Journal of the Air & Waste Management Association*, 59(11):1358–1369, 2009. doi:[10.3155/1047-3289.59.11.1358](https://doi.org/10.3155/1047-3289.59.11.1358).

Appendix A Scatter Plots Excluded From Chapter 5

A.1 MAIAC Aerosol Data

A.1.1 Study II - Regions of Varying Scale

Washington D.C

Figure A.1: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Washington D.C at urban scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.2: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Washington D.C at meso-scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.3: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations at continental scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Atlanta

Figure A.4: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Atlanta at urban scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.5: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Atlanta at meso-scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.6: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations at continental scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Indianapolis

Figure A.7: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Indianapolis at urban scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.8: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Indianapolis at meso-scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.9: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations at continental scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Salt Lake City

Figure A.10: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Salt Lake at urban scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.11: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Salt Lake at meso-scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.12: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations at continental scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Phoenix

Figure A.13: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Phoenix at urban scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.14: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in Phoenix at meso-scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.15: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations at continental scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Los Angeles

Figure A.16: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in LA at urban scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.17: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations in LA at meso-scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

Figure A.18: Linear relationship between MAIAC AODs and $PM_{2.5}$ concentrations at continental scale at 10 km (left), 3 km (center) and 1 km (right) spatial resolution.

A.2 MYD04 Aerosol Data

A.2.1 Study II - Regions of Varying Scale

Washington D.C

Figure A.19: Linear relationship between MYD04 AODs and PM_{2.5} concentrations in Washington D.C at urban scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.20: Linear relationship between MYD04 AODs and PM_{2.5} concentrations in Washington D.C at meso-scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.21: Linear relationship between MYD04 AODs and PM_{2.5} concentrations at country at 10 km (left) and 3 km (right) spatial resolution.

Figure A.22: Linear relationship between MYD04 AODs and PM_{2.5} concentrations in Atlanta at urban scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.23: Linear relationship between MYD04 AODs and PM_{2.5} concentrations in Atlanta at meso-scale at 10 km (left) and 3 km (right) spatial resolution.

Atlanta

Figure A.24: Linear relationship between MYD04 AODs and PM_{2.5} concentrations at country at 10 km (left) and 3 km (right) spatial resolution.

Indianapolis

Figure A.25: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in Indianapolis at urban scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.26: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in Indianapolis at meso-scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.27: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations at country at 10 km (left) and 3 km (right) spatial resolution.

Salt Lake City

Figure A.28: Linear relationship between MYD04 AODs and PM_{2.5} concentrations in SaltLake at urban scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.29: Linear relationship between MYD04 AODs and PM_{2.5} concentrations in SaltLake at meso-scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.30: Linear relationship between MYD04 AODs and PM_{2.5} concentrations at country at 10 km (left) and 3 km (right) spatial resolution.

Phoenix

Figure A.31: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in Phoenix at urban scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.32: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in Phoenix at meso-scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.33: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations at country at 10 km (left) and 3 km (right) spatial resolution.

Los Angeles

Figure A.34: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in LA at urban scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.35: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations in LA at meso-scale at 10 km (left) and 3 km (right) spatial resolution.

Figure A.36: Linear relationship between MYD04 AODs and $PM_{2.5}$ concentrations at country at 10 km (left) and 3 km (right) spatial resolution.

Appendix B Matlab Scripts

B.1 Grid AOD Data

```

1 clear all
2 close all
3 clc
4 multiWaitbar( 'CloseAll' );
5
6 resolution_choice = menu('Choose Resolution', 'MYD04 10 km','MYD04 3 km',...
7                           'MAIAC 1 km', 'MAIAC 3 km', 'MAIAC 10 km');
8 region = 'US';
9
10 GRD_10K = '-I11.5k/11.5k';
11 GRD_3K = '-I3.45k/3.45k';
12 GRD_1K = '-I1.15k/1.15k';
13 REG = '-R-125/-66/24/51';
14
15 %% MYD04 DATA
16
17 if resolution_choice < 3
18
19     if resolution_choice == 1
20         res = '10K';
21     elseif resolution_choice == 2
22         res = '3K';
23     end
24
25     for month = 1 : 4
26
27         multiWaitbar('Monthly progress..', 'color', 'b');
28
29         if month == 1
30             mname='January';
31         elseif month == 2
32             mname='April';
33         elseif month == 3
34             mname='July';
35         elseif month == 4
36             mname='October';
37         end
38
39         all_files=[];
40         JDays={};
41         hours={};
42
43         datafolder=strcat('D:\Master Thesis\Final Study\DATA\MODIS\',...
44                           'Raw data\', region, '-2008\ ', mname, '\ ', res);
45         cd(datafolder);
46
47         all_files = dir('*hdf');
48
49         for i = 1 : length(all_files)
50             JDays{i} = all_files(i).name(15:17);
51         end
52
53         unique_JDays=unique(JDays);
54
55         for i = 1 : length(unique_JDays)
56
57             multiWaitbar('Daily progress..', 'color', 'y');
58
59             indices=strmatch(unique_JDays(i),JDays);
60
61             files=all_files(indices);
62
63             all_data=[];
64
65             for k = 1 : length(files)
66

```

```

67         fname=files(k).name;
68
69         if k == 1
70             info=hdfinfo(fname);
71             scale_factor=info.Vgroup.Vgroup(1,2).SDS(1,11)...
72                 .Attributes(1,5).Value;
73         end
74
75         AOD=double(hdfread(fname, 'Optical_Depth_Land_And_Ocean'));
76         AT=double(hdfread(fname, 'Aerosol_Type_Land'));
77         LSF=double(hdfread(fname, 'Land_sea_Flag'));
78
79         lat = double(hdfread(fname, 'Latitude'));
80         lon = double(hdfread(fname, 'Longitude'));
81
82         valid_data = find(AOD >= 0 & AT > 0 & LSF > 0 );
83
84         AOD = AOD(valid_data);
85         AT = AT(valid_data);
86
87         lat = lat(valid_data);
88         lon = lon(valid_data);
89
90         hhmm = fname(19:22);
91         hour = str2num(hhmm(1:2))+str2num(hhmm(3:4))/60;
92         hour_vec= repmat(hour, length(AOD), 1);
93
94         all_data=[all_data; lon, lat, AOD, AT, hour_vec];
95
96     end
97
98     all_data(:,3)=all_data(:,3)*scale_factor;
99     savefolder=strcat('D:\Master Thesis\Final Study\DATA\MODIS\',...
100         'files2grid\' , region, '_2008\' , res);
101     cd(savefolder)
102     JD=unique(JDays(i));
103
104     filename_save=char(strcat(region, '_2008_', JD, '_', res));
105     save(filename_save, 'all_data', '-ascii')
106
107     %% GRID DATA
108     if resolution_choice == 1
109         res = '10K';
110         grid_out_folder = strcat('.../gridded/' , region,...
111             '_2008/10K/' , mname, '/');
112         fname_out_AOD = strcat('gridded_AOD.' , filename_save, '.nc');
113         fname_out_time = strcat('gridded_time.' , filename_save, '.nc');
114
115         grid_str_AOD = char(strcat('gmt xyz2grd', {' '},...
116             filename_save, {' '}, REG, {' '}, GRD_10K,...
117             ' -i0,1,2 -fg -G' , grid_out_folder, fname_out_AOD));
118
119         grid_str_time = char(strcat('gmt xyz2grd', {' '},...
120             filename_save, {' '}, REG, {' '}, GRD_10K,...
121             ' -i0,1,4 -fg -G' , grid_out_folder, fname_out_time));
122
123     elseif resolution_choice == 2
124         res = '3K';
125         grid_out_folder = strcat('.../gridded/' , region,...
126             '_2008/3K/' , mname, '/');
127
128         fname_out_AOD = strcat('gridded_AOD.' , filename_save, '.nc');
129         fname_out_time = strcat('gridded_time.' , filename_save, '.nc');
130
131         grid_str_AOD = char(strcat('gmt xyz2grd', {' '},...
132             filename_save, {' '}, REG, {' '}, GRD_3K,...
133             ' -i0,1,2 -fg -G' , grid_out_folder, fname_out_AOD));
134
135         grid_str_time = char(strcat('gmt xyz2grd', {' '},...
136             filename_save, {' '}, REG, {' '}, GRD_3K,...

```



```

137         '-i0,1,4 -fg -G', grid_out_folder, fname_out_time));
138
139     end
140
141     system(grid_str_AOD)
142     system(grid_str_time)
143
144     cd(datafolder)
145
146     multiWaitbar('Daily progress..', i/length(unique_JDays));
147 end
148 multiWaitbar('Monthly progress..', month/4);
149 end
150
151
152 %% MAIAC DATA
153
154 elseif resolution_choice > 2
155
156     cd('D:\Master Thesis\Final Study\DATA\MAIAC\Raw data')
157
158     buffer = fileread('AOTratio.dat');
159     fileFormat=['%n %n'];
160     conversion_table = textscan(buffer, fileFormat, 'delimiter', '\t', ...
161                                'headerlines', 6);
162
163     AOD_table_470=conversion_table(:,1);
164     AOD_ratio=conversion_table(:,2);
165
166
167     for month = 1 : 4
168
169         multiWaitbar('Monthly progress..', 'color', 'b');
170
171         if month == 1
172             mname='January';
173         elseif month == 2
174             mname='April';
175         elseif month == 3
176             mname='July';
177         elseif month == 4
178             mname='October';
179         end
180
181
182         all_files=[];
183         JDays={};
184         hours={};
185
186         datafolder=strcat('D:\Master Thesis\Final Study\DATA\MAIAC\', ...
187                           'Raw data\', region, '_2008\', mname);
188         cd(datafolder);
189
190         all_files = dir('*hdf');
191
192
193         for i = 1 : length(all_files)
194             JDays{i} = all_files(i).name(22:24);
195         end
196
197         unique_JDays=unique(JDays);
198
199         for i = 1 : length(unique_JDays)
200
201             multiWaitbar('Daily progress..', 'color', 'y');
202
203             indices=strmatch(unique_JDays(i), JDays);
204
205             files=all_files(indices);
206

```

```

207         all_data=[];
208
209         for k = 1 : length(files)
210
211
212
213             fname=files(k).name;
214
215             if k == 1
216                 info=hdfinfo(fname);
217                 scale_factor=info.Vgroup.Vgroup(1,1).SDS(1,1)...
218                     .Attributes(1,5).Value;
219             end
220
221             AOD_470=double(hdfread(fname, 'Optical.Depth.Land'));
222
223             tile = fname(11:16);
224
225             cd(strcat('D:\Master Thesis\Final Study\DATA\MAIAC',...
226                 'Raw data\US_2008\Coordinates'))
227
228             fname_coord=dir(strcat('*',tile, '.hdf'));
229             fname_coord=fname_coord.name;
230
231             lat = double(hdfread(fname_coord, 'lat'));
232             lon = double(hdfread(fname_coord, 'lon'));
233
234             cd(datafolder)
235
236             valid_data = find(AOD_470 >= 0);
237
238             AOD_470 = AOD_470(valid_data);
239             AOD_470=AOD_470*scale_factor;
240
241             interp_ratio=interp1(AOD_table_470, AOD_ratio, AOD_470);
242             AOD=AOD_470.*interp_ratio;
243
244             lat = lat(valid_data);
245             lon = lon(valid_data);
246
247             hhmm = fname(25:28);
248             hour = str2num(hhmm(1:2))+str2num(hhmm(3:4))/60;
249             hour_vec= repmat(hour, length(AOD), 1);
250
251             all_data=[all_data; lon, lat, AOD, hour_vec];
252         end
253
254         savefolder=strcat('D:\Master Thesis\Final Study\DATA\MAIAC',...
255             'files2grid\', region, '_2008\1K');
256         cd(savefolder)
257         JD=unique_JDays(i);
258         filename_save=char(strcat(region, '_2008-', JD, '_1K'));
259         save(filename_save, 'all_data', '-ascii')
260
261         %% GRID DATA
262         if resolution_choice == 3
263             grid_out_folder = strcat('.../gridded/', region,...
264                 '_2008/1K/', mname, '/');
265             fname_out_AOD = strcat('gridded.AOD.', filename_save, '.nc');
266             fname_out_time = strcat('gridded.time-', filename_save, '.nc');
267
268             grid_str_AOD = char(strcat('gmt xyz2grd', {' '},...
269                 filename_save, {' '}, REG, {' '}, GRD_1K,...
270                 '-i0,1,2 -fg -G', grid_out_folder, fname_out_AOD));
271
272             grid_str_time = char(strcat('gmt xyz2grd', {' '},...
273                 filename_save, {' '}, REG, {' '}, GRD_1K,...
274                 '-i0,1,3 -fg -G', grid_out_folder, fname_out_time));
275
276         elseif resolution_choice == 4

```

```

277         grid.out_folder = strcat('.../gridded/', region,...
278                                 '_2008/3K/', mname, '/');
279
280         fname_out_AOD = strcat('gridded.AOD.', filename_save(1:12), '3K.nc');
281         fname_out_time = strcat('gridded.time-', filename_save(1:12), '3K.nc');
282
283         grid.str_AOD = char(strcat('gmt xyz2grd', {' '},...
284                                 filename_save, {' '}, REG, {' '}, GRD_3K,...
285                                 '-i0,1,2 -fg -G', grid.out_folder, fname_out_AOD));
286
287         grid.str_time = char(strcat('gmt xyz2grd', {' '},...
288                                 filename_save, {' '}, REG, {' '}, GRD_3K,...
289                                 '-i0,1,3 -fg -G', grid.out_folder, fname_out_time));
290
291     elseif resolution_choice == 5
292         grid.out_folder = strcat('.../gridded/', region,...
293                                 '_2008/10K/', mname, '/');
294
295         fname_out_AOD = strcat('gridded.AOD.',...
296                                 filename_save(1:12), '10K.nc');
297         fname_out_time = strcat('gridded.time-',...
298                                 filename_save(1:12), '10K.nc');
299
300         grid.str_AOD = char(strcat('gmt xyz2grd', {' '},...
301                                 filename_save, {' '}, REG, {' '}, GRD_10K,...
302                                 '-i0,1,2 -fg -G', grid.out_folder, fname_out_AOD));
303
304         grid.str_time = char(strcat('gmt xyz2grd', {' '},...
305                                 filename_save, {' '}, REG, {' '}, GRD_10K,...
306                                 '-i0,1,3 -fg -G', grid.out_folder, fname_out_time));
307
308     end
309
310     system(grid.str_AOD)
311     system(grid.str_time)
312
313     cd(datafolder)
314
315     multiWaitbar('Daily progress..', i/length(unique_JDays));
316 end
317
318     multiWaitbar('Monthly progress..', month/4);
319 end
320 end
321
322 multiWaitbar('CloseAll');
323 cd 'D:\Master Thesis\Final Study'

```

B.2 Read PM_{2.5} Data

```

1 clear all
2 close all
3 clc
4 multiWaitbar( 'CloseAll' );
5
6 region = 'US';
7
8 data_folder='D:\Master Thesis\Final Study\DATA\EPA';
9 cd(data_folder)
10
11 file=dir('daily.88101.2008.csv');
12 filename=file.name;
13
14 buffer = fileread(filename);
15 fileFormat=['%n %*q %*q %*q %*n %n %n %*q'...
16            '%*q %*q %*q %q %*q %*q %*n'...
17            '%*n %n %*n %*n %*n %*q %*q %*q %*q'...
18            '%*q %*q %*q %*q'];
19 all_data = textscan(buffer, fileFormat, 'delimiter',...
20                    '\t', 'headerlines',1);
21
22 all_dates=cell2mat(all_data{4});
23
24 PM_data=[];
25
26 for month_loop = 1 : 4
27
28     multiWaitbar('Monthly progress..', 'color', 'b');
29
30     if month_loop == 1
31         month='01';
32     elseif month_loop == 2
33         month='04';
34     elseif month_loop == 3
35         month='07';
36     elseif month_loop == 4
37         month='10';
38     end
39
40     correct_month = find(strncmp(strcat('2008-',month,'-xx'),...
41                                cellstr(all_dates), 7));
42
43     state_ID=all_data{1}(correct_month);
44     PM=all_data{5}(correct_month);
45     lat_PM=all_data{2}(correct_month);
46     lon_PM=all_data{3}(correct_month);
47     date_PM=all_data{4}(correct_month);
48
49     ind_not_conus = ...
50     find(state_ID == 2 | state_ID == 15 | state_ID == 72 | state_ID == 78);
51
52     PM(ind_not_conus)=[];
53     lat_PM(ind_not_conus)=[];
54     lon_PM(ind_not_conus)=[];
55     date_PM(ind_not_conus)=[];
56
57     JDays=date2doy(datenum(date_PM));
58
59     PM_data_unfiltered=[lon_PM, lat_PM, PM, repmat(str2num(month),...
60             length(PM), 1), JDays];
61
62     %% AVERAGE MULTIPLE VALUES AT THE SAME SITE AND DAY
63
64     PM_sites_day = PM_data_unfiltered(:, [1:2 4:5]);
65
66     unique_PM_sites_day = unique(PM_sites_day, 'rows');
```

```

67
68     for j = 1 : length(unique_PM_sites_day(:,1))
69
70         multiWaitbar('Averageing Duplicate PM2.5 values..', 'color', 'g');
71
72         index = find(ismember(PM_sites_day, unique_PM_sites_day(j,:), ...
73                             'rows'), inf);
74
75         if length(index) > 1
76
77             PM_data_unfiltered(index(1), 3) = mean(PM_data_unfiltered(index, 3));
78             PM_data_unfiltered(index(2:end), :) = [];
79             PM_sites_day(index(2:end), :) = [];
80
81         end
82
83         multiWaitbar('Averageing Duplicate PM2.5 values..', ...
84                     j/length(unique_PM_sites_day(:,1)));
85
86     end
87
88     PM_data = [PM_data; PM_data_unfiltered];
89
90     multiWaitbar('Monthly progress..', month_loop/4);
91 end
92
93 savefolder = 'D:\Master Thesis\Final Study\MATLAB\saved.data\PM';
94 savename_PM = char(strcat('PM_data-', region, '_2008'));
95 cd(savefolder)
96 % save(savename_PM, 'PM_data')
97
98 multiWaitbar('CloseAll');
99 cd('D:\Master Thesis\Final Study')

```

B.3 Match Gridded AOD Data and PM_{2.5} Data

```

1 clear all
2 close all
3 clc
4 multiWaitbar( 'CloseAll' );
5
6 region = 'US';
7
8
9 for resolution_choice = 1 : 5
10
11     multiWaitbar('Resolution progress..', 'color', 'y');
12
13     for month_choice = 1:4
14
15         clearvars -except region_choice resolution_choice month_choice...
16             region test
17         multiWaitbar('Monthly progress..', 'color', 'r');
18
19
20         if month_choice == 1
21             month='01';
22             mname='January';
23         elseif month_choice == 2
24             month='04';
25             mname='April';
26         elseif month_choice == 3
27             month='07';
28             mname='July';
29         elseif month_choice == 4
30             month='10';
31             mname='October';
32         end
33
34         if resolution_choice == 1
35             res = '10K';
36             pixelsize = 11.5;
37             datafolder_AOD = strcat('D:\Master Thesis\Final Study\DATA\',...
38                 'MODIS\gridded\', region, '_2008\', res, '\', mname);
39             products=3;
40         elseif resolution_choice == 2
41             res = '3K';
42             pixelsize = 3.45;
43             products=3;
44             datafolder_AOD = strcat('D:\Master Thesis\Final Study\DATA\',...
45                 'MODIS\gridded\', region, '_2008\', res, '\', mname);
46         elseif resolution_choice == 3
47             res = '1K';
48             pixelsize = 1.15;
49             products=2;
50             datafolder_AOD = strcat('D:\Master Thesis\Final Study\DATA\',...
51                 'MAIAC\gridded\', region, '_2008\', res, '\', mname);
52         elseif resolution_choice == 4
53             res = 'MAIAC.10K';
54             pixelsize = 11.5;
55             products=2;
56             datafolder_AOD = strcat('D:\Master Thesis\Final Study\DATA\',...
57                 'MAIAC\gridded\', region, '_2008\', res(7:9), '\', mname);
58         elseif resolution_choice == 5
59             res = 'MAIAC.3K';
60             pixelsize = 3.45;
61             products=2;
62             datafolder_AOD = strcat('D:\Master Thesis\Final Study\DATA\',...
63                 'MAIAC\gridded\', region, '_2008\', res(7:8), '\', mname);
64         end
65
66         %% LOAD PM2.5 DATA

```

```

67
68     cd 'D:\Master Thesis\Final Study\MATLAB\saved_data\PM'
69
70     PM_file = char(strcat('PM_data-', region, '_2008'));
71     load(PM_file);
72
73     correct_month = find(PM_data(:,4) == str2num(month));
74
75     PM_data=PM_data(correct_month,:);
76
77     %% READ AOD DATA
78     cd(datafolder_AOD)
79
80     files = dir('*.nc*');
81
82     AOD_all=[];
83     SOPT_all=[];
84
85     lat_all=[];
86     lon_all=[];
87     JDays=[];
88     months=[];
89
90     for i = 1 : length(files)
91
92         multiWaitbar('Reading daily aerosol data..', 'color', 'g');
93
94         fname = files(i).name;
95
96         if i == 1
97             lon_org_vec = ncread(fname, 'lon');
98             lat_org_vec = ncread(fname, 'lat');
99
100             [lat_org, lon_org] = meshgrid(lat_org_vec, lon_org_vec);
101
102         end
103
104         if strcmp(fname(9:11), 'AOD')
105
106             AOD = ncread(fname, 'z');
107             valid_data = find(AOD >= 0);
108             AOD_all = [AOD_all; AOD(valid_data)];
109
110             lat_all = [lat_all; lat_org(valid_data)];
111             lon_all = [lon_all; lon_org(valid_data)];
112
113             JD=fname(21:23);
114
115             JDays = [JDays; repmat(str2num(JD), length(valid_data), 1)];
116             months = [months; repmat(str2num(month), length(valid_data), 1)];
117
118         elseif strcmp(fname(9:12), 'time')
119
120             SOPT = ncread(fname, 'z');
121             valid_data = find(SOPT >= 0);
122             SOPT_all = [SOPT_all; SOPT(valid_data)];
123
124         end
125
126         multiWaitbar('Reading daily aerosol data..', i/length(files));
127
128     end
129
130
131     AOD_data = [lon_all, lat_all, AOD_all, SOPT_all,...
132                months, JDays];
133
134     clear lon_all lat_all AOD_all SOPT_all months JDays
135
136

```

```

137
138
139 %% MATCH DATA
140
141 AOD_data_match=[];
142 PM_data_match=[];
143
144 unique_days=unique(PM_data(:,5));
145
146 for i = 1 : length(unique_days)
147
148     multiWaitbar('Matching daily pairs..', 'color', 'c');
149
150     JD = unique_days(i);
151
152     PM_data_temp = [];
153     AOD_data_temp = [];
154
155     id_PM = find(PM_data(:,5) == JD);
156     id_AOD = find(AOD_data(:,6) == JD);
157
158     PM_data_temp = PM_data(id_PM,:);
159     AOD_data_temp = AOD_data(id_AOD,:);
160
161
162     indices=[];
163     lat_diff=km2deg(1)*(pixelsize/2);
164     for j = 1 : length(PM_data_temp(:,1))
165         lon_diff = km2deg(1)/cosd(PM_data_temp(j,2))*(pixelsize/2);
166         lat_poly=[PM_data_temp(j,2)-lat_diff...
167                 PM_data_temp(j,2)+lat_diff...
168                 PM_data_temp(j,2)+lat_diff...
169                 PM_data_temp(j,2)-lat_diff...
170                 PM_data_temp(j,2)-lat_diff]';
171
172         lon_poly=[PM_data_temp(j,1)+lon_diff...
173                 PM_data_temp(j,1)+lon_diff...
174                 PM_data_temp(j,1)-lon_diff...
175                 PM_data_temp(j,1)-lon_diff...
176                 PM_data_temp(j,1)+lon_diff]';
177
178         IN = inpolygon(AOD_data_temp(:,1),AOD_data_temp(:,2),...
179                       lon_poly, lat_poly);
180
181         ix = find(IN == 1);
182
183         if length(ix) > 0
184             indices=[indices; repmat(j,length(ix),1), ix];
185         end
186     end
187
188     [m,n] = size(indices);
189
190     if m > 0
191         AOD_data_match=[AOD_data_match; AOD_data_temp(indices(:,2),:)]';
192         PM_data_match=[PM_data_match; PM_data_temp(indices(:,1),:)]';
193     end
194
195     multiWaitbar('Matching daily pairs..', i/length(unique_days));
196 end
197
198
199 savefolder = strcat('D:\Master Thesis\Final Study\MATLAB\',...
200                    'saved_data\Matched data\', region, '_2008');
201
202 savename_matched=char(strcat('Matched_data-', region, '-',...
203                             mname, '_2008-', res));
204
205 cd(savefolder)
206 save(savename_matched, 'AOD_data_match', 'PM_data_match')

```



```
207         multiWaitbar('Monthly progress..', month_choice/4);
208         end
209         multiWaitbar('Resolution progress..', resolution_choice/5);
210     end
211
212     multiWaitbar( 'CloseAll' );
213     cd 'D:\Master Thesis\Final Study\MATLAB'
```

B.4 Analyze Data for Study I

```

1 clear all
2 close all
3 clc
4 multiWaitbar( 'CloseAll' );
5 set(0, 'DefaultAxesFontSize', 14)
6
7 table_choice = menu('Save data in latex table format?', 'No', 'Yes');
8
9 region = 'US';
10
11 for month_choice = 1 : 4
12
13     if month_choice == 1
14         mname='January';
15     elseif month_choice == 2
16         mname='April';
17     elseif month_choice == 3
18         mname='July';
19     elseif month_choice == 4
20         mname='October';
21     end
22
23     for resolution_choice = 1 : 5
24
25         if resolution_choice == 1
26             res = '10K';
27             res_table = 'MYD04 10 km';
28         elseif resolution_choice == 2
29             res = '3K';
30             res_table = 'MYD04 3 km';
31         elseif resolution_choice == 3
32             res = '1K';
33             res_table = 'MAIAC 1 km';
34         elseif resolution_choice == 4
35             res = 'MAIAC_3K';
36             res_table = 'MAIAC 3 km';
37         elseif resolution_choice == 5
38             res = 'MAIAC_10K';
39             res_table = 'MAIAC 10 km';
40         end
41
42         datafolder = strcat('D:\Master Thesis\Final Study\MATLAB\', ...
43                             'saved_data\Matched data\', region, '_2008');
44         fname = char(strcat('Matched.data-', region, '-', ...
45                             mname, '_2008-', res));
46
47         cd(datafolder)
48         load(fname)
49
50         A_match=ones(length(AOD_data_match(:,3)),2);
51         A_match(:,1)=AOD_data_match(:,3);
52         d_match=PM_data_match(:,3);
53
54         x_match=inv(A_match'*A_match)*A_match'*d_match;
55         b_match=x_match(2);
56
57         i = find(PM_data_match(:,3)./AOD_data_match(:,3) > 700 & PM_data_match(:,3) > 3*b_match/2);
58         j= find(PM_data_match(:,3)./AOD_data_match(:,3) < 10);
59
60         fraction_removed = length([i;j])/length(AOD_data_match(:,1));
61
62         AOD_data_match([i;j],:)=[];
63         PM_data_match([i;j],:)=[];
64
65
66         AOD=AOD_data_match(:,3);

```

```

67     PM=PM.data_match(:,3);
68
69     number_of_points = length(PM);
70
71     % STATISTICAL SIGNIFICANCE
72     [~,p] = ttest(AOD, PM);
73
74     if p >= 0.001
75         p_str = strcat(' =', {' '}, num2str(p, '%.2f'));
76     else
77         p_str = strcat(' <$ 0.001');
78     end
79
80     % LINEAR REGRESSION
81     A=ones(length(AOD),2);
82     A(:,1)=AOD;
83     d=PM;
84
85     x=inv(A'*A)*A'*d;
86     a=x(1);
87     b=x(2);
88
89     AOD_reg=[0:0.01:max(AOD)];
90     PM_reg = a*AOD_reg+b;
91
92     % CORRELATION COEFFICIENT
93     R = corr(AOD, PM);
94
95     figure('Visible','off'); clf;
96     scatter(AOD, PM,'k')
97     hold on
98     plot(AOD_reg,PM_reg,'r')
99     xlabel('AOD');
100    ylabel('Daily mean PM-2.5 concentration [ $\mu\text{g}/\text{m}^3$ ]');
101    title(strcat(mname, {','}, {' '}, res.table))
102    legend('Matched AOD and PM-2.5 data points', strcat('PM-2.5=',...
103        num2str(a, '%.2f'), '\cdot AOD+', num2str(b, '%.2f')));
104
105
106    savefolder = strcat('D:\Master Thesis\Final Study\MATLAB\Images\',...
107        'Continental scale\All types combined\', region, '_2008');
108    savename = char(strcat(mname, '-', region, '_2008-', res, '.eps'));
109
110    cd(savefolder)
111    saveas(gcf,savename, 'psc2')
112
113
114    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
115    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
116    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
117
118
119    %% CREATE LATEX TABLE
120
121    if table_choice == 2
122
123        if month_choice == 1
124            k = 3;
125        elseif month_choice == 2
126            k = 8;
127        elseif month_choice == 3
128            k = 13;
129        elseif month_choice == 4
130            k = 18;
131        end
132
133        if resolution_choice < 3
134
135            if resolution_choice == 1 && month_choice == 1

```

```

137     table_1={'\textbf{Month}', ' & ',...
138             '\textbf{Statistics}', ' & ',...
139             '\textbf{MYD04 10 km}', ' & ',...
140             '\textbf{MYD04 3 km}','','\\ \hline'};
141
142     for i = 2 : 2 : 8
143         for j = 2 : 21
144             table_1{j,i}=strcat({' ',' '&'},{ ' '});
145         end
146     end
147
148     top={'\begin{tabular}{c l l l}' '' '' '' '' '' '' '' ''};
149     bottom={'\end{tabular}' '' '' '' '' '' '' '' ''};
150
151     table_1=vertcat(top,table_1);
152     table_1=vertcat(table_1,bottom);
153
154 end
155
156
157 if resolution_choice == 1
158     l = 5;
159 elseif resolution_choice == 2
160     l = 7;
161 end
162
163 if resolution_choice == 1
164
165     table_1{k,l} = strcat('\textbf{' , mname, '}');
166     table_1{k,l-2}='\textit{N}';
167     table_1{k+1,l-2}='\textit{R}';
168     table_1{k+2,l-2}='Slope';
169     table_1{k+3,l-2}='intersect';
170     table_1{k+4,l-2}='\textit{p}';
171
172     table_1{k,8}='\\';
173     table_1{k+1,8}='\\';
174     table_1{k+2,8}='\\';
175     table_1{k+3,8}='\\';
176     table_1{k+4,8}='\\ \\hline';
177
178 end
179
180 table_1{k,l}=num2str(number_of_points);
181 table_1{k+l,l}=num2str(R,'%.3f');
182 table_1{k+2,l}= num2str(a,'%.2f');
183 table_1{k+3,l}=num2str(b,'%.2f');
184 table_1{k+4,l}= p_str;
185
186 end
187
188 if resolution_choice > 2
189
190     if resolution_choice == 3 && month_choice == 1
191
192         table_2={'\textbf{Month}', ' & ',...
193                 '\textbf{Statistics}', ' & ',...
194                 '\textbf{MAIAC 10 km}', ' & ',...
195                 '\textbf{MAIAC 3 km}',' & ',...
196                 '\textbf{MAIAC 1 km}','','\\ \hline'};
197
198         for i = 2 : 2 : 10
199             for j = 2 : 21
200                 table_2{j,i}=strcat({' ',' '&'},{ ' '});
201             end
202         end
203
204         top={'\begin{tabular}{c l l l l}' '' '' '' '' '' '' '' '' ''};
205         bottom={'\end{tabular}' '' '' '' '' '' '' '' '' ''};
206

```

```

207         table_2=vertcat(top,table_2);
208         table_2=vertcat(table_2,bottom);
209
210     end
211
212
213     if resolution_choice == 5
214         l = 5;
215     elseif resolution_choice == 4
216         l = 7;
217     elseif resolution_choice == 3
218         l = 9;
219     end
220
221     if resolution_choice == 5
222
223         table_2{k,1} = strcat('\textbf{', mname, '}');
224         table_2{k,1-2}='\textit{N}';
225         table_2{k+1,1-2}='\textit{R}';
226         table_2{k+2,1-2}='Slope';
227         table_2{k+3,1-2}='intersect';
228         table_2{k+4,1-2}='\textit{p}';
229
230         table_2{k,10}='\\';
231         table_2{k+1,10}='\\';
232         table_2{k+2,10}='\\';
233         table_2{k+3,10}='\\';
234         table_2{k+4,10}='\\ \hline';
235
236     end
237
238     table_2{k,1}=num2str(number_of_points);
239     table_2{k+1,1}=num2str(R, '%.3f');
240     table_2{k+2,1}= num2str(a, '%.2f');
241     table_2{k+3,1}=num2str(b, '%.2f');
242     table_2{k+4,1}= p_str;
243
244     end
245 end
246 end
247 end
248
249 %% SAVE LATEX TABLES
250
251 if table_choice == 2
252
253     %% ALL DATA
254
255     tablefolder = strcat('D:\Master Thesis\Final Study\MATLAB\Tables\',...
256                         'Continental scale\All data\', region, '.2008');
257     cd(tablefolder)
258
259     tablename_1 = char(strcat(region, '_2008.no.type.separation.MYD04.tex'));
260     tablename_2 = char(strcat(region, '_2008.no.type.separation.MAIAC.tex'));
261
262     [m n] = size(table_1);
263     fileID_1 = fopen(tablename_1,'w');
264     fileID_2 = fopen(tablename_2,'w');
265     formatSpec = '%s\n';
266
267     for row = 1 : m
268         fprintf(fileID_1,formatSpec, char(strcat(table_1{row,:})));
269         fprintf(fileID_2,formatSpec, char(strcat(table_2{row,:})));
270     end
271
272     fclose(fileID_1);
273     fclose(fileID_2);
274
275     %% SOME DATA
276

```

```
277
278     tablefolder.some = strcat('D:\Master Thesis\Final Study\MATLAB',...
279                               'Tables\Continental scale\Some data\', region, '_2008');
280     cd(tablefolder.some)
281
282     fileID_1.some = fopen(tablename_1, 'w');
283     fileID_2.some = fopen(tablename_2, 'w');
284
285     rows_some = [1 2 3 4 7 8 9 12 13 14 17 18 19 22 23];
286
287     for i = 1 : length(rows_some)
288         fprintf(fileID_1.some, formatSpec, char(strcat(table_1{rows_some(i), :})));
289         fprintf(fileID_2.some, formatSpec, char(strcat(table_2{rows_some(i), :})));
290     end
291
292     fclose(fileID_1.some);
293     fclose(fileID_2.some);
294 end
295
296 cd 'D:\Master Thesis\Final Study\MATLAB'
```

B.5 Analyze Data for Study II

```

1 clear all
2 close all
3 clc
4 multiWaitbar( 'CloseAll' );
5 set(0, 'DefaultAxesFontSize', 14)
6
7 table_choice = menu('Save data in latex table format?', 'No', 'Yes');
8
9 region = 'US';
10 meso.size = 4;
11
12 fractions.removed.urban=0;
13 fractions.removed.meso=0;
14 fractions.removed.continental=0;
15
16 for city_choice = 1:6
17
18     multiWaitbar('City progress..', 'color', 'c');
19
20     if city_choice == 1
21         % WASHINGTON D.C
22         city_name = 'Washington D.C';
23         city_savename = 'Washington';
24         city_center.lat = 38.897;
25         city_center.lon = -77.006;
26         urban.size = 0.12;
27     elseif city_choice == 2
28         % ATLANTA
29         city_name = 'Atlanta';
30         city_savename = city_name;
31         city_center.lat = 33.764;
32         city_center.lon = -84.390;
33         urban.size = 0.2;
34     elseif city_choice == 3
35         % INDIANAPOLIS
36         city_name = 'Indianapolis';
37         city_savename = city_name;
38         city_center.lat = 39.777;
39         city_center.lon = -86.146;
40         urban.size = 0.08;
41     elseif city_choice == 4
42         % SALT LAKE CITY
43         city_name = 'Salt Lake City';
44         city_savename = 'SaltLake';
45         city_center.lat = 40.632;
46         city_center.lon = -111.899;
47         urban.size = 0.16;
48     elseif city_choice == 5
49         % PHOENIX
50         city_name = 'Phoenix';
51         city_savename = city_name;
52         city_center.lat = 33.572;
53         city_center.lon = -112.088;
54         urban.size = 0.156;
55     elseif city_choice == 6
56         % LOS ANGELES
57         city_name = 'Los Angeles';
58         city_savename = 'LA';
59         city_center.lat = 34.053;
60         city_center.lon = -118.240;
61         urban.size = 0.25;
62     end
63
64     for resolution_choice = 1:5
65
66         if resolution_choice == 1

```

```

67         res = '10K';
68         res_table = 'MYD04 10 km';
69     elseif resolution_choice == 2
70         res = '3K';
71         res_table = 'MYD04 3 km';
72     elseif resolution_choice == 3
73         res = '1K';
74         res_table = 'MAIAC 1 km';
75     elseif resolution_choice == 4
76         res = 'MAIAC_3K';
77         res_table = 'MAIAC 3 km';
78     elseif resolution_choice == 5
79         res = 'MAIAC_10K';
80         res_table = 'MAIAC 10 km';
81     end
82
83     PM_urban = [];
84     AOD_urban = [];
85
86     PM_meso = [];
87     AOD_meso = [];
88
89     PM_continental = [];
90     AOD_continental = [];
91
92     for month_choice = 1 : 4
93
94         if month_choice == 1
95             mname = 'January';
96         elseif month_choice == 2
97             mname = 'April';
98         elseif month_choice == 3
99             mname = 'July';
100        elseif month_choice == 4
101            mname = 'October';
102        end
103
104        datafolder = strcat('D:\Master Thesis\Final Study\MATLAB\', ...
105                            'saved.data\Matched data\', region, '_2008');
106        fname = char(strcat('Matched_data-', region, '-', mname, ...
107                            '_2008-', res));
108        cd(datafolder)
109        load(fname)
110
111        msc_lat = [city_center.lat - meso.size...
112                  city_center.lat + meso.size...
113                  city_center.lat + meso.size...
114                  city_center.lat - meso.size...
115                  city_center.lat - meso.size];
116
117        msc_lon = [city_center.lon - meso.size...
118                  city_center.lon - meso.size...
119                  city_center.lon + meso.size...
120                  city_center.lon + meso.size...
121                  city_center.lon - meso.size];
122
123        city_lat = [city_center.lat - urban.size...
124                   city_center.lat + urban.size...
125                   city_center.lat + urban.size...
126                   city_center.lat - urban.size...
127                   city_center.lat - urban.size];
128
129        city_lon = [city_center.lon - urban.size...
130                   city_center.lon - urban.size...
131                   city_center.lon + urban.size...
132                   city_center.lon + urban.size...
133                   city_center.lon - urban.size];
134
135        idx_urban = find(inpolygon(PM_data_match(:,2), ...
136                                  PM_data_match(:,1), city_lat, city_lon));

```



```

137         idx_meso = find(inpolygon(PM.data_match(:,2),...
138                                 PM.data_match(:,1), msc_lat, msc_lon));
139
140
141         PM_urban = [PM_urban; PM.data_match(idx_urban, :)];
142         AOD_urban = [AOD_urban; AOD.data_match(idx_urban, :)];
143
144
145         PM_meso = [PM_meso; PM.data_match(idx_meso, :)];
146         AOD_meso = [AOD_meso; AOD.data_match(idx_meso, :)];
147
148
149         PM_continental = [PM_continental; PM.data_match(:, :)];
150         AOD_continental = [AOD_continental; AOD.data_match(:, :)];
151
152     end
153
154     %% REMOVE OUTLIERS
155
156     % URBAN SCALE
157     A_urban=ones(length(AOD_urban(:,3)),2);
158     A_urban(:,1)=AOD_urban(:,3);
159     d_urban=PM_urban(:,3);
160
161     x_urban=inv(A_urban'*A_urban)*A_urban'*d_urban;
162     b_urban=x_urban(2);
163
164     i = find(PM_urban(:,3)./AOD_urban(:,3) > 700 & PM_urban(:,3) > 3*b_urban/2);
165     j= find(PM_urban(:,3)./AOD_urban(:,3) < 10);
166
167     AOD_urban([i;j],:)=[];
168     PM_urban([i;j],:)=[];
169     city_choice
170     fraction_removed_urban = length([i;j])/length(AOD_urban(:,1));
171     fractions_removed_urban=fractions_removed_urban+fraction_removed_urban;
172
173     % MESO-SCALE
174     A_meso=ones(length(AOD_meso(:,3)),2);
175     A_meso(:,1)=AOD_meso(:,3);
176     d_meso=PM_meso(:,3);
177
178     x_meso=inv(A_meso'*A_meso)*A_meso'*d_meso;
179     b_meso=x_meso(2);
180
181     i = find(PM_meso(:,3)./AOD_meso(:,3) > 700 & PM_meso(:,3) > 3*b_meso/2);
182     j= find(PM_meso(:,3)./AOD_meso(:,3) < 10); %25
183
184     AOD_meso([i;j],:)=[];
185     PM_meso([i;j],:)=[];
186
187     fraction_removed_meso = length([i;j])/length(AOD_meso(:,1));
188     fractions_removed_meso=fractions_removed_meso+fraction_removed_meso;
189
190     % CONTINENTAL SCALE
191     A_continental=ones(length(AOD_continental(:,3)),2);
192     A_continental(:,1)=AOD_continental(:,3);
193     d_continental=PM_continental(:,3);
194
195     x_continental=inv(A_continental'*A_continental)*A_continental'*d_continental;
196     b_continental=x_continental(2);
197
198     i = find(PM_continental(:,3)./AOD_continental(:,3) > 700 & PM_continental(:,3) > 3*b_continental/2);
199     j= find(PM_continental(:,3)./AOD_continental(:,3) < 10);
200
201     AOD_continental([i;j],:)=[];
202     PM_continental([i;j],:)=[];
203
204     fraction_removed_continental = length([i;j])/length(AOD_continental(:,1));
205     fractions_removed_continental=fractions_removed_continental+fraction_removed_continental;
206

```

```

207     %% PLOTS
208
209     %% URBAN SCALE
210
211     number_of_points_urban = length(PM_urban(:,3));
212
213     % STATISTICAL SIGNIFICANCE
214     [~,p_urban] = ttest(AOD_urban(:,3), PM_urban(:,3));
215
216     if p_urban >= 0.001
217         p_urban_str = strcat(' =', {' '}, num2str(p_urban, '%.2f'));
218
219     else
220         p_urban_str = strcat(' <$ 0.001');
221     end
222
223     % LINEAR REGRESSION
224     A_urban=ones(length(AOD_urban(:,3)),2);
225     A_urban(:,1)=AOD_urban(:,3);
226     d_urban=PM_urban(:,3);
227
228     x_urban=inv(A_urban'*A_urban)*A_urban'*d_urban;
229     a_urban=x_urban(1);
230     b_urban=x_urban(2);
231
232     AOD_reg_urban = [0:0.01:max(AOD_urban(:,3))];
233     PM_reg_urban = a_urban*AOD_reg_urban+b_urban;
234
235     % CORRELATION COEFFICIENT
236     if length(AOD_urban) > 0
237         R_urban=corr(AOD_urban(:,3), PM_urban(:,3));
238     else
239         R_urban = NaN;
240     end
241
242     figure('Visible','off'); clf;
243     scatter(AOD_urban(:,3), PM_urban(:,3),'k')
244     hold on
245     plot(AOD_reg_urban,PM_reg_urban,'r')
246     xlabel('AOD');
247     ylabel('Daily mean PM-2.5 concentration [ $\mu\text{g}/\text{m}^3$ ]');
248     title(strcat(city_name, ' urban scale',{' '}, res_table));
249     legend('Matched AOD and PM-2.5 data points', strcat('PM-2.5=',...
250         num2str(a_urban, '%.2f'),...
251         '\cdot AOD+', num2str(b_urban, '%.2f')));
252
253     savefolder = strcat('D:\Master Thesis\Final Study\MATLAB\Images\','...
254         'Varying scale');
255     savename = char(strcat(city_savename, '_2008-urban-scale-',...
256         res, '.png'));
257
258     cd(savefolder)
259     print('-dpng', savename);
260
261     %% MESO-SCALE
262
263     number_of_points_meso = length(PM_meso(:,3));
264
265     % STATISTICAL SIGNIFICANCE
266     [~,p_meso] = ttest(AOD_meso(:,3), PM_meso(:,3));
267
268     if p_meso >= 0.001
269         p_meso_str = strcat(' =', {' '}, num2str(p_meso, '%.2f'));
270     else
271         p_meso_str = strcat(' <$ 0.001');
272     end
273
274     % LINEAR REGRESSION
275     A_meso=ones(length(AOD_meso(:,3)),2);
276     A_meso(:,1)=AOD_meso(:,3);

```

```

277     d_meso=PM_meso(:,3);
278
279     x_meso=inv(A_meso'*A_meso)*A_meso'*d_meso;
280     a_meso=x_meso(1);
281     b_meso=x_meso(2);
282
283     AOD_reg_meso = [0:0.01:max(AOD_meso(:,3))];
284     PM_reg_meso = a_meso*AOD_reg_meso+b_meso;
285
286     % CORRELATION COEFFICIENT
287     if length(AOD_meso) > 0
288         R_meso=corr(AOD_meso(:,3), PM_meso(:,3));
289     else
290         R_meso = NaN;
291     end
292
293     figure('Visible','off'); clf;
294     scatter(AOD_meso(:,3), PM_meso(:,3), 'k')
295     hold on
296     plot(AOD_reg_meso,PM_reg_meso, 'r')
297     xlabel('AOD');
298     ylabel('Daily mean PM_{2.5} concentration [\mu g/m^3]');
299     title(strcat(city_name, ' meso-scale', ' { ' }, res_table))
300     legend('Matched AOD and PM_{2.5} data points', strcat('PM_{2.5}= ',...
301         num2str(a_meso, '%.2f'), '\cdot AOD+', num2str(b_meso, '%.2f'));
302
303     savefolder = strcat('D:\Master Thesis\Final Study\MATLAB\Images\',...
304         'Varying scale');
305     savename = char(strcat(city_savename, '_2008_meso-scale-',...
306         res, '.png'));
307
308     cd(savefolder)
309     print('-dpng', savename);
310
311     %% CONTINENTAL SCALE
312
313     number_of_points_continental = length(PM_continental(:,3));
314
315     % STATISTICAL SIGNIFICANCE
316     [~,p_continental] = ttest(AOD_continental(:,3), PM_continental(:,3));
317
318     if p_continental >= 0.001
319         p_continental_continental = strcat(' = ', { ' ' }, num2str(p_continental, '%.2f'));
320     else
321         p_continental_str = strcat(' <= 0.001');
322     end
323
324     % LINEAR REGRESSION
325     A_continental=ones(length(AOD_continental(:,3)),2);
326     A_continental(:,1)=AOD_continental(:,3);
327     d_continental=PM_continental(:,3);
328
329     x_continental=inv(A_continental'*A_continental)*A_continental'*d_continental;
330     a_continental=x_continental(1);
331     b_continental=x_continental(2);
332
333     AOD_reg_continental = [0:0.01:max(AOD_continental(:,3))];
334     PM_reg_continental = a_continental*AOD_reg_continental+b_continental;
335
336     % CORRELATION COEFFICIENT
337     if length(AOD_continental) > 0
338         R_continental=corr(AOD_continental(:,3), PM_continental(:,3));
339     else
340         R_continental = NaN;
341     end
342
343     figure('Visible','off'); clf;
344     scatter(AOD_continental(:,3), PM_continental(:,3), 'k')
345     hold on
346     plot(AOD_reg_continental,PM_reg_continental, 'r')

```

```

347 xlabel('AOD');
348 ylabel('Daily mean PM_{2.5} concentration [\mu g/m^{\{3\}}]');
349 title(strcat(city_name, ' continental scale,', {' '}, res_table))
350 legend('Matched AOD and PM_{2.5} data points', strcat('PM_{2.5}=','...
351         num2str(a_continental, '%.2f'), '\cdot AOD+', num2str(b_continental, '%.2f')));
352
353 savefolder = strcat('D:\Master Thesis\Final Study\MATLAB\Images',...
354                     '\Varying scale');
355 savename = char(strcat(city_savename, '_2008.continental_scale-',...
356                        res, '.png'));
357
358 cd(savefolder)
359 print('-dpng', savename);
360
361
362 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
363 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
364 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
365
366 %% CREATE LATEX TABLE
367
368 if table_choice == 2
369
370     if city_choice == 1
371         k = 3;
372     elseif city_choice == 2
373         k = 18;
374     elseif city_choice == 3
375         k = 33;
376     elseif city_choice == 4
377         k = 48;
378     elseif city_choice == 5
379         k = 63;
380     elseif city_choice == 6
381         k = 78;
382     end
383
384
385     if resolution_choice < 3
386
387         if resolution_choice == 1 && city_choice == 1
388
389             table_1={'\textbf{City}', ' & ', '\textbf{Region Size}',...
390                    ' & ', '\textbf{Statistics}', ' & ',...
391                    '\textbf{MYD04 10 km}', ' & ',...
392                    '\textbf{MYD04 3 km}', '\\ \hline'};
393
394             for i = 2 : 2 : 10
395                 for j = 2 : 91
396                     table_1{j,i}=strcat({' '}, '&', {' '});
397                 end
398             end
399
400             top={'\begin{tabular}{c 1 1 1 1}' ' ' ' ' ' ' ' ' ,...
401                '' '' ''};
402             bottom={'\end{tabular}' ' ' ' ' ' ' ' ' ' ' ' ' };
403
404             table_1=vertcat(top,table_1);
405             table_1=vertcat(table_1,bottom);
406
407         end
408
409     if resolution_choice == 1
410         l = 7;
411     elseif resolution_choice == 2
412         l = 9;
413     end
414
415

```

```

417         if resolution_choice == 1
418
419             table_1{k,1} = strcat('\textbf{' , city_name, '}');
420             table_1{k,3} = 'Urban scale';
421             table_1{k,1-2}='\textit{N}';
422             table_1{k+1,1-2}='\textit{R}';
423             table_1{k+2,1-2}='Slope';
424             table_1{k+3,1-2}='intersect';
425             table_1{k+4,1-2}='\textit{p}';
426
427             table_1{k+5,3} = 'Meso-scale';
428             table_1{k+5,1-2}='\textit{N}';
429             table_1{k+6,1-2}='\textit{R}';
430             table_1{k+7,1-2}='Slope';
431             table_1{k+8,1-2}='intersect';
432             table_1{k+9,1-2}='\textit{p}';
433
434             table_1{k+10,3} = 'Continental scale';
435             table_1{k+10,1-2}='\textit{N}';
436             table_1{k+11,1-2}='\textit{R}';
437             table_1{k+12,1-2}='Slope';
438             table_1{k+13,1-2}='intersect';
439             table_1{k+14,1-2}='\textit{p}';
440
441             table_1{k,10}='\\';
442             table_1{k+1,10}='\\';
443             table_1{k+2,10}='\\';
444             table_1{k+3,10}='\\';
445             table_1{k+4,10}='\\ \hline';
446
447             table_1{k+5,10}='\\';
448             table_1{k+6,10}='\\';
449             table_1{k+7,10}='\\';
450             table_1{k+8,10}='\\';
451             table_1{k+9,10}='\\ \hline';
452
453             table_1{k+10,10}='\\';
454             table_1{k+11,10}='\\';
455             table_1{k+12,10}='\\';
456             table_1{k+13,10}='\\';
457             table_1{k+14,10}='\\ \hline';
458
459         end
460
461         table_1{k,1}=num2str(number_of_points_urban);
462         table_1{k+1,1}=num2str(R_urban, '%.3f');
463         table_1{k+2,1}= num2str(a_urban, '%.2f');
464         table_1{k+3,1}=num2str(b_urban, '%.2f');
465         table_1{k+4,1}= p_urban_str;
466
467         table_1{k+5,1}=num2str(number_of_points_meso);
468         table_1{k+6,1}=num2str(R_meso, '%.3f');
469         table_1{k+7,1}= num2str(a_meso, '%.2f');
470         table_1{k+8,1}=num2str(b_meso, '%.2f');
471         table_1{k+9,1}= p_meso_str;
472
473         table_1{k+10,1}=num2str(number_of_points_continental);
474         table_1{k+11,1}=num2str(R_continental, '%.3f');
475         table_1{k+12,1}= num2str(a_continental, '%.2f');
476         table_1{k+13,1}=num2str(b_continental, '%.2f');
477         table_1{k+14,1}= p_continental_str;
478
479     end
480
481     if resolution_choice > 2
482
483         if resolution_choice == 3 && city_choice == 1
484
485             table_2={'\textbf{City}', ' & ', '\textbf{Region Size}',...
486                 ' & ', '\textbf{Statistics}', ' & ',...

```

```

487         '\textbf{MAIAC 10 km}', '&', ...
488         '\textbf{MAIAC 3 km}', '&', ...
489         '\textbf{MAIAC 1 km}', '\\ \hline'};
490
491
492     for i = 2 : 2 : 12
493         for j = 2 : 91
494             table_2{j,i}=strcat({' ', '&', {' '}));
495         end
496     end
497
498     top={'\begin{tabular}{c l l l l l}' ' ' ' ' ' ' ' ',...
499          ' ' ' ' ' ' ' '};
500     bottom={'\end{tabular}' ' ' ' ' ' ' ' ' ' ' ' ' ' };
501
502     table_2=vertcat(top,table_2);
503     table_2=vertcat(table_2,bottom);
504 end
505
506 if resolution_choice == 3
507     l = 11;
508 elseif resolution_choice == 4
509     l = 9;
510 elseif resolution_choice == 5
511     l = 7;
512 end
513
514 if resolution_choice == 5
515
516     table_2{k,l} = strcat('\textbf{' , city_name, '});
517     table_2{k,3} = 'Urban scale';
518     table_2{k,l-2}='\textit{N}';
519     table_2{k+1,l-2}='\textit{R}';
520     table_2{k+2,l-2}='Slope';
521     table_2{k+3,l-2}='intersect';
522     table_2{k+4,l-2}='\textit{p}';
523
524     table_2{k+5,3} = 'Meso-scale';
525     table_2{k+5,l-2}='\textit{N}';
526     table_2{k+6,l-2}='\textit{R}';
527     table_2{k+7,l-2}='Slope';
528     table_2{k+8,l-2}='intersect';
529     table_2{k+9,l-2}='\textit{p}';
530
531     table_2{k+10,3} = 'Continental scale';
532     table_2{k+10,l-2}='\textit{N}';
533     table_2{k+11,l-2}='\textit{R}';
534     table_2{k+12,l-2}='Slope';
535     table_2{k+13,l-2}='intersect';
536     table_2{k+14,l-2}='\textit{p}';
537
538     table_2{k,12}='\\';
539     table_2{k+1,12}='\\';
540     table_2{k+2,12}='\\';
541     table_2{k+3,12}='\\';
542     table_2{k+4,12}='\\ \hline';
543
544     table_2{k+5,12}='\\';
545     table_2{k+6,12}='\\';
546     table_2{k+7,12}='\\';
547     table_2{k+8,12}='\\';
548     table_2{k+9,12}='\\ \hline';
549
550     table_2{k+10,12}='\\';
551     table_2{k+11,12}='\\';
552     table_2{k+12,12}='\\';
553     table_2{k+13,12}='\\';
554     table_2{k+14,12}='\\ \hline';
555 end

```

```

557         table_2{k,1}=num2str(number_of_points_urban);
558         table_2{k+1,1}=num2str(R_urban, '%.3f');
559         table_2{k+2,1}= num2str(a_urban, '%.2f');
560         table_2{k+3,1}=num2str(b_urban, '%.2f');
561         table_2{k+4,1}= p_urban_str;
562
563         table_2{k+5,1}=num2str(number_of_points_meso);
564         table_2{k+6,1}=num2str(R_meso, '%.3f');
565         table_2{k+7,1}= num2str(a_meso, '%.2f');
566         table_2{k+8,1}=num2str(b_meso, '%.2f');
567         table_2{k+9,1}= p_meso_str;
568
569         table_2{k+10,1}=num2str(number_of_points_continental);
570         table_2{k+11,1}=num2str(R_continental, '%.3f');
571         table_2{k+12,1}= num2str(a_continental, '%.2f');
572         table_2{k+13,1}=num2str(b_continental, '%.2f');
573         table_2{k+14,1}= p_continental_str;
574
575     end
576
577 end
578 end
579 multiWaitbar('City progress..', city_choice/6);
580 end
581
582 mean_fraction_removed_urban=fractions_removed_urban/30
583 mean_fraction_removed_meso=fractions_removed_meso/30
584 mean_fraction_removed_continental=fractions_removed_continental/30
585
586 %% SAVE LATEX TABLES
587
588 if table_choice == 2
589
590     %% ALL DATA
591
592     tablefolder = strcat('D:\Master Thesis\Final Study\MATLAB\Tables\',...
593                         'Varying scale');
594     cd(tablefolder)
595
596     tablename_1 = char(strcat('varying-scale-US-2008-MYD04-continental.data.tex'));
597     tablename_2 = char(strcat('varying-scale-US-2008-MAIAC-continental.data.tex'));
598
599     [m n] = size(table_1);
600     fileID_1 = fopen(tablename_1, 'w');
601     fileID_2 = fopen(tablename_2, 'w');
602     formatSpec = '%s\n';
603
604     for row = 1 : m
605         fprintf(fileID_1,formatSpec, char(strcat(table_1{row,:})));
606         fprintf(fileID_2,formatSpec, char(strcat(table_2{row,:})));
607     end
608
609     fclose(fileID_1);
610     fclose(fileID_2);
611
612     %% SOME DATA
613
614     tablename_1.some = char(strcat('varying-scale-US-2008-MYD04-some.data.tex'));
615     tablename_2.some = char(strcat('varying-scale-US-2008-MAIAC-some.data.tex'));
616
617     fileID_1.some = fopen(tablename_1.some, 'w');
618     fileID_2.some = fopen(tablename_2.some, 'w');
619
620     rows_some = [1 2 3 4 7 8 9 12 13 14 17 18 19 22 23 24 27 28 29 32,...
621                 33 34 37 38 39 42 43 44 47 48 49 52 53 54 57 58 59 62 63 64 67,...
622                 68 69 72 73 74 77 78 79 82 83 84 87 88 89 92 93];
623
624     for i = 1 : length(rows_some)
625         fprintf(fileID_1.some,formatSpec, char(strcat(table_1{rows_some(i),:})));
626         fprintf(fileID_2.some,formatSpec, char(strcat(table_2{rows_some(i),:})));

```

```
627     end
628
629     fclose(fileID_1.some);
630     fclose(fileID_2.some);
631
632 end
633
634 multiWaitbar( 'CloseAll' );
635 cd 'D:\Master Thesis\Final Study\MATLAB'
```