

DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**DEVELOPMENT OF A REGIONAL-SCALE
PREDICTION METHOD FOR GROUND LEVEL
PARTICULATE MATTER CONCENTRATIONS
BASED ON AOD FROM DIFFERENT
SATELLITES IN TURKEY**

by
Gizem TUNA TUYGUN

July, 2021
İZMİR

**DEVELOPMENT OF A REGIONAL-SCALE
PREDICTION METHOD FOR GROUND LEVEL
PARTICULATE MATTER CONCENTRATIONS
BASED ON AOD FROM DIFFERENT
SATELLITES IN TURKEY**

**A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of
Philosophy in Environmental Engineering, Environmental Science Program**

**by
Gizem TUNA TUYGUN**

**July, 2021
İZMİR**

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled “**DEVELOPMENT OF A REGIONAL-SCALE PREDICTION METHOD FOR GROUND LEVEL PARTICULATE MATTER CONCENTRATIONS BASED ON AOD FROM DIFFERENT SATELLITES IN TURKEY**” completed by **GİZEM TUNA TUYGUN** under supervision of **PROF.DR. TOLGA ELBİR** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof.Dr. Tolga ELBİR

Supervisor

Assoc.Prof.Dr. Orhan GÜNDÜZ

Thesis Committee Member

Assoc.Prof.Dr. Okan FISTIKOĞLU

Thesis Committee Member

Prof.Dr. Sait C. SOFUOĞLU

Examining Committee Member

Prof.Dr. Ülkü ALVER ŞAHİN

Examining Committee Member

Prof.Dr. Özgür ÖZÇELİK

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Prof. Tolga ELBİR, my thesis supervisor, for his insightful guidance, full support, and great encouragement throughout my study. I gratefully appreciate advice and comments from my thesis committee members, Assoc. Prof. Orhan GÜNDÜZ and Assoc. Prof. Okan FISTIKOĞLU.

I also wish to thank to Prof. Abdurrahman BAYRAM, Prof. Mustafa ODABAŞI, Assist. Prof. Yetkin DUMANOĞLU, Assoc. Prof. Melik KARA, Hasan ALTIOK, Dr. Barış YAMAN, Ersan GÜNEL, Begüm KADIOĞLU, and Esra ÖZDEMİR for all support and moral motivation during preparation of my thesis.

I am notably thankful to my family, GÜL TUNA, EMRAH TUNA, AYŞİN TUYGUN, and ATILLA TUYGUN for their great understanding and support during my Ph.D. education.

Last but not most, I would like to express my most special thanks to my husband, ÖVÜNÇ TUYGUN for his endless understanding, support, and patient.

This work was financially supported by the Scientific and Technological Research Council of Turkey (TUBITAK) 2214A International Scholarship Programme for Ph.D. Students. This study is also supported by TUBITAK (Project No: 119Y005). I would like to thank TUBITAK for this financial support. I gratefully acknowledge NASA for making the MODIS aerosol products and MERRA-2 data publicly available. I also thank the Turkish Ministry of Environment and Urbanization for PM₁₀ data publicly available.

Gizem TUNA TUYGUN

DEVELOPMENT OF A REGIONAL-SCALE PREDICTION METHOD FOR GROUND LEVEL PARTICULATE MATTER CONCENTRATIONS BASED ON AOD FROM DIFFERENT SATELLITES IN TURKEY

ABSTRACT

Respirable particulate matter with aerodynamic diameters lower than $10\text{ }\mu\text{m}$ (PM_{10}) is closely associated with adverse impacts on the atmospheric environment and human health. Estimation of daily ground-level PM_{10} concentrations based on satellite aerosol optical depth (AOD) products has been widely used over large regions with the help of statistical regression models. This study was the first attempt in Turkey to estimate daily PM_{10} concentrations based on the newly released Collection 6.1 (C6.1) Moderate Resolution Imaging Spectroradiometer (MODIS) DTB AOD products from Terra and Aqua satellites across Turkey. This study developed two different space-time random forest (STRF) models that outperformed most models with strong predictive power to evaluate the differences between the time windows representing the mean PM_{10} around the overpass times of the satellites with the synergetic use of space-time information over Turkey. This model uses spatiotemporal information along with MODIS AOD, meteorological, and land-related data as input to improve the overall accuracy of PM_{10} estimation for 2008-2019. Solar radiation and dust mass surface concentrations were the most effective predictors for Terra and Aqua overpass times, respectively. The STRF model performed moderately well, with a moderate cross-validation (CV) correlation coefficient (R) of 0.71 (0.67) and a low root-mean-square error over the national scale with AOD. The model better estimated PM_{10} concentrations at both regional (R up to 0.80) and individual site scales (R up to 0.90). Moreover, the STRF model developed in this study is a first step to construct a high-quality PM_{10} dataset across Turkey that is important for air pollution studies.

Keywords: MODIS, AOD, PM_{10} estimation, Random Forest, Turkey

**TÜRKİYE'DE YER SEVİYESİNDEKİ PARTİKÜL MADDE
KONSANTRASYONLARININ BÖLGESEL TAHMİNİNDE FARKLI
UYDULARDAN TEMİN EDİLEN AEROSOL OPTİK DERİNLİĞİ
VERİSİNE DAYANAN BİR YÖNTEM GELİŞTİRİLMESİ**

ÖZ

Aerodinamik çapları 10 μm 'den küçük olan solunabilir partikül maddeler (PM_{10}), insan sağlığı ve atmosfer üzerindeki olumsuz etkilerle yakından ilişkilidir. Uydu bazlı aerosol optik derinliği (AOD) ürünlerine dayalı istatistiksel regresyon modelleri büyük alanlarda yer seviyesinde ölçülen günlük PM_{10} konsantrasyonlarının belirlenmesinde yaygın olarak kullanılmaktadır. Bu çalışma, Türkiye genelinde Terra ve Aqua uydularından elde edilen yeni geliştirilmiş Koleksiyon 6.1 (C6.1) Orta Çözünürlüklü Görüntüleme Spektroradiometresi (MODIS) DTB AOD ürünleri ile günlük PM_{10} konsantrasyonlarını tahmin etmek için Türkiye'de yapılan ilk çalışmadır. Çalışmada, zamansal ve mekânsal parametrelerin ortak kullanımı ile uyduların geçiş zamanlarındaki PM_{10} konsantrasyonlarının farklılığı dikkate alınarak Türkiye’de diğer modellere göre tahmin gücü daha yüksek olduğu bilinen zaman-mekân değişkenli iki farklı rastgele orman modeli (STRF) geliştirilmiştir. Bu model, 2008-2019 yılları arasında PM_{10} tahmininin doğruluğunu iyileştirmek için MODIS AOD, meteorolojik ve arazi bazlı verileri girdi olarak kullanmaktadır. Güneş radyasyonu ve toz yüzey konsantrasyonu sırasıyla Terra ve Aqua uydularının geçiş zamanı için en etkili tahmin parametreleri olarak belirlenmiştir. STRF modeli ile, Terra ve Aqua AOD için 0,71 (0,67) çapraz geçerlilik korelasyon katsayısı (R) ve kök ortalama kare hatası ile ulusal ölçekte orta düzeyde bir tahmin performansı elde edilmiştir. Modelin özellikle hem bölgesel (0.80’e varan korelasyon katsayısı) hem de tekil istasyon bazında (0.90’a varan korelasyon katsayısı) PM_{10} konsantrasyonlarını daha iyi tahmin ettiği belirlenmiştir. Ayrıca bu çalışmada geliştirilen STRF modeli, Türkiye genelinde hava kirliliği çalışmaları için önemli olan yüksek kaliteli bir PM_{10} veri seti oluşturmaının ilk adımıdır.

Anatar kelimeler: MODIS, AOD, PM_{10} tahmini, Rastgele Orman, Türkiye

CONTENTS

	Page
Ph.D. THESIS EXAMINATION RESULT FORM	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
ÖZ	v
LIST OF FIGURES.....	viii
LIST OF TABLES	xi
 CHAPTER 1 – INTRODUCTION	 1
1.1 Introduction	1
 CHAPTER 2 – LITERATURE REVIEW	 5
2.1 Fundamentals of Remote Sensing in Aerosol Pollution.....	5
2.2 Satellite sensors and platforms	11
2.3 MODIS AOD and Retrieval Algorithms	15
2.4 Ground-Based AOD observations	20
2.5 Ground-Level PM ₁₀ Measurements and Air Quality Directives	23
2.6 Relationship Between Ground-Level PM and Satellite Based AOD	26
2.7 Estimation methods of PM with satellite based AOD	33
2.7.1 Simulation-based models.....	33
2.7.2 Statistical Models	35
 CHAPTER 3 – MATERIALS AND METHOD	 47
3.1 Study Area	47
3.2 Data Collection and Integration	52
3.2.1 Satellite AOD Data	53
3.2.2 Ground-Based AOD Data.....	54
3.2.3 Ground Level PM ₁₀ Data	55
3.2.4 MERRA-2 Reanalysis Aerosol and Meteorological Fields.....	58

3.2.5 Auxiliary Data	60
3.2.6 Validation of MODIS AOD Data with AERONET Observations	61
3.2.7 Data Processing and Integration	65
3.2.8 Variable Selection.....	69
3.2.9 Construction and Validation of Estimation Model.....	71
CHAPTER 4 – RESULTS	80
4.1 Validation of MODIS AOD Products	80
4.1.1 Site Scale Validation Results.....	83
4.1.2 Seasonal Validation Results	93
4.2 Spatiotemporal Variation of AOD Over Turkey	100
4.3 Descriptive Statistics for Inputs of Estimation Models.....	112
4.4 Effect of input variables on PM ₁₀ estimation	119
4.4.1 Relationship Between AOD and PM ₁₀	119
4.4.2 Relationship Between PM ₁₀ and Auxiliary Variables	125
4.4.3 Variable Importance and Feature Selection for Estimation Model	129
4.5 Performance of The RF Models	133
4.5.1 National Scale Model	133
4.5.2 Regional and Site-Based Models.....	141
4.5.3 Seasonal Performance of RF Models	147
CHAPTER 5 – CONCLUSION AND RECOMMENDATIONS.....	149
5.1 Comparison with other studies	149
5.2 Conclusion and Recommendations	151
REFERENCES.....	157
APPENDICES	195

LIST OF FIGURES

	Page
Figure 2.1 Electromagnetic spectrum	6
Figure 2.2 Illustration of different paths of radiance received by the satellite sensor .	7
Figure 2.3 Illustration of effecting parameters on AOD	10
Figure 2.4 The geostationary and polar orbit satellites to retrieve AOD	13
Figure 2.5 Illustration of active and passive sensors.....	14
Figure 2.6 AOD retrieval algorithm of 10 km (red) and 3 km (blue) DT products ...	16
Figure 2.7 Enhanced AOD retrieval algorithm of 10 km DB product.....	17
Figure 2.8 CIMEL Sun-photometer	21
Figure 2.9 AERONET sites.....	22
Figure 2.10 The location of the AERONET stations in Turkey	23
Figure 2.11 PM estimation models widely used in the literature.....	42
Figure 3.1 Geographical regions in Turkey	48
Figure 3.2 Spatial distribution of multiannual average dust mass surface concentrations and AE values over Turkey	49
Figure 3.3 Spatial distribution of multiannual average meteorological parameters over Turkey	51
Figure 3.4 CORINE land cover over Turkey	52
Figure 3.5 Location of PM ₁₀ monitoring stations	57
Figure 3.6 Numbers of regional PM ₁₀ stations over Turkey.....	58
Figure 3.7 An illustration of the spatiotemporal collocation scheme of the satellite retrievals with the AERONET measurements.....	63
Figure 3.8 An example scheme of RF structure	73
Figure 3.9 Hyper grid search results for two dataset and min.node size	75
Figure 3.10 Schematic illustration of CV methods used in this study	77
Figure 3.11 Flow chart of modelling study.....	78
Figure 3.12 Illustration of 4 quadrants between PM ₁₀ and AOD for each satellite overpass time.....	79
Figure 4.1 Performance evaluation with EE over two AERONET sites in Turkey...	83
Figure 4.2 Validation results of the MODIS AOD products over the urban site.....	86

Figure 4.3 Validation of 3km-DT AOD over the coastal rural site	90
Figure 4.4 Validation results of 10km-DT AOD over the coastal rural site	91
Figure 4.5 Validation results of 10km-DTB AOD over the coastal rural site	92
Figure 4.6 Validation results of 10km-DB AOD over the coastal rural site.....	93
Figure 4.7 Seasonal validation results of MODIS AOD products over the coastal rural site	95
Figure 4.8 Seasonal EE values of MODIS AOD products over the coastal rural site.	96
Figure 4.9 Seasonal validation results of MODIS AOD products over the urban site	98
Figure 4.10 Seasonal EE values of MODIS AOD products over the urban site.....	99
Figure 4.11 Multiannual average (a) and regional (b) variation of NDVI over Turkey	101
Figure 4.12 Spatial distribution of multiannual average AOD products between 2008 and 2019 in Turkey	105
Figure 4.13 Multiannual seasonal distribution of DTB AOD over Turkey	108
Figure 4.14 Spatial distribution of multiannual seasonal average a) specific humidity and b) precipitation.....	109
Figure 4.15 Spatial distribution of multiannual seasonal average NDVI over Turkey (a) and in the geographic regions of Turkey (b).....	110
Figure 4.16 Multiannual seasonal average DTB AOD from Terra and Aqua satellites in the geographic regions of Turkey.....	111
Figure 4.17 Relationship between PM ₁₀ and AOD over Turkey.....	120
Figure 4.18 The R between PM ₁₀ and AOD over 213 PM ₁₀ sites.....	122
Figure 4.19 Data coverage of TAOD and AAOD data over 213 PM ₁₀ sites	124
Figure 4.20 Correlation matrix of the variables for the TAOD model variable selection	126
Figure 4.21 Correlation matrix of the variables for the AAOD model.....	127
Figure 4.22 Variable importance for TAOD and AAOD model before (a) and after (b) variable selection.....	130
Figure 4.23. Variable importance scores of for regional RF models	133
Figure 4.24 National RF model results for each satellite overpass time (black and red	

dotted lines represent 1:1 line and regression line, respectively).....	135
Figure 4.25 Variation of the observed and predicted daily PM ₁₀ concentrations by national TAOD and AAOD RF models	135
Figure 4.26 Quantile regression results for the Terra overpass time (black and red dotted lines represent 1:1 line and regression line, respectively).....	139
Figure 4.27 Quantile regression results for the Aqua overpass time (black and red dotted lines represent 1:1 line and regression line, respectively).....	140
Figure 4.28 Regional TAOD RF model performance.....	143
Figure 4.29 Regional AAOD RF model performance	144
Figure 4.30 Site scale TAOD RF model performance.....	145
Figure 4.31 Site scale RF model performance for Aqua overpass time.....	146
Figure 4.32 Illustration of regional estimated and observed PM ₁₀	147

LIST OF TABLES

	Page
Table 2.1 Summary information about widely used satellites and sensors.....	13
Table 3.1. Summary of AOD data used for validation	64
Table 3.2 List of data used in modelling study	69
Table 3.3 RF hyper parameters for grid search.....	74
Table 4.1 Overall retrieval accuracy between AOD products in two sites	81
Table 4.2 Regional statistics of DTB AOD over Turkey	106
Table 4.3 Summary statistics for the dependent and selected independent variables over the Terra overpass time	113
Table 4.4 Summary statistics for the dependent and selected independent variables over the Aqua overpass time	113
Table 4.5 Regional descriptive statistics of dependent and independent variables for the TAOD model.....	115
Table 4.6 Regional descriptive statistics of dependent and independent variables for the AAOD model	116
Table 4.7 Seasonal descriptive statistics of the dependent and independent variables for the TAOD model	118
Table 4.8 Seasonal descriptive statistics of the dependent and independent variables for the AAOD model.....	119
Table 4.9 Relationship between AOD and PM ₁₀ based on different quadrants.....	122
Table 4.10 VIF results for each national RF model	128
Table 4.11 Spatiotemporal predictive power of national TAOD and AAOD RF models	137
Table 4.12 Seasonal performance of STRF models.....	148
Table 5.1 Previous studies in the literature	150

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Air pollution has significant effects on human health, particularly in urban areas, due to population growth, fossil fuel consumption, rapid industrialization, and other anthropogenic activities (EEA, 2020). The inhalable PM with a diameter lower than 10 μm are defined as PM_{10} and responsible for worldwide increasing health effects (Feng et al., 2019; Choubin et al., 2020). The epidemiological research indicated the long-term and short-term health effects of PM_{10} and finer PM on public health (Pope et al., 2002; Dominici et al., 2006; Brook et al., 2010; Stafoggia et al., 2013). Public attention has been increasing about the health effects of PM, mainly over the highly polluted countries in the world (Liu et al., 2019; Marchetti et al., 2019; Choubin et al., 2020). According to the World Health Organization (WHO), about 7 million people die per year due to PM pollution. Ninety-one percent of the world's populations live in regions where air quality does not meet the criteria depicted the WHO guideline (WHO, 2020).

Anthropogenic activities (emissions from power, transportation, industry, and waste incineration) and different natural sources (dust storms, volcanic eruptions, sea spray, wildfires, etc.) can cause high-level PM concentrations (Shtein et al., 2018). High-level PM concentrations can also significantly affect the temperature, solar radiation, precipitation cycle, and energy balance of the earth's atmosphere; thus, it also has acute effects on the environment and ecological systems (Sorek-hamer et al., 2016; Banerjee et al., 2017; Ranjan et al., 2021). The accurate information on the spatiotemporal variations of natural and anthropogenic PM at a regional and national scale is essential for effective air quality management. However, limited monitoring sites in regional areas make it difficult (Li, 2020).

Satellite aerosol products with favorable characteristics including broad coverage, high spatial resolutions, and long-term records provide valuable aerosol information. Aerosol optical depth (AOD) is a widely used dimensionless geophysical product,

representing the degree to which aerosols attenuate light in the atmospheric column. AOD from various satellites and sensors have been widely used to estimate both ground-level PM₁₀ or PM_{2.5} concentrations over large areas as demonstrated in the literature (Meng et al., 2016; Chen et al., 2018; Ahmad et al., 2019; Park et al., 2019; Staggofia et al., 2019; Jin et al., 2020; Meng et al., 2021; Wei et al., 2021a). AOD can be measured by both satellites and ground-based sun photometers at different network stations over the world (NASA, 2018a). Since the coverage of the ground-based stations is minimal in most of the countries, satellites have a significant role in obtaining more extensive spatial coverage of aerosol data. However, satellite-based PM₁₀ estimations with AOD has still problems producing sensitive, consistent, and reliable PM estimations over different regions since spatial and temporal variability in the relationship between AOD and surface PM is location-specific.

With the continuous improvements in aerosol retrieval algorithms, AOD products from various satellite sensors have been commonly used in PM₁₀ estimation purposes. However, AOD from the MODIS aboard NASA's Terra and Aqua satellites are the most used data source for PM estimation or prediction (Stafoggia et al., 2017; Chen et al., 2019a; Park et al., 2019; Nabavi et al., 2019; Wei et al., 2019a; Wei et al., 2020a; Yan et al., 2021). These studies preferred different AOD products, collections, and algorithms. However, AOD products were produced from different aerosol retrieval algorithms. The accuracy of the products could show significant regional and seasonal variations (He et al., 2017; Wei et al., 2018). However, numerous studies have indicated that AOD products have been effectively used to estimate PM concentrations on large scales due to the considerable association between AOD and PM concentrations (Hu et al., 2014a; Guo et al., 2017a; Xiao et al., 2017).

Simple linear regression models with only AOD data were used in early studies to estimate ground-level PM concentrations (Chu et al., 2003; Engel-Cox et al., 2004). However, only AOD data cannot directly provide accurate estimation for surface PM concentrations. Geographical parameters, meteorological and land use information should be included in PM prediction models (Gupta & Christopher, 2009) to integrate the physical insights into the dynamics of PM concentrations. There is a broad range

of developed model with traditional regression approaches to analyze the relationship between ground-based PM and satellite-based AOD (Ma et al., 2016; Xiao et al., 2017; Zeydan & Wang, 2019). Instead of the conventional statistical models, recent advances with the development of advanced statistical techniques such as artificial intelligence (AI) and machine learning (ML) models have become popular due to their highly flexible application and higher prediction abilities (Hu et al., 2017). They can also be the best options to solve the complex nonlinear relationship between AOD, PM, and auxiliary variables and generally achieve superior predictive performance (Di et al., 2016; Chen et al., 2018; Wei et al., 2019a; Just et al., 2020; Park et al., 2020; Wei et al., 2020a; Wei et al., 2021a; Zhang et al., 2021).

Although several studies have estimated daily PM₁₀ or PM_{2.5} concentrations with AOD over China (Hu et al., 2017; Chen et al., 2018; Wei et al., 2019a), the USA (Jin et al., 2020; Just et al., 2020; Ghahremanloo et al., 2021), and Europe (Staggofia et al., 2017; 2019), there is only one study (Zeydan & Wang, 2019) that estimated 24-h mean PM_{2.5} concentrations in Turkey. This study used MODIS AOD data at 3 km and 10 km spatial resolutions from both Terra and Aqua satellites with ground-level PM_{2.5} data over Turkey. Five years (2011–2015) of PM_{2.5} measurements from seven monitoring stations in Ankara and three years (2013–2015) of PM_{2.5} measurements 13 monitoring stations in the Marmara Region were obtained during the heating season. Zeydan & Wang (2019) focused on the linear and nonlinear regression models to find the relationship between PM_{2.5} and AOD data. They used old collection (C6) MODIS AOD data and simple linear regression for PM_{2.5} estimations. Since Turkey is located in the Mediterranean Basin, the region is affected by the long-range transport of desert dust, causing relatively high PM levels in specific seasons (Toros et al., 2013; Erdun et al., 2015; Ozdemir et al., 2020). Despite removing the PM_{2.5} concentrations affected by free tropospheric long-range transport using back trajectory modeling analysis, the R-value of the regional model, including 20 monitoring sites, reached 0.59. It indicated that it is difficult to obtain a high PM estimation accuracy with the traditional regression model in the region due to the complexity of the data.

In this study, an STRF model was developed to solve the nonlinear and complex relationship between the PM_{10} and AOD using meteorological and land-related data over Turkey. To the best of our knowledge, this is the first proposed study to estimate intra-daily PM_{10} concentrations using the C6.1 MODIS AOD product in Turkey. Since the MODIS sensor is onboard two satellite platforms (Terra and Aqua), which have different overpass times (10:30 LST and 13:30 LST), two different national models were developed to estimate intra-daily PM_{10} concentrations around the overpass times of Terra and Aqua satellites. These estimations are helpful to determine the effect of satellite overpass time on estimated PM_{10} concentrations that possibly originate from different sources, rather than using daily mean PM_{10} estimates. The main objectives of this study are: (i) to present an overview on the accuracy of MODIS Terra, and Aqua AOD products, (ii) to analyze the spatiotemporal variation of AOD and PM_{10} concentrations over Turkey, (iii) to estimate intra-daily (mean PM_{10} around the Aqua and Terra overpasses) PM_{10} concentrations based on AOD and various land-related and meteorological parameters using STRF model, which includes spatiotemporal information to address temporal variations and spatial heterogeneities in Turkey, and (iv) to compare the estimation performances of each STRF model for different time windows. Regional and seasonal differences of model performances were also discussed.

CHAPTER TWO

LITERATURE REVIEW

In this section, the essential background of current data and methods is given in detail. A literature review was mainly constructed to clarify the methodology which can deal with satellite data for PM estimation. Section 2.1 introduces the fundamentals of remote sensing in aerosol pollution. Section 2.2 gives background information properties of satellite sensors and platforms to address knowledge about them. Section 2.3 introduces MODIS AOD products and retrieval algorithms to address the suitability of MODIS AOD for estimation of PM₁₀ pollution according to our defined purposes. Section 2.4 and 2.5 explain the ground based AOD and PM₁₀ measurements. Section 2.6 and 2.7 discuss the importance of PM₁₀ estimation with AOD and identify the potential relationships, affecting factors, and methods. The information of the current methodology and review of modeling approaches are discussed in Section 2.7.

2.1 Fundamentals of Remote Sensing in Aerosol Pollution

Remote sensing is a method to collect information at a distance from the earth's surface without a physical sampling by different instruments placed on a satellite platform (NASA, 2016). In other words, the principle of remote sensing is mapping the features of the earth's surface using reflected electromagnetic energy, which is varying by different objects (NASA, 2016).

Since the remote sensing process includes interaction between incident radiation and the targets of interest, the primary necessity of remote sensing is electromagnetic radiation to brighten the target. All electromagnetic radiation has principal properties and shows predictable acts based on wave theory (NRC, 2015). To understand remote sensing, the frequency and wavelength of electromagnetic radiation are especially significant. The wavelength and frequency of electromagnetic radiation are powerful to understand the extracted information from remote sensing data, and the two properties are inversely related to each other (NRC, 2015).

The UV or ultraviolet portion of the spectrum representing the shortest wavelengths is practical for most remote sensing applications. Some materials, including rocks and minerals, can fluoresce or emit visible light when UV radiation brightens (NRC, 2015). The visible wavelengths cover the narrow ranges from approximately 0.4 to 0.7 μm . The range approximately from 0.7 μm to 100 μm is covered by the near infrared (NIR) region. The electromagnetic spectrum is given in Figure 2.1.

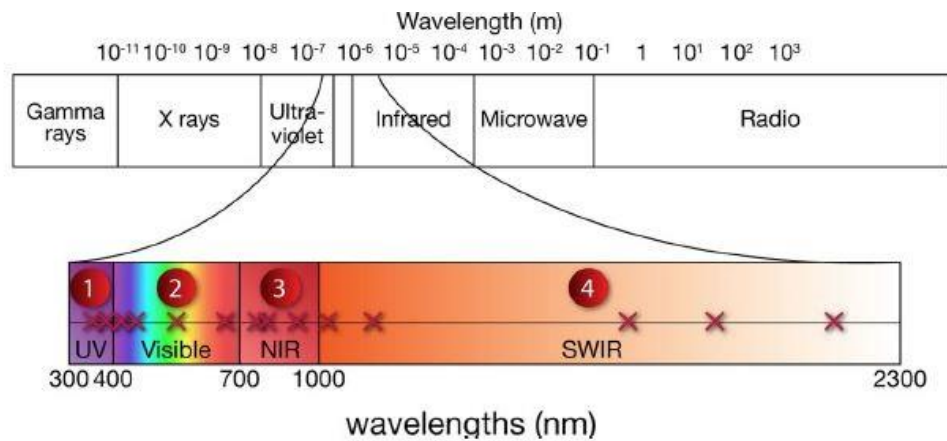


Figure 2.1 Electromagnetic spectrum (NASA, 2021a)

Earth-observing systems consist of the satellite (platform) and the instrument (sensor). The radiation whose intensity is influenced by the atmosphere, or the surface (land, water, ice) is measured by a sensor. Thus, information about both the atmospheric and surface conditions is collected by satellites (NASA, 2016). Radiation travels through the atmosphere before radiation used for remote sensing. A specific atmospheric component must be measured in the spectral range in which it absorbs, emits, or scatters radiation using spectral regions in the UV, Visible, Infrared, to microwave by the instrument (NASA, 2016).

When the electromagnetic radiation from the Sun passes through the atmosphere, it is reflected by the earth, the atmospheric contribution to the satellite signal occurs. The satellite sensor detects it when it passes through the atmosphere again, as seen in Figure 2.2 (Hadjimitsis & Clayton, 2009). In the case of atmospheric effects, the origin for any attempt to perform atmospheric correction to satellite data is the setting up of an equation describing all the processes with the different atmospheric variables and

processes that contributes to the attenuation of the signal received by a satellite sensor (Hadjimitsis & Clayton, 2009). Mainly interaction processes including scattering and absorption occur through the atmosphere before the Sun's energy reaches the surface. The incoming light and radiation can be affected by mechanisms of absorption and scattering with the help of particles and gases in the atmosphere. The intensity of these processes increases or decreases depending on the wavelength before reaching the sensor (Hadjimitsis & Clayton, 2009). Energy interactions in the atmosphere and the earth's surface are given in Figure 2.2.

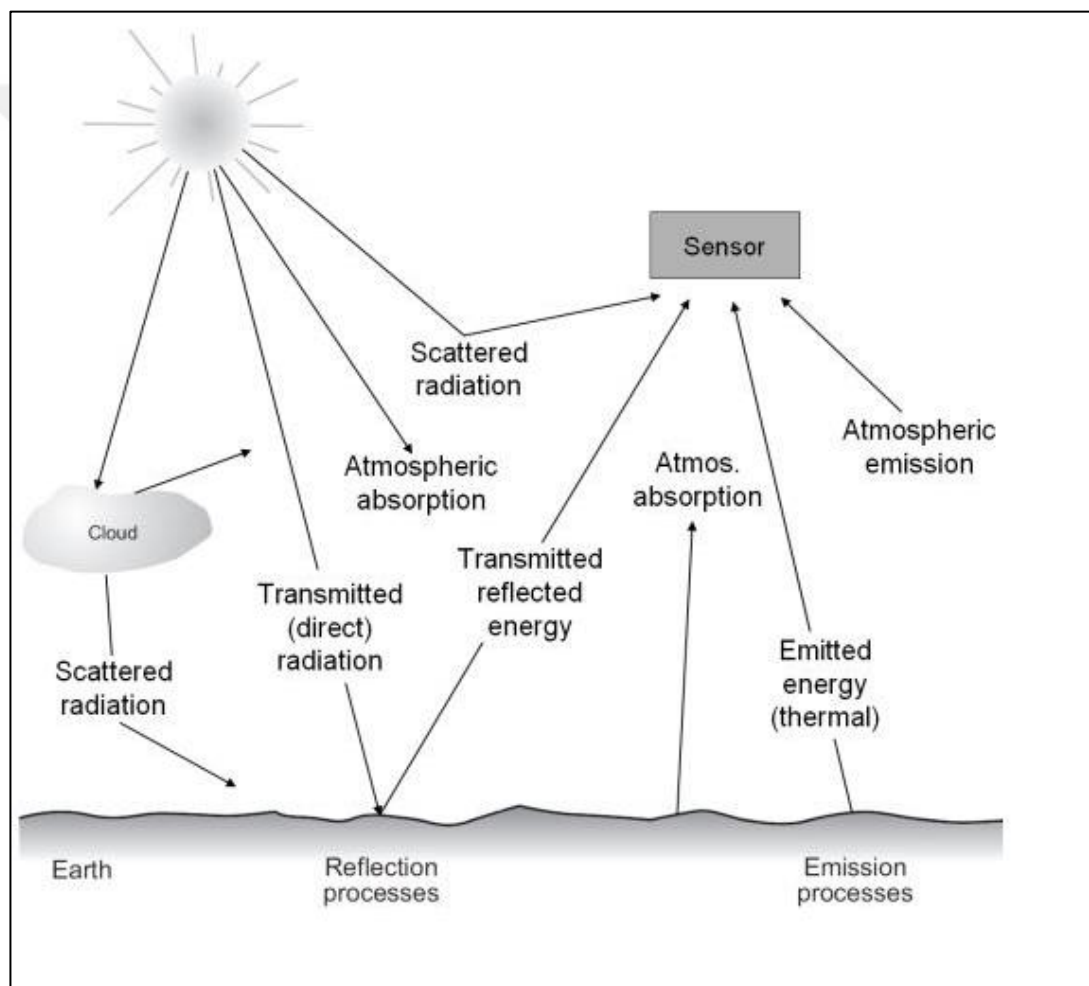


Figure 2.2 Illustration of different paths of radiance received by the satellite sensor (Hadjimitsis & Clayton, 2009)

Solar backscatter based on aerosol-induced variations in reflectance is used for aerosol remote sensing (Martin, 2008). When particles exist in the atmosphere, they

interact with electromagnetic radiation and redirect the radiation from its original paths, scattering (NRC, 2015). The scattering amount depends on the abundance of particles or gases in the atmosphere, the radiation wavelength, and the radiation travel distance through the atmosphere. Three types of scattering exist, including Rayleigh, Mie, and Nonselective. Tiny particles (diameters lower than 1 μm) such as air molecules, dust particles compared to the wavelength of the radiation cause Rayleigh scattering. Shorter wavelengths of energy are scattered by Rayleigh scattering much more than longer wavelengths. The particles about the same size with the wavelength of radiation cause Mie scattering in the lower portions of the atmosphere. Dust, pollen, smoke, haze, and water vapor cause Mie scattering which tends to influence longer wavelengths than Rayleigh scattering. The particles that are much larger than the wavelength (e.g., water droplets, fog, large dust particles, etc.) of the radiation cause nonselective scattering. All wavelengths are scattered equally in nonselective scattering. This type of scattering causes fog and clouds to appear white since blue, green, and red lights are scattered in approximately equal quantities.

Sensors typically measure aerosol amount using the aerosol optical depth (AOD) (Engel-Cox et al., 2004). AOD is the dimensionless parameter to determine the aerosol amount, including sea salt, desert dust, and fine particles from biomass burning, distributed within an air column from the instrument to the top of the atmosphere (Hadjimitsis & Clayton, 2009). It is typically ranging from 0 to 5 (Hadjimitsis & Clayton, 2009). Small values (<0.1) represent a crystal-clear sky with good visibility, larger AOD than 1 represent severe haze conditions (NOAA, 2021). The maximum value of 5 indicates a high density of aerosol amount exists in the atmosphere. People would have difficulty seeing the Sun, even at mid-day.

To dark scenes and small aerosol optical depth, the total reflectance (R_{tf}) measured at a specific wavelength can be used as an approximation of the atmospheric reflectance due to molecular scattering (R_{m}), the atmospheric reflectance due to aerosol scattering (R_{a}), and the surface reflectance (R_{s}):

$$R_{\text{tf}}(\theta) \approx R_{\text{m}}(\theta) + R_{\text{a}}(\theta) + R_{\text{s}}(\theta) \quad (2.1)$$

$R_a(\theta)$ is a function of the aerosol optical depth (τ_a) representing the measure of light extinction by aerosols in the atmospheric column, of the aerosol single scattering albedo representing the scattered fraction of incident radiation, and of the aerosol phase function $P(\theta)$ indicating the angular distribution of scattered radiation (Martin, 2008):

$$R_a(\theta) \propto \tau_a w P(\theta) \quad (2.2)$$

In addition to τ_a , $R(\theta)$ at different wavelengths and angles over land also contain information on aerosol optical properties, including the aerosol shape, composition, and size distribution (Martin, 2008). Although it is primarily for absorbing species at short visible and ultraviolet wavelengths, external information on the relative vertical aerosol profile is needed due to atmospheric scattering (Herman et al., 1997). Absorption due to dust and organic carbon can also decrease or reverse aerosol scattering effects mainly at short wavelengths (Herman et al., 1997). Non-absorbing aerosols are generally not visible over bright surfaces such as ice, snow, and clouds. In contrast, absorbing aerosols are observed over such surfaces by exploiting the wavelength dependence of aerosol absorption, mainly intense ultraviolet and blue wavelengths (Herman et al., 1997).

The major challenge in aerosol remote sensing is to exclude cloudy scenes affecting R_{tf} and determine R_s . To minimize cloud contamination, high spatial resolution is essential (Martin, 2008). Additionally, different atmospheric constituents that can scatter light should be considered to calculate AOD. The aerosol amount is obtained by subtracting the optical depth due to Rayleigh scattering, water vapor, and other wavelength-dependent trace gases, as shown in Figure 2.3.

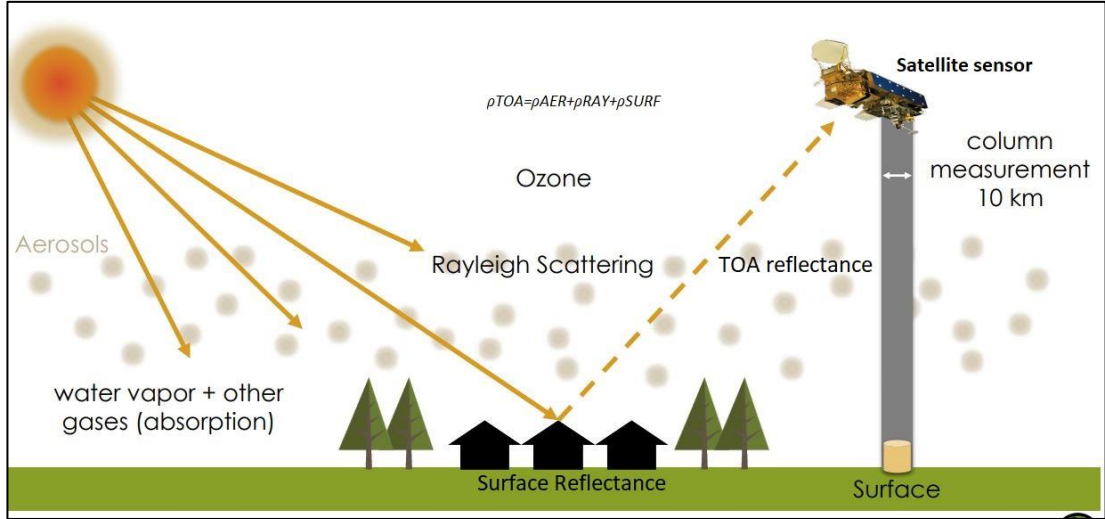


Figure 2.3 Illustration of effecting parameters on AOD (NASA, 2018b)

Spectral AOD, typically from 440 nm to 870 nm, can be used to estimate the size distribution of aerosols (Eck, 1999). The negative slope (or first derivative) of logarithmic scale AOD is described as the Angstrom parameter (α). It is calculated using two or more wavelengths with a least-square fit (Eck, 1999). Angstrom parameter more significant than 2.0 indicates fine mode particles (e.g., sulfates and smoke particles), while values of α near zero indicate the presence of coarse particles such as desert dust (Eck, 1999).

$$\alpha = \frac{d \ln \tau_a}{d \ln \lambda} \quad (2.3)$$

where α is the Angstrom parameter, τ_a is the aerosol optical depth, and λ is the wavelength.

Aerosol columnar properties, composition, morphology, and vertical distribution have been extensively identified by satellites (Kloog et al., 2015). Fast progress was observed in the research of AOD in the last 20 years (Wei et al., 2020b). AOD is a critical parameter to understand the atmospheric processes and regional air quality since it can quantify aerosol amount in the atmosphere (Ghotbi et al., 2016). Observations from satellite sensors for retrieving AOD have become the standard technique for monitoring aerosol loadings over large areas (Wei et al., 2020b). On the other hand, the main problem still exists to obtain high-resolution AOD due to the

constraints in algorithmic synthesis, the low-signal to-noise ratio, selection of aerosol model, downscaling issues, and cloud screening schemes (Damascena et al., 2021). There are two main methods mainly used to measure AOD: ground sun photometers and satellite sensors. Ground-based stations with sun photometers are used to measure accurate AOD but point measurements can only represent a limited area with high temporal resolution. Remote sensing has advantages to monitor AOD over global scale (Gupta et al., 2018; Bright & Gueymardb, 2019; Wei et al., 2020b), even over on the regions covering one or more countries (Nichol & Bilal, 2016; Bilal et al., 2018; Almazroui, 2019; Che et al., 2019; Huang et al., 2019; Aldabash et al., 2020) or on an urban scale (Butt et al., 2017; Tian et al., 2018; Bilal et al., 2019; Zhang et al., 2019). However, AOD retrievals from satellite measurements are challenging over land to determine the aerosol pollution due to the underlying heterogeneous surface albedo. Ground-based AOD is not affected by surface albedo. Thus, it is used for evaluation, testing, and calibration of satellite retrieval algorithms to increase the prediction accuracy of AOD (Wei et al., 2019b). Therefore, ground based AOD measurements are complementary to satellite AOD. Improvements and advances in retrieval algorithms of satellite AOD with the help of ground based AOD measurements and designing with new satellite sensors have enhanced the capability of AOD for better analyzing the PMs concentrations for environmental health assessment, improved air quality management, and climate change studies (Nichol & Bilal, 2016).

2.2 Satellite sensors and platforms

AOD retrievals from satellites have been presented since the 1980s with the free accessibility of diverse satellite sensors (Wei et al., 2020b). Earth-observing remote sensing instruments are named according to the satellite and the instrument (sensor). Satellites are also characterized by paths followed by a satellite referred to as orbit (NASA, 2009). Distance between the earth's surface and the satellite determines how frequently the satellite moves around the earth can be described as orbit height (NASA, 2009). The orbit of the satellite and the viewing and scanning properties of the instrument affects the measurement's spatial and temporal resolution (EC, 2007). Most satellite instruments are looking down the earth (nadir viewing). Randomly selected

starting points in a satellite's orbit can be described as nadir points. Each orbit sensors cover a track on the earth surface, and their width depends on the viewing properties of the instrument (swath width). Swath width ranges from less than 100 to almost 3000 km, and it affects the temporal resolution of the satellite (larger the swath – higher the temporal resolution) (NASA, 2016).

Recently, two types of satellites, including geostationary and polar-orbiting satellites, measure AOD (Li et al., 2009). Both have different limitations and advantages according to each other. Polar-orbiting satellites retrieve AOD with a high spatial resolution (nearly global coverage). However, due to their design limitations to produce high temporal and spatial resolution simultaneously, one or two AOD can be retrieved over a region (Li et al., 2009).

Various satellite instruments have been placed onboard polar-orbiting platforms to produce AOD products, operated by NASA and some international partners (Bibi et al., 2015). PARASOL (Polarization and Anisotropy of Reflectance's for Atmospheric Sciences coupled with Observations from a Lidar), CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations), MODIS (Moderate Resolution Imaging Spectroradiometer), MISR (Multi-angle imaging spectroradiometer), and OMI (Ozone Monitoring Instrument) can be used to obtain AOD products. Recently, VIIRS (Visible Infrared Imaging Radiometer Suite) aboard Suomi–NPP has been started to adapt and improve MODIS. The infrared sounders of Advanced Infrared Radiation Sounder (AIRS) aboard Aqua can also generate AOD (NASA, 2018a).

High-frequency AOD is provided by Geostationary satellites over a constant location, flying in earth-synchronous orbits that are useful for continuous monitoring during the daytime. The effect of the cloud can be decreased due to its high temporal resolution. The GOES (Geostationary Operational Environmental Satellites) is a series of geostationary satellites operated by the United States (Zhang et al., 2011). Different countries have been trying to retrieve AOD on different locations of the world with several next-generation geostationary satellites, including the Spinning Enhanced Visible and Infrared Imager (SEVIRI) aboard the Meteosat Second Generation (MSG)

(15 min, 3 km) and Himawari-8/9 (10 min, 2 km) (Fernandes et al., 2015). They can provide hourly or less than hourly information on major aerosol events, including forest fires, dust transport, and smoke. However, obtaining images with high spatial resolution is restricted due to their high orbit altitude and limited storage. Higher uncertainties in AOD retrievals than those used in polar orbit satellites are the other disadvantages (Paciorek & Liu, 2009). The illustration of the Geostationary (a) and Polar Orbit (b) satellites is given in Figure 2.4. Summary information about widely used satellites and sensors is given in Table 2.1.

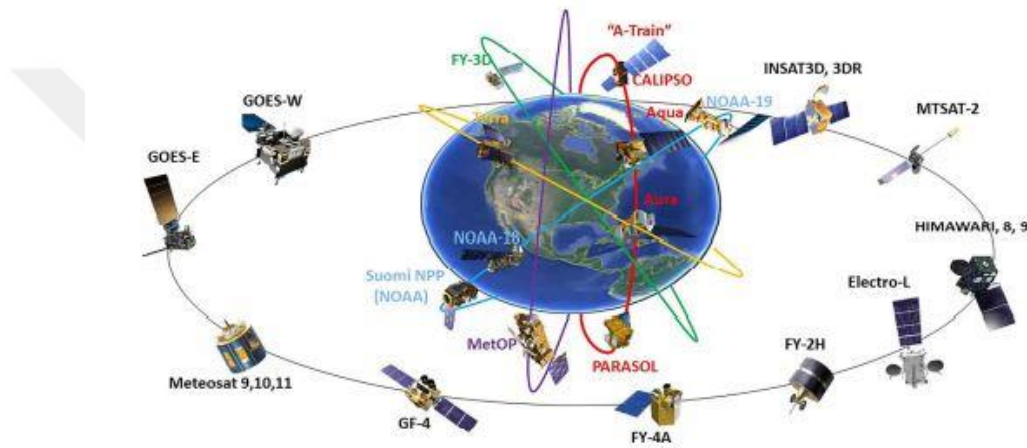


Figure 2.4 The geostationary and polar orbit satellites to retrieve AOD (Wei et al., 2020c)

Table 2.1 Summary information about widely used satellites and sensors

Satellite	Sensor	Type of Orbit	Measurement bands	Start year
Terra	MODIS	Sun-synchronous	VIS, IR	1999
	MISR		VIS	1999
Aqua	MODIS		VIS, IR	2002
Terra & Aqua	MAIAC		VIS	2000
Aura	OMI		VIS, UV	2003
CALIPSO	CALIOP		LiDAR (532 nm, 1064 nm)	2006
Soumi NPP	VIIRS		VIS, IR	2011
Himawari-8	AHI	Geostationary	-	2014
GOES-16	ABI			2016
COMS	GOCI			2010

A satellite tracking on the earth is subdivided into ground pixels comprised of a two-dimensional array of individual images and arranged in columns and rows. An intensity value and a location address of each ground pixel are added in the two-dimensional image (NASA, 2016). The size of ground pixels from atmospheric measurements varies from 0.5 to 1000 km (NASA, 2016).

Also, two primary types of remote sensing instruments exist as active and passive sensors. To brighten the objects, active sensors can provide their source of energy. Active sensors also emit radiation in the direction of the target to be investigated, and the sensor measures the backscattered or reflected radiation from the target. Reflected or emitted natural energy (radiation) by the object or scene is observed by passive sensors. Passive sensors most commonly measure the reflected sunlight radiation (NASA, 2021a). Passive sensors are MODIS, MISR, OMI, and VIIRS, whereas active sensors are LIDAR (NASA, 2016). The illustration of passive and active sensors is given in Figure 2.5.

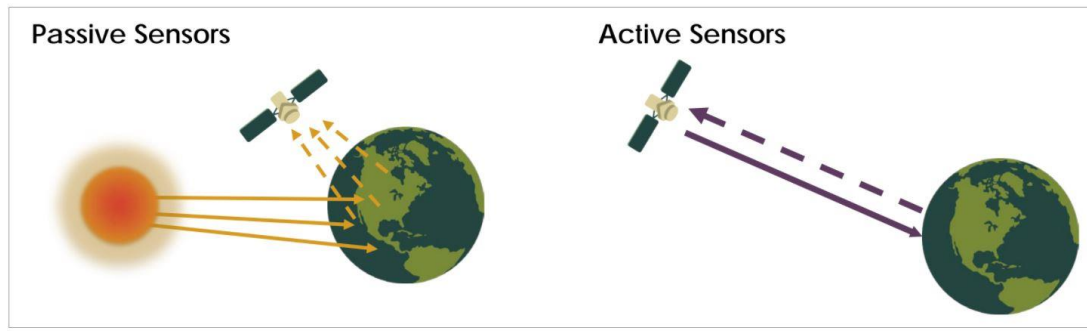


Figure 2.5 Illustration of active and passive sensors (NASA, 2021a)

Passive sensors have been widely used to retrieve AOD in the literature (Che et al., 2019; Huang et al., 2019; Wei et al., 2019b; Wei et al., 2020b). The complex surface condition and the uncertainty caused by the cloud mask (Kacenelenbogen et al., 2011), the calibration methods (Ignatov & Stowe, 2002), and retrieval algorithms (Kokhanovsky et al., 2007) affect the accuracy of AOD products from different sensors and cause differences among them. The calibration system is better in MODIS to retrieve AOD with more accuracy and low uncertainty (Li et al., 2009).

2.3 MODIS AOD and Retrieval Algorithms

Among various passive satellite sensors, MODIS is the earliest operating instrument. It collects information on Earth's surface and atmosphere operated on the Terra (~10:30 LST) and Aqua (13:30 LST). It has been successfully used after launched in 1999 and 2002 on the Terra and Aqua satellites, respectively. Terra and Aqua are both polar-orbiting satellites with wide swaths and scan the Earth's entire surface with high temporal resolution (every 1–2 days). Therefore, they have been most widely used for studying atmospheric aerosols on local, regional, and global scales (Wang et al., 2013; Lee et al., 2014; Van Donkelaar et al., 2015; Zhang & Li, 2015; Lee et al., 2016a; You et al., 2016; Sever et al., 2017; Farahat, 2019; Almazroui, 2019; Rupakheti et al., 2019; Wei et al., 2019a) to improve understanding of the Earth atmosphere over land and ocean. Another reason for the wide use of MODIS sensor is that it has high-quality AOD retrievals having high correlation with AERONET measurements (Bibi et al., 2015).

MODIS has a wide spectral range with wavelength bands ranging from 0.41 to 15 μm (Remer et al., 2005). MODIS with several AOD retrieval algorithms has also been extensively used for aerosol studies. Two types of independent algorithms, namely Dark Target (DT) and Deep Blue (DB), generate data over bright (e.g., desert, and metropolitan region) and the dark (e.g., vegetated region) surfaces, respectively. The primary operational DT algorithm uses MODIS data measured over dark surfaces to retrieve AOD (Kaufman et al., 1997; Levy et al., 2013). Kaufman et al. (1997) firstly proposed the DT algorithm for densely vegetated areas based on the presence of a dark surface in two noticeable channels (0.47 and 0.66 μm) and the proximate transparency of the air at a generally long-wavelength 2.12 μm to get a proper assessment of the atmosphere scattering as shown in Figure 2.6.

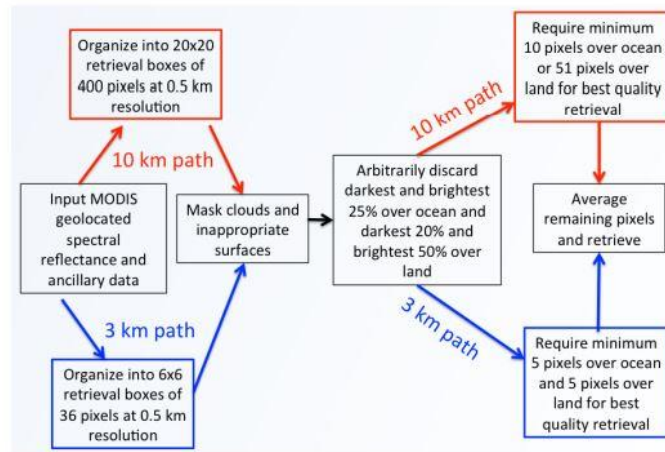


Figure 2.6 AOD retrieval algorithm of 10 km (red) and 3 km (blue) DT products (NASA, 2018a)

The surface reflectance ratios were modified by Levy et al. (2007, 2013), considering the effects of underlying surface type modified to increase the accuracy of the DT retrievals. DT algorithm has two types of retrievals over land and ocean. The algorithm over ocean (Levy et al., 2013) is an algorithmically independent DT approach, but it is similar over land. The water, dark surfaces, is considered, and the aerosol retrieval algorithm neglects the surface reflectance. Since the ocean surface is darker than most land surfaces, retrievals of the algorithm over ocean show relatively lower uncertainty due to the uniform surface compared to land (Levy et al., 2013).

The surface reflectance of bright desert surfaces is stable and dark in the deep blue channels (Hsu et al., 2004, 2013). Considering the limitations of the invisible channels over bright surfaces Deep Blue (DB) algorithm was developed with deep blue channels. MODIS DB algorithm tends to the issue on a superficial level brightness along these lines, however using the attributes of the darker surface in two blue channels (0.412 and 0.470 μm) and absorption by dust in a red channel (e.g., 0.670 μm) as given in Figure 2.7. However, improvement is required for several components of the old retrieval algorithm, including the usage of a static surface database to determine the surface reflectance. This is mainly significant over non-vegetated and mixed vegetated surfaces, which have substantial seasonal variations in land cover. A hybrid approach that uses the combination of primarily calculated surface reflectance data for defining the surface reflectance for aerosol retrievals was developed to solve

this problem. With the aerosol data generated by the enhanced Deep Blue algorithm, AOD can be obtained in the arid and semiarid areas and the whole surfaces (Hsu et al., 2013).

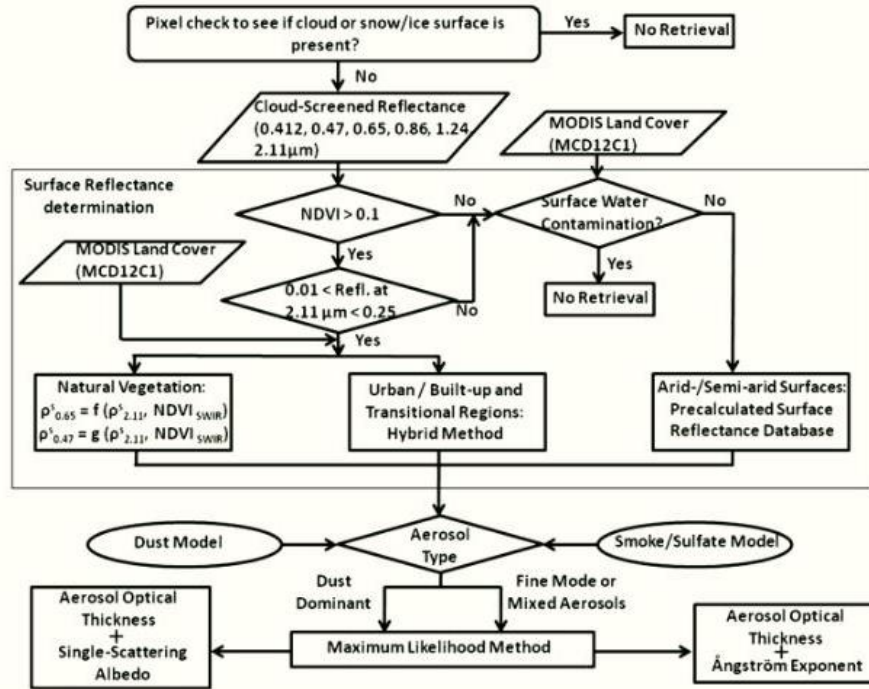


Figure 2.7 Enhanced AOD retrieval algorithm of 10 km DB product (Hsu et al., 2013)

MODIS merged AOD product including DT and DB AOD (DTB) based on the Normalized Difference Vegetation Index (NDVI) was generated by a combination of the DT and DB algorithms (Levy et al., 2013). NDVI, derived from satellites, measures the surface reflectance using the distinct wavelengths of visible near infrared sunlight reflected by the plants and gives a quantitative estimation of vegetation growth. The range of NDVI is -1 to $+1$. The higher value of NDVI represents the dense vegetation, while lower NDVI values represent the sparse vegetation over land (Jin et al., 2020). Spatial coverage over land is increased with DTB product, specifically in central transition regions with low vegetated surfaces and adequately dark surfaces (Levy et al., 2013). In the generation of the DTB dataset, the following steps are used:

- if $NDVI > 0.3$ DT retrievals are selected,
- if $NDVI < 0.2$ DB retrievals are selected,

- if $0.2 \leq \text{NDVI} \leq 0.3$, the average of DT and DB AOD or the available retrieval providing the recommended quality assurance ($\text{QAC} = 3$ for DT and $\text{QAC} \geq 2$ for DB) are selected.

MODIS provides long-term and global aerosol data with two operational satellites. The Level-2 daily products are provided by 10-km spatial resolution (MxD04_L2), where O is for Terra and Y is for Aqua) and 3-km (MxD/04_3K) resolutions, and Level-3 daily (MxD08_D3), eight-day (MxD08_E3), and monthly (MxD08_M3) that are spatiotemporally aggregated from the MxD04_L2 products at a $1^\circ \times 1^\circ$ resolution (Platnick et al., 2015). Three datasets, including DT, DB, DTB AOD, were provided by aerosol products at 10-km spatial resolution; however, only DT was provided by aerosol products at 3-km resolution.

Although the AOD with 3 km resolution product almost similar to the AOD with 10-km resolution product retrieved by the DT algorithm, fine spatial resolution is expected to pass other products representing the specific characteristics such as better variation of aerosol plumes and gradients and closer retrievals to clouds, coastlines, and small islands that the coarse-resolution aerosol product misses (Munchak et al. 2013; Remer et al. 2013). However, due to its high resolution, the 3-km DT AOD may be overestimated over urban areas with noisy calculation error where the 10-km DT frequently avoids (Munchak et al., 2013).

The first and well-documented products of MODIS are the Collection 4 (C4) products (Levy et al., 2013). The second AOD products are C5 as C5.1 and C5.2 and C6 as C6.0 and C6.1. They performed better than C4. Reprocessing of Collection 6.1 (C6.1) started on the 28th of September 2017 and completed on the 7th of March in 2018 (NASA, 2021b). The new C6.1 aerosol product was released after applying the significant improvements based on the new updated Level 1B calibrated radiance products in all aerosol retrieval algorithms. NASA Ocean Biology Processing Group (OBPG) developed additional calibration corrections by applying to the MODIS top-of-atmosphere (TOA) radiance product. Changes differ over land and ocean to generate C6.1 DT aerosol product. The mask for sediment was changed to produce

more accurate retrievals over ocean. For low-AOD conditions fraction of cloud, the retrieval pixel numbers, and standard deviation were included in the quality control process. Quality assurance confidence (QAC) information is included in all MODIS products designed to assist users and to make data more understandable and usable. Each product's QAC information is provided as a separate dataset. QAC of retrievals deteriorated to the lowest (QAC = 0) if more than fifty percent of coastal pixels or twenty percentages of water pixels exist in the retrieval box at 10×10 km over land (Gupta et al., 2016). The estimation model for surface reflectance in urban areas precisely where fractions of urban areas are more than 20% (Mattoo, 2017). The DB algorithm has been more affected by different improvements, including reduced mistakes on heterogeneous topographical surfaces, detection of heavy smoke and the enhanced bug fixes, and revised seasonal and regional aerosol optical models compared to DT. The same merging scheme was applied to generate the C6.1 DTB.

The aerosol model used in the inversion process or surface reflectance mainly affects the variance in collocated AOD retrievals (Levy et al., 2013). The variances during high aerosol loading are common due to an error in the aerosol scheme (Levy et al., 2013). The selection of which fine-dominated aerosol model is fixed based on location and season. The coarse-dominated model (dust) is considered globally fixed (Levy et al., 2007). Single Scattering Albedo (SSA) is defined as the scattering ratio to the extinction number of aerosols. Extinction is the sum of scattering and absorption. SSA varies with aerosol types from absorbing to non-absorbing (from 0.85 to 0.99). Levy et al. (2013) used the same methodology as described by Levy et al. (2007) and performed a cluster analysis of the entire AERONET climatology through 2010. The season and location are critical parameters for the selection of the fine-dominated aerosol model. He et al. (2015) showed the significant impacts of the aerosol particle shapes on fine-sized aerosol optical properties. Therefore, the particle shapes can also cause uncertainty in the MODIS AODs.

An expected error (EE) envelope representing approximately one standard deviation of the matchups (at least 67% of data points) is defined for each satellite algorithm with the validation studies. The overall EE for each AOD is determined with

the combination of absolute error dominating at low AOD and relative error dominating at high AOD (Levy et al., 2013). The EE for 10-km DT AOD at 0.55 μm was defined as $\pm (0.05 + 0.15 \times \text{AOD}_{\text{AERONET}})$ over land (Levy et al., 2013) where $\text{QAC} = 3$ and $+(0.04 + 0.10 \times \text{AOD}_{\text{AERONET}}, -(0.02 + 0.10 \times \text{AOD}_{\text{AERONET}}))$ over ocean to reflect consistent asymmetry where $\text{QAC} \geq 1$, while the EE for 3-km DT AOD at 0.55 μm was defined as $\pm (0.05 + 0.20 \times \text{AOD}_{\text{AERONET}})$ over land (Gupta et al., 2018). Dark or bright pixels removed from the 10 km product might still exist at the 3 km product since there are more pixels for selecting from the deselection process at 10 km. These pixels possibly make the 3km-DT product noisier than the 10-km product (Gupta et al., 2018). Therefore, the EE over land for the 10 km-DT is relatively stricter than the 3-km product (Nichol & Bilal, 2016). The EE is approximately $\pm(0.03+0.20 \times \text{AOD}_{\text{AERONET}})$ for DB retrievals at a 10-km resolution over land. The EE for DTB retrievals is the same as DT (Wei et al., 2019b).

MAIAC (Multi-Angle Implementation of Atmospheric Correction) is a new algorithm applied to MODIS images (Lyapustin et al., 2011). This algorithm simultaneously retrieves aerosol over land and surface bidirectional reflectance. The MAIAC algorithm with improved ability to analyze aerosol plumes and making retrievals over deserts can provide the aerosol information at a 1-km spatial resolution over both dark and bright surfaces using spatial analysis and time-series of MODIS measurements. The accuracy evaluation of MAIAC-derived AOD has been carried out, indicating that its quality is dependent on the surface properties and can achieve a comparable performance over DT AOD (Lyapustin et al., 2011; Martins et al., 2017). Currently, MAIAC AOD data in C6.1 is not available.

2.4 Ground-Based AOD observations

Sun and sky scanning radiometers with automatic tracking (CIMEL Electronique CE-318 Sun-sky radiometer) were used as the instrument globally to measure AOD at ground level (Holben et al., 1998). CIMEL sun photometer located southern part of Turkey is shown in Figure 2.8. This instrument measures the direct radiation every 15 minutes with a 1.2° whole field of view at 340, 380, 440, 500, 675, 870, 940, and 1020 nm wavelengths (Eck et al., 2005). To retrieve columnar water vapor, only 940 nm is

used, AOD is retrieved by all other wavelengths (Holben et al., 2001; Eck et al., 2010). Sun photometers provide spatially limited, temporally continuous, and freely accessible AOD database with low uncertainties (0.01-0.02) and high temporal resolution (15 min) (Sayer et al., 2013). Different products, including aerosol microphysical and radiative properties to analyze the aerosol type, source, mixtures, and transport in a region, are also provided.



Figure 2.8 CIMEL Sun-photometer (Kubilay & Koçak, 2009)

The AERONET (AErosol RObotic NETwork), which is a ground-based remote sensing aerosol network, was established by NASA and LOA-PHOTONS (CNRS) (Holben et al., 1998; Dubovik & King, 2000; O'Neill et al., 2005). Then different networks such as AEROCAN, SKYNET, Baseline Surface Radiation Network (BSRN) Global Atmosphere have established in addition to AERONET (Levin et al., 2008). National networks also exist, such as the Chinese Sun Haze Meter Network (CASHNET), Australia Bureau of Meteorology (BOM), German Weather Service (DWD), Japan Meteorological Agency, U.S. (Levin et al., 2008). The instruments in AERONET have been installed in the 1990s, and currently, this network has more than 1200 different stations covering the whole world, as shown in Figure 2.9 (NASA, 2020).

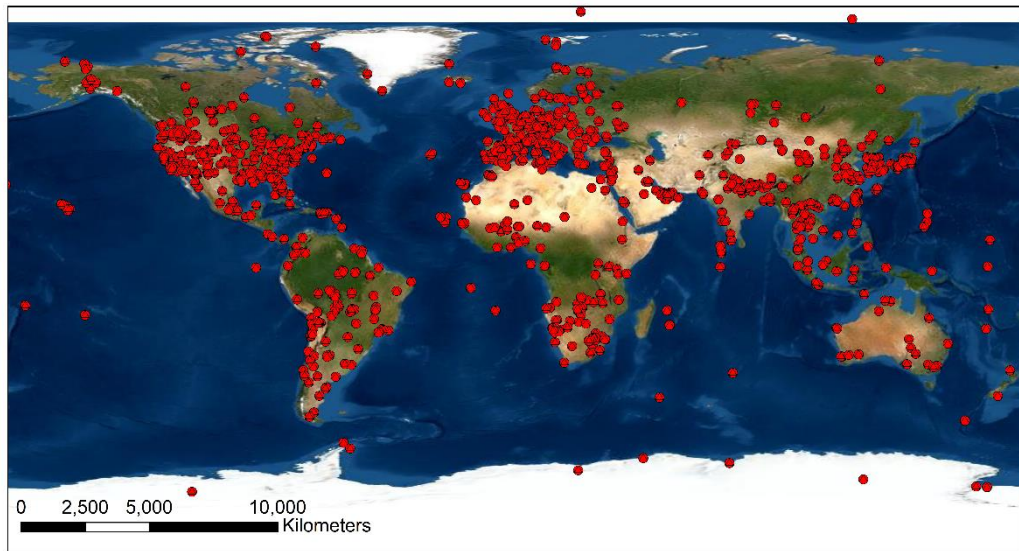


Figure 2.9 AERONET sites

The data processing algorithms have progressed from Version 1.0 to Version 2.0 and, recently, Version 3.0. The Version 3 databases are currently available on the AERONET website. Version 2 data can be obtained via the website until 2018, and data after 2018 can be obtained by special request. Version 3 AOD products are provided with three data quality levels, including Level 1.0 (unscreened), Level 1.5 (cloud-screened and quality controlled), and Level 2.0 (quality-assured). The last version (Version 3) database provides more accurate results than Version 2 when the sites located in different regions with long-term measurements and aerosol characteristics are analyzed. They indicated mainly reasonable variations in the number of available days, AOD, and AE (Giles et al., 2019). Detailed assessment of data quality is discussed by Giles et al. (2019).

The AERONET provides a high-accuracy ground-level aerosol database with standard data processing. They are also used for satellite calibration despite their asymmetrical distribution with high-cost instruments mainly in less populated areas. Therefore, AOD from AERONET is commonly used for the validation of satellite AOD. These studies performed at the hundreds of stations on a global scale (Gupta et al., 2018; Bright & Gueymardb, 2019; Wei et al., 2019b; Wei et al., 2020b), or focusing on the regions covering one or more than one country (Nichol & Bilal, 2016; Bilal et al., 2018; Almazroui, 2019; Che et al., 2019; Huang et al., 2019; Aldabash et al., 2020)

or in multiple stations on an urban scale (Butt et al., 2017; Tian et al., 2018; Bilal et al., 2018b; Zhang et al., 2019).

There are five AERONET stations in Turkey. The IMS-METU-ERDEMLI station located in the southern region of Turkey is the only station that continuously monitors AOD. The other four Turkish stations (TUBITAK_UZAY_Ankara, Tuz_Golu, Tuz_Golu_2, and Tuz_Golu_3) were used for various short-term scientific studies in the past. Among these, the TUBITAK_UZAY_Ankara station is the only station that generates data between 2009 and 2012. A mobile station deployed at three different locations around Tuz Lake produced less amount of discontinuous AOD data between 2009 and 2017. Locations of national AERONET stations in Turkey are shown in Figure 2.10.



Figure 2.10 The location of the AERONET stations in Turkey

2.5 Ground-Level PM₁₀ Measurements and Air Quality Directives

PM consists of a mixture of solid and/or liquid materials of various sizes ranging from a few nm in diameter to around 100 μm (about the thickness of human hair). The

presence of aerosol particles in the lower atmosphere is significant in air quality and public health considerations (Horne & Dabdub, 2017). PM is typically defined by size in air-quality applications. The smaller particles have more adverse health effects. Commonly measured values for PM are particles PM_{10} and $PM_{2.5}$ (Kloog et al., 2012). Reliable concentration data is provided by ambient monitoring stations located in fixed locations. However, monitoring sites are restricted due to several limitations, such as the cost and inadequate infrastructure (power, cooling, and security).

Many countries have established their air quality monitoring networks to obtain PM concentrations at larger spatial and temporal scales (Tian & Chen, 2010). Protection of air pollution in Turkey is governed by the Ministry of Environment and Urbanization (MEU). Many organizations such as WHO (World Health Organization), EPA (Environmental Protection Agency), and the European Union (EU) determine air quality standards worldwide. Previous EU legislation governing air quality (EC 1996, 1999, 2004) are flexible than recent legislation (EC, 2008). The current directive sets a 24 hour-limit value for PM_{10} of $50 \mu g/m^3$. The annual average should not exceed $40 \mu g/m^3$ with a maximum number of exceedances of the daily limit of 35 times per year (Council Directive 99/30/EC). According to Council Directive (96/62/EC), the occurrence of exceeding the limit values and possible reasons should be reported by EU member states. Turkey adopted the Daughter Directive into the Air Quality regulation in June 2008 as an EU candidate country (Kendall et al., 2010). Meeting the criteria of PM_{10} limit values regulated by EC legislation or any other international guidelines requires an improved understanding of pollution concentrations, particularly in developing countries such as Turkey. Before adopting the EC Air Quality Framework Directives in 2008, the capacity and reporting of air quality measurement in Turkey were very restricted (Kendall et al., 2011).

After monitoring air quality was committed to the Ministry of Environment and Urbanization (MEU) in Turkey, air quality monitoring stations were established in 81 provinces between 2005-2007 to accurately measure air pollution and create clean air plans (MEU, 2019). In 2008, the number of monitored cities increased to more than 81, and PM_{10} firstly became one of the significant measured air pollutants.

Specifically, stations located throughout all provinces in Turkey have been operated by the MEU. However, air quality measurements are also provided by some municipalities. Air quality monitoring stations belonging to Izmir Metropolitan Municipality, Istanbul Metropolitan Municipality, Refik Saydam Hygiene Center, and Kocaeli Dilovası Organized Industrial Zone were integrated into the system (MEU, 2019). Within the scope of the protocol signed between the MEU and the Ministry of Health on March 16, 2009, the air quality measurement stations and mobile measurement stations in Ankara were transferred to the MEU.

Clean Air Centers (CACs) have also been established in 8 regions (Istanbul, Samsun, Erzurum, Izmir, Adana, Konya, Ankara, and Diyarbakır), providing an effective, efficient, and quality assured air quality monitoring network instead of managing the stations from a single center. All provinces have air quality monitoring stations, and the number of stations has been increasing rapidly. The air quality monitoring network has 355 air quality stations in Turkey (MEU, 2019). PM₁₀ concentrations are continuously determined by automatic beta radiation attenuation technique according to EC Directive 2008/30/EC (EC, 2008). Beta attenuation or Tapered Element Oscillating Microbalance (TEOM) mass monitors are used to measure PM₁₀ mass concentration certified as equivalent to the reference manual gravimetric method. Strict quality assurance or quality control (QA/QC) procedures were applied, as defined in the legislation for the air quality monitoring European network (2008/50/EC).

Hourly average measurements obtained from the stations were monitored through a special network. They are transferring to the Data Management Center located in Gölbaşı/Ankara of the Laboratory, Measurement and Monitoring Department under the General Directorate of Environmental Impact Assessment (EIA) Permitting and Inspection of the Ministry of Environment and Urbanization. In this center, a validation study is carried out on the data, considering the calibration and alarm information of the devices. Accordingly, monthly, and annual reports are prepared with the evaluated data. The raw data obtained from the monitoring network are published on https://sim.csb.gov.tr/STN/STN_Report/StationDataDownload. After

the data validation studies are carried out at the end of each month, the verified data are transferred to the website and freely available online. However, verified data is not good enough to be analyzed for publication.

2.6 Relationship Between Ground-Level PM and Satellite Based AOD

Aerosols are composed of solid and liquid suspended particles originating from fog, mist, industrial emissions, dust, anthropogenic emissions, and organic particles (Ranjan et al., 2020). With the significant improvements of remote sensing technologies, monitoring atmospheric aerosols with satellites has become more popular (Mhawish et al., 2017). Satellite-based AOD has been used as a proxy to assess PM concentrations with varying performance over different regions (Gupta & Christopher, 2009; Xiao, 2011; Park et al., 2019). Therefore, it has been more widely used with various studies, academic disciplines, decision-making, and health exposure assessment (You et al., 2016; Mhawish et al., 2017).

Only AOD is not a good predictor of PM, according to Gupta & Christopher (2009), since AOD represents columnar aerosol loading which indicates a quantitative estimation of the aerosol amount in the atmosphere. In contrast, PM represents the suspended solid or liquid particle concentration at the surface. A typical mismatch between spatially averaged temporally restricted satellite AOD and temporally averaged point measurements of PM is also a type mismatch. Temporal averaging methods and geographical location affect the relationship between AOD and ground-level PM. The correlation between PM and AOD strongly depends on the vertical distribution of aerosols and numerous meteorological factors, including temperature, humidity, and wind components, since meteorological conditions can particularly impact the aerosol dispersion processes (Gupta & Christopher, 2009). It is widely known that the relationship between AOD and PM is also affected by elevation, NDVI, land cover type and can change in both space and time (Meng et al., 2021).

Also, aerosol sizes, composition, and properties are not stable (Hoff & Christopher, 2009). These factors may affect the relationship between PM and AOD to a different

degree under various conditions. The AOD could be used as a parameter to estimate ground-level PM concentrations in some regions. However, it is not a uniform parameter to detect aerosol amount (Liu, 2015).

Numerous studies have reported a relation between AOD and PM₁₀ or PM_{2.5} concentrations over different regions (Wang & Christopher, 2003; Chu et al., 2003; Engel-Cox et al., 2004; 2006; Al-Saadi et al., 2005; Koelemeijer et al., 2006; Gupta & Christopher, 2009; Kloog et al., 2015; Kong et al., 2016; Strinberg et al., 2018; Nabavi et al., 2019). The PM₁₀ and AOD showed a high correlation in a single site over northern Italy (Chu et al., 2003). Wang & Christopher (2003) studied with Terra and Aqua AOD and compared with 1-hr and 24-hr averaged PM_{2.5} concentrations from seven different sites. Linear correlation coefficients between AOD and PM_{2.5} obtained from all seven sites for hourly data were found as 0.70. Kong et al. (2016) found high correlations between the PM_{2.5}, PM₁₀ concentration, and AOD, but the linear regression results showed significant differences for seasons.

AOD (τ) is described as the integral of aerosol extinction through an atmospheric column from the ground to the top of the atmosphere (Tian & Chen, 2010). Theoretically, τ can be calculated by:

$$\tau = \int_0^{\infty} \sigma_{ext}(h) dh \quad (2.4)$$

Where $\sigma_{ext}(h)$ is the aerosol extinction coefficient at the vertical elevation, h is the vertical height. The radiative transfer model can calculate AOD based on recorded radiances using a sensor (Kaufman et al., 1997). According to the satellite and ground-based measurements majority of aerosols exist in the lower troposphere (mainly the planetary boundary layer (PBL) where the particles are more equally distributed and the particle mass distribution below the PBL tends to be more homogeneous due to the active mixing (Tian & Chen, 2010). The aerosol properties usually do not show considerable variation across different altitudes in the PBL, mainly when convection

is evident (Tian & Chen, 2010). Hence, the PBL height (PBLH) can approximately represent the scale height (Liu et al., 2005).

The impact of water vapor on AOD can be determined using a strongly non-linear function of the boundary layer humidity causing the growth of the particles conducted by several researchers (Liu et al., 2005; van Donkelaar et al., 2006; Koelemeijer et al., 2006). The hygroscopic growth of the particle size can occur exponentially when humidity is high. Humidity increases their extinction efficiencies significantly and causes the overestimation of aerosol mass (Tian & Chen, 2010).

Thus, AOD can be corrected (AOD*) to obtain a better correlation with PM mass concentration measured near the ground by the following equation (Koelemeijer et al., 2006):

$$AOD^* = \frac{AOD}{f(Humidity) \times PBLH} \quad (2.5)$$

where *PBLH* denotes the planetary boundary layer height, and *f(Humidity)* is a columnar humidity function. The effects of water vapor and atmospheric column height on the optical extinction efficiency are removed by this correction. Koelemeijer et al. (2006) found improvements in the correlation between PM and AOD when the AOD is corrected by the mixing layer height and relative humidity.

Physical models typically develop based on physical relationships between ground-level PM concentrations and satellite AOD products with humidity and altitude corrections using the following equation (Zhang & Li, 2015). If PM is well mixed in the boundary layer and the sky is cloud-free, the AOD-PM relationship can be described as:

$$PM = \frac{4\rho_{eff}}{3 f(RH) \times Q_{ext,dry}} \times \frac{f(PBL)}{H_{PBL}} \times AOD = \frac{f(PBL)}{H_{PBL}} \times \frac{1}{S} \times AOD \quad (2.6)$$

where eff is particle effective radius, r is particle density (g/m^3), $f(RH)$ is the ratio of RH and coefficients of dry extinction, $Q_{ext, dry}$ is particle extinction efficiency; $f(PBL)$ is the AOD fraction in the boundary layer, PBL is the boundary layer height; and S is particle specific extinction efficiency (m^2/g^{-1}) at relative humidity (RH). The vertical distribution of aerosols is necessary, and the PBLH provides this information. The magnitude of hygroscopic growth with RH depends on particles' size and chemical species (Gupta & Christopher, 2009). However, ground-level PM is not directly affected by RH since it is measured under dry conditions.

A robust linear correlation between AOD and $PM_{2.5}$ under low PBLH conditions was reported by Gupta et al. (2006). A very similar number of particles were observed by the satellites with ground-level PM measurements if the low PBLH exists. Chudnovsky et al. (2013) also shown that a shallow PBLH in winter leads low correlation between AOD and $PM_{2.5}$ in Boston, USA. Smaller PM concentrations are observed in the upper atmosphere since the particles are mainly restricted within the shallow boundary layer (Strinberg et al., 2018).

The effect of RH is argumentative in the literature. Some studies showed no significant effect of RH on the relationship between AOD and $PM_{2.5}$ (Chudnovsky et al., 2013). However, different studies reported that AOD could represent PM mass concentration better under dry conditions (Gupta & Christopher, 2009; Strinberg et al., 2018; Nabavi et al., 2019). Strinberg et al. (2018) showed that the relationship between AOD and PM_{10} is non-linear and strongly affected by the PBLH, RH , wind direction, and wind speed in Berlin. Increased AOD with RH causes overestimation of the dry PM concentrations near the ground. Nabavi et al. (2019) also used RH and PBLH to normalize AOD and correct PM, respectively. The correlation coefficient between PM and AOD increased from 0.15 to more than 0.50.

Accordingly, determining the most significant meteorological conditions affecting the relationship between AOD and PM is essential. Quality of satellite AOD can also determine the underestimation or overestimation of ground-level PM concentrations. However, using AOD as a proxy for PM concentrations can be misleading in physical

models. The highly variable and often non-linear relationship between these two parameters is affected by several factors that have the most significant importance for estimating PM_{10} by AOD and (Engel-Cox et al., 2006). Current studies have used several meteorological parameters (e.g., PBLH, RH, temperature, wind speed and direction, precipitation, cloud cover, etc.) since these factors are mainly related to PM concentrations (Zhang et al., 2018; Park et al., 2019).

Wind speed and direction can affect the particle type and amount by transporting additional emissions over different regions and cities (Zhang et al., 2021) or the PM dispersion in the atmosphere (Xiao, 2011). Furthermore, Benas et al. (2013) significantly improved PM estimation results by integrating temperature into PM estimation since temperature plays a significant role in the boundary-layer dynamics. It controls the ground-level PM concentrations by dilution and trapping of pollutants to manage the photochemical reactions. The vertical distribution of aerosols in the atmospheric layer near the ground was restricted by forming surface temperature inversions (Wu, 2014). Secondary organic aerosols (SOA) are a significant contributor to PM directly related to air temperature (Fine et al., 2008; Nabavi et al., 2019). The information, including particle composition (Huang et al., 2018; Ghahremanloo et al., 2021), size distribution (Yan et al., 2017; Park et al., 2019), and vertical profile, could be used to improve the modeling of the AOD –PM relationship (Zhang et al., 2018). A range of factors can influence AOD and ground-level $PM_{2.5}$ and strengthen or weaken their relationship (Zhang & Li, 2015; Zheng et al., 2017). Therefore, integrating additional variables such as geographical and meteorological variables is required to estimate better $PM_{2.5}$ concentrations (Chu et al., 2016; Guo et al., 2017a; Chen et al., 2018). The complex topography, NDVI, some land-use types including urban, rural or industrial, transportation network such as roadway or railway, forest area, and the air mixture including natural and anthropogenic aerosol sources, are significant challenges addressed (Staggofia et al., 2019). For example, covariation of free tropospheric aerosols (e.g., desert dust) and ground-level $PM_{2.5}$ may result in a strong relationship between the two measurements without any physical connection. The most recent studies included the aerosol types from reanalysis to explain the relationship between AOD and the ratio of $PM_{2.5}/PM_{10}$. On the other hand, satellite

snapshots usually are exposed to cloud contamination due to covering large areas (Kamarul Zaman et al., 2017). Clouds restrict the surface, and ground-level aerosols cannot be detected under cloudy conditions by satellites for intensive cloud cover regions. Therefore, the incorporation of fractional cloud cover improved the prediction of PM (Kamarul Zaman et al., 2017).

Previous studies also showed that land-related data such as NDVI, elevation, traffic density, land cover as the percentage of built-up area or forest had been widely used to analyze the complex relationship between PM and AOD due to their strong relationship with an accumulation of particles and indirect effect on particle emission (Park et al., 2019; Staggofia et al., 2019; Wei et al., 2019a).

The relationship between PM and AOD should be analyzed regionally, considering its region-specific conditions. The studies of the temporal variations of the relationship between PM_{2.5}-AOD have notably focused on the seasonal variation (Wei et al., 2019a; Yang et al., 2019a). To make a comprehensive analysis of the PM-AOD relationship, comparisons for different cities and periods are required over the large study area and long-time period (Yang et al., 2019a). Regional analysis usually gives more detail better than national ones (Liu, 2015).

There are also different challenges to estimating PM based on AOD over global, continental, regional or local scale. Firstly, obtaining satellite-estimated PM with fine spatial and temporal resolutions is challenging. Recent improvements of satellite-estimated PM concentrations have separately achieved with a spatial resolution of 1 km (Staggofia et al., 2017; Huang et al., 2018; Wei et al., 2019a; Wei et al., 2021a) and a temporal resolution of monthly (Huang et al., 2018), daily (Brokamp et al., 2018; He & Huang, 2018; Staggofia et al., 2019), and hourly (Guo et al., 2017a; Zhang et al., 2016; Zhang et al., 2019). However, using a single satellite AOD product, it is hard to derive PM concentrations with fine spatiotemporal resolutions. AOD products provided by polar orbit satellites such as MODIS, MISR, and VIIRS retrieve AOD with fine spatial resolution, but they cannot estimate hourly PM concentrations. The AOD product is produced with several aerosol retrieval algorithms by MODIS sensor

(DT, DB, MAIAC) onboard the two satellite platforms (Terra and Aqua). Recently developed MAIAC algorithm was used in different regions for PM estimation (Staggofia et al., 2017; Huang et al., 2018; Wei et al., 2019a; Wei et al., 2021a) due to its advantages with a relatively high spatial resolution (~ 1 km) for PM estimations including, extended time coverage and improved accuracy over bright surfaces. However, Nabavi et al. (2019) found that the normalized 10-km DTB AOD shows a higher relationship with PM measurements than 1-km MAIAC AOD over Tehran. AOD from geostationary satellites such as Advanced Himawari Imager (AHI), GEOS, and reanalysis AOD (MERRA-2 and CAMS) are also used with very high temporal resolutions as well as their coarse spatial resolution for PM estimation.

Limited or no AOD data over surfaces covered by snow and cloud, high surface reflectance, or high aerosol loading affect the availability of valid AOD. Thus, the temporally and spatially continuous satellite-based PM estimations are prevented (Engel-Cox et al., 2004). Notably, obtaining spatially continuous AOD is difficult, mainly over the regions that have dense cloud cover (Park et al., 2019). Clouds restrict the surface and make detection of aerosols on the ground impossible. AOD under cloudy conditions could not be validated since the AERONET stations, which provide point based continuous AOD measurements, do not also provide AOD information under cloudy conditions. However, the ground-level PM concentrations are measured continuously regardless of sky conditions (Park et al., 2020). Cloud obscuration is not negligible. Thus, the incorporation of fractional cloud cover improved the estimation of PM (Kamarul Zaman et al., 2017). It is also known that satellite-estimated daily PM concentrations with data gaps tend to underestimate the concentrations compared with ground-level measurements (Xiao et al., 2017; Stafoggia et al., 2019). A few studies have tried to solve this challenge by combining satellite-derived and model-simulated AOD to fill gaps (Xiao et al., 2017; Stafoggia et al., 2019; Jiang et al., 2021, Tuna Tuygun et al., 2021). Xiao et al. (2017) combined MAIAC with chemical transport model (CTM) simulations in Yangtze River Delta, China, and Stafoggia et al. (2019) filled missing MAIAC AOD with Copernicus Atmosphere Monitoring Service (CAMS) in Italy.

Due to the complex relation between AOD and PM with errors in all the input parameters, uncertainties in satellite PM estimates are inevitable (Liu, 2015). Additionally, AOD can be used as a measure for PM concentrations in some regions, but it cannot be a suitable parameter in others (Liu, 2015). More complex approaches are currently used to support meteorological and land use information to further examine this relationship. Models for PM estimation are mainly divided into two types: simulation-based models and statistical models. Simulation-based models are based on chemical transport models (van Donkelaar et al., 2010). In contrast, statistical models are based explicitly on statistical regression and advance machine learning models (Lin et al., 2015).

2.7 Estimation methods of PM with satellite based AOD

2.7.1 *Simulation-based models*

The satellite AODs have been compared with ground-level PM concentrations. The ratio of $PM_{2.5}/AOD$ has been addressed in most papers (van Donkelaar et al., 2010; 2016; Yang et al., 2019a); however, fewer studies worked on PM_{10}/AOD (Steinberg et al., 2018; Tuna Tuygun & Elbir, 2020). Numerical models and forecasting systems require accurate emission information of aerosols. Meteorological models are required to provide information about atmospheric conditions for the estimation period as input information for transport, chemistry, and deposition. These simulations can mainly capture the average spatial trends and the dynamic range in PM compared with ground-level measurements. However, possible underestimates and overestimates due to the uncertainties in emission inventories and meteorology can occur. Ground-level PM and satellite AOD datasets are appropriate for improving and evaluating these simulations. MODIS AOD was used by Mathur (2008) during the 2004 Alaska fires to enhance the pollution characterization from outside of the modeling domain. PM concentrations were estimated over each model grid with satellite AOD using a previous relationship between PM and AOD.

The CTMs such as GEOS-Chem (Goddard Earth Observing System-Chem) and GOCART (Georgia Tech/Goddard Global Ozone Chemistry Aerosol Radiation and Transport) are simulation-based models to obtain $PM_{2.5}/PM_{10}$ concentrations from satellite AOD. They can simulate the η (PM/AOD), the scale factor of $PM_{2.5}/PM_{10}$ concentration and AOD, and then applies η to calculate satellite-based PM concentrations without using PM measurements (van Donkelaar et al. 2006; Jin et al., 2020). η is a function of several factors, including aerosol properties, vertical distribution of aerosols, and relative humidity (Jin et al., 2020) and mainly characterize the spatial and temporal PM-AOD relationships (Su et al., 2018).

Liu et al. (2004) first proposed simulation-based PM-AOD modeling with GEOS-Chem and GOCART models to calculate annual average $PM_{2.5}$ concentrations using MISR AOD over the U.S. in 2001. A good relationship was found with the correlation coefficient between the estimated and observed annual mean $PM_{2.5}$ concentrations around 0.8. Then, this technique was improved by van Donkelaar et al. (2006) to obtain global (of $1^\circ \times 1^\circ$ spatial resolution) $PM_{2.5}$ concentrations using MISR and MODIS AOD from 2001 to 2002. The first long-term $PM_{2.5}$ datasets was constructed on a global scale from 2001 to 2006 with a higher spatial resolution ($0.1^\circ \times 0.1^\circ$) by van Donkelaar et al. (2010). The most recent global $PM_{2.5}$ estimates at $0.01^\circ \times 0.01^\circ$ from 2001 to 2010 were presented with a combined geophysical-statistical method using multiple information including satellite-, simulation, and monitor-based datasets by van Donkelaar et al. (2016). Validation results of the annual mean satellite-derived $PM_{2.5}$ estimations with the geographically weighted regression (GWR) showed a high R-square (0.81) in cross-validation (CV) in North America. In contrast, the simulation-based model without GWR adjustment showed a lower accuracy with an R-square of 0.65 in CV. Seasonal validation results in North America also indicated that the spatial correlation coefficients ranged from 0.37 to 0.68 (van Donkelaar et al., 2015), which are not appropriate for air pollution and epidemiological studies, specifically over urban areas.

The Community Multi-Scale Air Quality (CMAQ) modeling system to predict $PM_{2.5}$ concentrations were used by Zhao et al. (2013) in Eastern China during 2010.

The correlation coefficients between predictions and ground-level PM_{2.5} were 0.56 and 0.64 at two monitoring stations. National PM_{2.5} concentrations were also predicted. However, PM_{2.5} was underestimated significantly in spring and summer. The simulations of CTMs tend to have a high bias due to the improper emission inventories and missingness of mechanisms.

Thus, considering the advantage of ground-level air quality monitoring network and satellite data, numerous statistical models were developed to estimate more accurate ground-level PM concentrations. The Multiple Linear Regression (MLR) and the CMAQ were used by Cordero et al. (2013) to estimate PM_{2.5} concentrations. They showed that the MLR obtained a more accurate estimation of PM_{2.5} concentrations than the CMAQ model. Incorporating MLR with CMAQ also produced more accurate predictions.

2.7.2 Statistical Models

The statistical regression approach is the most popular method to estimate PM concentrations due to its superficial characteristics and reasonable accuracy. However, this method requires a large amount of ground-level PM data for developing and validating the model results.

Different mathematical forms identified the PM-AOD relationship, including the linear regression (LR) model (Gupta & Christopher, 2009), the MLR model (Zeydan & Wang, 2019), the GWR model (Yang et al., 2019b), the linear mixed-effect (LME) model (Stafoggia et al., 2017), the multiple stage mixed models (Ma et al., 2016; Stafoggia et al., 2017), and the geographically and temporally weighted regression (GTWR) model (He & Huang, 2018).

First studies were used simple linear regression models to determine the quantitative relation between satellite AOD and ground-level PM concentrations. A linear relationship between AOD and PM was identified with the linear regression models in different regions (Wang & Christopher, 2003; Engel-Cox et al., 2004; Al-

Saadi et al. 2005; Zhang et al., 2009; Sorek-Hamer et al., 2013). PM is used as the dependent variable, and AOD is the independent variable in this model. The expression of a simple two-variable linear equation is calculated by:

$$PM = C + M \times AOD \quad (2.7)$$

where PM is the ground-level PM concentration, AOD is the columnar aerosol amount measured at 550 nm, C is the intercept, and M is the coefficient of AOD (Gupta & Christopher, 2009).

However, these studies also presented that the PM-AOD relationship showed high spatial heterogeneity. For example, Engel-Cox et al. (2004) showed that the linear relationship between AOD and PM was weak over the western US, indicating that this relationship depended on location and various factors. Since PM concentrations are influenced by various factors other than AOD, an alternative method to eliminate the negative effect of meteorology on the PM-AOD relationship was developed to use meteorological parameters as ancillary explanatory variables in MLR (Liu et al. 2005; Gupta & Christopher 2009).

The MLR model was developed to estimate PM concentrations, including meteorological parameters into the modeling stage as follows:

$$PM = C_1 + C_2 \times AOD + C_3 \times V_3 + \dots + C_n \times V_n \quad (2.8)$$

where PM represents the ground-level PM concentrations, AOD is the columnar aerosol amount measured at 550 nm, C₁ is the intercept of MLR, whereas C₂ - C_n represents the coefficients of corresponding independent variables including AOD and V₃ - V_n (such as PBLH, RH, temperature, precipitation, wind speed, etc.).

Gupta & Chiristopher (2009) used an MLR model to estimate surface PM_{2.5} in the south-eastern US during 2004-2006 in 85 monitoring stations with MODIS AOD and additional PBLH, wind speed, and temperature factors RH. They showed that satellite

derived PM_{2.5} concentrations were improved with the inclusion of the additional variables. However, MLR, a simple model including appropriate variables, is usually an unsuccessful method to estimate accurate PM concentrations with high bias.

Advanced statistical models considering spatiotemporal information such as LME (Lee et al., 2011; Xie et al., 2015) and GWR model (Yang et al., 2019) have been recently developed to determine space-time variation of the PM-AOD relationship and enhance the accuracy of satellite-based PM estimations. These models have become increasingly complex, from simple linear regression to advanced statistical methods, including LME, generalized additive models (GAM), GWR, hierarchical models, and Bayesian models (Huang et al., 2018).

The mixed-effects model (MEM) is the second widely preferred approach to PM estimation after MLR. MEM has a powerful estimation performance by integrating fixed and random relationships between dependent and independent variables at various spatiotemporal scales (Hu et al., 2014b; Lee et al., 2016b). The following equation represents the model:

$$PM_{(i,j)} = (\alpha + u_j) + (\beta + v_j) \times AOD_{(i,j)} + s_i + \varepsilon_{(i,j)} \quad (u_j, v_j) \sim N[(0,0), \Sigma] \quad (2.9)$$

where $PM_{(i,j)}$ is the ground-level PM concentration at location i on day j , $AOD_{(i,j)}$ is the AOD at location i on day j , α and u_j are fixed, and random intercepts corresponding to fixed β and random v_j slopes, respectively, s_i is the random intercept at the location i , $\varepsilon_{i,j}$ is the error term at the location i on day j , and Σ is the variance-covariance matrix for the day-specific random effects. Lee et al. (2016b) developed MEM for every season to estimate daily PM_{2.5} concentrations to detect the effects of temporal variations of land use and meteorological data on PM_{2.5} concentrations over the western United States.

As a MEM, LME models showed good performances since this approach allows the day-to-day variations to generate satellite-derived PM concentrations. However, LME modeling restricted PM estimations only in a small area since it assumes that

space demonstrates little spatial variability within the study region, which is a significant affecting factor of the PM-AOD relationship (Chu et al., 2016). Stafoggia et al. (2017) developed a spatiotemporal LME model to predict daily PM₁₀ for each 1-km² of Italy. Particularly, AOD and ground-level PM₁₀ concentrations were calibrated using a multivariate linear mixed effect model, with random intercepts by day and slopes by AOD, and fixed effects for the other spatial and spatiotemporal parameters. The model predicted 65% of PM variability with a slight bias on held-out monitors.

The GWR is another widely used technique to identify local variation to deal with the non-stationarity spatial relationship of PM-AOD over a large area due to their simplicity and versatility (Xiao et al., 2017; Yang et al., 2019). The GWR model was mainly proposed to consider local variations so that the coefficients in the model could be estimated in specific locations (Brunsdon et al., 1996). GWR can be described as:

$$PM_{(i,j)} = C_{0,(i,j)} + C_{1,(i,j)} AOD_{(i,j)} + C_{2,(i,j)} V_{2,(i,j)} + C_{3,(i,j)} V_{3,(i,j)} + \dots + C_{n,(i,j)} V_{n,(i,j)} \quad (2.10)$$

where $PM_{(i,j)}$ is the ground-level PM concentration at location i on day j , $C_{0,(i,j)}$ is the intercept at location i on day j , $AOD_{(i,j)}$ is the AOD at 550 nm at location on a day j , and $C_{2,(i,j)}$ to $C_{n,(i,j)}$ are coefficients of corresponding independent variables $V_{2,(i,j)}$ to $V_{n,(i,j)}$ at a location on day j (Brunsdon et al., 1996). However, this model's main challenge is ignoring the temporally varying relationship of PM-AOD that is probably dependent on days closed to the estimation day. Therefore, temporal autocorrelation in the data was not considered in daily GWR, and the model fails to build up a model for days with fewer PM-AOD paired samples (He & Huang 2018).

A daily GWR model and a two-stage (LME + GWR) model were also implemented to analyze spatial and temporal weighting, which developed a reliable association between surface PM and satellite AOD (He & Huang 2018). Most recently, the GTWR model, a space-time regression model developed by Huang et al. (2010), has been used for PM estimations (Chu et al. 2015; Bai et al. 2016; Guo et al. 2017b; Guo et al.,

2021). Differing from the two-stage model, GTWR separately calibrates the relationship between PM and AOD with two processes. The GTWR model can consider temporal information from the current and previous days into spatial non-stationarity through a spatiotemporal weighting mechanism combining the spatial and temporal distances. Guo et al. (2021) proposed a study across China to calibrate PM_{2.5} concentrations during 2015–2018 using several auxiliary data with a novel GTWR model. Sample-based (CV) of the GTWR model showed a reasonable performance with a R^2 equal to 0.67, a RMSE of 10.32 $\mu\text{g}/\text{m}^3$, and a mean prediction error of 6.56 $\mu\text{g}/\text{m}^3$.

In addition to the regression models discussed above, different models are proposed based on the spatial variation. The land-use regression (LUR) model uses population, land cover, and traffic information to estimate PM concentrations. LUR is mainly used to determine air pollution in a densely populated area. Mirzaei et al. (2018) used the LUR model to estimate surface PM_{2.5} in Alberta during the wildfire season. Their results indicated that model performance was highly associated with ground-level concentrations. A nonlinear regression model named GAM allows nonlinear regression between independent and dependent variables (Liu et al., 2007). GAM connects the dependent variable with independent variables and allows exponential family distribution for combining characteristics of a Generalized Linear Model with additive models. It is considered a non-parametric extension of Generalized Linear Models and an extension of classical linear models. However, those models tend to show low estimation performance. A most recent study (Shtein et al., 2019) used the advanced geographically weighted GAM to estimate PM₁₀ and PM_{2.5} concentrations. The GAM approach provides information about the weights for each learner to change spatially and by concentration compared to the standard method for ensemble averaging that uses linear regression to obtain fixed weights for each learner. The CV R^2 ranges from 0.73 to 0.76 for PM₁₀ and from 0.79 to 0.81 for PM_{2.5}.

Estimation studies have mainly focused on PM_{2.5} concentrations. However, the number of studies interested in PM₁₀ concentrations around the world has been increasing. The traditional physical method with corrected AOD with PBLH and RH

was used to determine 10-km resolution PM_{10} concentrations over Europe (Koelemeijer et al., 2006; Emili et al., 2010). PM_{10} concentrations in Athens, Greece, were predicted by Benas et al. (2013) with a nonlinear algorithm. Nordio et al. (2013) developed the LME model in northern Italy to estimate PM_{10} concentrations. Linear and nonlinear multiple regression models were proposed to estimate PM_{10} concentrations in Tehran, Iran (Sotoudeheian & Arhami, 2014). The MEM with day-specific random effects was used to estimate PM_{10} and $PM_{2.5}$ concentrations in London, UK (Beloconi et al., 2016). Ghotbi et al. (2016) tried to estimate PM_{10} concentrations in Tehran, Iran using the WRF model. Kamarul Zaman et al. (2017) developed the MLR and artificial neural network (ANN) models to estimate PM_{10} concentrations in Malaysia.

PM with excellent spatial and temporal heterogeneities is affected by several factors consisting of meteorological parameters, human activities, and topography. These heterogeneities explain and construct the PM–AOD relationship difficult with traditional physical and statistical regression approaches due to their weak data-mining abilities causing poor estimation of PM concentrations. In addition, remote sensing of aerosol still deals with high estimation uncertainties and multiple missing values over bright and heterogeneous surfaces (Wei et al., 2021b).

In addition to the traditional statistical models, machine learning, a subdivision of artificial intelligence (AI), is recently one of the best methods to determine the complex nonlinear relationship between AOD, PM, and other related variables. They particularly indicate an outstanding predictive performance (Breiman, 2001; Zhan et al., 2017). Most AI methods have been adopted simply in PM predictions or estimation despite their more vital data mining ability ignoring the spatiotemporal variations of PM (Breinmen, 2001). Moreover, a computer's performance highly affects deep learning and is not efficient computationally.

However, machine learning (ML) algorithms are capable of particular abilities. They can solve the pattern instead of memorizing training data. ML used standard manner by making estimations using new data compared to statistical regression

models. This is the most crucial advantage of ML compared to traditional models (Breiman, 2001). It also uses no assumption widely used in different statistical methods related to data distributions (Breiman, 2001). In recent years, more studies have focused on the use of satellite-based products to estimate surface PM₁₀ or PM_{2.5} concentrations (Chen et al., 2018; He & Huang, 2018; Shtein et al., 2018; Ahmad et al., 2019; Park et al., 2019; Stafoggia et al., 2019; Jin et al., 2020; Wei et al., 2020a; Jiang et al., 2021) using unlimited ML methods. Due to the flexibility of combining remote sensing, land use, meteorological, and CTM inputs with the capabilities of handling nonlinear relations between variables, the machine learning approaches showed superior or comparable performance to traditional statistical models. They have can automatically extract advanced representations of nonlinear interactions among variables with a strong learning capacity. ML also has a high capacity for handling many predictors (Di et al., 2016). To handle big datasets, statistical analysis has become an increasingly important tool to analyze complexity in spatiotemporal data, relationships quantitatively, and dependencies uprooting statistical machine learning methods.

In addition, previous studies reported relatively good performances on estimation/prediction of PM based on traditional parametric regression models (e.g., LME, GWR) (Fang et al., 2016; Ma et al., 2016; Staggofia et al., 2017), ML as a non-parametric regression had advantages to getting higher predictive power (Hu et al., 2017). Machine learning models mainly include the random forest (RF, Chen et al., 2018; Brokamp, et al., 2018; Wei et al., 2019a), the extreme gradient boosting model (XGBoost; Chen et al., 2019a; Ma et al., 2020; Yan et al., 2020; Zhang et al., 2021), back-propagation and generalized regression neural networks (BRNN and GRNN; Li et al., 2017), artificial neural network model (ANN; e.g., Kamarul Zaman et al., 2017), back-propagation neural network (e.g., Gupta & Christopher, 2009), the geo-intelligent deep belief network (e.g., Li et al., 2017; Yan et al., 2020; Yan et al., 2021), and support vector regression (SVM; e.g., Weizhen et al., 2014; Mehdipour et al., 2018), extremely randomized trees (ERT, Wei, et al., 2020a; Wei et al., 2021a). All statistical models used for PM estimation are illustrated in Figure 2.11.

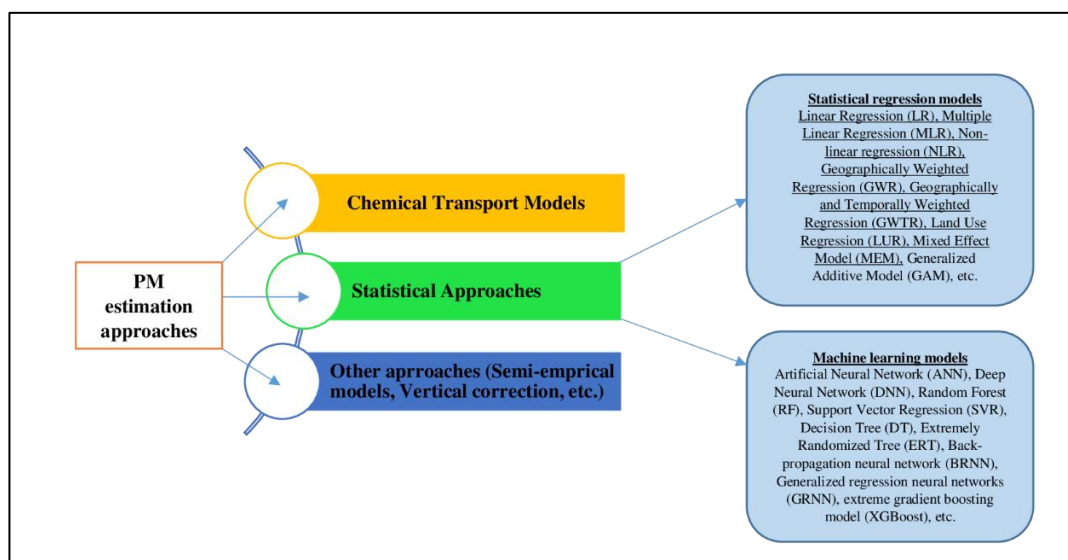


Figure 2.11 PM estimation models widely used in the literature

For example, Di et al. (2016) developed an ANN model with AOD, CTM outputs, the absorbing aerosol index (AAI), land use, and meteorological parameters to estimate daily $PM_{2.5}$ in the USA. Results indicated that ANNs have great potential ($R^2 \sim 0.84$) to estimate surface $PM_{2.5}$. Li et al. (2017) presented a GRNN model for entire China from February 2013 to December 2014. The GRNN model presents excellent advantages in $PM_{2.5}$ estimation compared with conventional models such as the linear regression model and GWR model. Mehdipour et al. (2018) indicated that SVM showed good performance in explaining the variability of $PM_{2.5}$ concentrations in Iran ($R^2 \sim 0.88$). The extremely randomized trees (ERT; Geurts et al., 2006) are tree-based ensemble learning approaches. It is preferred to handle sophisticated regression problems and develop a consistent relationship between PM and AOD. Cut points in this model were randomly selected and split into nodes. All training samples were used to construct and grow trees rather than the bootstrap approach. ERT model efficiently solved variance problems, and high predictive performance was obtained. Wei et al. (2021a) improved the estimated near-surface $PM_{2.5}$ with the space-time extremely randomized trees (STET) model. The $PM_{2.5}$ estimates are substantially consistent with ground-level measurements on the national scale, with an out-of-sample cross-validation coefficient of determination (CV- R^2) of 0.88, RMSE of $16.52 \mu g/m^3$, and a mean absolute error (MAE) of $10 \mu g/m^3$. Wei et al. (2021b) also developed the STET model to estimate ground-level PM_{10} concentrations. The MAIAC AOD,

meteorological parameters, land cover, elevation, population, and emission datasets were included in the STET model to create a high-quality PM₁₀ dataset for China from 2015 to 2019. The model achieved high performance with out of sample CV-R² of 0.86 and RMSE of 24.28 µg/m³.

In contrast to many other ML algorithms, the RF approach is user-friendly and more flexible to use, including fewer limited assumptions with minimal tuning parameters to get excellent performance (Breinmen, 2001). Many single and multimodel evaluation studies have indicated the accuracy of the RF model. (Hu et al., 2017; Huang et al., 2018; Park et al., 2019; Wei et al., 2019a; Park et al., 2020b; Yang et al., 2020; Ghahremanloo et al., 2021; Meng et al., 2021; Lu et al., 2021; Jiang et al., 2021). Random forests as a type of machine learning algorithms have been widely preferred for nonparametric regression studies with competitive accuracy (Hu et al., 2017). Hu et al. (2017) used the RF model to predict daily ground-level PM_{2.5} concentrations in the USA, with CV-R² of 0.80. Park et al. (2019, 2020) used an RF to estimate daily PM_{2.5} and PM₁₀ over South Korea, integrating 32 parameters including MODIS AOD. Park et al. (2019) achieved good performance with R² values of 0.78 and 0.73 for PM₁₀ and PM_{2.5}, respectively. The RF models with spatial information produced better estimation results with R² of 0.88 and 0.90 for PM₁₀ and PM_{2.5}, respectively (Park et al., 2020).

Huang et al. (2018) also developed a high-performance RF model with the MAIAC AOD, meteorological parameters, land cover, and five PM_{2.5} species, including black carbon, organic carbon, sea salt, dust, sulfate from MERRA-2 to estimate monthly PM_{2.5} concentrations in North China Plain. A recent study by Stafoggia et al. (2019) developed a novel approach over Italy, applying an RF model with the gap-filled AOD with atmospheric ensemble models as the main predictor to estimate daily PM₁₀ and PM_{2.5} concentrations. RF models captured most PM variations, with mean CV R² of 0.75 and 0.80 for PM₁₀ and PM_{2.5}, respectively.

RF model performances were also compared with traditional regression models. Wei et al. (2019a) estimated PM_{2.5} concentrations in China based on the AOD data

from MAIAC with a currently developed STRF model. The 24-h mean $\text{PM}_{2.5}$ concentrations were estimated in 2016 with meteorological and other ancillary variables. The MLR model, the GWR model, and the two-stage model were developed to compare results with the 10-fold CV approach. High and robust accuracy ($R^2 \sim 0.85$, RMSE of $15.57 \mu\text{g}/\text{m}^3$) was found with a sample-based CV of the STRF model and surpassed the traditional regression models. Ghahremanloo et al. (2021) developed an RF model to estimate daily $\text{PM}_{2.5}$ concentrations with MAIAC AOD and additional control parameters, including the daily total column densities of dust, AE, sea salt, sulphur dioxide (SO_2), sulphate (SO_4), organic carbon (OC), and black carbon (BC) to solve the regional differences of $\text{PM}_{2.5}$ over Texas. The performance of the RF model was stable throughout the whole study area and showed a high ability to estimate $\text{PM}_{2.5}$ concentrations. The performance comparison of traditional regression models (MLR and MEM) was evaluated with the RF model. RF model outperformed the MLR and MEM over Texas.

Recently, the ability of ensemble models to predict from multiple base learners combining their predictive power and create a final prediction that outperforms the forecasts from each base learner with integrating predictions from numerous base learners combining their predictive power make them a popular approach. The primary assumption of this approach is that each model has its advantages and limitations, and an ensemble model that incorporates the PM predictions from a heterogeneous set of base learners generated from different modelling approaches is beneficial (Shtein et al., 2019; Yazdi et al., 2020). In addition to ensemble models, to integrate the respective advantages of different models, combined models (stacking methods based on different statistical models) have been increasingly developed and used to improve the accuracy of PM estimation (Xiao et al., 2017; Chen et al., 2019b; Feng et al., 2020).

One learner may better predict high PM concentrations and another low concentration, or one learner may be better in one region and another in a different area. Therefore, the ensemble approach can be favorable for the areas characterized by complicated air pollution problems due to its diverse geo-climatic areas and the

complex sources of air pollution from anthropogenic and natural sources (Shtein et al., 2019). LME, RF, XGBoost, and a CTM were used to improve the estimation of daily PM₁₀ and PM_{2.5} concentrations over Italy, developing a GW-GAM ensemble model incorporating the combined predictions from different models. Yazdi et al. (2020) predicted daily PM_{2.5} levels in the Greater London area using an ensemble ML approach using AOD, land use, and meteorological data. The ensemble model included predictions from an RF, GBM, and a k-nearest neighbor (KNN) approach. The performance of the ensemble model was high, with a CV-R² of 0.828. However, of the three machine learners, the RF outperformed the GBM and KNN.

Chen et al. (2019b) combined AdaBoost, XGBoost, and RF as a stacking model to estimate the hourly PM_{2.5} concentrations from Himawari 8 AOD data in Central and Eastern China. The three machine learning models were stacked with MLR models in the current study, and the accuracy evaluation showed that the proposed stacking model performed better than the single models ($R^2 \sim 0.85$).

According to PM estimation studies in the literature, statistical models, notably including the spatiotemporal information, predominates simulation-based and simple PM-AOD models such as MLR with better statistical performance. Also, the statistical model developed for one region could be easily applied for another region with the same variables used in a statistical model differing from a physics-based model that requires complex technical information for a specific region. Consequently, machine learning is the best option to estimate high accuracy PM concentrations. However, model performances vary significantly by region.

Studies mainly focus on the national level PM prediction (Staggofia et al., 2017; Chen et al., 2018; Park et al., 2019; Staggofia et al., 2019; Wei et al., 2019a; Guo et al., 2021; Wei et al., 2021a). Some of them only focused on specific regions in China (Huang et al., 2018; Chen et al., 2019a), the USA (Ghahremanloo et al., 2021), and Europe (Staggofia et al., 2017, 2019). However, there is no study to determine the variation of the national model performances over different regions in an entire country. Mainly site scale performances were discussed at national level estimation

studies. Furthermore, determining the regional differences is required as well as national model performances. Heterogeneity in the aerosol characteristic, surface type, and meteorological factors including temperature, RH, PBLH differs at various spatial scales with diverse land cover types may cause a weak relationship between PM and satellite based AOD and affect the estimation of spatial model performances (Chen et al., 2018). On the one hand, air quality monitoring stations are rarely distributed over the large areas of suburban and rural regions, causing very few PM₁₀ measurements. The satellite AOD retrievals may also be inadequate due to the contamination of clouds and the limitation of the aerosol retrieval algorithm over different areas. Therefore, it is essential to develop a robust estimation model over a different geographical region.

Most importantly, few studies have estimated long-term PM₁₀ at a national or regional scale worldwide using satellite based AOD (Mang et al., 2016; Kamarul Zaman et al., 2017; Chen et al., 2018; Park et al., 2019; Wei et al., 2021b). However, long-term data have not been used for PM₁₀ estimation over Turkey, with lower aerosol loading than in China. One short-term PM_{2.5} estimation study with the old collection (C6) MODIS AOD was conducted over Turkey using MLR (Zeydan & Wang, 2019). However, this study is limited to ten stations from two different regions.

This study estimated intra-daily concentrations of PM₁₀ over Turkey during the past 12 years with the most recent ground monitoring data, AOD, meteorological, land-related, and aerosol type information using the most popular ML approach. Model development was performed for each geographical region separately to consider regional heterogeneities. It is also known that estimation models showed different performances under different PM₁₀ levels (Park et al., 2019). Quantile regression (QR) with RF was also conducted in this study to test model performances affected by different PM₁₀ concentrations.

CHAPTER THREE

MATERIALS AND METHOD

3.1 Study Area

Turkey, located between 35° and 42° north and between longitude 26° and 44° east, on the border of two continents of Europe and Asia, and covers an area of 780,580 square kilometers (the 36th largest in the world) and recently its population has been increasing (Ustaoglu & Aydinoglu, 2019). The number of people living in urban areas has risen dramatically since 1996. Turkey's population is approximately 84 million, and the annual population growth rate is 5.5% in 2020 (TSI, 2021). The population count of province/district centers and towns/villages indicated that 93% of Turkey's population in 2020 live in urban areas, and 7% live in cities (TSI, 2021). The rapid expansion of metropolitan and urban areas is observed with economic development, growth in population, and migration from rural areas to urban regions in Turkey.

Turkey is subdivided into seven geographic regions and 81 provinces (Figure 3.1). Thirty provinces are specified as metropolitan municipalities with a population of more than 750,000. Around two-thirds of the population lives in the three largest cities (İstanbul, Ankara and İzmir) and other large and medium-size cities such as Bursa, Gaziantep, Adana, Malatya, Erzurum, Trabzon, Kayseri, Antalya, Kocaeli, Konya, Eskisehir, Mersin, Diyarbakir, and Samsun (Ustaoglu & Aydinoglu, 2019).

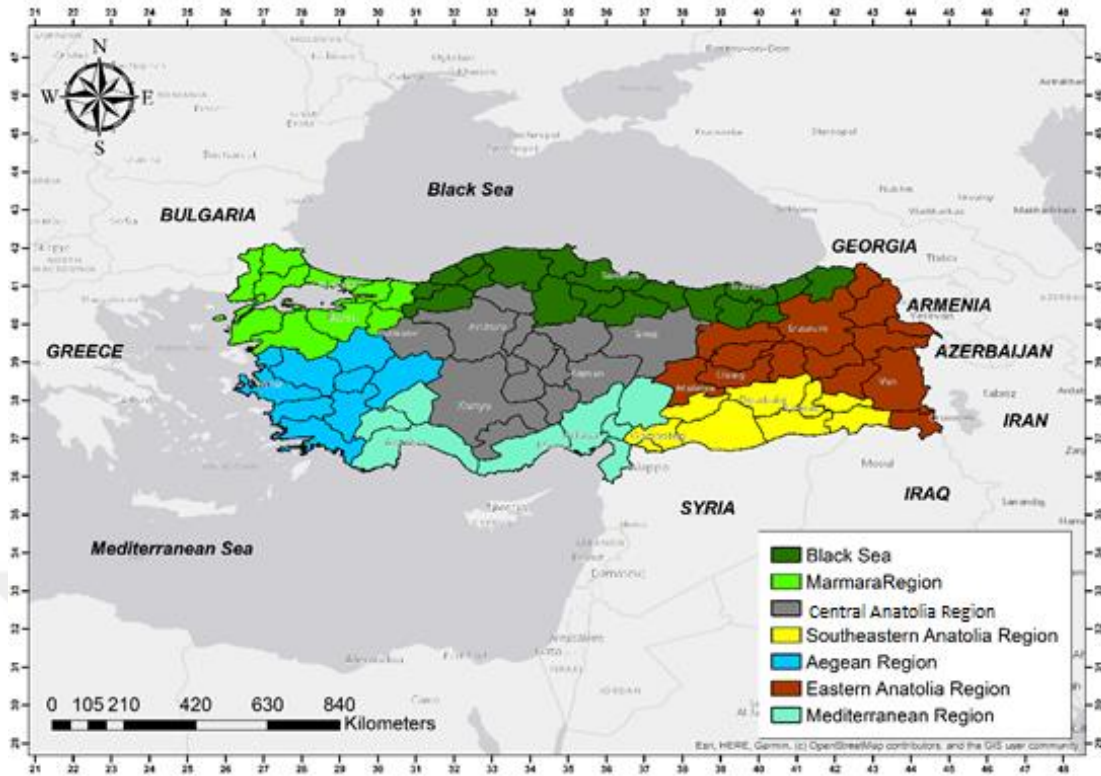


Figure 3.1 Geographical regions in Turkey

Turkey has a significant location for air circulation between Europe, Asia, and Africa. Coastal borders surround it with the Mediterranean Sea, the Aegean Sea, and the Black Sea and by the land borders of Greece, Bulgaria, Armenia, Azerbaijan, Georgia, Syria, Iraq. Especially in summer, air quality can be decreased by the Mediterranean lower troposphere since the Mediterranean is a crossroad of the long-range transport of aerosols (Georgoulas et al., 2016). It is also known that, in the February–April period, African dust contributed to 80% of the PM_{10} concentrations in the southeastern part of the Mediterranean (Lelieveld et al., 2002). The spatial distribution of multiannual average Dust mass surface concentrations and Angstrom Exponent (AE) values from MERRA-2 is illustrated in Figure 3.2. Higher dust mass concentrations were observed in Turkey's Eastern and south-eastern parts, with the lower AE values (~ 0.7) representing the coarse particles. Long-range transport of pollution contributes to aerosol pollution in the urban areas over these regions. According to Figure 3.2, higher AE values (>1) observed over the western and north-western parts of Turkey represent fine particles that mainly originate from anthropogenic sources. Turkey is also affected by marine aerosols, mainly composed

of sea salt and sulfate aerosols from the Black Sea and the Mediterranean Sea (Ozdemir et al., 2020).

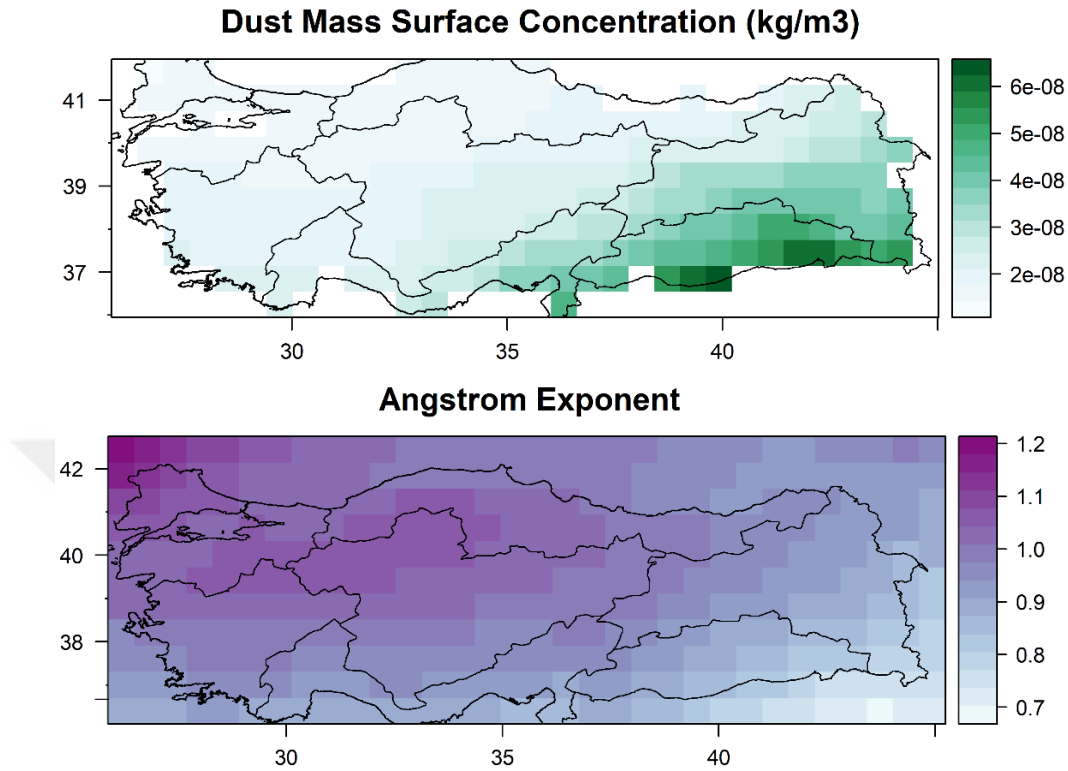


Figure 3.2 Spatial distribution of multiannual average dust mass surface concentrations and AE values over Turkey

The climate in Turkey is varying by region. The Mediterranean climate is mainly observed in the Mediterranean and Aegean coastal regions. The coastal region climate is moderate with more significant precipitation, and the continental climate is typical in the Black Sea region. The inner part has Anatolia Plateau with hot summers and cold winters and limited rainfall characterized by the steppe climate (Toros et al., 2013). Spatial distributions of multiannual average precipitation, temperature, solar radiation, and total area fraction are illustrated in Figure 3.3. The spatial distribution of regimes becomes more complex from west to east and from coastal to inner parts. Extended period of winter-spring precipitation can affect the Aegean, southeastern Mediterranean, and southwest Anatolia regions. Continental Anatolia regions including Central Anatolia, Eastern Anatolia, and the southern part of the Black Sea regions can modify the formation of precipitation by local topographic conditions (Kutiel & Türkes 2005). The eastern part of Turkey receives much less precipitation

in winter than other parts of the country (Türkmen, 1998). The Black Sea region receives the most significant amount of precipitation, and it is the only region that receives precipitation throughout the year. From north to south, solar radiation, and temperature increase. However, cloudy conditions are most common in the northern part of Turkey. Mediterranean and south-eastern regions are characterized by relatively higher temperatures compared to other regions. Spatial distribution of solar radiation is closely associated synoptic-scale atmospheric circulation forms of the cloudy conditions. In summary, dominant cloud and precipitation regimes in the north and northwest explain the low solar radiation over northern region. Cloud cover can enter towards the south in the central regions from northern part of Turkey. Therefore, spatial distribution of solar radiation over Turkey can be defined by a reverse seesaw between the southwest and the southeast.

Local sources and topographical and meteorological conditions can seriously affect the spatial and temporal variation of PM (Erdun et al., 2011). The main affecting factors for air quality in Turkey are industrialization, urbanization, massive construction activities, and high fuel consumption (Elbir et al., 2000; Muezzinoglu et al., 2003; Elbir, 2004; EEA, 2019). Turkey contributes to more than half of the industrial PM₁₀ emissions in Europe (EEA, 2019). The air pollution profile of Turkey has changed with the political decisions and actions with the significant improvements in the air quality in some metropolitan cities and residential areas. However, some regions in Turkey have still air pollution problems due to geographical locations, characteristics of land use, meteorological factors, industrial, nonindustrial, and fugitive sources (Elbir & Muezzinoglu, 2004; Elbir et al., 2010; Kara et al., 2014; Kara et al., 2015; Tuna Tuygun et al., 2017). There is at least one air quality monitoring station in all provinces, although the number of stations increases yearly (307 in 2019). EEA (2019) also reported that concentrations above the PM₁₀ annual limit value (40 µg/m³) in 2017 were monitored at 7 % of all the reporting stations (207stations), and most of them were in Turkey (121).

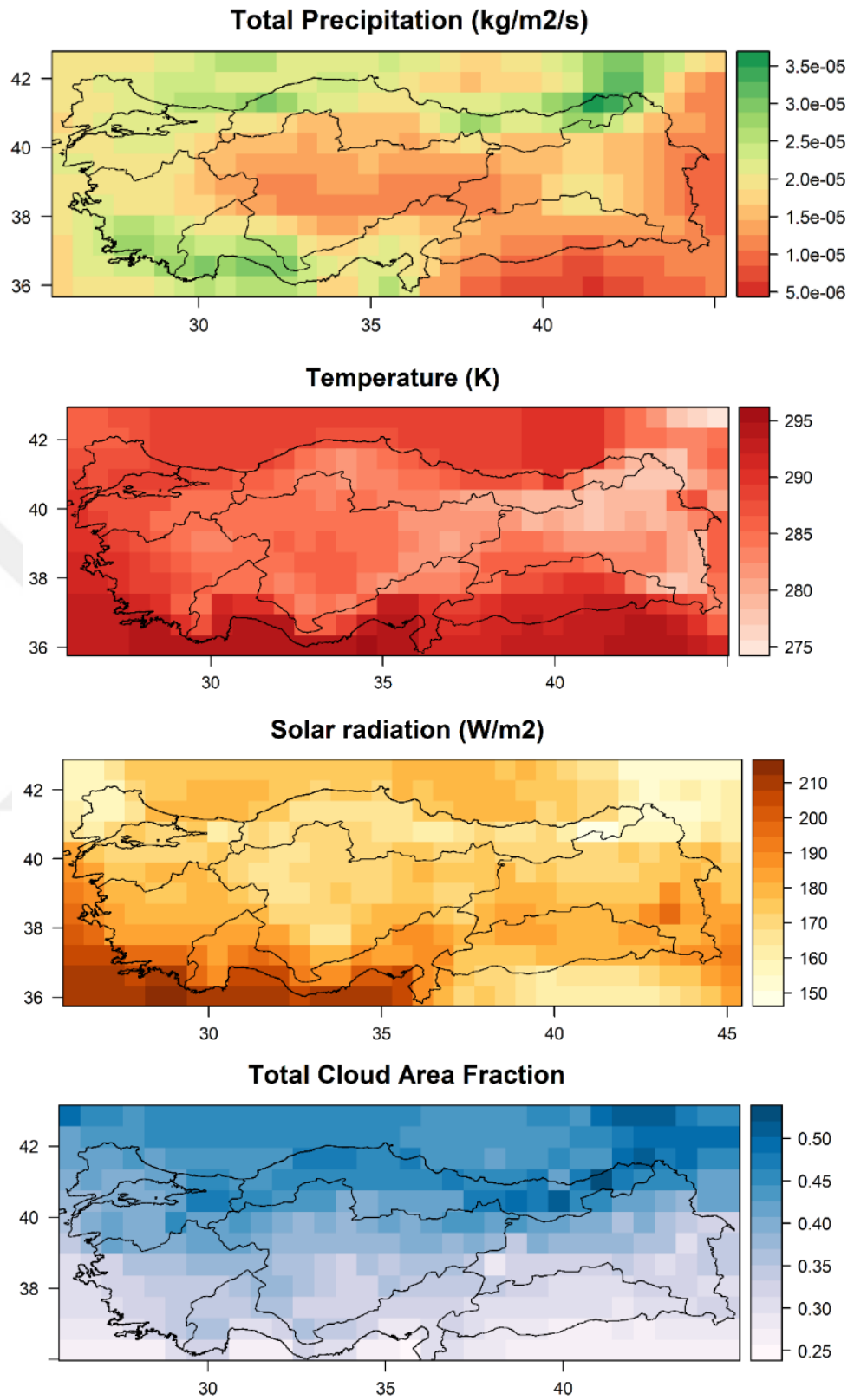


Figure 3.3 Spatial distribution of multiannual average meteorological parameters over Turkey

Turkey is also the 17th largest country regarding its population and has extensive area coverage globally (Ustaoglu & Aydinoglu, 2019). Fifteen percent of the land area is dominated by forest cover, particularly representing the coastal and mountainous areas. In contrast, agricultural areas cover forty-four percentages of the country (Aldabash et al., 2020). As seen in Figure 3.4 vegetation cover is densely in northern, southern, and western parts of Turkey. Agricultural areas are mostly situated in Inner part. Vegetated areas located in and around urban areas can influence the mixture of particles. The emission of Biogenic Volatile Organic Compounds (BVOCs, e.g., isoprene and monoterpenes) leads to secondary aerosol formation. Depending on the vegetation type, this effect can vary significantly; agricultural areas, for example, can have pollution levels similar to the urban background.

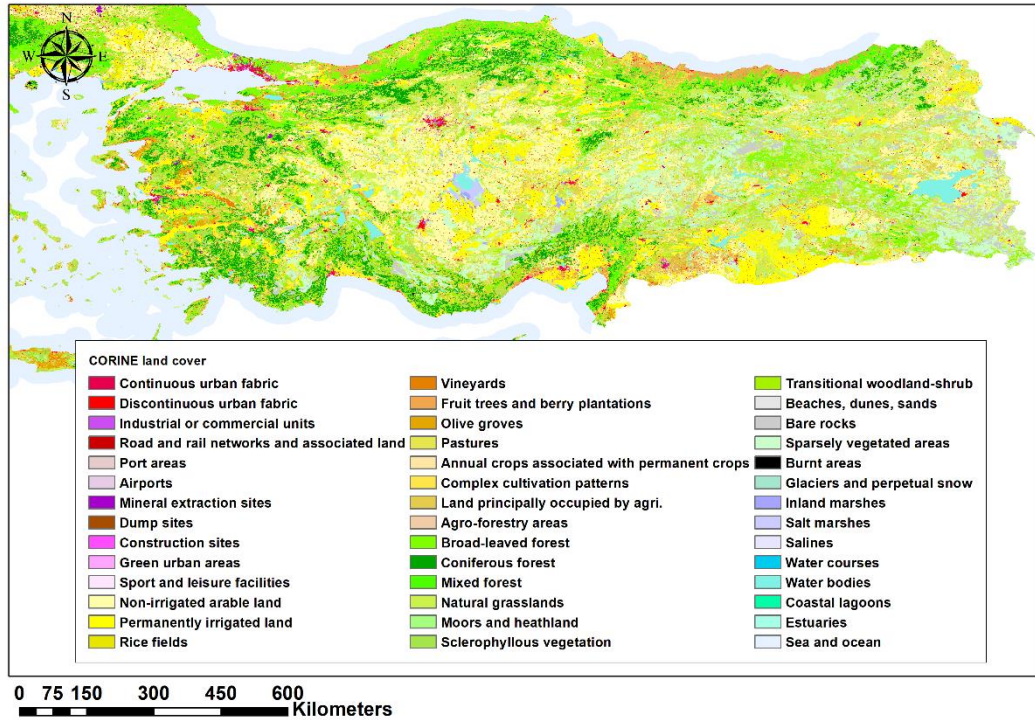


Figure 3.4 CORINE land cover over Turkey

3.2 Data Collection and Integration

Data from different sources were collected including the daily PM₁₀ concentrations, MODIS AOD data, assimilated meteorological and aerosol fields characterizing atmospheric properties, and land-related variables including elevation and NDVI.

3.2.1 Satellite AOD Data

An appropriate product should be selected to analyze the long-term aerosol trend over a region since various processes due to deserts, industrialized and populated countries cause spatiotemporal trends of AOD worldwide (Georgoulas et al., 2016). Level 2 MODIS AOD products from January 2008 to December 2019 over Turkey were obtained from the Level 1 Atmosphere Archive and Distribution System (LAADS) website (<https://ladsweb.modaps.eosdis.nasa.gov/>) as an HDF (Hierarchical Data Format) file and archived. The HDF files are multi-objective files. They restore the data from 53 different parameters in a special format (SDS-Scientific Data Set) with their cellular structure in a matrix form.

Both C6.1 MOD04 (Terra) and MYD04 (Aqua) were used to extract AOD at 550 nm (AOD550) with a spatial resolution of both $3 \text{ km} \times 3 \text{ km}$ and $10 \text{ km} \times 10 \text{ km}$ products for validation. In this study, both MOD/MYD04_3K and MOD/MYD04_10K AOD were extracted at 550 nm (SDS: Optical_Depth_Land_And_Ocean) for the Dark Target (DT) algorithm. In addition to DT AOD, Deep Blue (DB) AOD (SDS: Deep_Blue_Aerosol_Optical_Depth_550_Land_Best_Estimate) was also obtained for the 10 km product. It should be noted that DT and DB AOD only contain high-quality retrievals (DT QAC= 3 over land, QAC=1,2,3 over ocean, and DB QAC= 2,3). Quality assessment was performed on these best quality AOD retrievals due to their higher confidence with free of cloud contamination. The C6.1 MODIS Level-2 operational aerosol products also contain a new Scientific Data Set (SDS) “AOD_550_Dark_Target_Deep_Blue_Combined” for Terra (MOD04) and Aqua (MYD04) representing the merged aerosol product (DTB) based on the DT and DB aerosol retrieval algorithms. As neither DT nor DB AOD consistently outperforms each other in a certain region, a merged data product including DT and DB algorithms was generated by Sayer et al. (2014). DTB AOD (SDS: “AOD_550_Dark_Target_Deep_Blue_Combined”) were also extracted for 10 km product. AOD retrievals obtained from the SDS “AOD_550_Dark_Target_Deep_Blue_Combined” were filtered for highest quality

assurance flag (QAC = 3) using the SDS “AOD_550_Dark_Target_Deep_Blue_Combined_QA_Flag”. The equatorial overpass time of Terra is about 10:30 local solar time (LST), and Aqua is about 13:30 LST, respectively. Terra and Aqua C6.1 AODs at 550 nm obtained with DT, DB, and DTB algorithms were simultaneously evaluated. The scan starts of two MODIS sensors (Terra and Aqua) vary between 07:00-12:00 UTC over Turkey.

Negative (but small magnitude) AOD values due to the uncertainty in surface reflectance assumptions of the DT algorithm (Levy et al., 2013) were also included in validation analyses. They also protect negative bias in long-term low AOD statistics and are allowed by NASA to avoid more bias over land. They have minimal impact on the results, as discussed in Jin et al. (2020), since they are close to zero. MODIS is not sensitive over land to retrieve AOD better than ± 0.05 in spotless conditions. NASA recommends individual value analysis to convert negative values to zero. Therefore, in the estimation model, all negative values are assumed as zero.

3.2.2 Ground-Based AOD Data

The AERONET Version 3 (V3) Level 2 (L2) AOD product was selected and used in this study since it is particularly preferred in scientific research. Cloud-screened and quality-assured (V3, L2) AERONET data were downloaded from the website (<http://aeronet.gsfc.nasa.gov>) to validate the MOD/MYD04 AOD products. The AERONET provides observations with seven channels (0.340–1.020 μm) every 15 min that is more accurate than satellite data with an uncertainty of ~ 0.01 – 0.02 in the absence of cloud contamination (Holben et al., 1998).

Two different (coastal rural, and urban) AERONET sites in Turkey were used in this study. The coastal IMS-METU-ERDEMLI site (hereafter coastal rural site) is located in Erdemli district in the southern coastal part of Turkey (36.57°N and 34.26°E). Typically, three aerosol sources exist over Erdemli district, including dust from the Middle East and Sahara deserts located in the southern part, anthropogenic particles from industrialized regions located in the north-eastern part, and marine

aerosols from the Mediterranean Sea (Tutsak & Kocak, 2019; Ozdemir et al., 2020). The coastal rural region also has a typical Mediterranean climate, with dry and hot summers and warm and wet winters. It has the highest rainfall in winter, mainly in December and January (Ozdemir et al., 2020).

The inner urban TUBITAK_UZAY_Ankara (hereafter urban site) site is temporarily instrumented as discussed in Section 2.4. The site was in Ankara, the inner part of Turkey, and the instrument was mounted on the roof of a building (39.89°N and 32.78°E). Ozdemir et al. (2020) reported that aerosol sources at this site are predominantly marine particles in all seasons except in winter. However, continental aerosols are dominant aerosol types over the site in winter due to the urban/ industrial activities over the region. The site is also significantly affected by dust aerosols in summer. The inner region has a hot-summer Mediterranean climate which closely borders a hot summer Mediterranean continental climate. Due to its elevation and inland location, Ankara has cold, somewhat snowy winters and hot, dry summers. The high topography of Ankara causes long-term cloudy weather and snow in winter. The high level of cloud cover is challenging for satellite aerosol retrieval over the urban site. Rainfall occurs mainly during the spring and autumn (Ozdemir et al., 2020). Operational times, geographical properties, pollution sources, and surface characteristics are different over two AERONET sites in Turkey.

Some data gaps were found in AERONET AOD datasets. It is known that frequent calibrations (annually) demanded by the network protocol cause gaps in time series, as mentioned in Raptis et al. (2020). Also, cloudy skies and instrument malfunction can cause gaps and multi-day gaps in time series, respectively. The AERONET observations are available from January 2008 to December 2019 at the coastal rural site and from December 2009 to April 2012 at the urban site.

3.2.3 Ground Level PM_{10} Data

PM_{10} concentrations were analyzed in this study due to the greater availability of PM_{10} data in Turkey. PM_{10} concentrations can be more appropriate to analyze since

larger particles could represent local conditions and are not easily transported. Additionally, PM_{2.5} does not contain the more significant fraction of particles that are nevertheless accountable for light extinction and are thus measured by the satellite.

The hourly ground-based PM₁₀ concentrations for 2008–2019 were obtained for 307 PM₁₀ monitoring stations over Turkey. Data were retrieved for each station from the public website of MEU (https://sim.csb.gov.tr/STN/STN_Report/DataBank), which is the government agency in charge of the collection of air pollution data in Turkey (MEU, 2020). This website contains historical air quality measurements since 2005. An ambient suspended particulate monitor system (Model MP101M of Environment S.A), and the instrument's measuring principle is based on Beta Gauge monitoring. It is approved by the U.S. Environmental Protection Agency (EPA) are used to measure PM₁₀ concentrations (Karaca, 2012).

In Turkey, PM₁₀ measurements are shared with quality control. However, it is not enough to remove all complex data from the dataset. Additional quality control of all measurement data should be performed before the data is used. Hourly PM₁₀ concentrations were used after data cleaning. However, there is no validation procedure for data cleaning in Turkey; the U.S. EPA Ambient Data Reduction and Audit Procedures were used for validation of measurements (<https://www.in.gov/idem/airquality/2377.htm>, Chapter 12). The codes were written in R programming language, including the following quality control procedure for PM₁₀ data:

- Negative and zero PM₁₀ measurements should be removed,
- Rates of change in PM₁₀ concentrations greater than 50 µg/m³ per hour should be detected,
- Particulate concentrations that remain unchanged for 5 hours may indicate a monitor malfunction,

At sites with both PM₁₀ and PM_{2.5} monitors, PM_{2.5} concentrations should be less than PM₁₀ concentrations.

The data was checked, and some apparent outliers were removed. Missing data (values recorded as NA, *Not Available*) and values lower than the $2 \mu\text{g}/\text{m}^3$ were also removed (Jiang et al., 2021). PM₁₀ data that does not meet the quality procedure given above were removed from the dataset. Ninety-four stations were not included in this study due to the inadequate data number and low data quality. Mobile and test stations were also not included in the study. Finally, 213 ground monitoring stations were included in this study, shown in Figure 3.5. The distribution of the stations is not symmetrical, with several clusters located in the most populated urban areas (mainly western and north-western parts of Turkey), while decentralized stations were in the eastern and south-eastern of Turkey (Figure 3.5). Rural areas also have little coverage. Station numbers according to the region are given in Figure 3.6. According to the geographical regions, the distribution of monitoring stations shows that 34% of monitoring stations are in the Marmara region. Only eight monitoring stations measuring PM₁₀ are in the South-eastern Anatolia region.

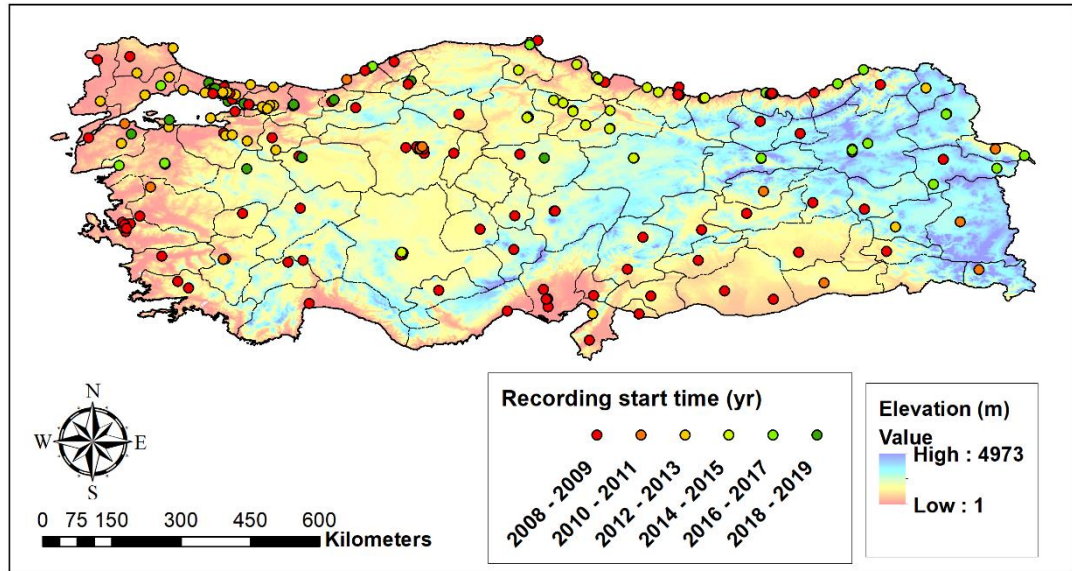


Figure 3.5 Location of PM₁₀ monitoring stations

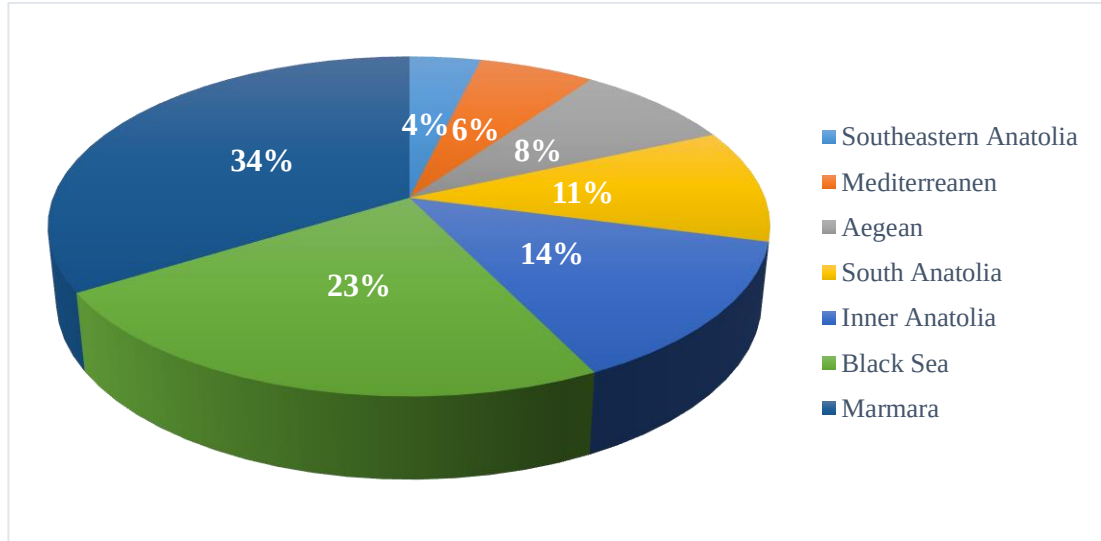


Figure 3.6 Numbers of regional PM₁₀ stations over Turkey

3.2.4 MERRA-2 Reanalysis Aerosol and Meteorological Fields

Meteorological parameters have crucial roles in the relationship between PM₁₀ and AOD. They are commonly used as temporal predictors in estimating the PM₁₀ concentrations (Meng et al., 2016). Since there are no distinct guidelines to select predictors in PM estimation, the related meteorological variables were considered according to many previous studies (Su et al., 2018; Stagoffia et al.; 2019; Nabavi et al., 2019; Wei et al., 2020a; Jiang et al., 2021) that have indicated those factors having substantial influence on PM and its relationship with AOD. The overall relationship between Terra and Aqua AOD and PM₁₀ measurements are not close across Turkey (overall $R \sim 0.10$ and 0.18), suggesting the influence of meteorological and other auxiliary quantities; thus, these variables were also introduced into the modeling part to enhance the relationship.

Hourly data of planetary boundary layer height (PBLH, m), solar radiation (SRAD, Wm^{-2}), pressure (P, Pa), total cloud area fraction (CLOUD, unitless), total precipitable water vapor (TQV, kg/m^2), total precipitation (PREC, $\text{kg/m}^2/\text{s}$), u-component (u, m/s) v-component (v, m/s) of the wind, specific humidity (SHUM, kg water vapor/kg air), dew point temperature at 2 m (DEWP, K), the temperature at 2 m (T, K), evaporation (EVAP, kg/m^2) and surface roughness (ZOM, m) were used in this study covering the period of 2008-2019. The hourly averaged assimilated meteorological variables were derived from the Modern-Era Retrospective Analysis for Research and Applications,

version 2 (MERRA-2) (<https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2>). MERRA-2 provides a multidecadal reanalysis for aerosol and meteorological observations assimilated within a global data assimilation system (Gelaro et al., 2017).

The PBLH, P, T, and SRAD were selected since they are associated with atmospheric stability affecting the vertical distribution of pollution. T at 2 m also affects chemical composition and physical properties of aerosols (Price et al., 2016). Wind components represent the dispersion. Specific humidity and total precipitable water representing the vertical distribution of atmospheric water vapor can cause hygroscopic growth of aerosol, potentially impacting AOD. Precipitation effectively cleanses the atmosphere locally, and total cloud cover can affect the AOD with increased humidity that favors the hygroscopic growth of aerosols and leads to higher AOD values. Precipitation, evaporation, and surface pressure can also encourage the production or removal of PM₁₀ (Steinberg et al., 2018). Cloud cover leads to reduced photochemical reaction related to PM₁₀ production and precipitation remove PM₁₀ from the atmosphere, leading to decreased PM₁₀ dry mass concentration. RH can also affect the particle size and the extinction of radiation in the atmosphere. However, this study did not include RH since PM₁₀ concentrations are measured at ground monitoring stations with controlled constant humidity and temperature. The increased humidity does not affect PM₁₀ measurements as significantly as AOD. Barnes et al. (2014) showed that spatially and temporally varying surface roughness can significantly affect the ground-level air quality.

The PM–AOD relationship might also depend on the compositions of aerosols (Yang et al., 2019; Jin et al., 2020). MERRA-2 also includes a multi analysis technique in the aerosol data assimilation used to assimilate AOD considering the reflectance from different sensors such as the MODIS on Terra (2000–present) and Aqua (2002–present), the Advanced Very-High-Resolution Radiometer (AVHRR) (1979–2002), the MISR (2000–2014) and AOD measurements from the AERONET network (1999–2014) (Randles et al., 2017). In this study, hourly (M2T1NXAER 5.12.4) MERRA-2 aerosol diagnostic dataset including the total AE parameter and the surface mass

concentrations of dust (DUSMASS, kg/m³), sea salt (SSMASS, kg/m³), SO₄ (SO₄MASS, kg/m³) were obtained closest coordinates at the PM₁₀ monitoring station.

3.2.5 Auxiliary Data

The geographical data of land parameters, including the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) at 250 m resolution, was used to obtain the elevation to represent the dispersion effects of topography. Since the land surface types vary significantly across Turkey and the terrain variations can affect the diffusion of air pollutants (Wei et al., 2019a). The MODIS C6 monthly NDVI products (MOD/MYD13A3) at 1 km resolution were also used to define surface types. NDVI is commonly used to represent vegetation change status and vegetation coverage. NDVI ranges from -1 to 1, with negative values indicating urban water area, rocks, or soil are indicated with near-zero values and positive values representing the green and densely vegetated area (Didan et al., 2015). Land areas can classify with NDVI values as lands with no vegetation cover (NDVI<0.2), partially vegetation cover (0.2≤NDVI≤0.3), moderately vegetation cover (0.3<NDVI<0.5), and densely vegetation cover (NDVI≥0.5). Thus, a climatological mean value of NDVI is computed for a 50 km radius around each PM₁₀ monitoring station using the monthly MODIS Collection 6 NDVI data for 2008–2019 to consider the contribution of regional biogenic emissions from oxidation (Jin et al., 2020).

It is known that species of PM₁₀ and the corresponding characteristics (like light extinction) show space and time variations, and there are apparent spatio-temporal variations in the relationship between PM₁₀ and other predictor variables. Thus, many studies have considered these problems (Huang et al., 2018; Ghahremanloo et al., 2021). Remote sensing applications, more significantly, machine learning estimation models were dependent on the quantities of predictor variables, and they also regard both spatial and temporal information (Hu et al., 2017; Brokamp et al., 2018; Huang et al., 2018).

The sample size and concentrations of ground based PM₁₀ vary by region. Therefore, PM models with location information, including longitude and latitude, were developed in this study. To control the temporal variability of PM₁₀, such as seasonality and long-term effect, some temporal predictors were also considered, including the day of the year, day of the week, calendar month, the month of day, and hour of the day in the literature (Chen et al., 2018; Wei et al., 2020a; Jiang et al., 2021). The spatial variables of longitude (Lon) and latitude (Lat) and the temporal variables of cumulative days (cdate) representing the long-term trends and Day of the Year (DOY) representing the seasonality effects were included in the model. The cdate as a covariate provides the information to the RF model to fit different spatial patterns each day, underpinning that the experimental parameters are different from day to day (Hengl et al., 2018). Lanzaco et al. (2016) and Chen et al. (2018) showed that the use of day of the year (DOY) as an informative temporal feature could successfully replace remaining unknown parameters that regulate daily PM_{2.5}/PM₁₀ concentration. Hence, DOY named as jday in study was used as input to resolve the significant temporal variation of PM₁₀ in Turkey.

3.2.6 Validation of MODIS AOD Data with AERONET Observations

The most recent updated algorithms were discussed in detail in Section 2.3. Validation of previous collections of C4, C5, C5.1, and C6, has been studied from local areas to global scales (Li et al., 2007; Levy et al., 2013; Remer et al., 2013; Georgoulias et al., 2016; Mhawish et al., 2017; Sayer et al., 2015; Wei et al., 2018; Wei et al., 2019b). However, The C6.1 of MODIS AOD has been used in different regions (Almazroui, 2019; Rupakheti et al., 2019). However, many researchers focused on China (Tian et al., 2018; Wei et al., 2018; Bilal et al., 2019; Che et al., 2019; Zhang et al., 2019) or use Level-3 gridded products (10× 10) for global validation (Bright & Gueymard, 2019; Wei et al., 2019b), the performance of the new C6.1 of MODIS AOD is still uncertain over specific regions that have mixed surface characteristics. However, the aerosol datasets obtained from different aerosol retrieval algorithms have variations in regions (He et al., 2017; Wei et al., 2018). The aggregated global and regional statistics, particularly the general structure for the

overall accuracy of different AOD products, are provided. However, a particular uncertainty is caused due to the different sample sizes of the inhomogeneous global distribution of AERONET sites. It means that sites are not combined with weighted equally in calculated metrics. Site-based validation is required. An assessment of the last version of the MODIS (C6.1) 3 km-DT product is still limited in the literature.

Since the accuracy of the satellite AOD compared to the AERONET AOD may cause PM estimation error (Wang et al., 2019), it is significant to determine the quality of the AOD products and consider the C6.1 MODIS AOD performance over Turkey. The C6.1 aerosol product has not been thoroughly evaluated in Turkey. The relative performance differences and advancements among different satellite derived AODs have not been evaluated. Therefore, the first goal of this study is to extensively validate the recent C6.1 MODIS AOD, mainly considering their relative performances.

It is known that AOD retrievals can be correlated to geographical conditions (Wei et al., 2019b) and surface types (Wang et al., 2019). It is indicated that surface types affect the quality of DT AOD products with 3 km and 10 km spatial resolutions (He et al., 2017). It is also concluded that the DB algorithm performs better than the DT algorithm over bright areas. It mainly depends on the accuracy of surface reflectance estimation over different land cover types (Wei et al., 2019b). Thus, the performance of all products was determined under two different surface types, including urban and coastal rural AERONET sites over Turkey. The urban site is also located over a partially vegetated surface ($NDVI \sim 0.30$). In contrast, the coastal rural site is located over the moderately vegetated surface ($NDVI > 0.40$) (Tuna Tuygun et al., 2020). Detailed analysis of aerosol types over AERONET sites in Turkey was studied in our previous study (Ozdemir et al., 2020). Improvements of C6.1 MODIS DT products were also discussed in detail by Tuna Tuygun et al. (2020). Analysis was performed with direct comparisons between MODIS AOD retrievals and AERONET ground-based observations based on product and site type. The comparisons were made between 2008 and 2019 at the coastal rural site and from 2009 to 2012 at the urban site using C6.1 releases. Terra and Aqua C6.1 AODs with the recommended QAC for the DT ($QAC = 3$ for land and $QAC = 1, 2, 3$ for ocean), DB ($QAC \geq 2$), and merged DTB

products (QA = 3) at a 10-km resolution are concurrently collected for validation. Both Terra and Aqua AOD products were validated in this study since more rapid degradation of the MODIS-Terra instrument causes distinct differences between the Terra and Aqua satellites (Levy et al., 2018). Particularly in local regions, significant variations in spectral response and data imaging time have still been observed between the Terra and Aqua satellites (Levy et al., 2013, Gupta et al., 2018; Levy et al., 2018).

Satellites simultaneously provide a snapshot of a larger region; however, AERONET provides a repeated point measurement over time. Collocation of both the AERONET measurements and the MODIS retrievals is required with a spatiotemporal window to consider both the spatial and temporal variation of aerosol distribution and sufficiently match the spatially averaged satellite AOD retrievals with the temporally averaged ground-based measurements. The spatiotemporal approach was firstly proposed by Ichoku et al. (2002) to increase the number of matched datasets for the sites without long-term measurements and decrease the effectiveness of various underlying aerosol properties. They suggested comparing spatially averaged satellite retrievals in a 5×5 grid box centered sun photometer measurements of AOD within a time window of ± 30 min of satellite overpass time as illustrated in Figure 3.7 for 10 km product. This approach has been adopted in several validation studies for MODIS aerosol products (Ichoku et al., 2002; Remer et al., 2013; Tao et al., 2015; Tian et al., 2018; Wei et al., 2019b; Wang et al., 2019; Jin et al., 2020). Those studies have used the average MODIS AOD in a window of three or five pixels to match the AERONET measured AOD.

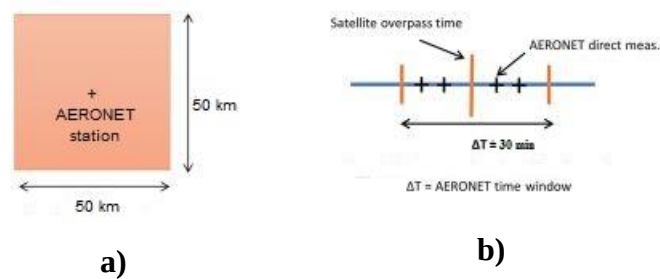


Figure 3.7 An illustration of the spatiotemporal collocation scheme of the satellite retrievals with the AERONET measurements (Jethva et al., 2019)

Differences in aerosol characteristics are minor in the 5×5 -pixel windows. It was also considered that the 10-km averaged AOD over 3×3 pixels is not appropriate to reflect the regional aerosol characteristics since the lower statistical sample may reduce the correlation (Ichoku et al., 2002). Since AERONET does not have measurements at 550 nm, satellite-retrieved AOD was interpolated from 500 nm to 550 nm based on the Angstrom exponent (440–675) (Tian et al., 2018) using Equation (3.1)

$$AOD_{550} = AOD_{500} \times \left(\frac{550}{500}\right)^{-\alpha} \quad (3.1)$$

where,

AOD_{550} : Aerosol optical depth at 550 nm

AOD_{500} : Aerosol optical depth at 500 nm

α : Angstrom Exponent ($AE_{440-675}$)

Data used for validation is summarized in Table 3.1.

Tablo 3.1. Summary of AOD data used for validation

Product	Scientific Data Set (SDS) name	Contents	Spatial and temporal resolution
MOD/ MYD04_L2	Optical_Depth_Land_And_Ocean	DT over land (QA = 3) and ocean (QA = 1,2,3)	Daily, 10 km
	Deep_Blue_Aerosol_Optical_Depth_550_Land_Best_Estimate	DB over land (QA $\geq$ 2)	
	AOD_550_Dark_Target_Deep_Blue_Combined	DTB over land and ocean (QA=1,2,3)	
	AOD_550_Dark_Target_Deep_Blue_Combined_QAF_Flag	QAF = 1 to 3	
MOD/ MYD04_3K	Optical_Depth_Land_And_Ocean	DT over land (QA = 3) and ocean (QA = 1,2,3)	Daily, 3 km
AERONET	AOD_{500}	Version 3 Level 2	15 min, point

Statistical indicators, including the correlation coefficient (R), the relative mean bias (RMB), and the root mean square error (RMSE), were used to quantify the accuracy of MODIS AODs against AERONET ground measurements. Expected errors

were calculated in terms of the part (ratio) that remained within the expected error values (=EE), below (<EE), and above (>EE). The validation results and uncertainties for the matched data were mainly analyzed to evaluate the quality of the MODIS AOD estimates with the expected error (EE) envelope containing at least 67% of the matchups over land and ocean. The percentage of the matchup pairs falling within the EE envelope, which is defined as $[\pm (0.05 + 0.15 \text{ AOD}_{\text{AERONET}})]$ and $[+0.04/-0.02 + 0.1 \text{ AOD}_{\text{AERONET}}]$ for DT AOD at a 10-km resolution over land and ocean, respectively (Levy et al., 2013). The performance of mixed land and ocean AOD retrievals was evaluated with the same EE over land (Wei et al., 2019b). The EE is also similar to the DT product for DTB AOD (Wei et al., 2019b). The EE is approximately $[\pm (0.03+0.20 \text{ AOD}_{\text{AERONET}})]$ for DB retrievals at a 10-km resolution over land.

As the performances of AOD products differ seasonally, all statistical evaluations were also calculated based on seasons. This matchup and validation process was carried out on a station-based using MATLAB 2019b and R 3.5.2 programming languages. The statistical equations used in the study were given below:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (AOD_{\text{MODIS}} - AOD_{\text{AERONET}})^2} \quad (3.2)$$

$$RMB = \overline{AOD_{\text{MODIS}}} / \overline{AOD_{\text{AERONET}}} \quad (3.3)$$

3.2.7 Data Processing and Integration

Spatiotemporal regression matrix was prepared by integrating geographical covariates, including AOD, meteorology, aerosol fields, and available additional temporal and spatial covariates for the area in PM₁₀ estimation. MODIS-Terra and MODIS-Aqua overpass in the late morning and early afternoon, respectively. It was also investigated whether the AOD-PM₁₀ relationship varies when limited data with the time window in which the MODIS instruments passing over Turkey separately. The dataset from all available observations to only observations was obtained between

7-10 LST for Terra and 11:00–13:00 LST for Aqua to evaluate the AOD-PM₁₀ relationship. Towards midday, explained variability of PM with AOD increases. The mixing layer height during midday is generally well-mixed during conditions with accurate AOD retrievals. The observed atmospheric path is longer at higher solar zenith angles (low sun). Horizontal gradients in aerosol properties or atmospheric conditions influence the AOD in a larger extend (Schaap et al., 2008). Therefore, correlation between AOD and PM is expected to be higher under small solar-zenith angles (Schaap et al., 2008).

Thus, AOD from both satellites was evaluated with a separate estimation model. Two main PM₁₀ estimation models were developed for each time window: (1) Intra-daily AAOD model using the Aqua AOD product to estimate PM₁₀ around the Aqua overpass time. This model estimates the mean PM between 10:00–12:00 LST; (2) the Intra-daily TAOD model estimates PM₁₀ around the Terra overpass time. This model uses Terra AOD product as a predictor and estimates the mean PM between 07:00–10:00 LST. The observed PM₁₀ concentrations as a target variable were not log-transformed since they significantly affected model performance.

Re-processing data with different spatial intervals and temporal resolutions for integration is required. A 10-km grid was created precisely matched to the combined AOD pixels. Within a spherical distance of 50 km from each PM₁₀ monitoring site valid AOD retrievals were extracted from the MODIS Level-2 swaths (in 10 km resolution). A mean AOD for a given date and station was calculated if more than five valid AOD measurements are available within a search radius of 50 km. This radius was selected to consider the differences in the sampling of AOD and ground-level PM₁₀. The AOD observations are pixel based (spatially averaged) instantaneous measurements, whereas the near-surface PM₁₀ measurements are point measurements. Air parcels from at least a 50 km radius are advected over the monitoring station even with very low wind speeds. Using a radius threshold of 50 km increases the number of days with valid AOD retrievals.

The MERRA-2 dataset used to analyze relationships between PM₁₀ and AOD has a spatial resolution close to 50 km. AOD values were obtained similar to validation; the average AOD was calculated within 5×5 pixels sampling window (at least 5 available pixels of 25 pixels) centered on each PM₁₀ monitoring site. The MODIS 10 km grid was used to integrate all the datasets. When there was more than one monitoring site in the same 10 km $\times$ 10 km MODIS grid cell reporting PM₁₀ on the same day (and thus would share all covariate values), all available PM₁₀ measurements were averaged.

The aerosol crosses a pixel in 1-hr when wind speed is approximately 10 km/h. It means that AOD can correspond to hourly ground-level measurements. Both hourly and daily PM concentrations have been used for comparison in the literature (Seo et al., 2015; Guo et al., 2017a; Chen et al., 2018; Nabavi et al., 2019; Wei et al., 2019a; Wei et al., 2020a). Hourly estimations can be more valuable since they may be used to determine the effect of PM on health outcomes at different hours of the day, probably originating from different sources, rather than using a 24-hours mean PM estimate. The use of hourly or daily mean PM measurements influences the relation between PM and AOD (Engel-Cox et al., 2006). However, Nabavi et al. (2019) compared two methods, and they found no significant differences in model performance using hourly or daily averages. In this study, the single value or average of PM₁₀ concentrations within ± 30 min (~ 1 -hr) of the overpass times of Terra and Aqua satellites was used (Seo et al., 2015; Guo et al., 2017a; Yao & Palmer, 2021) since different times of the day are characterized by different PM concentrations due to variability in the influencing factors such as anthropogenic activities such as industry, heating, and transportation, meteorological conditions and air mass properties, and long-range dust transport or dust from local sources (Tuna Tuygun & Elbir, 2020).

MERRA-2 aerosol fields and meteorological data were matched to each MODIS grid if this MODIS grid fell into a given MERRA-2 grid. Output from the 0.625° by 0.5° grid cell containing each PM₁₀ monitoring station was extracted. Since the accumulation effects are significant in air pollution (Sun et al., 2021), 24 hours accumulated precipitation was calculated with hourly data used in the model as used

in Park et al. (2019) and Sun et al. (2021). Meteorological values were assigned to the study grid without interpolation (Jin et al., 2020; Just et al., 2020). All synchronous meteorological data at PM₁₀ observation times were used for this investigation. The same collocation method was used for aerosol data obtained from MERRA-2.

Values of AOD, meteorological conditions, aerosol types, NDVI, and elevation were extracted at the location of each monitoring station. Spatial predictors (e.g., latitude, longitude, and elevation) were grid cell levels changing from cell to cell but assumed to be fixed over time. All data used for estimation modeling was summarized in Table 3.2. The 25 selected independent variables, including spatial and temporal covariates, were paired with the intra-daily PM₁₀ concentrations for each station on each day. After discarded invalid data, 153145 and 138744 paired intra-daily ground-level PM₁₀ concentrations and all predictor variables involving the days during 2008- 2019 for Terra and Aqua, respectively. These data were integrated into a standard grid as the training datasets and then were extracted according to the latitude and longitude points. Only the grids containing all the independent variable data and daily PM₁₀ measurements were considered in the variable selection and modeling process.

Table 3.2 List of data used in modelling study

Data Type	Parameter	Abbreviation	Unit	Spatial and Temporal Resolution	Source
Meteorology	Planetary Boundary Layer Height	PBLH	m	Hourly - 50 km	MERRA-2 Reanalysis https://gmao.gsfc.nasa.gov/reanalysis/
	Solar Radiation	SRAD	W/m ²		
	Total Precipitable Water Vapor	TQV	kg/m ²		
	Total Precipitation	PREC	kg/m ² /s		
	Specific Humidity	SHUM	kg water vapor/kg air		
	Evaporation	EVAP	m		
	Temperature at 2m	TEMP	K		
	Dewpoint Temperature at 2m	DEWTEMP	K		
	Surface Pressure	P	hPa		
	Total Cloud Area Fraction	CLOUD	unitless		
	u wind component	u	m/s		
	v wind component	v	m/s		
	Surface roughness	ZOM	m		
Reanalysis Aerosol	Sea Salt Surface Mass Concentration	SSSMAS	kg/m ³	Hourly - 50 km	MERRA-2 Reanalysis https://gmao.gsfc.nasa.gov/reanalysis/
	Dust Surface Mass Concentration	DUSMAS			
	Sulphate Surface Mass Concentration	SO4SMAS			
	Angstrom Exponent	AE	Unitless		
Ground-based measurement	PM ₁₀ Concentration	PM ₁₀	µg/m ³	Hourly - point	Ministry of Environment and Urbanization (MEU) https://www.havaizleme.gov.tr/
Satellite product	Terra and Aqua C6.1 MODIS AOD ₅₅₀ Dark_Target_Deep_Blue_Combined (QA=3)	AOD _{MODIS}	Unitless	Daily – 10 km	Terra and Aqua MODIS Level 1 Atmosphere Archive and Distribution System (LAADS) website https://ladsweb.modaps.eosdis.nasa.gov/
Auxiliary data product	Normalized Differences Vegetation Index	NDVI	Unitless	Monthly-1 km	https://ladsweb.modaps.eosdis.nasa.gov/
	GMTED2010 Terrain Elevation	Elevation	m	250 m	https://www.usgs.gov

3.2.8 Variable Selection

AOD is a significant predictor that can provide PM concentrations in a region without ground-based monitoring. Two main random forest models for Terra and Aqua overpass times according to the availability of daily Terra AOD (TAOD) and daily Aqua AOD (AAOD) named as TAOD and AAOD models. According to two essential metrics, the most important predictors for each model were selected to obtain more

simple models; some variables were also removed based on the predictive measure of association to avoid repeating the information. The predictive performance of the models can affect the largely redundant variables. However, feature selection can improve the interpretability and scalability for updating and iteratively improving estimation models in the future. Firstly, predictive features with strong correlations were determined using Pearson's correlation coefficient (PCC) to reduce data redundancy and model complexity. The PCCs between the predictive and dependent variables PM_{10} concentration were computed to represent each predictive factor's importance roughly.

The predictor variables' collinearity is required to be considered not to duplicate the information transmitted in the predictor variables. Collinearity among the predictor variables was detected with the variance inflation factor (VIF) method. VIF is an index widely used to measure the level of collinearity of multiple variables in a regression analysis (He & Huang, 2018; Wei et al., 2019a; Wei et al., 2020a). The VIF approach was applied to determine multivariable collinearity problems which level the predictors are independent of each other. Recommended threshold values for VIF range from 5 to 10 (Wei et al., 2019a). The VIF value of less than 10 indicates no collinearity problem, and values between 10 and 100 indicate high collinearity. In this study, variables with a VIF value higher than 10 are considered to indicate a severe collinearity problem, as reported by He & Huang (2018).

The role of the input variables in each model was finally explored based on the analysis of feature importance in the RF model. Results from PCC, VIF, and RF variable importance scores were evaluated together to find the most influential independent variables. Contrary to traditional statistical regression models, the RF for regression randomly creates bootstrap samples from the training data and builds regression trees for each bootstrap sample. A random sample of all predictors was selected at each node of the regression tree to choose the best split among them (Liau & Wiener, 2002). Firstly, all independent variables were included in the preliminary RF models, and variable importance was measured. Increasing in node purities (Inc.NodePurity) was calculated for each model. In this study, variables with

Inc.NodePurity <8,000,000 were excluded from the final RF models. The remaining variables below the node purity limit used in this study were not included since they did not increase the predictive ability of the models as discussed in Hu et al. (2017) and Chen et al. (2018).

A low correlation was found between AOD and PM₁₀, which will be discussed in detail in subsection 3.2. Although nearly all variables showed a weak correlation with PM₁₀, they were still effective on the final RF model accuracy and performance. Since the basis of variable importance determination calculated by RF determine the importance scores of variables used in the RF model by using the Mean Squared Error (MSE) of out-of-bag (OOB)/held-back data which do not use linear relation between the variables. The importance metrics of each variable were then calculated after the model fitting process. The features selected at the higher nodes of the tree provide a more robust estimation in the final RF model. Finally, according to the variable selection procedure, some variables with low correlations and importance scores were removed from the final RF model to simplify the model. Then, the most critical variables were used to predict PM₁₀ in the final RF models.

3.2.9 Construction and Validation of Estimation Model

Previous studies that estimated PM concentrations with ML techniques frequently used the RF model (Hu et al., 2017; Chen et al., 2018; Wei et al., 2019a; Zhao et al., 2019; Yang et al., 2020), which showed performance than other ML models (Hu et al., 2017; Chen et al., 2018; Ghahremanloo et al., 2021). RF is the traditional ML model based on decision trees and commonly used in PM estimation since it has many abilities such as handling several input variables without dimensionality reduction, robustness to outliers, user-friendliness, and requiring the only a few parameters to tune.

RF was applied in this study to develop estimation models during 2008–2019 and estimate PM₁₀ levels incorporating MODIS AOD, land use data including elevation, NDVI, simulated aerosol, and meteorological fields from MERRA-2. To compare the

performances of AOD from different satellites, two main random forest models were developed according to the availability of daily Terra-MODIS AOD (TAOD) and Aqua-MODIS AOD (AAOD). Ground-based PM₁₀ measurements were used as the dependent variable, and all mentioned variables were used as independent variables to training the RF model.

It is also essential to keep several variables to split at in each node (mtry) high possibly (e.g., $\sqrt{p}$ where p is the number of independent variables) since a common value for mtry could result in time being systematically missed from the selection. Selection of multiple poor classifiers and form a robust classifier is a basic of RF bagging. The method has relatively simple mathematical formulation. An emphasis is put on iteratively by training the algorithm until a strong learner is produced. A training dataset is obtained from the original dataset using a bootstrap sampling method with replacement. Predictions are made based on data outside the training dataset to train each tree in the forest. Predictions in RF are generated as an estimate from several decision trees based on bootstrap samples (bagging). The average of predictions of individual trees is the final prediction (Breiman, 2001; Biau & Scornet, 2016; Hengl et al., 2018):

$$\theta^B_{(x)} = \frac{1}{B} \cdot \sum_{b=1}^B t_b^*(x), \quad (3.4)$$

where B is the total number of trees, b is the individual bootstrap sample, and t_b^* is the individual learner, i.e., the individual decision tree:

$$t_b^*(x) = t(x; z_{b1}^*, \dots, z_{bK}^*), \quad (3.5)$$

where z_{bk}^* ($k = 1 \dots K$) is the k -th training sample with pairs of values for the target variable (y) and covariates (x): $z_{bi} = (x_k, y_k)$.

To evaluate the model performance predictions from all trees are averaged and then compared with their corresponding measurements. A random subset of the predictors

is selected at each candidate split in the learning process to develop each tree in the forest, with the assumption that, a few variables will be selected in many of these trees if they are very predictive for the dependent variable. The RF structure is shown in Figure 3.8.

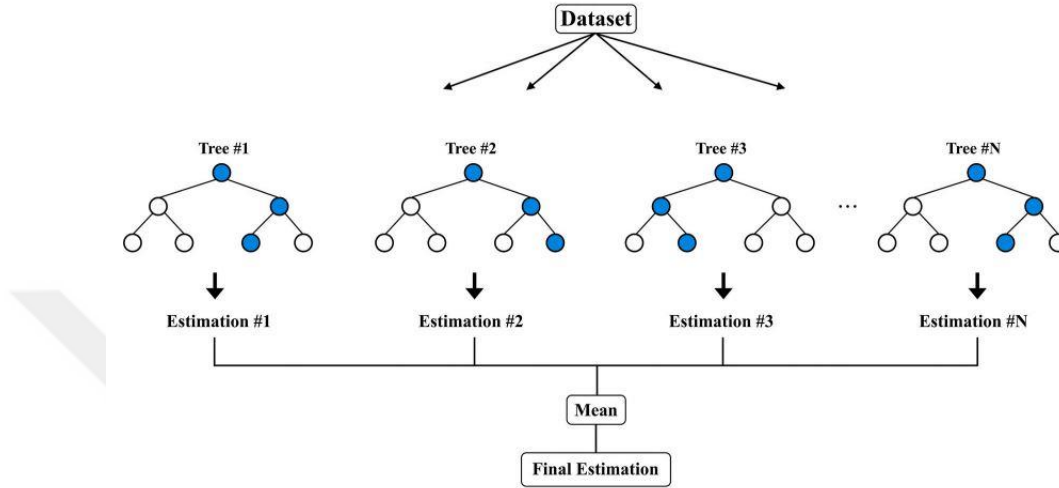


Figure 3.8 An example scheme of RF structure (Ghahremanloo et al., 2021)

The most significant tuning parameters of the RF model are (Probst & Boulesteix, 2018):

- `mtry` - number of variables to possibly split at in each node
- `min.node.size` - minimal terminal node size
- `sample.fraction` - fraction of observations to sample in each tree
- `num.trees` - number of trees. The number of trees in RF does not really need to be fine-tuned, it is recommended to set it to a computationally feasible large number (Hengl et al., 2018).

Parameters of `mtry`, `ntree`, and `node.size` generated tuning parameter combinations were tuned to be optimized simultaneously using a grid search method with the 10-fold CV with the training data and MSE as performance measure until achieving the best estimation accuracy. Table 3.3 shows the hyperparameter ranges of RF models and the optimized grid search results for each satellite with 25 independent data and a variable selection step. Changing the values of `ntree` and `mtry`, the best parameters of the model were found until obtaining the best model performance. Results were

compared using the different tuning settings and repeated CV were carried out to find a low RMSE value for each model. Results indicated that model settings showed a slight difference for each model. Best resampling results across tuning parameters for each model were given in Figure 3.9, and decreasing error rates with the number of trees, node size, and mtry were given in Figure 3.9. The numbers of tree (ntree) and min.node size for each satellite were selected as 1000 and 5, respectively. Both satellite products showed higher performance when mtry was set higher value (20 for Terra and 18 for Aqua, respectively).

Table 3.3 RF hyperparameters for grid search

Parameter	Range	RF TAOD optimum value	RF AAOD optimum value
ntree	100, 500, 1000	1000	1000
mtry	from 6 to 26 by 2	20	18
min node size	5,10	5	5

After removing the redundant variables with the variable selection process, mtry, ntree, and min.node size were defined as 6, 1000, and 5, with the smallest RMSE in this study for both satellite datasets. In this study, the ranger package of the R statistical program was used to implement the RF model, which is fast and specifically appropriate for high dimensional data and has several parameters that can be tuned in the R statistical program (Wright & Ziegler, 2017). R is the freely available software, and its extensive libraries offering built-in functions of supervised ML algorithms. These various geospatial tools can read and write vector and raster data, and extensive and flexible visualization options R, an open-source scripting language, is commonly used to analyze data and develop statistical tools (R Development Core Team, 2020). The graphical user interface RStudio was used with R in this study. The graphical user interface RStudio was used with R in this study.

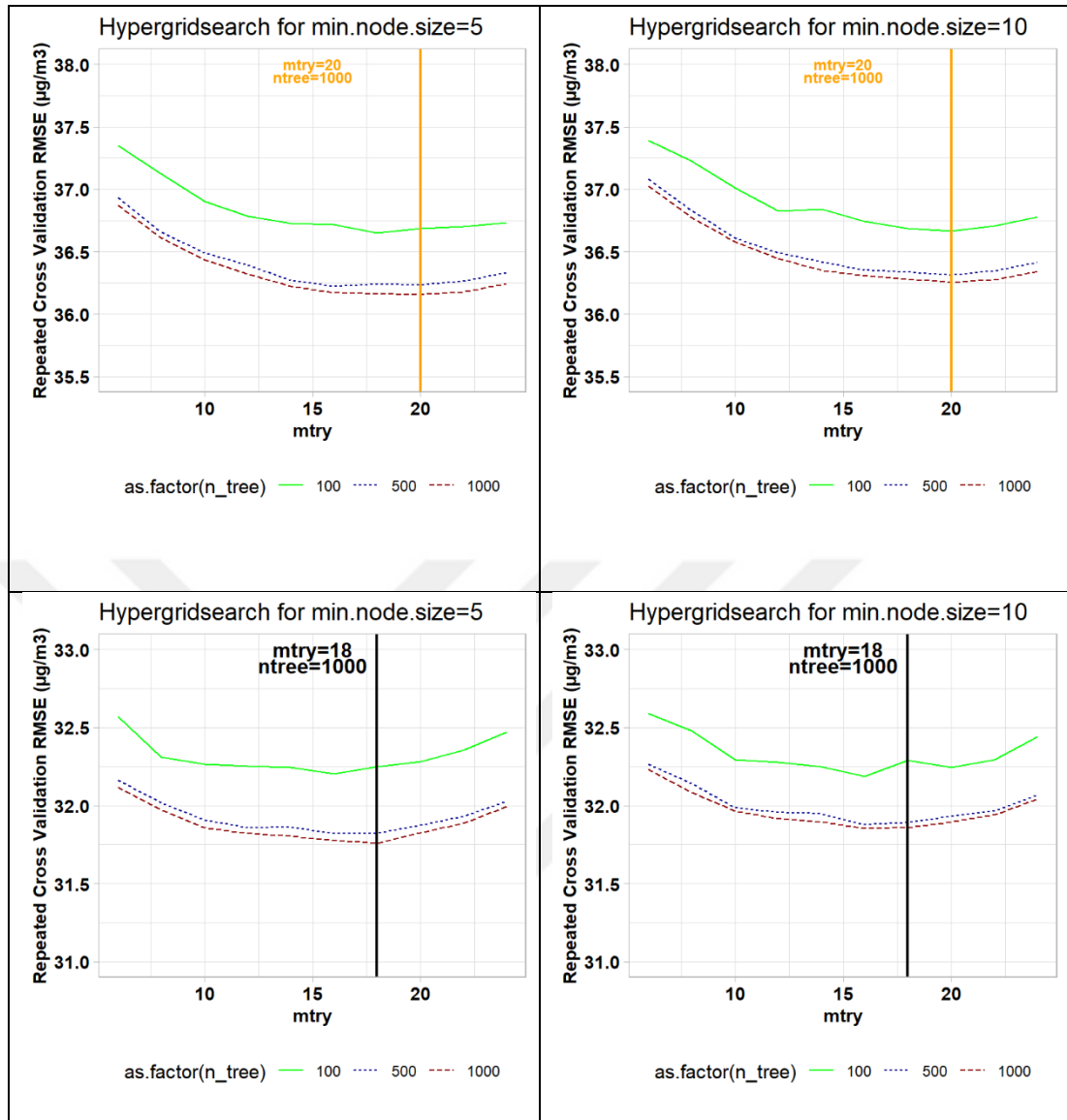


Figure 3.9 Hyper grid search results for two dataset and min.node size

The final best RF models after variable selection process with AOD from MODIS-Terra and Aqua, developed using ground measurements during 2008-2019, were applied to estimate the daily concentrations of PM₁₀ over Turkey. PM₁₀ concentrations were affected by different conditions such as population, pollution sources, meteorology (e.g., precipitation, wind speed, etc.) (Zhang et al., 2009). To consider the differences in seasonal and regional conditions, model development and PM₁₀ estimation were performed by each region (7 geographical regions) and season (spring, summer, autumn, and winter) separately.

Spatial and temporal variability of satellite-based PM_{10} estimations must be controlled against ground-level PM_{10} for model performance validation. To determine the predictive abilities of RF models, three methods, i.e., sample-, spatial-, and temporal-based 10-CV were chosen for validation of the estimated ground-level PM_{10} concentrations (Wei et al., 2019a). CV of the training data gives them confidence and an idea of how the model will perform unseen data. First, ninety percent of monitoring data from each station and month were randomly selected as the training set. In contrast, the monitoring data from the remaining stations were used as the testing set. The correlation between ground-level PM_{10} concentrations and predictions were checked. A commonly used validation method is the sample-based CV (Rodriguez et al., 2010), involving randomly chosen ninety percentages of the data for training and the remaining ten percentages for model validation. A commonly used validation method is the sample-based CV (Rodriguez et al., 2010), involving randomly chosen ninety percentages of the data for training and the remaining ten percentages for model validation.

However, this study mainly aims to determine the model performance in areas where ground monitoring stations are unavailable. The spatial- and temporal-based CV were also required. A 10-fold CV approach was applied to the monitoring stations to identify potential over-fitting in the second part. The training dataset was built by a randomly selected region (90%), and the rest of the region (10%) was used as the test dataset. Finally, controlling the correlation between ground-level and estimated PM_{10} concentrations in held-out regions was done. Model evaluation was done with the testing data only. This method can provide those different spatial points create the training and testing samples from various locations where the surface and atmospheric conditions may be distinct. It would influence the modeling results. A time-based CV was also conducted to analyze the models' reliability and stability on the temporal scales. Similar to the sample and spatial-based CV 10% days of the year (time-CV) were randomly dropped. Previous studies also implemented regular temporal 10-fold CV by randomly choosing 10% days of data each time (Wei et al., 2019a). This procedure was repeated 10 times until all subsets were estimated. The CV scheme was designed to ensure that data were not over fitted in any of these conditions (Shtein et

al., 2019). An illustration of the CV methods used in this study is given in Figure 3.10. The dataset also validated the model performance for ten days in January, April, July, and October that were randomly selected as the testing set. The remaining dataset was used as the training set as suggested by Yang et al. (2020). National model probably introduces unexpected points that have a particularly different performance from other observations within the relatively homogeneous measurement points and decrease the overall accuracy for model validation, especially for a large study area with various land cover types. In addition to two separate national RF models, tested alternative regional RF models according to geographical regions of Turkey were developed since it is difficult to develop a robust model for several regions probably due to the limited number of PM₁₀-AOD paired samples. To estimate ground-level PM₁₀ for the entire Turkey, 14 regional STRF models were developed for each geographical region and satellite separately with the same independent variables.

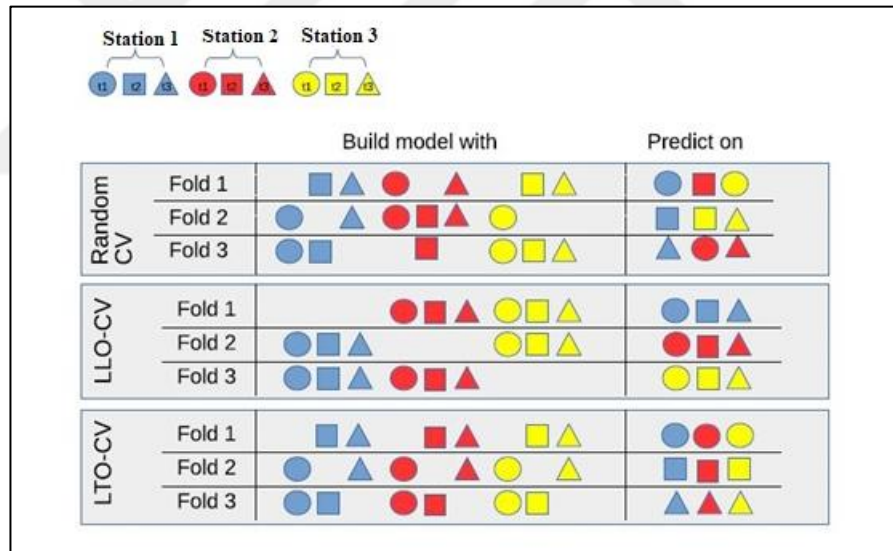


Figure 3.10 Schematic illustration of CV methods used in this study (Meyer et al., 2018)

The linear regression was applied between ground-level and estimated PM₁₀ to control the differences between datasets. Performance of each random forest model was evaluated using the R², RMSE, MAE, slope, and intercept.

Flow chart of the modeling study is given in Figure 3.11.

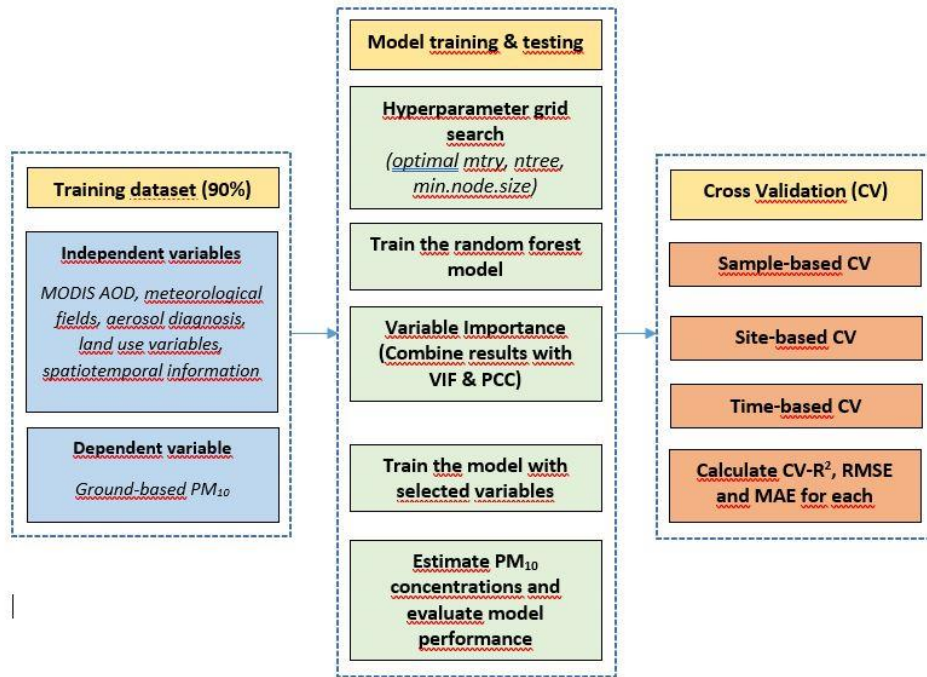


Figure 3.11 Flow chart of modelling study

Considering that many models currently have poor capture capabilities for high values (Wei et al., 2020a), the primary model was used to calculate the R^2 for the high-concentration sets using the 95th and 98th percentiles of the monitoring data. Model performance of the low values was also considered by calculating R^2 for low-concentration sets. Predictions on quantiles with a quantile regression forest were carried out with “quantreg” option in the ranger package. This algorithm estimates the quantiles of the target variable at prediction points (Hengl et al., 2018).

Furthermore, as similar to Loria-Salazar (2018), the quadrant method was used to explore modeling performance considering the relationship between surface PM₁₀ and AOD. Two datasets were divided into four quadrants in this study. These quadrants were determined using a threshold PM₁₀ concentration of 50 $\mu\text{g}/\text{m}^3$ (daily PM₁₀ standard value from EEA) and AOD of 0.2 (AOD of 0.2 was used in Ozdemir et al. (2020)). These values were considered as boundaries to determine clean and polluted periods. Quadrant one (Q-1): PM₁₀ < 50 $\mu\text{g}/\text{m}^3$ and AOD < 0.2 representing low pollution both at the surface and in the column, Quadrant two (Q-2): PM₁₀ > 50 $\mu\text{g}/\text{m}^3$ and AOD < 0.2 representing high pollution at the surface and low in the column,

Quadrant three (Q-3): $PM_{10} > 50 \mu g/m^3$ and $AOD > 0.2$ representing high pollution at the surface and in the column and Quadrant four (Q-4): $PM_{10} < 50 \mu g/m^3$ and $AOD > 0.2$ indicating low pollution at the surface and high in the column. Quadrants are illustrated for each satellite over Turkey in Figure 3.12.

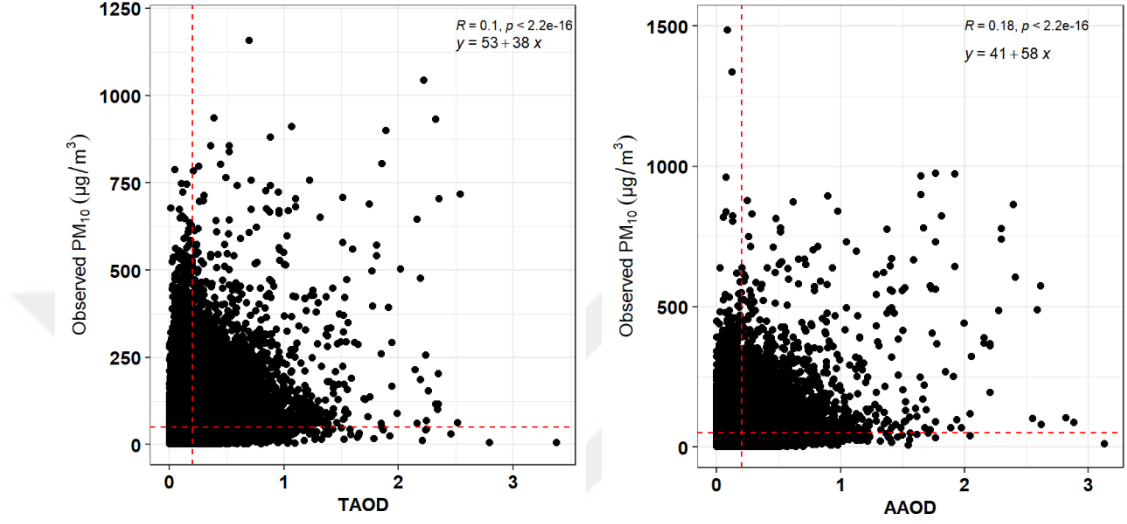


Figure 3.12 Illustration of 4 quadrants between PM_{10} and AOD for each satellite overpass time

CHAPTER FOUR

RESULTS

4.1 Validation of MODIS AOD Products

To investigate the performance of daily MODIS C6.1 AOD retrievals from both Terra and Aqua platforms in Turkey, validation results of the DT, DB, and DTB products with the ground level AERONET AOD measurements between 2008 and 2019 in the IMS-METU-ERDEMLI and TUBITAK_UZAY_Ankara stations are compared in Table 4.1. Results indicated that performances of AODs were not consistent over two sites and showed both regional and seasonal differences. Differences in operation period are the main reason for differences between collocated data numbers. Figure 4.1 also showed the expected error values of each site and product. Site-scale validation indicates the best representative algorithm at each site over Turkey. The accuracy of the MODIS retrievals is affected by two main parameters, including estimation of the surface reflectance and the aerosol model that can be caused underestimation during both polluted and clear conditions (Tian et al., 2018). It is known that the lower NDVI can cause a higher contribution of surface reflectance and decrease the quality of retrievals over the urban area (Che et al., 2019; Huang et al., 2019).

Firstly, differences were found in data numbers according to the satellite over two different AERONET sites. The statistics of the data number indicated that Terra satellite produces a high amount of data over the coastal rural site compared to Aqua except for the DB algorithm. However, the number of data retrieved by the Aqua satellite was generally smaller due to more clouds in the sky in the afternoon over the inland urban site. Differences between the topography and cloud cover properties over the two regions may cause this difference, as reported by Che et al. (2019).

Considerably high differences were found in data numbers between DT and DB retrievals over the coastal rural site due to the DT algorithm only retrieving the ocean pixels. The highest number of collocated data were obtained by DT-3km product for

both sites with the Terra satellite. In contrast, the relatively higher data number (22%) was obtained by the DTB algorithm with the Aqua satellite over the urban site. It was found that high (mainly higher than 0.80) correlations were obtained over the coastal rural site except for DB AOD products. In contrast, the urban site where AOD levels were ~2 times lower than the coastal site showed moderate correlations ranging from 0.58 to 0.78.

Table 4.1 Overall validation results of MODIS AOD products

AERONET site	Product	n	R	RMSE	RMB	mean AOD_{MODIS}	mean AOD_{AERONET}
Rural coastal	Terra-3km-DT	1375	0.88	0.067	1.19	0.25	0.21
	Aqua-3km-DT	1426	0.91	0.062	1.19	0.24	0.20
	Terra-10km-DT	1109	0.86	0.062	1.07	0.23	0.21
	Aqua-10km-DT	1163	0.90	0.053	1.02	0.21	0.20
	Terra-10km-DB	311	0.69	0.036	0.47	0.08	0.16
	Aqua-10km-DB	229	0.77	0.035	0.48	0.07	0.15
	Terra-10km-DTB	977	0.85	0.063	1.05	0.23	0.22
	Aqua-10km-DTB	1096	0.90	0.054	1.01	0.21	0.21
Urban	Terra-3km-DT	145	0.77	0.075	1.54	0.30	0.19
	Aqua-3km-DT	81	0.69	0.08	1.48	0.24	0.16
	Terra-10km-DT	109	0.66	0.079	1.54	0.29	0.19
	Aqua-10km-DT	80	0.78	0.065	1.46	0.24	0.16
	Terra-10km-DB	80	0.59	0.073	0.85	0.15	0.18
	Aqua-10km-DB	88	0.62	0.071	0.95	0.14	0.15
	Terra-10km-DTB	117	0.60	0.089	1.43	0.26	0.18
	Aqua-10km-DTB	99	0.58	0.089	1.26	0.21	0.17

AOD products retrieved by the DT algorithm with 3-km and 10-km spatial resolutions showed poor performance over the urban site with high RMB (~1.46-1.54), indicating a higher overestimation of AOD. Differences between mean AOD values from MODIS and AERONET also indicated overestimating the site (Table 4.1). The correlation of this product was higher than other products, while the percentage of data within EE values did not meet the criteria (67%) as given in Figure 4.1. A significant part of the retrievals (>50%) was higher than the EE values for the DT and DTB AOD products. NDVI was relatively low at this site that is not appropriate for the DT algorithm. Overestimation observed for DT products is mainly due to the highly absorbing aerosols over complex urban areas (Tuna Tuygun et al., 2020). The results

indicated that further improvements are required to estimate satellite based AOD over inland urban areas in Turkey. However, DB AOD performed better over the urban site with low RMB (~ 0.85 - 0.95) and relatively similar mean AOD values with AERONET AOD. It is known that urban sites included in this study can expose a mixture of aerosol (Ozdemir et al., 2020) and current satellite algorithms still have lower accuracy with complex surface characteristics and unclear aerosol types. Neither the Terra nor the Aqua 10km-DT or 3km-DT products achieved a satisfactory performance (at least 67% of the retrievals within the EE). Results suggested that the overestimation existing in the C6 DT over complex surfaces was not effectively reduced in the current C6.1 DT products.

Apparent underestimation of the DB AOD over the rural coastal site with low RMB values (0.47 - 0.48) indicated an inaccurate assumption in the aerosol model. However, the satellite retrievals and ground measurements showed good consistency with the correlations of 0.85 - 0.91 for Terra and Aqua DT and DTB AOD products at the coastal rural site. It indicated the suitability of DT AOD over this site to determine regional aerosol variation. A high data fraction within the EE was found, ranging from 79% to 90.3% for the DT and DTB AOD products (Figure 4.1).

In general, comparing the validation performance of AODs over urban and coastal rural sites, the DB AOD performed better than the DT AODs in the urban site. However, the validation results of DT were obviously outperformed DB in the coastal region with low elevation. For the inland urban site with high elevation, DB performed notably better than DT. The performance of DT, DB, and DTB varies with the land cover types and NDVI over Turkey. However, the Aqua retrievals showed relatively higher percentages of data samples within the EE compared to the Terra retrievals.



Figure 4.1 Performance evaluation with EE over two AERONET sites in Turkey

Detailed sited-based validation was conducted to analyze the impact of the surface type on the quality of AOD retrievals over Turkey due to the diverse location of the AERONET stations. All possible over ocean retrievals ($QAC = 1,2,3$) and the highest-quality ($QAC = 3$) 10-km DT and 3km-DT AOD retrievals at 550 nm over land and ocean from 2008 to 2019 were analyzed for validation over the coastal rural site. All possible and the highest-quality retrievals ($QAC = 3$) of DTB AOD products were also used. The DB AOD products used for validation include all possible retrievals ($QAC = 2,3$) over land. The urban station is an inland station. Thus, there is no ocean pixel over the site. All products include all possible highest quality retrievals ($QAC=3$) over land. The validation results and uncertainties for the matched data were analyzed to evaluate the quality of the MODIS AOD estimates with the EE envelope containing at least

67% of the matchups over land. With the EE values, many AOD collocation pairs (N), R, RMSE, and RMB were shown in Figure 4.2. AODs are generally lower than 0.5 indicating good air quality in the urban site. The dotted and black solid lines in Figures 4.2-4.6 represent the EE envelopes and 1:1 line, respectively. The percentage of data falling within, above, and below the EE are shown by =EE, >EE, and <EE, respectively.

Total available collocated data numbers showed no variation by AOD product except for the Terra DB AOD product. Satellite- and ground-based retrievals agreed moderately ($R = 0.58-0.77$) over the rural coastal site as given in Figure 4.2. However, many DT and DTB AOD retrievals were overestimated ($RMB = 1.26-1.54$), with ~36.4%-62.4% of the matchups falling above the EE envelope. Overestimation is more evident in 10km-DT AOD products. The lowest overestimation (36.4% matchups falling above the EE envelope) was found for Aqua 10km-DTB AOD product, followed by the DB AOD products over the site.

Correlations in DB were relatively low compared to DT, but the other statistical indicators were superior to DT and DTB over the urban site. At the same time, slight overestimation was observed in both DT and DTB AODs, while the DB AODs showed better estimation with the RMB of 0.99. Considering the results, the DB AOD was the best AOD product compared to the DT AOD in the urban site. Furthermore, the percentages within the EE for DB AOD (67.5% for Terra and 68.2% for Aqua) was much more than for DT and DTB AOD. The DB algorithm can follow the actual variation in aerosol amount as measured by AERONET with good performance and slight underestimation with ~22.5% and 13.6% of the matchups falling below the EE envelope for Terra and Aqua, respectively.

Levy et al. (2013) introduced that the retrievals of DT and DB are diverse over different areas due to the assumptions in the algorithms. A merged dataset was developed by Sayer et al. (2014) since each algorithm has its advantages and disadvantages and cannot notably surpass each other. The DTB retrievals showed

similar performance to the DT retrievals, but numerical differences were found at the urban site. The DTB products were superior the DT-10km products with a higher percentage falling within EE (41.9%-54.5% for Terra and Aqua, respectively). The DB algorithm outperforms the DT and DTB algorithms in partially vegetated urban surfaces, suggesting that it is not always best to use DTB products over low aerosol loading urban areas (Wei et al., 2019b). However, it is also known that the DTB AOD retrievals performed well for the European sites located over moderately vegetated surfaces compared to the partially and densely vegetated surfaces in terms of the large percentage within the EE and small RMSE (Bilal et al., 2018).

Zhang et al. (2019) reported a positive error for the DT 3 km product due to incorrect surface reflectance estimation. This effect is more apparent at 3 km than 10 km over the urban site. Most of the bright urban surfaces in the DT retrieval box are removed during the pixel selection process. However, the overestimation of the 10km-DT product was more evident at the urban site. Nevertheless, 36.5% and %47.5 of the 10km-DT retrievals within the EE are still lower than the expected target (67%) (Gupta et al., 2018). A variation of the surface reflectance of 0.01 cause deviation in AOD retrievals (0.1) (Kaufman et al., 1997). The effects of surface reflectance are huge during the extremely low aerosol loadings, resulting in significant bias (Bilal & Nichol, 2017). A reasonable assumption of the surface reflectance relationship between the visible and the shortwave infrared band over the urban areas needs to be updated for Turkey. Lower percentages of collocation below the EE for the DT and DTB products indicated that the MODIS retrievals had ignorable underestimation over the urban surfaces.

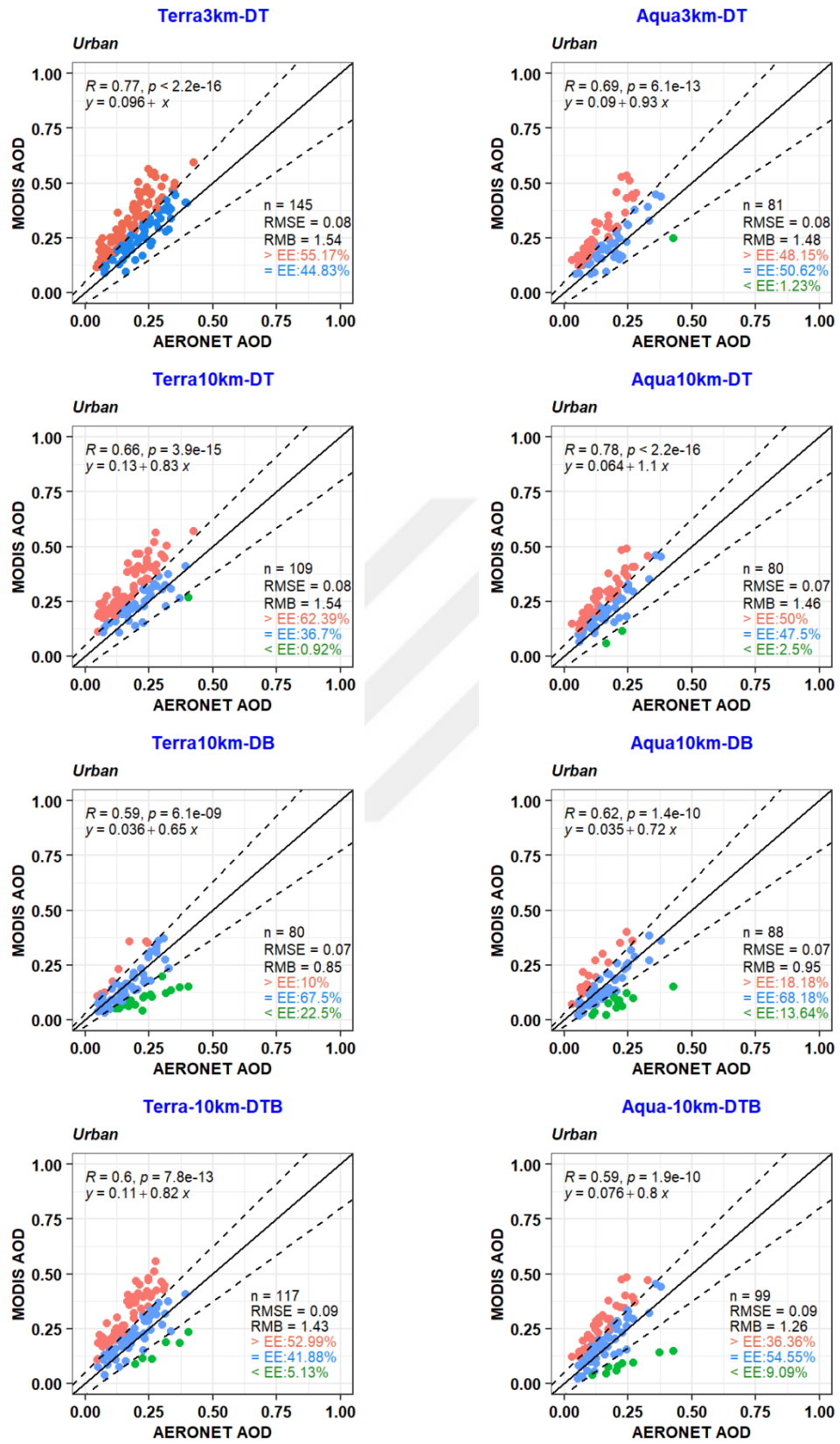


Figure 4.2 Validation results of the MODIS AOD products over the urban site

To investigate the performance of daily MODIS AOD retrievals over land and ocean from both Terra and Aqua platforms, validation results of MODIS AOD products with the ground level AERONET AOD measurements between 2008 and 2019 over the coastal rural site are compared in Figures 4.3-4.6. The validation results and uncertainties for the matched data were analyzed to evaluate the quality of the MODIS AOD estimates with the EE envelope containing at least 67% of the matchups over both land and ocean. The performance of mixed land and ocean MODIS AOD retrievals evaluated with the same EE over land. The EE boundaries illustrated as dotted lines, several AOD collocation pairs (N), R, RMSE, RMB, and 1:1 line as a solid line is given in Figures 4.3-4.6. Terra and Aqua products were individually analyzed due to differences found between two platforms to determine the differences in PM₁₀ estimation performance. Overall, the MODIS land pixels outperformed ocean pixels regarding the relatively higher percentage of matchup pairs within EE. Approximately 10% differences were found between the EE ranges of land and ocean for both satellites. However, both land and ocean algorithms achieved a satisfactory performance of at least 67% of matchups falling within the EE range (Levy et al., 2010). Ocean pixels were slightly tended to overestimate compared to land pixels.

According to the scatterplots in Figures 4.3-4.6, whether land, ocean, or land and ocean, the R were generally high between MODIS AOD and AERONET AOD (ranging between 0.77 and 0.92). The data distribution was concentrated with low RMSE (ranging between 0.05 and 0.07), and consistent RMB values (ranging between 1.0 and 1.19) were found. However, ocean pixels showed higher performance in terms of relatively higher R and lower RMSE values. Using both land and ocean mixed pixels improved the retrieval performance over the coastal site. Thus, ocean pixels over the coastal site were included in this study.

Significantly low percentages of overestimation and underestimation were observed for the DT and DTB products for both land and ocean pixels. The collocated number of the DT products over ocean was approximately 1.5-2 times higher than over land. However, the collocated numbers were relatively similar, or lower data numbers were found over the ocean for the DTB AOD products due to the quality flag filtering

process (only QAC=3). Small percentages of retrievals below EE for the DT and DTB products over land. The DT products mainly overestimate the AOD over the moderately vegetated coastal area. The land surface involving various land types with different surface reflectance could be the main reason. Also, the sources, size, and components of aerosols show great spatial and temporal variation. It is known that the DT errors mainly derive from surface reflectance assumptions and aerosol models used in Look up Tables (LUTs) (Bilal et al., 2017). The general assumption is that surface reflectance affects the retrievals during low aerosol loadings, while the aerosol model affects parameters for the AOD retrievals during high aerosol loadings (Bilal & Nichol, 2017). It can be summarized that the DT retrievals are appropriate for high aerosol loadings and low surface reflectance (Wang et al., 2019). In this study, the DT and DTB algorithms showed more accurate performance and reasonable accuracy with low uncertainty over the moderately vegetated coastal site, as reported by Wei et al. (2018).

Apart from the urban site, the DT algorithm can be considered favorable over the coastal rural site, and the fractions within the EE are all above 80%. The best performance was found with the Aqua 10 km-DT and 10km-DTB products, with the data percentages within the EE achieving 90.2%. The 10 km-DT retrievals had a relatively small number of matching observations than the 3km-DT retrievals. However, all performance metrics were similar found for both DT products. At the same time, the DTB products performed almost similar to the DT products. Overestimation was more stable in the 10km-DT and DTB products with optimum RMB values (~ 1) than the 3km-DT products. However, the DB retrievals have only a slightly lower R (0.7-0.77) value relative to DT (R ~ 0.8 -0.9). The other criteria were also predominant to DB, mainly with the RMB values indicating near 1.0, and the percentage of retrievals within EE, which fulfilled 67%.

Moreover, underestimated AOD values ($< EE \sim 62$ -67%) were found with the DB AODs with the data percentages within the EE. It was also low ($\sim 35\%$) at the coastal rural site, which is located moderately vegetated with low elevation area. In conclusion, problems may be occurred in the aerosol model selection or development

of surface reflectance over these areas (Bilal et al., 2016), resulting in common bias and inferior performance in the DB products. The underestimation (RMB~0.48) can be seen in Figure 4.6.

Results indicated that the DB AODs could be used over partially vegetated and low aerosol loading areas instead of the operational merged DTB AOD product. The DT algorithm sometimes does not perform with reasonable accuracy over these surfaces. The DT AODs in the DTB products can cause low accurate AOD and increase the merged product's uncertainty. Wei et al. (2019b) also indicated that the DTB products have almost a similar performance except for bright surfaces with the DT products over most areas worldwide (Wei et al., 2019b). Monthly NDVIs are mainly higher than 0.3 in autumn and summer. In most areas, similar to the surface of the coastal rural site in Turkey could be the main reason. The DTB AOD retrievals over this region primarily include the DT values due to the relatively high vegetated surface type.

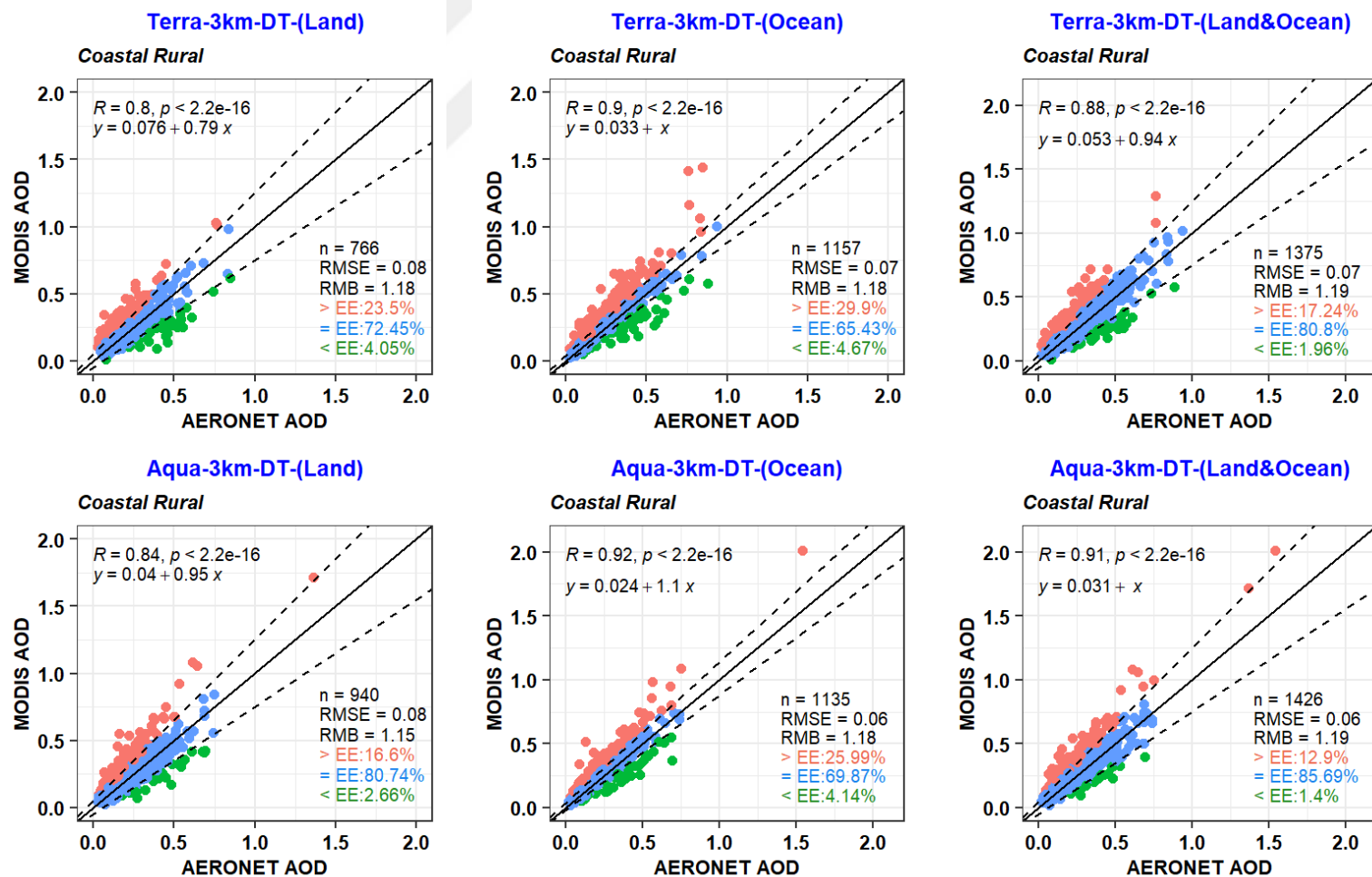


Figure 4.3 Validation of 3km-DT AOD over the coastal rural site

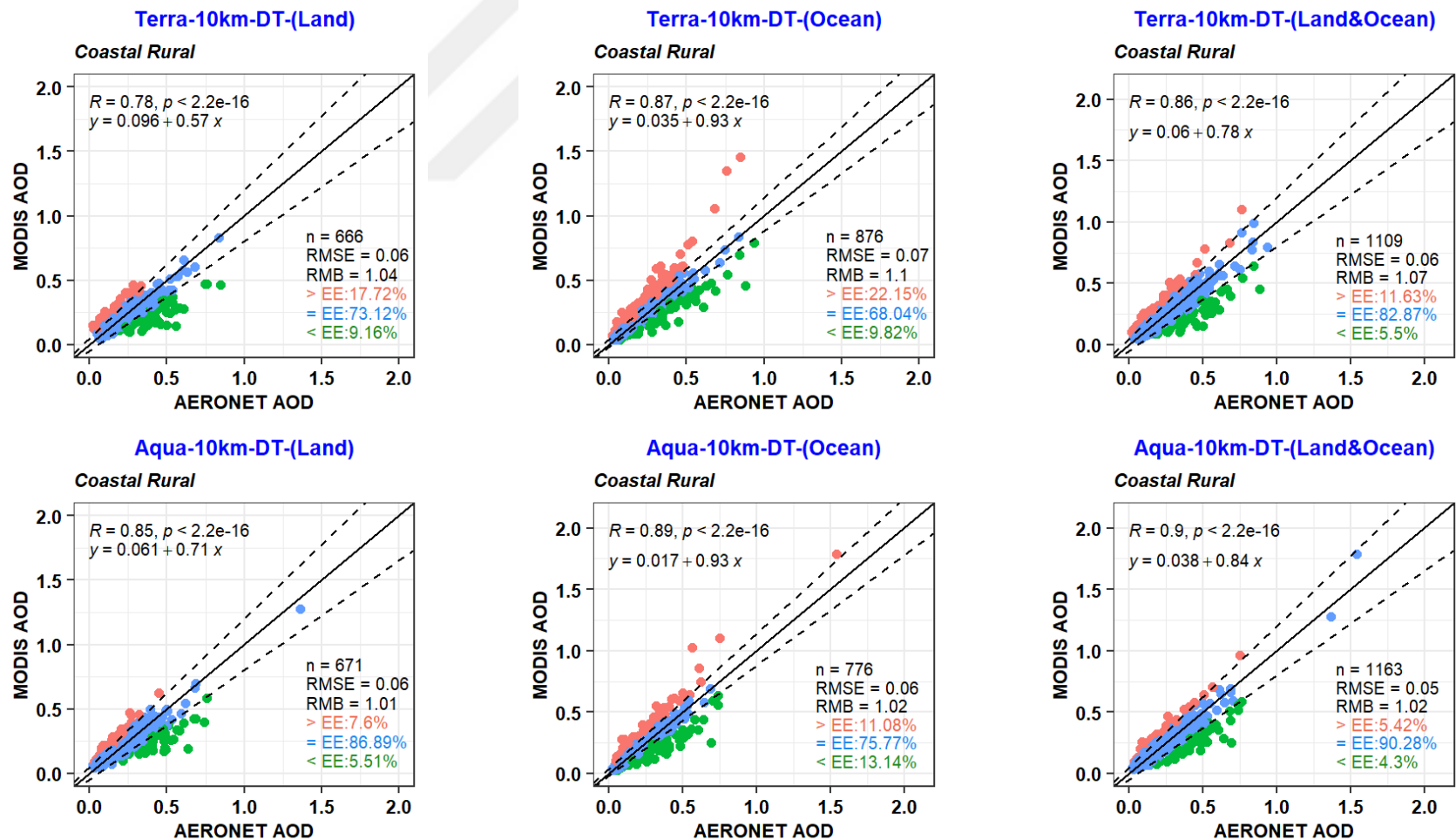


Figure 4.4 Validation results of 10km-DT AOD over the coastal rural site

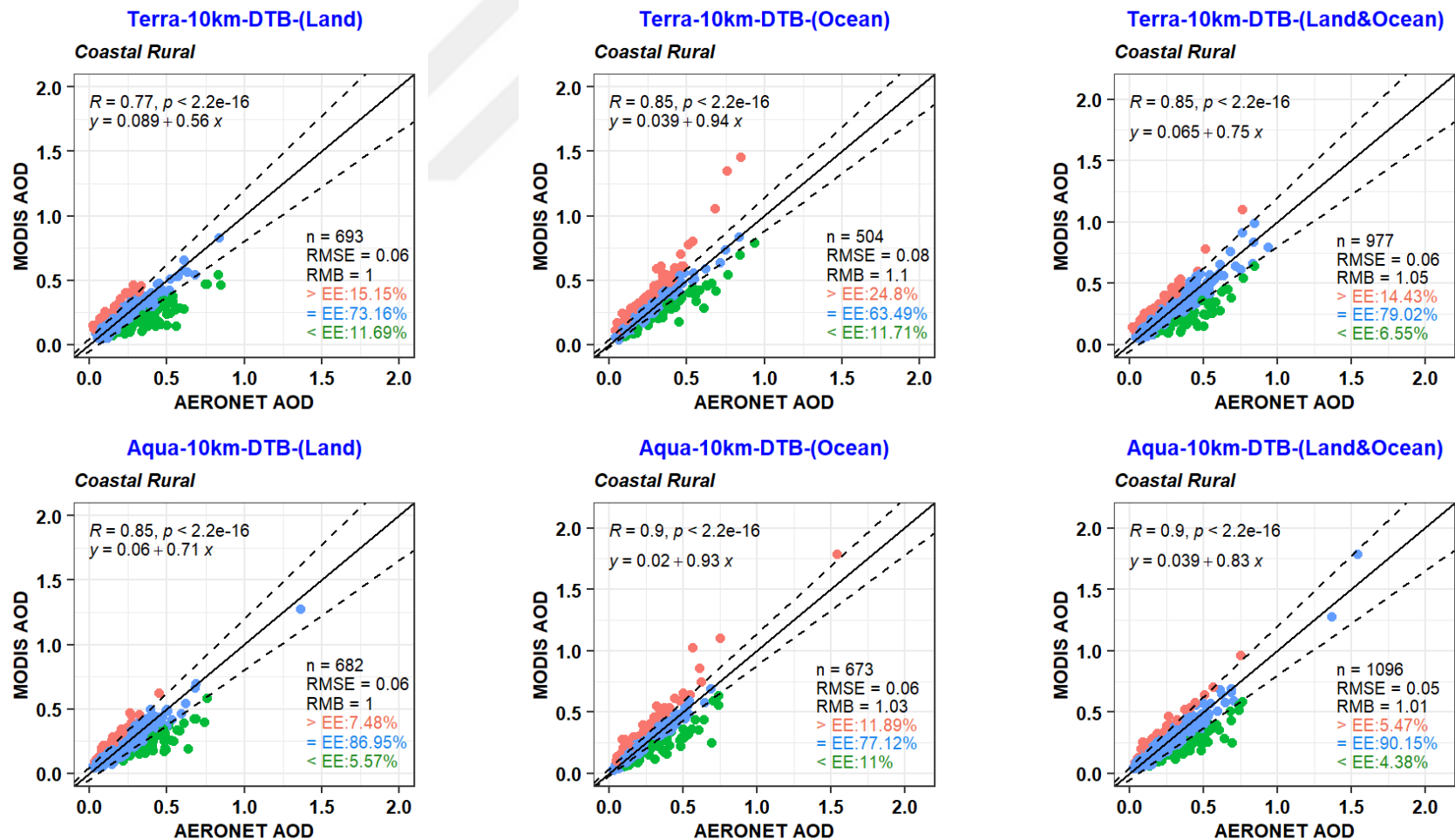


Figure 4.5 Validation results of 10km-DTB AOD over the coastal rural site

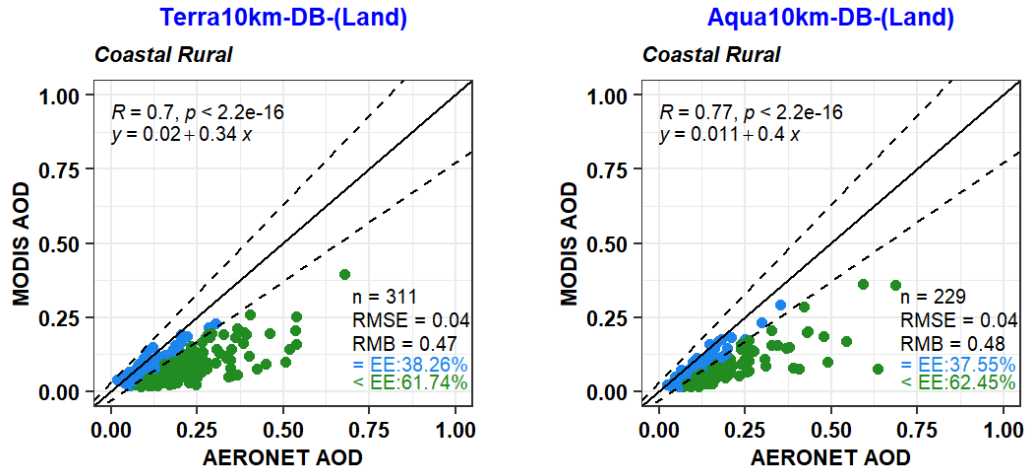


Figure 4.6 Validation results of 10km-DB AOD over the coastal rural site

4.1.2 Seasonal Validation Results

Different numbers of collocated observations for each month and season can also cause different temporal trends and performances over the sites. It should be noted that the relatively few collocations were also available in winter (only $n = 4-5$ over 4 years) due to cloud cover and short operating time in the urban site. There was no evaluation for the winter season. Figures 4.7 and 4.8 show seasonal scatter plots of the AODs obtained from the MOD/MYD04 retrievals against the AERONET observations. The colored solid lines represent seasonal regression lines with a 95% confidence interval.

As illustrated in Figure 4.7, the seasonal performance of DT and DTB AODs showed similar performance with R^2 ranging from 0.83 to 0.91, with a relatively higher R^2 in cold seasons (autumn and winter) and a slightly lower R^2 in warm seasons (spring and summer). The number of collocated data was highest in summer and lowest in winter. Although the limited number of matchups was available in winter due to the problem in the DT retrieval algorithm over cloudy, snowy regions or ice surfaces, the DT AOD from the Aqua satellite in winter showed good performance with a higher ratio of within EE matchups ($\sim 90\%$), and relatively low overestimation ($RMB \sim 1$). Since pixels with high reflectance were discarded by the rigorous mask algorithm in winter, fewer pixels decreased the uncertainty in the surface reflectance scheme remaining from a bright target. Therefore, the DT and DTB AODs performed better

regarding the data percentage within the EE, RMSE, and RMB in winter as reported in the literature (Li et al., 2005). Since the vegetation shows seasonal changes due to its regular cycle and grows up in the warm season, with reasonable data point falling within EE, summer and autumn performed slightly better than spring. In spring, overestimation was more evident with the data percentages above EE (~20%) for the DT and DTB AOD products. Data percentages falling within EE were higher than 80%, mainly in summer with high R and lower RMB values (~0.92-0.95). Overestimation was more evident in the 3 km-DT products for all seasons. Overestimation at high AOD levels was observed mainly in spring. Marine aerosols mixed with fine anthropogenic aerosols in large proportions were the dominant type in spring over the coastal rural site. Spring and summer are also characterized by differences in aerosol types and larger AOD than the other seasons (Ozdemir et al., 2020). Intense dust events were mainly detected in spring. These natural processes lead to more substantial diurnal AOD variation causing differences between the Terra and Aqua retrieval performances. Figure 4.8 also indicated that overestimation or underestimation changed by the seasons and algorithms. Aqua showed a higher data percentage within EE for all seasons compared to the Terra satellite.

The seasonal performance of the DB products had the R values ranging from 0.41 to 0.90, with a relatively higher R in autumn and spring and a lower R in summer and winter. Lower correlations (~0.41-0.61) and RMB (0.36-0.38) values are indicating an apparent underestimation of AOD with the higher percentage of data below EE (~65%). However, the performance of the AOD was lowest in summer. Underestimation was more significant than winter, with a few percentages of data falling within EE (~13%). The highest percentages of data were below EE (~86%) as seen in Figure 4.8. Best validation results were obtained in the spring season with higher R (0.81-0.90 for Terra and Aqua), higher RMB (~0.6), and data percentage of falling EE (~53% and 45% for Terra and Aqua, respectively).

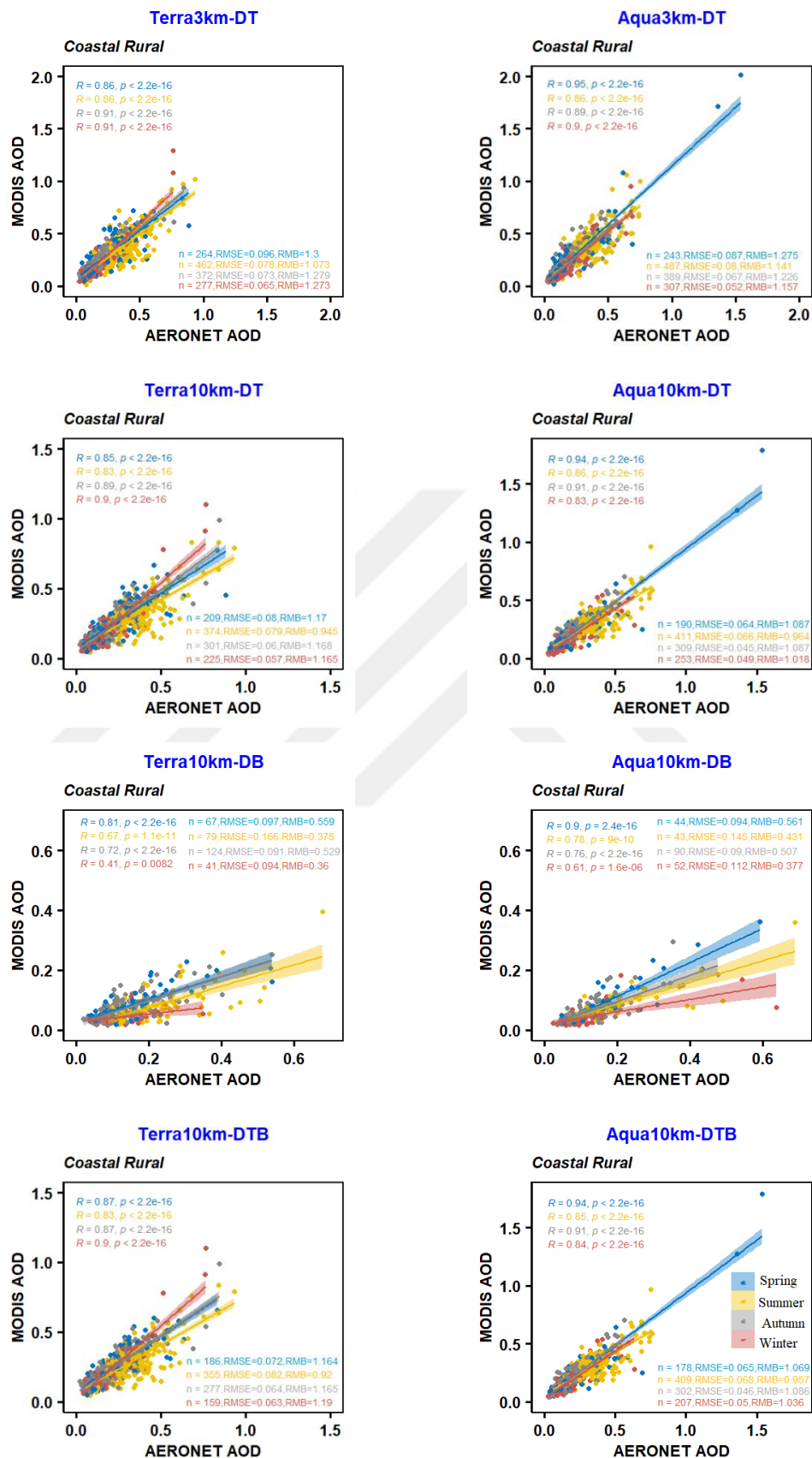


Figure 4.7 Seasonal validation results of MODIS AOD products over the coastal rural site

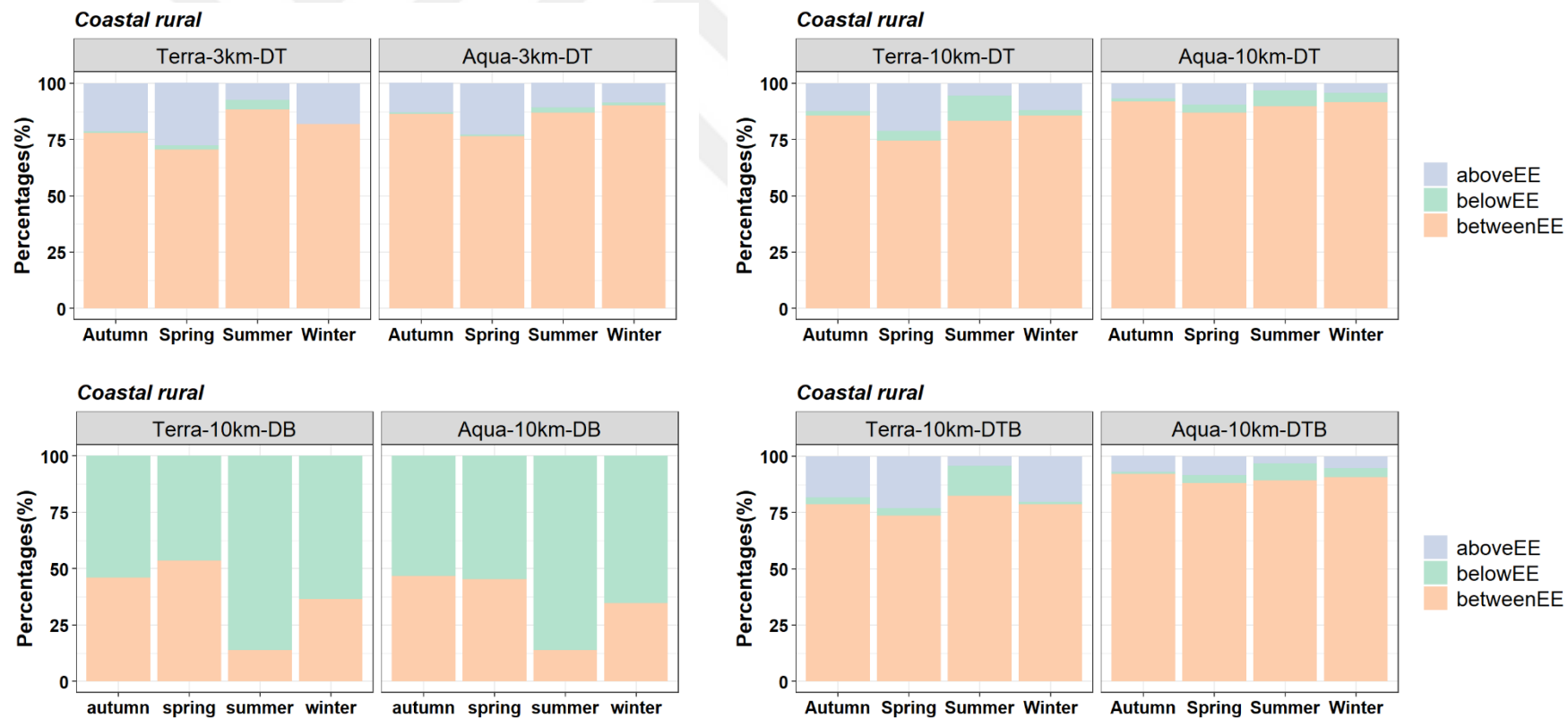


Figure 4.8 Seasonal EE values of MODIS AOD products over the coastal rural site

Figure 4.9 shows the scatter plots of the AODs from MxD04 and against the AODs from the AERONET network. Figure 4.10 shows the seasonal EE percentages at the urban site in spring, summer, and autumn. Large estimation errors and overestimations were observed at the urban site for the DT and DTB products. However, the DB AOD products fulfilled the EE requirements with higher percentages of data falling within EE ranging from 74% and 86% except autumn. In autumn overestimation and underestimation equally affect the performance. Maximum overestimation was observed in summer for the MOD04/MYD04 DT and DTB AOD products with higher RMB and data percentage above EE (70%-80%), while it was minimum in spring over the urban site. The errors in the estimated aerosol scheme and surface reflectance during high and low aerosol loadings may cause this situation. Previous studies reported similar results over the similar regions (Tian et al., 2018; Rupakheti et al., 2019; Wei et al., 2019b). Xia et al. (2013) also indicated that the MODIS tended to overestimate AOD in summer. High AOD values with large deviations were observed in summer over the urban site due to the variations of aerosol sources over the region (Ozdemir et al., 2020). While AOD was lowest, vegetation was highest in spring over the urban site. This can be attributed to considerable uncertainty and instability of aerosol models for retrieving AOD used in LUTs (Huang et al., 2019). Relatively higher correlations were found in summer and slightly lower correlations were found in autumn for all products. Overestimation was less significant for the DT and DTB products in spring with slightly lower RMB values (~ 1.2 - 1.3). However, the DT and DTB products fulfilled the requirements of EE with the high percentage of collocations falling within EE ($>70\%$) just for spring. The reason for the seasonal variation of AOD characteristics might be due to the differences in surface reflectance, aerosol type, and local climate (Zhang et al., 2019). Data from both satellites (MOD04 and MYD04) had similar seasonal variations with obvious seasonal and regional features.

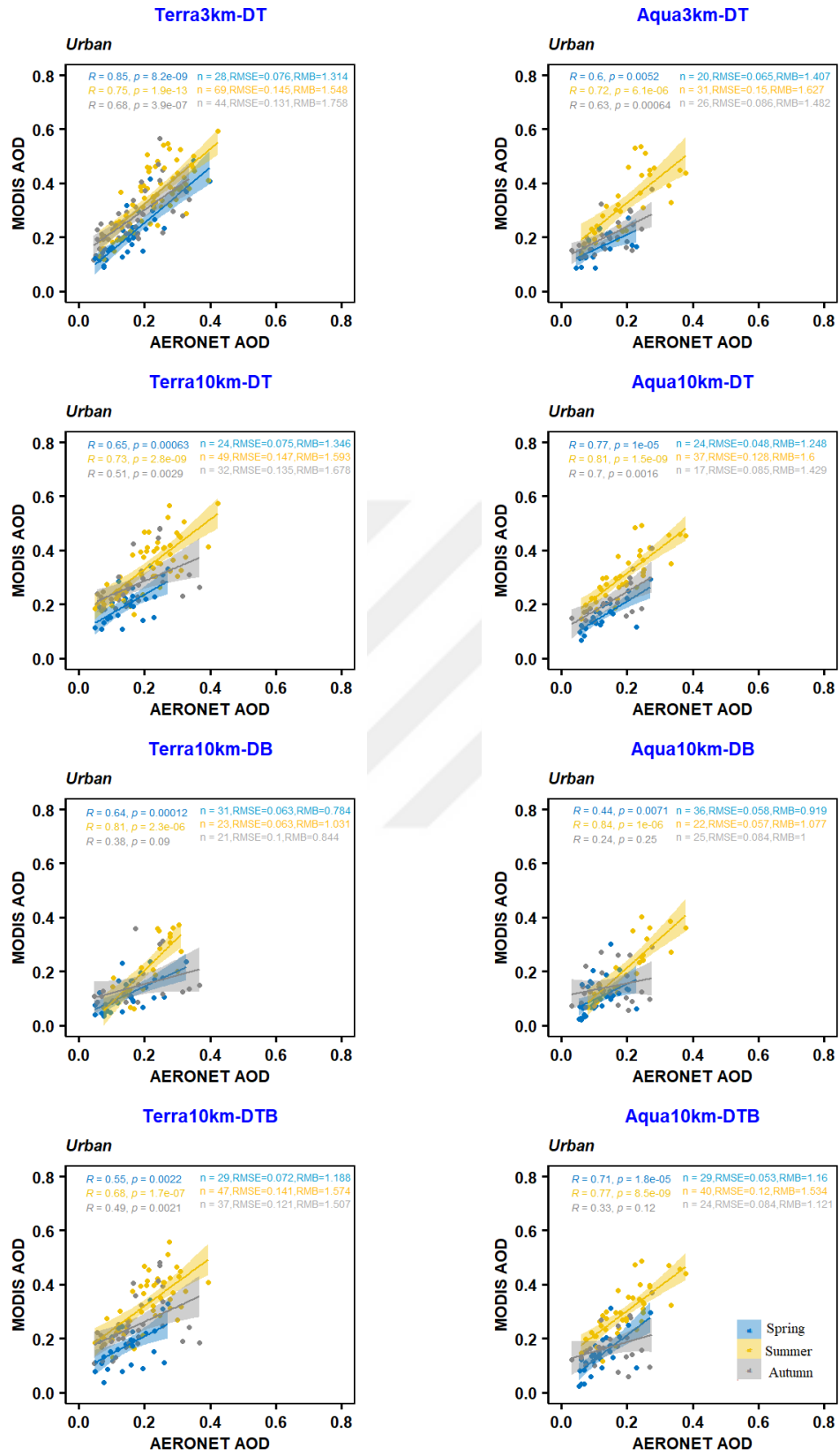


Figure 4.9 Seasonal validation results of MODIS AOD products over the urban site

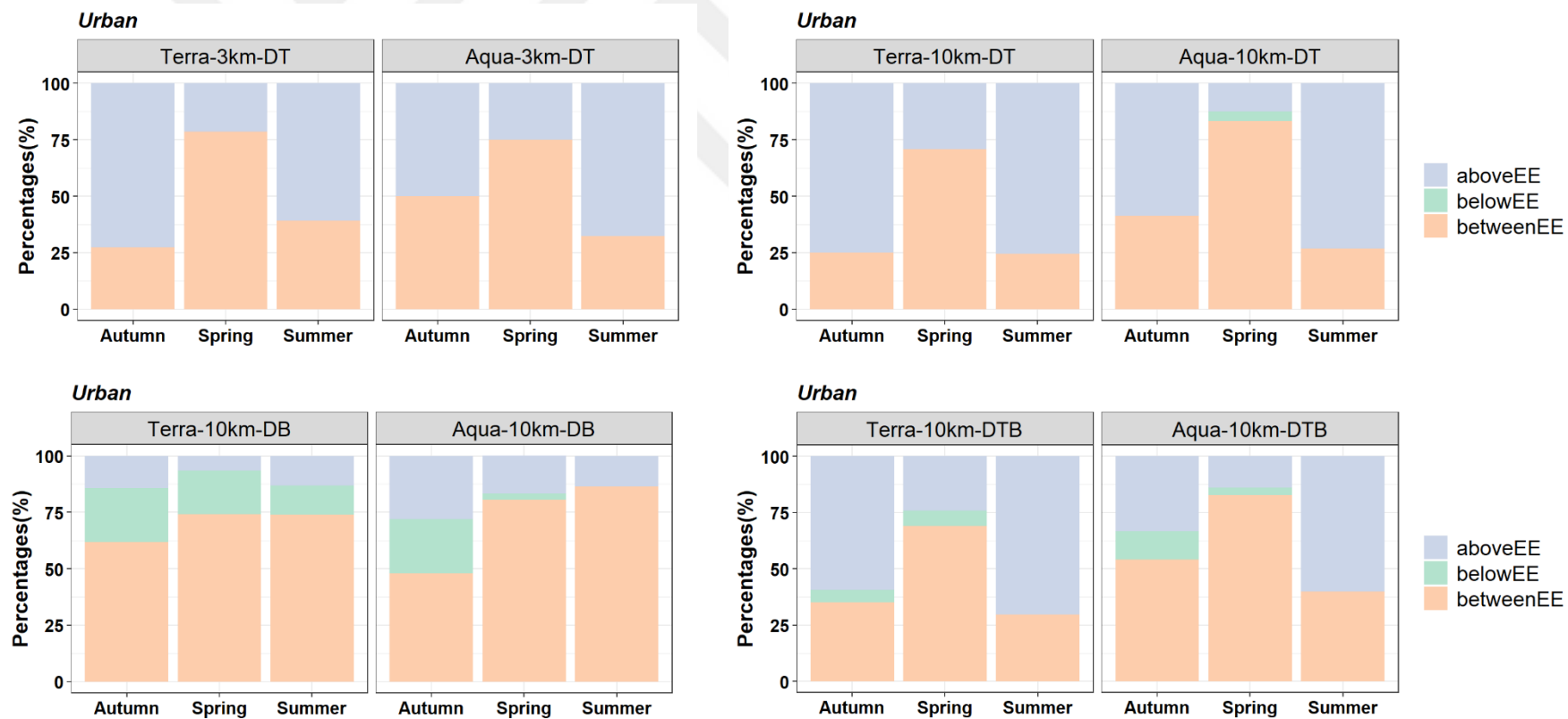


Figure 4.10 Seasonal EE values of MODIS AOD products over the urban site

4.2 Spatiotemporal Variation of AOD over Turkey

In this section, temporal, and spatial variation of multiannual and seasonal average AODs are represented in different regions covering entire Turkey based on MODIS different algorithms of Terra and Aqua satellites over land and ocean. Data were analyzed for the 12-years long-term period from 2008 to 2019. Note that the number of days used to calculate AOD averages varies considerably for each region and AOD product. Thus, the average DT AOD in the low vegetated surfaces and DB AOD in the vegetated surfaces can provide limited representation. Aerosol loadings are not also spatially and temporally homogeneous over land. Over land, aerosol datasets retrieved by DT and DTB products showed similar spatial distributions with low aerosol levels were widely distributed over the Black Sea region. It is known that surface characteristic is the most critical parameter to analyze AOD, and the multiannual spatial distribution map of NDVI on geographic regions is illustrated in Figure 4.11. Spatial variation of NDVI showed substantial geographical distribution decreasing from the west to the east of the country. Considering the mean NDVI values, the Marmara and Black Sea regions were identified as the highly vegetated surfaces with NDVI values larger than 0.5. The regions of Aegean, South-eastern, and Mediterranean have moderately vegetated surfaces (NDVI~0.3-0.5), while the regions of Central Anatolia and Eastern Anatolia consist mainly of partially vegetated areas (NDVI~0.3-0.5). The standard deviation of NDVI was higher in the Black Sea and Mediterranean regions (0.14-0.16), while low variations in NDVI (~0.07-0.09) were observed in arid regions (Inner, Eastern, and South-eastern Anatolia regions).

Figure 4.12 indicates the spatial patterns of AOD from DT, DB, and the merged DTB datasets. The spatial patterns of the difference between Terra and Aqua satellites are also given in each product (DT, DB, and DTB). AOD is less than 0.1 represents the crystal-clear sky with maximum visibility, while larger values than 1 represent very hazy conditions with minimum visibility (NASA, 2021a). AOD values in the range of 0.2-0.3 indicate moderate aerosol pollution in the atmosphere. An AOD value larger than 0.3 indicates dust transport or local source continental aerosols characterized by large AOD values in the literature. Particularly, the AOD values above 0.5 indicate

very intense aerosol pollution over a region (Georgoulas et al., 2016; Glantz et al., 2019; Wei et al., 2019b; Filonchyk et al., 2020).

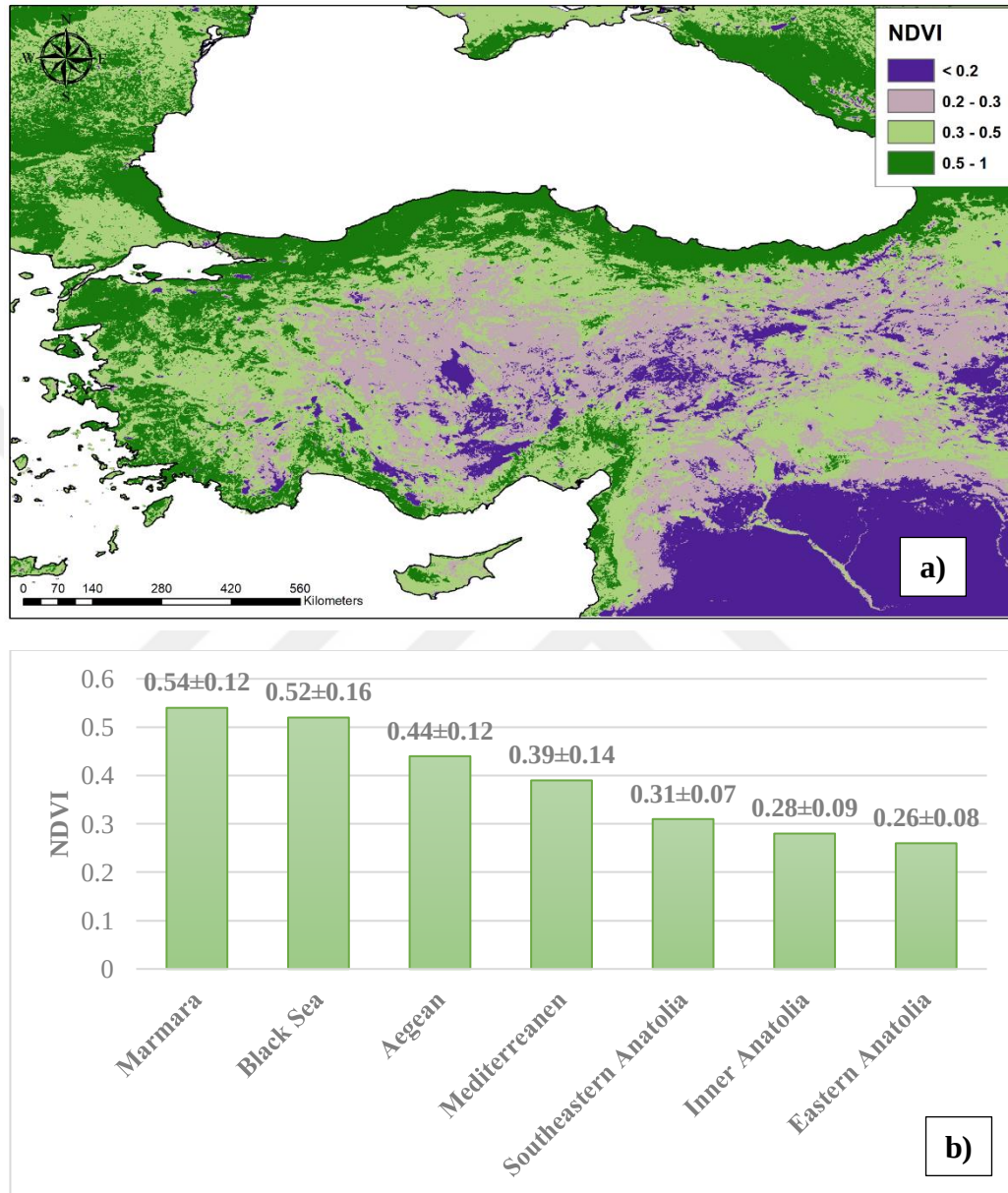


Figure 4.11 Multiannual average (a) and regional (b) variation of NDVI over Turkey

The Terra AOD products are relatively higher than the Aqua AOD products. However, minor differences were found between Terra and Aqua AODs on their spatial distributions. The average AOD generally remains stable over the regions and ranges from 0.10 to 0.20, mainly with the 10-km DT and DTB products over land.

There are some differences over the region regarding different aerosol loadings and surface characteristics. By contrast, the range of AOD values over land is more extensive compared to AODs over ocean. Significant differences were observed in the averages of AOD values over land and ocean. The AOD over land was much higher than over ocean, especially in the western and north-western regions of Turkey. This is mainly due to the urban and industrial aerosols that are existing over land. Compared to the land surfaces, the distribution of AOD is more stable with much lower spatial variability over ocean surfaces. It indicates the clear sky atmosphere (~ 0.1), mainly over the Black Sea and the Aegean Sea, far away from human activities' influence. Meanwhile, dust aerosols which are the most prominent particle constituent by mass, and emissions from natural sources are dominant aerosol types over the southern part of Turkey, cause higher aerosol loading over the Mediterranean Sea (0.2-0.3). Relatively high AODs were also found over the Marmara Sea due to the heavy maritime traffic over the region.

The spatial distribution of AOD over Turkey is defined by different AOD levels, including low, moderate, and high values. AOD values larger than 0.3 were identified over the inner, western, north-western, and southeastern parts of Turkey. The main reason is the most densely populated and industrialized regions located in the north-western and western regions of Turkey with low elevation and a high number of anthropogenic aerosols from local sources. However, large AOD values, mainly over the southern part of Turkey, were explained by high concentrations from dust transport over arid regions and local sources (Ozdemir et al., 2020). The regions with the smallest AOD values up to 0.1 were in the densely vegetated northern parts and moderately vegetated regions (NDVI ~ 0.3 -0.5), as seen in Figure 4.12. Global validation of the MODIS DT AOD products against AERONET AOD was presented by Levy et al. (2010). They indicated that ignorable bias and good correlation were found for AOD lower than 0.15 over the stations located in moderately vegetated areas.

There were differences between the 10 km-DT and 3 km-DT products over land, while no differences were found over ocean surfaces. Differences between the 10 km-DT and 3km-DT datasets over land were more evident in the regions that have low

vegetated urban surfaces. Higher AODs (> 0.3) were observed in the inner region of Turkey, particularly around the Tuz Lake. Occasionally some higher AODs (0.2-0.3) were observed over southern and eastern parts. Large AOD values are mainly observed over mainly industrialized and urbanized areas, including the western part of Marmara (north-western), Aegean (western), and Central Anatolia (inner part) regions. However, low AOD values (~ 0.1 -0.2) were widely distributed throughout the rest of the areas. This indicated that the increasing human activities caused more aerosol emissions over these regions, and the DT algorithm also overestimated the surface reflectance over urban areas. The DT product showed many missing values in the most arid areas over the Middle East located southeastern part of Turkey due to the limitations of the DT algorithm. Most of the DB AOD values are less than DT and DTB AOD values, and the relationship between DB and DT varies with the NDVI and land cover types. Meanwhile, the variation during high aerosol loading conditions was more significant than that during low aerosol loading conditions. High aerosol loadings in the DT and DB retrievals occur in the southern part of Tuz Lake values (mainly ranging from 0.3 to 0.4 and up to 1 in a few pixels), with a similar spatial structure for all products. The dissimilarity in selection of the aerosol model and the surface reflectance assumptions notably causes the variations in retrieval performance of both aerosol models in LUTs (Sayer et al., 2014). It is evident that most of the DB AODs in the eastern part were more extensive than those of DT AODs. The differences between the Terra and Aqua AOD datasets over Turkey are similar to the findings of Georgoulas et al. (2016).

Overall, significant differences (~ 0.1 -0.2) were found between the DT and DB AODs. Notably positive differences were found over the whole of Turkey except for the DT-algorithm-limited over bright surfaces. The DT AODs were much higher than DB AODs, mainly over the Inner and eastern Anatolia regions (~ 0.2). Substantial differences were not observed between the DTB and DT AODs over most regions, indicating that primarily the DT AODs were selected over moderately and densely vegetated surfaces in DTB AODs. Differences were found in mainly partially vegetated arid areas. The DTB algorithm primarily selects DB retrievals over these areas since the DT algorithm is not applicable. However, substantial differences,

particularly over semi-arid with high elevation areas, were found, such as in the eastern part of Turkey. The restrictions of each algorithm over different surface conditions likely responsible for the apparent differences in AOD from the three algorithms. The selection of appropriate AOD products is hard to study for Turkey's whole territory with the various spatial distribution of mountainous, coastal, industrial, urban, suburban, and arid regions resulting in various meteorological conditions affecting aerosols loading in each region.

According to the validation results, the DTB algorithm showed better performance with reasonable expected errors for both satellites (>85%) over coastal rural sites with moderately vegetated surfaces. At the same time, it was not a superior DB AOD product over the urban site. It is shown that the validation results vary commonly with the change of land cover and aerosol types (fine or coarse particles) among different regions over Turkey. AERONET sites are limited in Turkey with only two geographical regions, and the urban station represents the conditions only in the inner part of Turkey. This site is located in a high elevated area, and it is known that AOD retrieval algorithms have limitations to retrieve AOD over the high elevated regions. However, Turkey also has urban locations located in coastal and low elevated parts, and land cover and NDVI conditions can vary according to the regions. Thus, it is hard to select one AOD algorithm to developed national models and estimate PM_{10} concentrations. According to the results of this study, it is recommended to use DTB products. Because it is selected which AOD algorithm will be appropriate for a region based on NDVI value, considering all diversities and limitations over the region and validation results. Also, the DTB product provides better spatial coverage of AOD overland (i.e., to include pixels from DT (DB) for areas where DB (DT) fails to retrieve). Therefore, the merged DTB products were selected for the PM_{10} estimation study over Turkey. In summary, the difference in AOD products between Terra and Aqua in these AERONET stations was small. However, the data number and accuracy of each satellite showed regional-specific results. Therefore, both satellites were used for the model development stage to identify the differences in model performances.

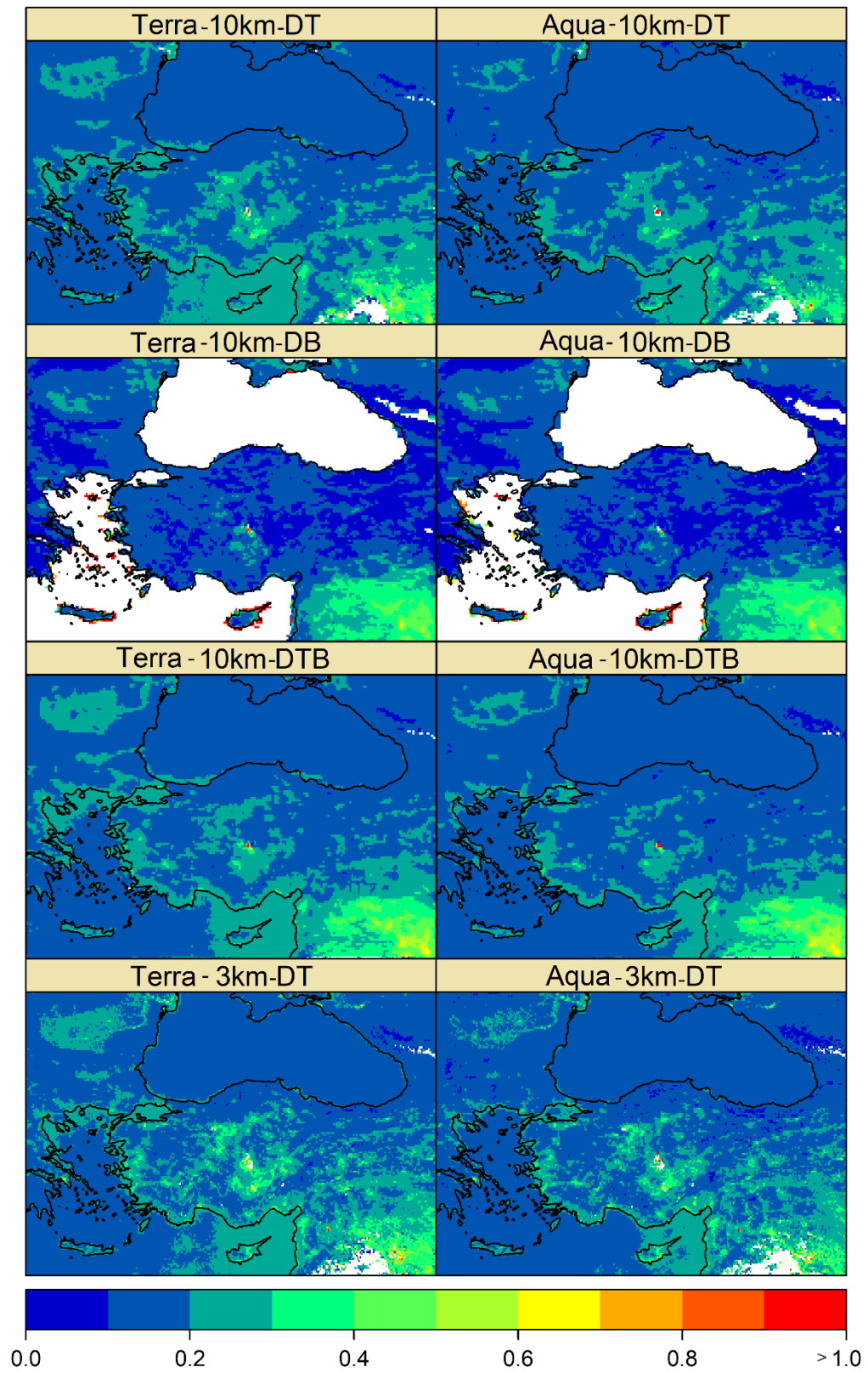


Figure 4.12 Spatial distribution of multiannual average AOD products between 2008 and 2019 in Turkey

Spatial distribution of the multiannual and seasonal average AOD values at the sub-regions of Turkey are given in Figures 4.13 and 4.14. According to Table 4.2, mean AOD values indicated similar AOD levels over seven geographical regions in the range of 0.154-0.202. The Terra DTB AODs were higher than the Aqua DTB AODs in all regions. Clearly, AOD showed low spatial differences over the regions. However, the Black Sea and the eastern Anatolia regions showed relatively lower AODs as 0.154 and 0.180, respectively. The AOD values higher than 0.2 were particularly observed over the regions of Mediterranean, Aegean, and Southeastern Anatolia. Standard deviations were also similar, while maximum variation was observed in the Central Anatolia region. This is probably due to the maximum values (up to 2) over the region due to pixels around the Tuz Lake. The lowest values of multiannual average AOD were observed on the eastern part, including sparsely populated areas.

Table 4.2 Regional statistics of DTB AOD over Turkey

Region	Terra-10km-DTB				Aqua-10km-DTB			
	Min	Max	Mean	Std	Min	Max	Mean	Std
Aegean	0.114	0.361	0.202	0.040	0.105	0.355	0.188	0.045
Marmara	0.125	0.424	0.191	0.036	0.116	0.335	0.175	0.035
Black Sea	0.097	0.306	0.154	0.030	0.092	0.278	0.139	0.027
Central Anatolia	0.092	1.955	0.190	0.062	0.076	1.000	0.173	0.115
Southeastern Anatolia	0.093	0.300	0.206	0.032	0.099	0.278	0.190	0.030
Eastern Anatolia	0.069	0.282	0.180	0.029	0.058	0.276	0.161	0.029
Mediterranean	0.100	0.375	0.200	0.044	0.085	0.360	0.191	0.047

Figure 4.13 shows seasonal mean AOD from 2008 to 2019 in Turkey. The spatial distribution patterns of seasonal average AOD in the seven geographical regions are also shown in Figure 4.14. The pixels with white background represent the regions that have no data. A long-term absence of the DTB retrievals can be observed in the eastern

part of Turkey, where the surface is mainly bright in winter. DB AODs in DTB AOD do not cover the regions covered with snow most of the time, particularly in the eastern part. Meanwhile, lower AOD coverage was found in the regions with high elevation such as eastern part of Turkey.

The distribution of seasonal AOD was also analyzed to determine the temporal variation of AOD in Turkey. According to Figure 4.13, during the entire period of observations over Turkey, AOD showed clear seasonal variations in all regions depending on the variation of NDVI and meteorological factors. Multiannual seasonal average AOD values over Turkey changed during four seasons (summer, spring, autumn, and winter). Significant differences (~ 0.1) were found in spring and summer between the Terra and Aqua satellites over ocean surfaces, mainly on the Black Sea. Differences were also identified over the Mediterranean Sea in autumn between Terra and Aqua. The highest AOD values over the region of the Black Sea are observed during spring, which has been attributed to solitary Saharan dust transport and infrequent agricultural fires (Floutsi et al., 2016).

The spatial variation of seasonal average AOD was similar to the spatial variation of multiannual average AOD in Turkey. It is known that emissions of anthropogenic and natural sources and meteorological parameters affecting AOD variation such as humidity (hygroscopic growth) and precipitation (effect on particle settling) vary in the country. Multiannual seasonal variations of specific humidity and total precipitation as the most effective meteorological parameters are given in Figure 4.14.

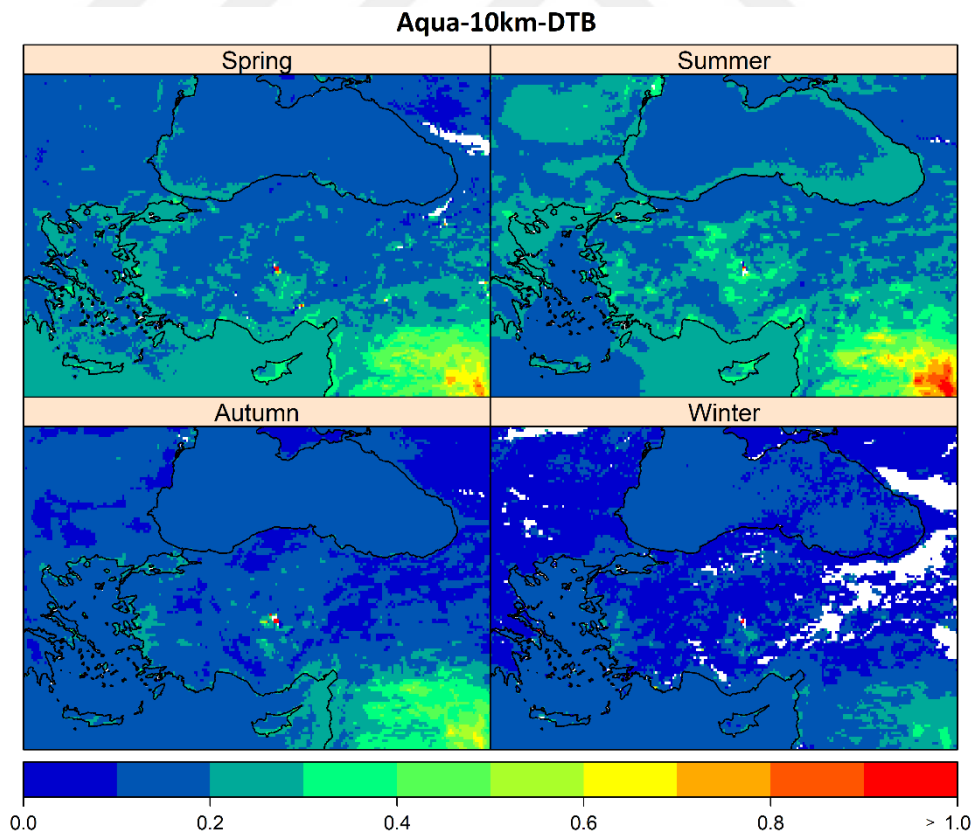
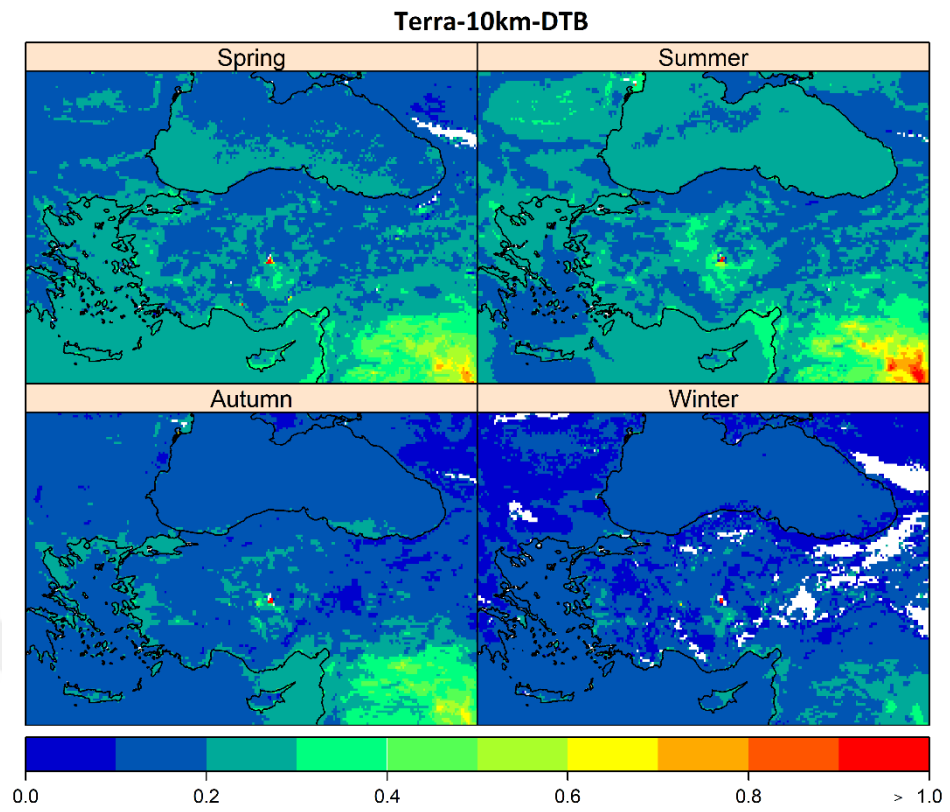


Figure 4.13 Multiannual seasonal distribution of DTB AOD over Turkey

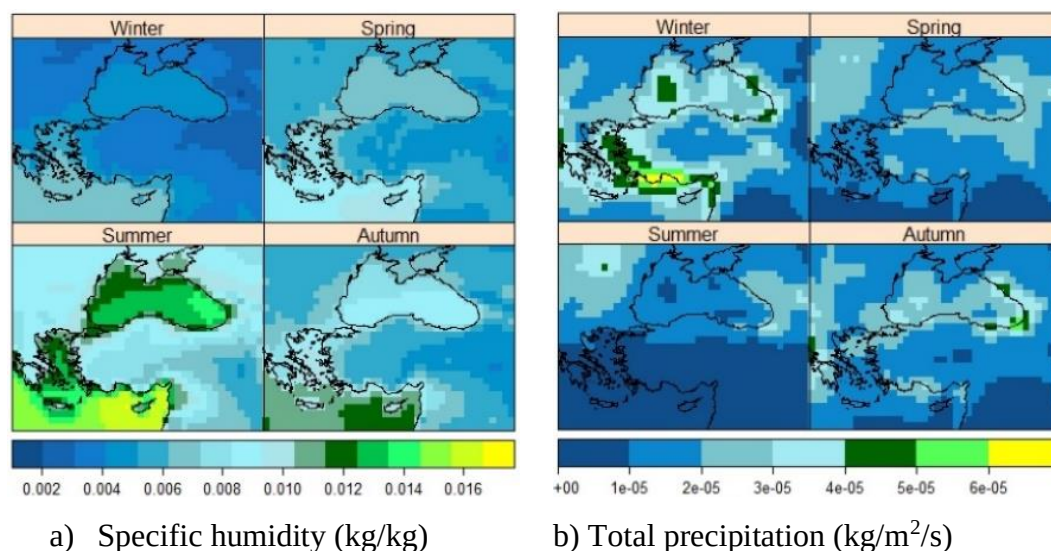


Figure 4.14 Spatial distribution of multiannual seasonal average a) specific humidity and b) total precipitation

Figure 4.14 shows that summer and spring were the seasons with large AOD values in Turkey, probably due to dust loadings from deserts of arid regions around Turkey. Many biomass-burning aerosols transported from natural surfaces in spring and summer periods also cause large AOD values (Ozdemir et al., 2020). Another reason for high AOD values in summer is higher humidity increasing atmospheric aerosols capacity with hygroscopic growth (Tutsak & Koçak, 2020). Higher AOD (up to 0.3) over ocean surfaces is also due to the humidity effect. All regions of the country had low AODs in winter due to higher precipitation and low humidity (Figure 4.14). NDVI also showed seasonal variation. Therefore, the multiannual data was divided into four seasons to analyze the spatial distribution of NDVI better understanding the relationship with aerosol sources. Figure 4.15 shows the spatial distribution of multiannual (2008 - 2019) seasonal mean NDVI in different seasons over Turkey. The NDVI in winter is lower than in other seasons (spring, summer, and autumn), and the lowest values appear in the eastern part of Turkey. The NDVI in summer is most extensive, and the value is up to 0.5-0.6 in the Marmara and the Black Sea regions. Figure 4.15 indicates that NDVI in spring and summer is similar except for southeastern Anatolia.

According to seasonal variation of aerosols, the highest AODs were observed in spring and summer whereas the lowest AODs were observed in autumn or winter in

most regions of Turkey. Almost the same average values in nearly all regions were available in the seasons of spring and summer, while a few differences were observed in the Inner and Southeastern Anatolia regions. Depending on the diverse aerosol sources and surface properties, significant seasonal AOD variation was found in seven geographical regions.

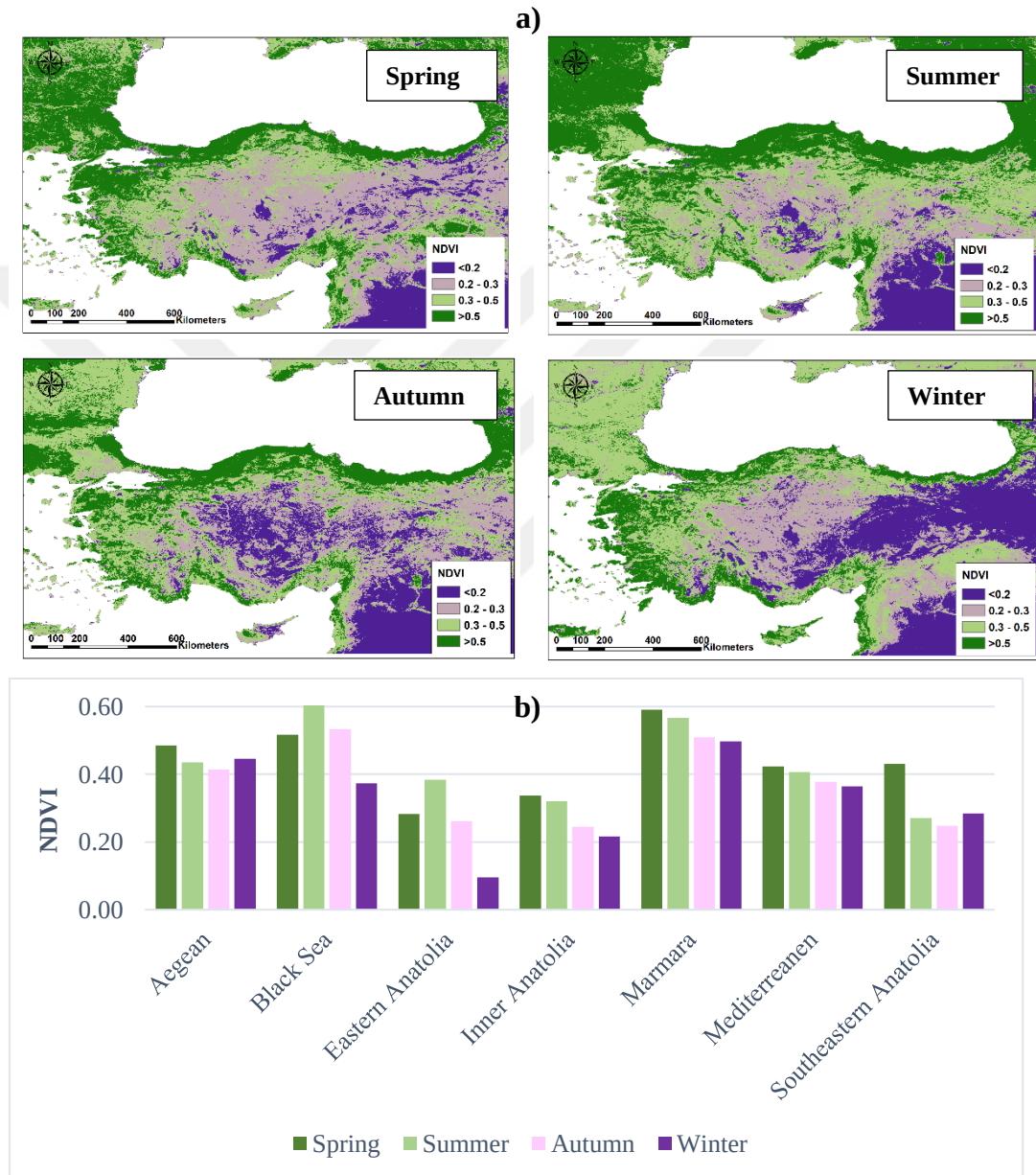


Figure 4.15 Spatial distribution of multiannual seasonal average NDVI over Turkey (a) and in the geographic regions of Turkey (b)

Anthropogenic aerosol emissions in different regions mainly depend on the seasonal variation of human activity consisting of industry, transport, residential

heating, and agriculture. Natural dust aerosols in the Mediterranean and southeastern regions are more frequent in spring and summer (Figure 4.16). Transportation of dust aerosol during dust storms can cause increasing in aerosol loading and affect the aerosol distribution of other adjacent regions of the country. Similar to the spatial distribution of multiannual AOD, low spatial AOD variability in all seasons was observed over Turkey by minimal AOD values in winter. In the eastern Anatolia region AOD was similar in autumn with winter. This is probably associated with a high density of precipitation in autumn, which causes lower aerosol levels. The NDVI values in summer and spring are larger than those of in autumn and winter. This indicates that AOD is increasing with NDVI increase and the AOD values are related to the seasonal variation of NDVI. There is a relation between vegetation cover and aerosol loading over Turkey. Vegetation can decrease the aerosols in the atmosphere by sorption. In contrast, aerosols by changing the surface solar radiation also affect vegetation growth. This affects the variation in NDVI, indicating that there are reliable relationships between AOD and NDVI over Turkey.

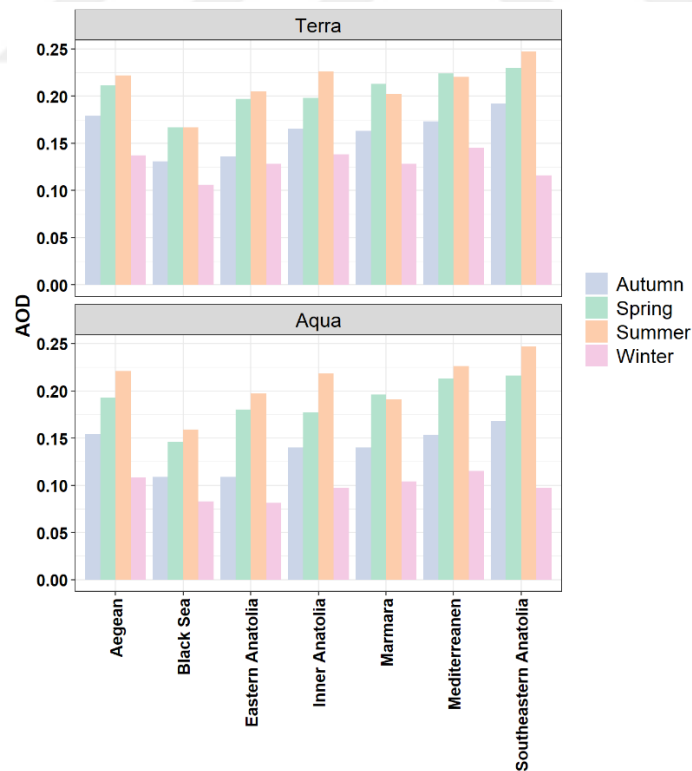


Figure 4.16 Multiannual seasonal average DTB AOD from Terra and Aqua satellites in the geographic regions of Turkey

4.3 Descriptive Statistics for Inputs of Estimation Models

Tables 4.3-4.4 present the descriptive statistics of modeling dataset indicating the minimum, mean, maximum values and quantiles of the grid-averaged daily PM₁₀ concentrations and grid-located independent variables from matched samples in the entire datasets over Turkey for Terra and Aqua overpass times, respectively. The multi-annual mean observed PM₁₀ concentration between 2008 and 2019 was observed as 61.17 µg/m³ and 51.73 µg/m³, which was 1.53 and 1.29 times higher than the National Ambient Air Quality Standard annual mean value of 40 µg/m³. In contrast, the multi-annual mean AOD were 0.21 and 0.19 for the datasets over Terra and Aqua overpass time, respectively. The minimum and maximum values of PM₁₀ and AOD for both datasets were similar. The number of low PM₁₀ concentrations and AOD were significantly higher than high values (PM₁₀>50, and AOD>0.2) according to the quantile values. Moreover, AOD mainly ranged between 0 and 0.25, with the mean value of ~0.2 over Turkey. Maximum values up to 3 can be observed in both satellite overpass time. Generally, it can be defined that 0.01 of AOD represents notably clean, and 0.4 indicates very hazy conditions (NOAA, 2020). Average AOD over Turkey for both satellite overpass times showed relatively clean conditions during the study period, similar to Texas (Ghahremanloo et al., 2021). The frequency distributions of PM₁₀ and AOD were similar, indicating that the distribution of the two parameters is related to each other. Higher PM₁₀ and AOD values were obtained at the Terra overpass time over Turkey. The multiannual mean of PM₁₀ concentrations over Turkey was much lower than that of other countries like China, with an annual average of 52-62 µg/m³ (Wei et al., 2019a, 2020a). Summary statistics for independent variables included in the final RF model after variable selection were also given in Table 4.3 and 4.4.

Table 4.3 Summary statistics for the dependent and selected independent variables over the Terra overpass time

Parameters	Min	Lower Quantile (25%)	Median (50%)	Mean	Upper Quantile (75%)	Max
PM ₁₀ (µg/m ³)	2.01	31.18	48.73	61.17	74.84	1158.56
TAOD	0	0.11	0.18	0.21	0.27	3.38
SRAD (W/m ²)	73.81	597.25	743.75	697.95	816.75	961.00
P (hPa)	77571	88807	94719	93374	99417	103185
u (m/s)	-10.95	-1.96	-0.23	-0.25	1.48	11.78
v (m/s)	-13.45	-2.44	-0.61	-0.45	1.43	16.38
ZOM (m)	0.003	0.041	0.067	0.223	0.233	2.375
DUSMASS (kg/m ³)	0.00E+00	4.07E-09	8.51E-09	1.74E-08	1.80E-08	2.14E-06

Table 4.4 Summary statistics for the dependent and selected independent variables over the Aqua overpass time

Parameters	Min	Lower Quantile (25%)	Median (50%)	Mean	Upper Quantile (75%)	Max
PM ₁₀ (µg/m ³)	2.01	27.50	42.14	51.73	63.18	1486.45
AAOD	0.00	0.10	0.16	0.19	0.25	3.13
SRAD (W/m ²)	100.10	563.00	714.20	671.10	793.50	954.20
P (hPa)	77315.00	89568.00	95509.00	93940.00	99597.00	103092.00
u (m/s)	-11.22	-1.62	0.23	0.18	2.01	12.28
v (m/s)	-15.86	-2.96	-1.03	-0.76	1.29	16.15
ZOM (m)	0.00	0.04	0.07	0.23	0.24	2.38
DUSMASS (kg/m ³)	0.00E+00	3.77E-09	8.08E-09	1.55E-08	1.68E-08	1.88E-06

The annual mean PM_{10} concentrations over the country differ for the individual region and season due to the combination of the different PM_{10} emission sources such as traffic, combustion, industry, mineral dust, sea salt, and meteorological, geographical, or economic factors (Houthuijs et al., 2001; Erdun et al., 2015). As, RF model developed in each region and season in this study, a descriptive summary of grid-located data from 2008 to 2019 at regional and seasonal scale RF model is illustrated in Tables 4.5 and 4.6 for the Terra and Aqua overpass times, respectively. The annual means of observed PM_{10} concentrations were higher than the country mean level in the Aegean, Central Anatolia, Marmara, and Southeastern regions for both satellite overpass times. In contrast, lower annual mean PM_{10} concentrations were observed in the Black Sea and Eastern Anatolia regions. Standard deviations also showed similar results, higher in the Central Anatolia, Mediterranean, and Southeastern regions.

There were notable differences between data numbers in regional scale RF model. The number of collocated samples with AOD for the TAOD model in the Marmara region was 38126, three times greater than those found in the region ($N=12355$) due to ground-level PM_{10} monitoring stations are unevenly spatially distributed, and much fewer monitors are operated in southeastern Anatolia. Meanwhile, the mean value of the and AOD (0.214 vs. 0.194) over the southeastern region was slightly higher than in the Marmara region. However, there was no difference between the two regions in the overall mean PM_{10} concentrations. Statistics showed a similar pattern in the AAOD model dataset.

Meanwhile, data number was higher in the Marmara region than in the TAOD model ($N=38464$). At the same time, it was lower in the southeastern region ($N=10557$) due to the differences in meteorological conditions (e.g., cloud cover, precipitation, etc.) over the two regions. Both lower AOD and PM_{10} levels were observed over the Black Sea region. All distinctions in data numbers and diverse conditions over regions and satellite overpass time affect the model performances to estimate PM_{10} concentrations for different regions and times in a day.

Table 4.5 Regional descriptive statistics of dependent and independent variables for the TAOD model

Region/Variable	Aegean (N=19998)	Black Sea (N=24935)	Central Anatolia (N=21561)	Marmara (N=38126)	Mediterranean (N=17750)	Eastern Anatolia (N=18420)	Southeastern Anatolia (N=12355)	Overall (N=153145)
TAOD								
Mean (SD)	0.232 (0.120)	0.160 (0.125)	0.217 (0.131)	0.191 (0.129)	0.234 (0.151)	0.208 (0.121)	0.214 (0.160)	0.204 (0.134)
Median [Min, Max]	0.210 [0, 1.37]	0.136 [0, 2.35]	0.193 [0, 2.24]	0.167 [0, 3.38]	0.204 [0, 2.54]	0.190 [0, 1.86]	0.182 [0, 2.34]	0.180 [0, 3.38]
PM₁₀ (µg/m³)								
Mean (SD)	66.8 (48.9)	56.1 (46.3)	64.3 (55.1)	63.7 (47.2)	61.6 (54.1)	50.9 (44.8)	63.8 (58.2)	61.2 (50.2)
Median [Min, Max]	55.9 [2.01, 937]	44.2 [2.04, 883]	49.5 [2.03, 913]	51.1 [2.03, 840]	49.5 [2.01, 758]	39.4 [2.01, 807]	49.4 [2.03, 1160]	48.7 [2.01, 1160]
DUSMASS (µg/m³)								
Mean (SD)	13.7 (23.1)	13.8 (25.1)	17.6 (33.2)	11.8 (23.7)	24.8 (49.7)	21.7 (29.6)	29.9 (52.8)	17.3 (33.4)
Median [Min, Max]	7.65 [2.34E-10, 615]	7.01 [1.34E-10, 594]	8.32 [2.92E-07, 921]	5.95 [9.62E-15, 525]	12.0 [1.13E-07, 2010]	13.0 [4.56E-09, 699]	15.3 [3.76E-10, 2140]	8.51 [9.62E-15, 2140]
SRAD (W/m²)								
Mean (SD)	717 (138)	705 (155)	674 (147)	681 (162)	707 (148)	738 (138)	676 (151)	698 (152)
Median [Min, Max]	757 [110, 947]	752 [73.8, 940]	718 [96.6, 906]	733 [85.8, 931]	752 [152, 934]	780 [170, 961]	722 [98.8, 921]	744 [73.8, 961]
ZOM (m)								
Mean (SD)	0.140 (0.0890)	0.518 (0.595)	0.0494 (0.0334)	0.278 (0.306)	0.128 (0.136)	0.0763 (0.106)	0.251 (0.274)	0.223 (0.340)
Median [Min, Max]	0.0690 [0.0286, 0.382]	0.162 [0.0154, 2.38]	0.0405 [0.0130, 0.238]	0.186 [0.00346, 1.23]	0.0646 [0.0213, 0.566]	0.0502 [0.00659, 0.640]	0.0642 [0.0186, 1.06]	0.0668 [0.00346, 2.38]
Elevation (m)								
Mean (SD)	463 (328)	738 (465)	1090 (348)	287 (240)	528 (410)	1660 (382)	748 (167)	727 (552)
Median [Min, Max]	385 [110, 1180]	626 [64.2, 1890]	1170 [93.2, 1610]	189 [37.3, 1070]	434 [19.8, 1340]	1580 [1080, 2370]	803 [231, 930]	640 [19.8, 2370]
u (m/s)								
Mean (SD)	-1.03 (2.23)	0.291 (1.94)	-0.223 (2.47)	-0.878 (2.95)	-0.177 (2.53)	0.472 (2.31)	0.653 (2.48)	-0.248 (2.56)
Median [Min, Max]	-1.08 [-8.86, 10.6]	0.284 [-8.13, 11.2]	-0.271 [-8.09, 10.4]	-1.07 [-11.0, 10.7]	-0.339 [-10.2, 11.4]	0.515 [-9.41, 11.8]	0.744 [-10.8, 8.89]	-0.231 [-11.0, 11.8]
v (m/s)								
Mean (SD)	-1.45 (3.09)	-0.704 (2.39)	-0.448 (2.98)	-0.722 (3.00)	0.0707 (2.46)	0.779 (2.70)	-0.121 (2.69)	-0.454 (2.87)
Median [Min, Max]	-1.68 [-11.3, 13.8]	-1.13 [-9.11, 16.4]	-0.584 [-11.2, 14.7]	-1.10 [-11.9, 15.9]	0.248 [-13.5, 11.9]	0.839 [-8.66, 12.8]	-0.0803 [-12.9, 11.3]	-0.609 [-13.5, 16.4]
NDVI								
Mean (SD)	0.456 (0.0635)	0.617 (0.143)	0.286 (0.0744)	0.535 (0.105)	0.431 (0.0868)	0.355 (0.119)	0.316 (0.139)	0.452 (0.156)
Median [Min, Max]	0.457 [0.218, 0.660]	0.646 [0.145, 0.870]	0.267 [0.100, 0.572]	0.528 [0.205, 0.882]	0.442 [0.146, 0.621]	0.334 [0.125, 0.769]	0.254 [0.172, 0.815]	0.449 [0.100, 0.882]
P (hPa)								
Mean (SD)	96700 (3560)	93600 (5610)	88900 (2940)	98700 (3050)	94500 (5170)	83200 (3830)	92400 (3340)	93400 (6320)
Median [Min, Max]	97900 [85800, 102000]	93900 [80400, 103000]	88800 [83200, 97500]	100000 [90000, 103000]	96400 [84400, 102000]	82900 [77600, 90700]	92300 [84800, 101000]	94700 [77600, 103000]

Table 4.6 Regional descriptive statistics of dependent and independent variables for the AAOD model

Region/Variable	Aegean (N=17747)	Black Sea (N=22797)	Central Anatolia (N=17192)	Marmara (N=38464)	Mediterranean (N=17193)	Eastern Anatolia (N=14794)	Southeastern Anatolia (N=10557)	Overall (N=138744)
AAOD								
Mean (SD)	0.222 (0.122)	0.150 (0.123)	0.190 (0.120)	0.176 (0.124)	0.226 (0.159)	0.194 (0.120)	0.199 (0.156)	0.189 (0.133)
Median [Min, Max]	0.197 [0, 1.20]	0.125 [0, 2.82]	0.166 [0, 1.78]	0.153 [0, 2.41]	0.193 [0, 3.13]	0.173 [0, 2.04]	0.167 [0, 2.88]	0.164 [0, 3.13]
PM₁₀ (µg/m³)								
Mean (SD)	57.4 (39.7)	46.3 (35.8)	52.5 (43.5)	52.2 (38.4)	52.8 (50.3)	47.0 (43.6)	55.9 (54.0)	51.7 (42.5)
Median [Min, Max]	49.1 [2.07, 880]	38.2 [2.04, 896]	40.9 [2.01, 777]	43.2 [2.03, 819]	43.6 [2.03, 1490]	36.3 [2.03, 874]	43.7 [2.03, 974]	42.1 [2.01, 1490]
DUSMASS (µg/m³)								
Mean (SD)	12.3 (18.0)	11.6 (21.1)	16.2 (29.7)	10.8 (19.6)	24.0 (41.8)	18.5 (24.7)	27.3 (47.0)	15.5 (28.5)
Median [Min, Max]	7.28 [2.04E-07, 386]	6.09 [1E-08, 674]	8.29 [7.31E-05, 557]	5.65 [1.96E-13, 536]	8.1 [1.08E-05, 864]	1.7 [2.14E-06, 610]	14.7 [6.65E-06, 1880]	8.08 [1.96E-13, 1880]
SRAD (W/m²)								
Mean (SD)	721 (141)	666 (156)	645 (148)	686 (156)	667 (154)	658 (143)	610 (155)	671 (154)
Median [Min, Max]	765 [124, 954]	709 [100, 933]	685 [130, 897]	740 [103, 942]	712 [119, 911]	695 [141, 917]	645 [127, 894]	714 [100, 954]
ZOM (m)								
Mean (SD)	0.138 (0.0885)	0.542 (0.600)	0.0493 (0.0334)	0.274 (0.303)	0.132 (0.137)	0.0754 (0.104)	0.260 (0.275)	0.233 (0.347)
Median [Min, Max]	0.0685 [0.0286, 0.382]	0.298 [0.0154, 2.38]	0.0401 [0.0138, 0.238]	0.185 [0.00341, 1.23]	0.677 [0.0237, 0.508]	0.508 [0.00702, 0.63]	0.0707 [0.0186, 1.06]	0.0676 [0.00341, 2.38]
Elevation (m)								
Mean (SD)	446 (312)	693 (443)	1100 (342)	276 (233)	472 (388)	1640 (382)	747 (173)	674 (537)
Median [Min, Max]	385 [110, 1180]	593 [64.2, 1890]	1170 [93.2, 1610]	181 [37.3, 1070]	434 [19.8, 1340]	1580 [1080, 2370]	803 [231, 930]	513 [19.8, 2370]
u (m/s)								
Mean (SD)	0.187 (2.38)	0.203 (1.99)	0.297 (2.57)	-0.636 (3.15)	0.586 (2.69)	0.964 (2.38)	1.17 (2.47)	0.182 (2.69)
Median [Min, Max]	0.207 [-8.87, 12.2]	0.230 [-7.54, 8.94]	0.310 [-8.58, 10.2]	-0.861 [-11.2, 11.8]	0.447 [-10.2, 11.1]	1.10 [-8.03, 12.3]	1.25 [-9.50, 10.7]	0.233 [-11.2, 12.3]
v (m/s)								
Mean (SD)	-1.61 (3.23)	-1.68 (2.26)	-0.614 (3.11)	-1.48 (3.08)	0.983 (2.72)	0.641 (2.82)	0.192 (2.71)	-0.764 (3.06)
Median [Min, Max]	-1.88 [-11.0, 13.3]	-2.04 [-9.11, 16.1]	-0.739 [-11.3, 14.6]	-1.99 [-10.5, 14.8]	1.01 [-15.9, 11.6]	0.577 [-10.4, 12.0]	0.232 [-12.3, 12.6]	-1.03 [-15.9, 16.1]
NDVI								
Mean (SD)	0.459 (0.0620)	0.625 (0.139)	0.282 (0.0722)	0.534 (0.105)	0.437 (0.0844)	0.353 (0.119)	0.317 (0.142)	0.460 (0.155)
Median [Min, Max]	0.458 [0.218, 0.660]	0.653 [0.185, 0.870]	0.264 [0.127, 0.572]	0.526 [0.270, 0.882]	0.449 [0.164, 0.62]	0.333 [0.0249, 0.769]	0.254 [0.172, 0.815]	0.458 [0.0249, 0.882]
P (hPa)								
Mean (SD)	96800 (3360)	94100 (5400)	88900 (3010)	98800 (2970)	95100 (4990)	83300 (3750)	92400 (3350)	93900 (6200)
Median [Min, Max]	97900 [86500, 102000]	94800 [80600, 103000]	88800 [83000, 97300]	90000 [89900, 103000]	900 [84400, 102000]	3000 [77300, 90600]	92200 [84700, 100000]	95500 [77300, 103000]

The data in Tables 4.7 and 4.8 show that the mean daily PM₁₀ concentrations were highest in the winter, lowest in summer, and moderate in spring and autumn, but AOD showed an increasing trend from winter to summer. In summer, the number of collocated samples was 55,894 (AAOD model) and 61,026 (TAOD model), approximately five times greater than those found due to inappropriate surface characteristics in winter to retrieve AOD.

The influence of dust transportation from the deserts in North Africa and Middle East during spring and summer was well known over Turkey (Kabatas et al., 2014; Agacayak et al., 2015). Dust from the desert rises to the upper layers of the atmosphere and travels long distances. Therefore, the country is frequently exposed to dust storms due to its proximity to African and Arabian deserts. This situation causes high AOD values reported in the summer and spring (Tutsak & Koçak, 2019; Ozdemir et al., 2020), although low PM₁₀ concentrations were observed at the surface due to the high aerosol loadings outside the surface layer (Tuna Tuygun & Elbir, 2020). Meanwhile, wet deposition due to precipitation lowered AOD in winter (Ozdemir et al., 2020). Both autumn and spring had similar PM₁₀ concentrations. The partial contribution of intense Saharan dust advection and deposition episodes that affected this area in spring and autumn is one reason for higher PM₁₀ and AOD levels (Tuna Tuygun & Elbir, 2020). Notably, there was a seasonal anti-correlation between AOD and PM₁₀ over Turkey. Similar results were also reported in Spain (Segura et al., 2017) and Athens (Raptis et al., 2020).

In addition to PM₁₀ concentrations and AOD, the values of model variables presented their own seasonal and regional characteristics (Tables 4.7-4.8). The P was high in winter and mainly over inland regions (Inner and Eastern Anatolia), while it was lower in summer, mainly over coastal regions of Turkey. Daily SRAD values ranges between 738 and 961 W/m² with the mean of 698 W/m² and 100-954 W/m² with the mean of 671 W/m² over Turkey. Higher solar radiation was observed in the Aegean region, while lowest values were observed in the south-eastern part. The DUSMASS concentrations reached 2140 µg/m³ and 1880 µg/m³ in spring in the southern and eastern regions of Turkey. Lower mean DUSMASS value was obtained for the AAOD

model dataset ($15.5 \mu\text{g}/\text{m}^3$) than the TAOD model dataset ($17.3 \mu\text{g}/\text{m}^3$). Seasonal statistics including selected variables used in this study are shown in Tables 4.7-4.8.

Table 4.7 Seasonal descriptive statistics of the dependent and independent variables for the TAOD model

Season/Variable	Autumn (N=46900)	Spring (N=34440)	Summer (N=61026)	Winter (N=10779)	Overall (N=153145)
TAOD					
Mean (SD)	0.181 (0.130)	0.218 (0.153)	0.224 (0.124)	0.146 (0.114)	0.204 (0.134)
Median [Min, Max]	0.154 [0, 2.54]	0.189 [0, 2.79]	0.207 [0, 2.24]	0.123 [0, 3.38]	0.180 [0, 3.38]
PM₁₀ ($\mu\text{g}/\text{m}^3$)					
Mean (SD)	69.3 (54.6)	61.5 (49.3)	48.5 (32.9)	96.3 (80.7)	61.2 (50.2)
Median [Min, Max]	55.5 [2.01, 743]	50.1 [2.03, 1160]	42.3 [2.01, 937]	74.4 [2.01, 840]	48.7 [2.01, 1160]
DUSMASS ($\mu\text{g}/\text{m}^3$)					
Mean (SD)	16.3 (30.9)	22.5 (43.6)	15.0 (21.9)	18.8 (52.9)	17.3 (33.4)
Median [Min, Max]	7.89 [2.27E-09, 863]	8.96 [2.52E-11, 921]	8.81 [9.62E-15, 699]	7.70 [1.07E-08, 2140]	8.51 [9.62E-15, 2140]
SRAD (W/m^2)					
Mean (SD)	581 (118)	751 (98.9)	805 (66.4)	435 (90.8)	698 (152)
Median [Min, Max]	587 [103, 846]	770 [101, 952]	813 [170, 961]	427 [73.8, 695]	744 [73.8, 961]
ZOM (m)					
Mean (SD)	0.215 (0.337)	0.230 (0.320)	0.234 (0.367)	0.174 (0.236)	0.223 (0.340)
Median [Min, Max]	0.0611 [0.00377, 2.32]	0.0694 [0.00398, 2.29]	0.0671 [0.00548, 2.38]	0.0660 [0.00346, 1.86]	0.0668 [0.00346, 2.38]
Elevation (m)					
Mean (SD)	759 (569)	626 (481)	790 (585)	558 (403)	727 (552)
Median [Min, Max]	650 [19.8, 2370]	483 [19.8, 2370]	715 [19.8, 2370]	483 [19.8, 2130]	640 [19.8, 2370]
u (m/s)					
Mean (SD)	-0.264 (2.46)	0.299 (2.55)	-0.595 (2.56)	0.0460 (2.62)	-0.248 (2.56)
Median [Min, Max]	-0.191 [-9.92, 10.7]	0.281 [-10.8, 11.4]	-0.595 [-11.0, 8.89]	0.0689 [-10.2, 11.8]	-0.231 [-11.0, 11.8]
v (m/s)					
Mean (SD)	-0.145 (2.60)	0.146 (2.82)	-1.29 (2.79)	1.00 (3.18)	-0.454 (2.87)
Median [Min, Max]	-0.184 [-13.5, 13.2]	-0.0209 [-12.5, 13.6]	-1.62 [-11.4, 13.8]	0.980 [-12.9, 16.4]	-0.609 [-13.5, 16.4]
NDVI (m/s)					
Mean (SD)	0.412 (0.160)	0.508 (0.128)	0.460 (0.159)	0.395 (0.125)	0.452 (0.156)
Median [Min, Max]	0.398 [0.139, 0.848]	0.504 [0.106, 0.882]	0.446 [0.153, 0.880]	0.408 [0.100, 0.729]	0.449 [0.100, 0.882]
P (hPa)					
Mean (SD)	93300 (6480)	94600 (5700)	92400 (6510)	95600 (5120)	93400 (6320)
Median [Min, Max]	93900 [77600, 103000]	95800 [77600, 103000]	92400 [77600, 102000]	96500 [78000, 103000]	94700 [77600, 103000]

Table 4.8 Seasonal descriptive statistics of the dependent and independent variables for the AAOD model

Season/Variable	Autumn (N=42845)	Spring (N=29473)	Summer (N=55894)	Winter (N=10532)	Overall (N=138744)
AAOD					
Mean (SD)	0.159 (0.129)	0.201 (0.146)	0.218 (0.123)	0.125 (0.111)	0.189 (0.133)
Median [Min, Max]	0.131 [0, 2.88]	0.169 [0, 3.13]	0.199 [0, 2.28]	0.102 [0, 2.41]	0.164 [0, 3.13]
PM₁₀ (µg/m³)					
Mean (SD)	55.9 (47.5)	51.8 (41.6)	44.7 (31.6)	72.0 (60.7)	51.7 (42.5)
Median [Min, Max]	45.3 [2.03, 1490]	42.4 [2.01, 974]	38.8 [2.03, 896]	55.8 [2.08, 815]	42.1 [2.01, 1490]
DUSMASS (µg/m³)					
Mean (SD)	14.7 (28.0)	19.5 (38.2)	13.6 (19.5)	17.5 (36.4)	15.5 (28.5)
Median [Min, Max]	7.63 [1E-07, 866]	7.97 [1E-08, 856]	8.40 [1.96E-13, 610]	8.48 [3.51E-05, 1880]	8.08 [1.96E-13, 1880]
SRAD (W/m²)					
Mean (SD)	556 (120)	726 (105)	777 (77.6)	423 (89.6)	671 (154)
Median [Min, Max]	560 [103, 848]	746 [152, 950]	789 [203, 954]	415 [100, 701]	714 [100, 954]
ZOM (m)					
Mean (SD)	0.225 (0.343)	0.243 (0.328)	0.247 (0.377)	0.167 (0.215)	0.233 (0.347)
Median [Min, Max]	0.0630 [0.00378, 2.32]	0.0756 [0.00398, 2.29]	0.0681 [0.00542, 2.38]	0.0684 [0.00341, 1.86]	0.0676 [0.00341, 2.38]
Elevation (m)					
Mean (SD)	709 (552)	577 (458)	733 (574)	487 (377)	674 (537)
Median [Min, Max]	607 [19.8, 2370]	434 [19.8, 2370]	618 [19.8, 2370]	427 [19.8, 2130]	513 [19.8, 2370]
u (m/s)					
Mean (SD)	0.208 (2.51)	0.668 (2.63)	-0.154 (2.80)	0.506 (2.74)	0.182 (2.69)
Median [Min, Max]	0.302 [-9.48, 10.4]	0.657 [-9.50, 12.2]	-0.131 [-11.2, 9.06]	0.461 [-10.2, 12.3]	0.233 [-11.2, 12.3]
v (m/s)					
Mean (SD)	-0.488 (2.69)	-0.147 (2.98)	-1.59 (3.14)	0.792 (3.01)	-0.764 (3.06)
Median [Min, Max]	-0.616 [-11.0, 12.6]	-0.440 [-11.1, 14.7]	-2.19 [-11.3, 10.7]	0.646 [-15.9, 16.1]	-1.03 [-15.9, 16.1]
NDVI					
Mean (SD)	0.422 (0.161)	0.516 (0.126)	0.469 (0.159)	0.410 (0.122)	0.460 (0.155)
Median [Min, Max]	0.414 [0.139, 0.848]	0.512 [0.116, 0.882]	0.455 [0.153, 0.880]	0.428 [0.0249, 0.729]	0.458 [0.0249, 0.882]
P (hPa)					
Mean (SD)	93800 (6350)	95100 (5480)	93000 (6430)	96300 (4910)	93900 (6200)
Median [Min, Max]	95500 [77600, 103000]	96500 [77300, 103000]	94700 [77500, 102000]	97500 [78400, 103000]	95500 [77300, 103000]

4.4 Effect of input variables on PM₁₀ estimation

4.4.1 Relationship Between AOD and PM₁₀

Estimating PM by satellite data is highly dependent on the relationship between satellite AOD and ground-level PM₁₀ concentration (Wang & Christopher, 2003; Gupta et al., 2006). The accuracy of estimated PM₁₀ over any region can affect the robustness of the linear relationship between AOD and PM₁₀ concentrations, which strongly depend on AOD from satellite and other factors such as vertical size distribution of aerosols and local meteorological conditions.

A simple linear regression model was used to analyze the relationship between PM₁₀ and AOD as the first step of this study. Similar to studies (Guo et al., 2009; Kloog et al., 2015) in the literature, AOD and PM₁₀ are not strongly correlated over

Turkey as seen in Figure 4.17. However, Aqua AOD represents aerosol variation better than Terra AOD over Turkey ($R=0.18$).

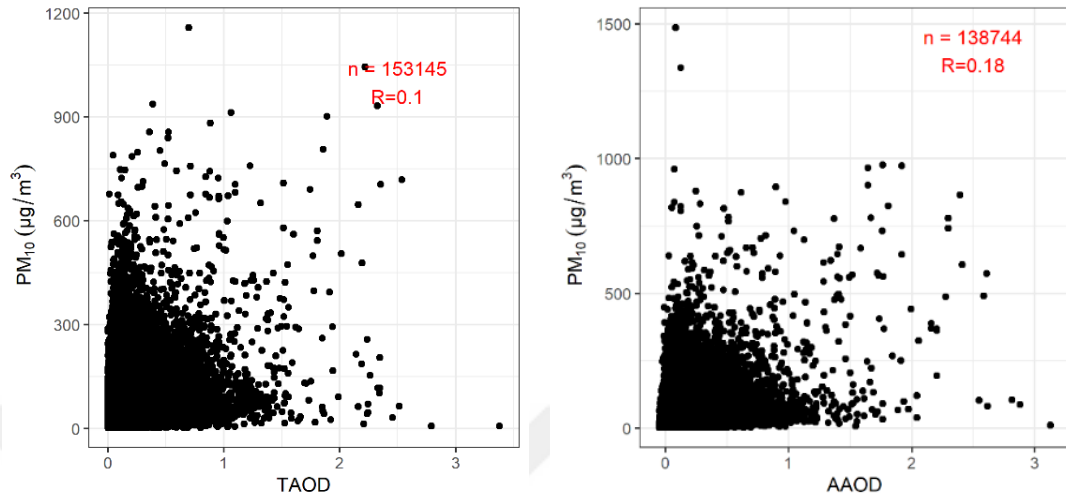


Figure 4.17 Relationship between PM₁₀ and AOD over Turkey

Figure 4.18 shows the correlation coefficients between PM₁₀ and AOD over 213 PM₁₀ sites. TAOD over all PM₁₀ monitoring stations ($n=213$) in Turkey had R ranging from 0.01 to 0.62 with the highest R (0.62) in the Black Sea region and lowest (0.01) for regions of Marmara and Central Anatolia. The higher R values were found for the coastal regions. The R for the AAOD is found between 0.01-0.6. The highest R (0.6) was found over the Southeastern Anatolia and the Black Sea regions. The relationship between MODIS AOD and PM₁₀ is poor in the stations in highly urbanized areas. The vertical distribution of aerosol affects the relationship between PM₁₀ and AOD since the AOD increases with elevated aerosol layers, but they are not determined by ground-based observations (Schafer et al., 2008). Each region has stations that have correlations lower than 0.1. Most of them are in the Marmara region. The low R values of the Marmara region could be due to dense anthropogenic activities over the major urban areas. MODIS AOD was validated against AERONET AOD in two different geographical regions in Turkey; however, no validation study has been performed for the remaining regions since there are not any AERONET stations in these regions. The low correlations over partially vegetated and high elevation areas can be due to relatively weak agreement between MODIS and AERONET AOD in those regions.

The correlation ranges were similar for two satellites. However, correlations with AAOD relatively higher than the TAOD. The higher correlations (>0.4) for AAOD were found mainly over the stations located in the Southeastern and Black Sea regions. Higher AOD-PM₁₀ correlations were obtained in the stations located in the southern part of Turkey, mainly affected by dust transport. In summary, the relationship between AOD and PM₁₀ found in a region cannot directly be applied to another region since the sources of aerosols showed regional variation. The AOD-PM₁₀ relationship may also vary due to region-dependent factors. It indicates that the auxiliary variables are required over Turkey to quantify this relationship and estimate PM₁₀ concentrations. Their inclusion in the AOD-PM₁₀ relationship can enhance the PM₁₀ estimation model accuracies over Turkey. Meanwhile, increasing the correlation between AOD and PM₁₀ with AAOD indicated that the overpass times of the satellites affect the relationship between PM₁₀ and AOD over Turkey.

The relationship between AOD and PM₁₀ was also investigated based on the quadrant method. This investigation considers which level of surface PM₁₀ pollution can be explained by satellite based AOD. It was also applied to investigate the effect of the relationship between AOD and PM₁₀ on the PM estimation models. Investigating the seasonal relationship between AOD and PM₁₀ is a widely used approach in statistical models to estimate surface concentrations with AOD. However, it is not enough to determine the atmospheric conditions affecting the relationships between AOD and PM₁₀. Each dataset was broken into four quadrants, and correlations between the two datasets are given in Table 4.9. The highest correlations were found for Terra (0.230) and Aqua (0.315) satellites with high AOD (>0.2) and high PM₁₀ ($50\mu\text{g}/\text{m}^3$) conditions. It means that satellite AOD can provide reliable information about polluted days in Turkey.

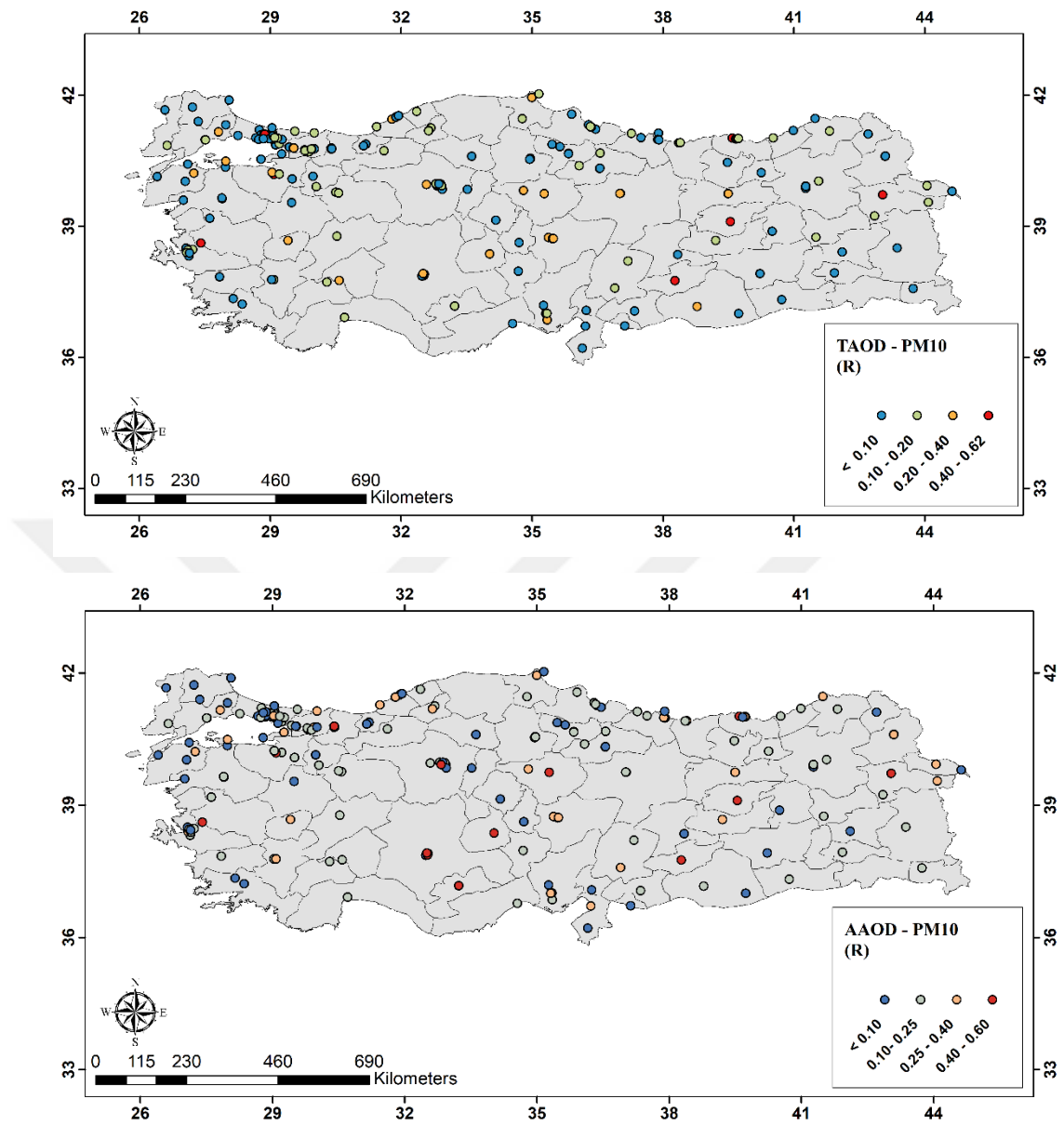


Figure 4.18 The R between PM₁₀ and AOD over 213 PM₁₀ sites

Table 4.9 Relationship between AOD and PM₁₀ based on different quadrants

Quadrant	TAOD			AAOD		
	n	R	<i>p-value</i>	n	R	<i>p-value</i>
Q-1	47536	0.061	<i>2.20E-16</i>	56493	0.066	<i>2.20E-16</i>
Q-2	39483	0.053	<i>2.20E-16</i>	29999	0.070	<i>2.20E-16</i>
Q-3	34296	0.230	<i>2.20E-16</i>	23990	0.315	<i>2.20E-16</i>
Q-4	31546	0.032	<i>1.19E-08</i>	28262	0.065	<i>2.20E-16</i>

The availability of AOD showed differences with the overpass times of two satellites since the monitoring frequency varies according to the location of the geographical regions (Filonchyk et al., 2019). The data availability rate was ranging between 12.0% and 62.5% for the Terra satellite. The availability rate ranged from 12.4% to 58.4% for the Aqua satellite. Data coverage of each satellite differed by region. As discussed in the validation section, cloud and ice cover and surface characteristics can affect the number of AOD retrievals over different regions. However, high data coverages over the station (>50%) were mainly obtained at the Terra overpass time. Only three stations had data availability higher than 50% at the Aqua overpass time. As seen in Figure 4.19, comparatively less valid TAOD were found at the station in the Marmara region. In contrast, relatively low AAOD availability rates were found over the Southeastern Anatolia, Central Anatolia, and Black Sea regions. The meteorological (e.g. cloud cover, precipitation) and geographical conditions (e.g. elevation) can cause this difference between the coastal and inland regions.

It is also known that the number of data samples from the DT algorithm in DTB AOD product is less than that from the DB algorithm in low-vegetated regions ($NDVI < 0.2$). However, the number of the DT algorithm increases over the moderately vegetated regions. The differences in the number of data samples between the two algorithms (DT and DB) can change the data availability ratios over different region in Turkey.

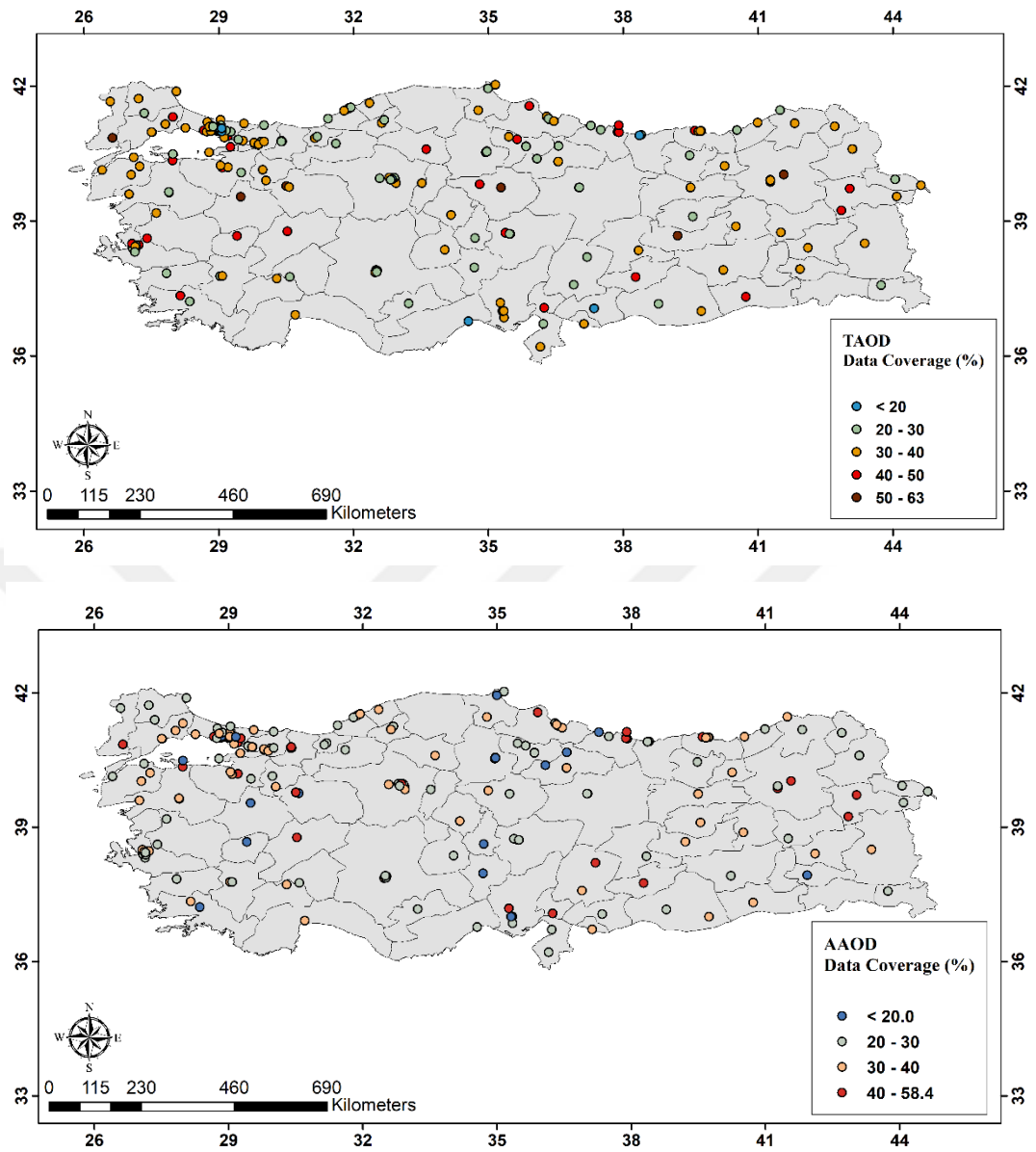


Figure 4.19 Data coverage of TAOD and AAOD data over 213 PM₁₀ sites

4.4.2 Relationship Between PM_{10} and Auxiliary Variables

Figures 4.20 and 4.21 show that all the independent variables were weakly correlated with the ground-level PM_{10} concentrations at p -value < 0.05 , and there were positive relationships with PM_{10} for CLOUD, u and v components of wind, DUSMASS, SO4MASS, and Elevation while remaining variables were negatively related to the ground-level PM_{10} concentrations. The positive correlation of PM_{10} with DUSMASS and SO4MASS is acceptable since they are components of PM_{10} concentrations (Ghahremanloo et al., 2021). The relatively strong negative relationship between PM_{10} and SRAD showed that particulate matter could increase the quantities of radiation absorbed, diffused into space, and subject to some other optical effects and change the solar radiation (Prendez et al., 1995). Short wave radiation is remarkably changed when it travels a polluted atmosphere and is reduced by anthropogenic emissions (Prendez et al., 1995). In general, vertical dispersion of PM is blocked with low PBLH and causes an increase in the PM_{10} concentrations (Guo et al., 2016; Yang et al., 2017). However, no correlation was found between PM_{10} and PBLH, mainly over the Aqua satellite overpass time that requires further research. Of all explanatory variables, SRAD ($r=-0.3$) for the Terra overpass time and DUSMASS ($r=0.3$) for the Aqua overpass time showed the strongest correlation with PM_{10} . Meanwhile, DUSMASS ($r=0.21$) and SRAD ($r=-0.19$) were other variables with a relatively high correlation with PM_{10} concentrations for the Terra and Aqua overpass times, respectively. Dust storms originating mainly from North Africa and the Arabian Peninsula also considerably affect the country. PM_{10} concentrations are notably higher than the national daily limit value of $50 \mu\text{g}/\text{m}^3$ (Koçak et al., 2007) during these events. TEMP also has a higher correlation with PM_{10} ($r=0.18$) at the Terra overpass time. Higher temperatures can increase the PM_{10} concentrations causing conversion of NO_2 into nitrate and SO_2 into sulfate (Lin et al., 2019). Higher temperature also causes more biogenic volatile organic compounds (BVOCs) from mainly vegetated areas (Di et al., 2016). A negative correlation between NDVI and PM_{10} concentrations was reported in the literature (Yang et al., 2019). Since the Marmara region, which has a highly vegetated surface, has high PM_{10} concentrations. Impacts of the independent variables on PM_{10} concentrations were higher in the Terra overpass time. Twelve of

independent variables showed higher correlations than ± 0.1 , whereas there were nine variables in the Aqua overpass time.

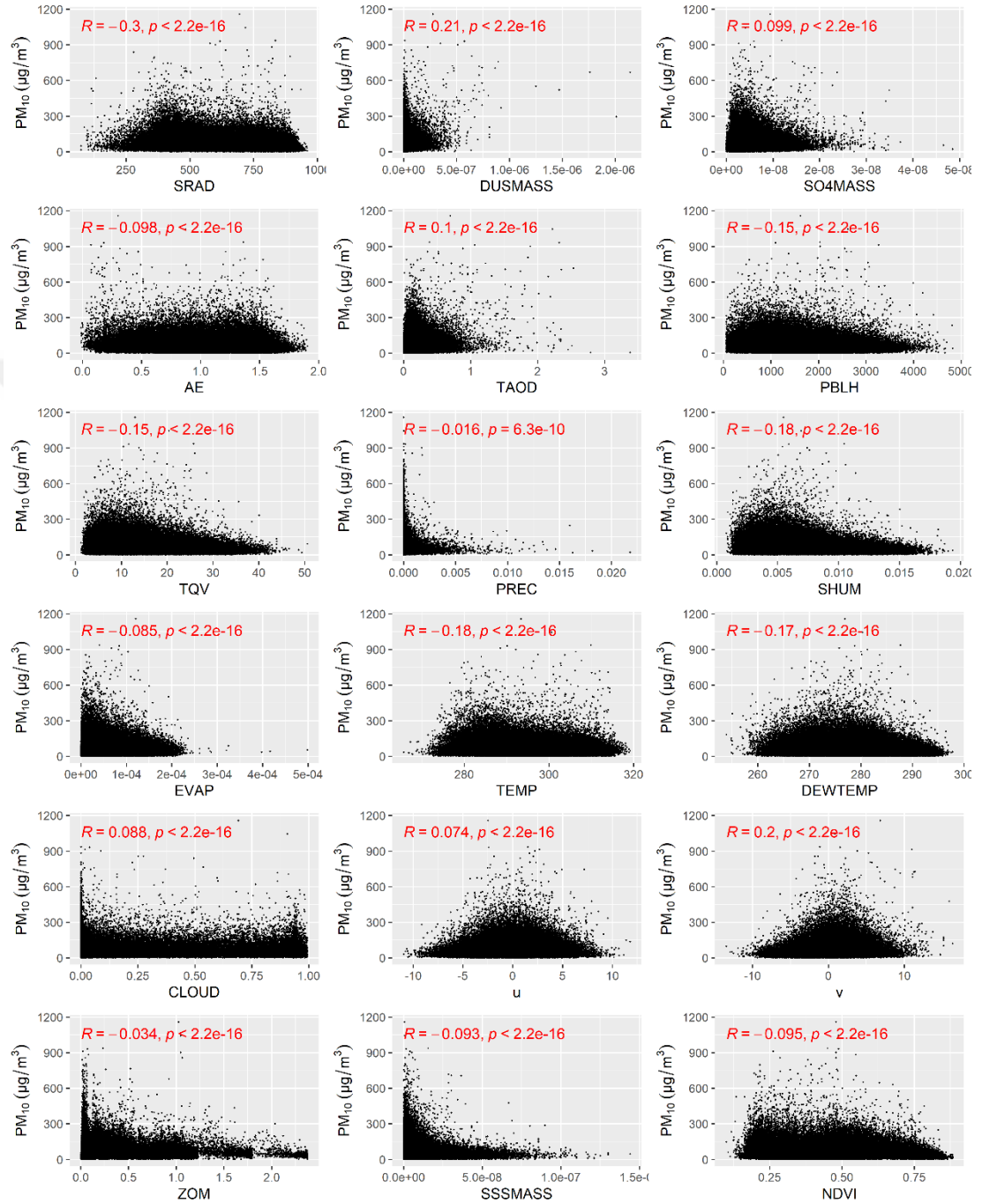


Figure 4.20 Correlation matrix of the variables for the TAOD model

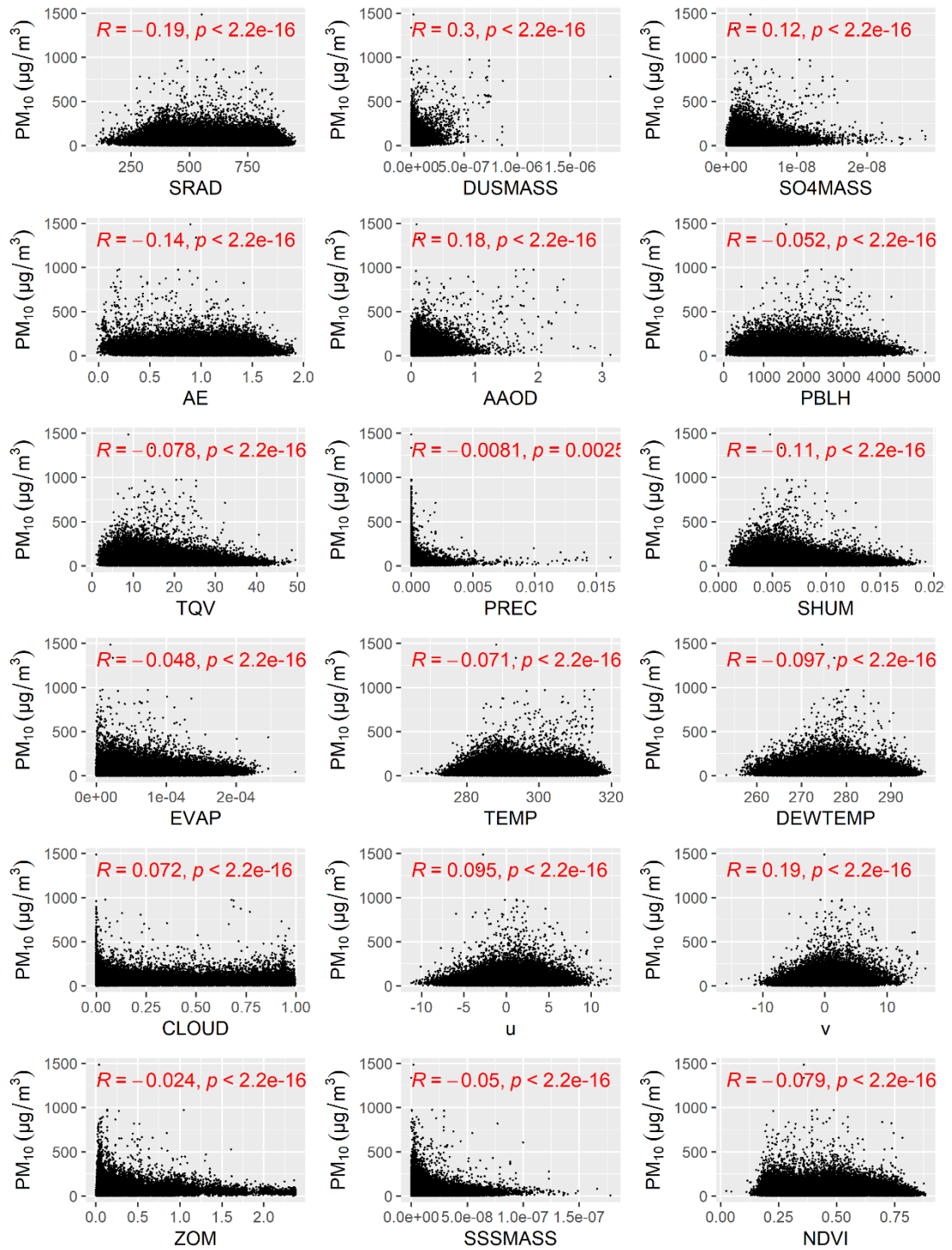


Figure 4.21 Correlation matrix of the variables for the AAOD model

The VIF indicating collinearity between variables was also calculated for all independent variables, and the results are shown in Table 4.10. Including cdate, Year, and jday into the model simultaneously would cause large collinearity since there is a strong relationship between the parameters of cdate and Year with a correlation coefficient of 1.0. Thus, cdate and jday were selected for the final RF model. SHUM and DEWTEMP also showed high collinearity (VIF~31-38) with a correlation coefficient of 0.97. Therefore, SHUM and DEWTEMP were removed from the final RF model which is developed after variable selection step. There was no remarkable collinearity problem among the remaining independent variables. P (~10.66-10.93) indicated high collinearity. However, it is included in the final RF model due to the importance score in the RF variable selection step.

Table 4.10 VIF results for each national RF model

Variable	Terra overpass time	Aqua overpass time
cdat	19716709.52	19431136.10
Year	19614692.92	19357494.37
jday	84872.81	86113.74
SHUM	36.26	37.28
DEWTEMP	31.32	31.41
P	10.93	10.66
TQV	8.50	8.38
Elevation	7.75	8.01
TEMP	6.23	5.78
PBLH	5.04	4.62
SRAD	4.34	3.91
NDVI	2.89	2.86
AOD	2.42	2.45
EVAP	2.17	2.38
SSSMAS	1.99	2.37
SO4MAS	1.81	1.99
DUSMAS	1.76	1.88
AE	1.75	1.79
Lon	1.75	1.79
Lat	1.64	1.62
ZOM	1.61	1.62
v	1.55	1.57
u	1.34	1.36
CLOUD	1.32	1.31
PREC	1.09	1.09

4.4.3 Variable Importance and Feature Selection for Estimation Model

Finally, all independent variables in the TAOD and AAOD models for estimation of spatiotemporal PM₁₀ concentrations were selected for the model in terms of their importance scores. The variables that would not significantly affect the model performance in terms of R were removed from the final RF model.

The RF determined the variable importance by calculating the decreasing overall variance of the estimated PM₁₀ at each node. Firstly, 25 variables commonly used to estimate PM₁₀ concentrations were tried to evaluate in terms of their importance, as shown in Figure 4.22. The scores of twelve features such as PREC, TEMP, EVAP, PBLH, SO4MASS, SSSMASS, AE, TQV, DEWTEMP, SHUM, CLOUD, and Year were smaller than 8,000,000 and 4% in the RF model. For the estimation of PM₁₀ concentrations, these twelve variables played insignificant roles with weak contributions than others. Figure 4.22 illustrates the superior 12 independent variables that were the most contributing variables by variable selection in the RF models to estimate PM₁₀ concentrations. SRAD, DUSSMASS, Lon, Lat, jday, cdate, ZOM, Elevation, AOD, u, v, NDVI, and P were commonly identified as contributing variables by the RF models to estimate the ground-level PM₁₀. Variable importance scores and contributions varied with satellite; however, results indicated the same variables (Figure 4.22).

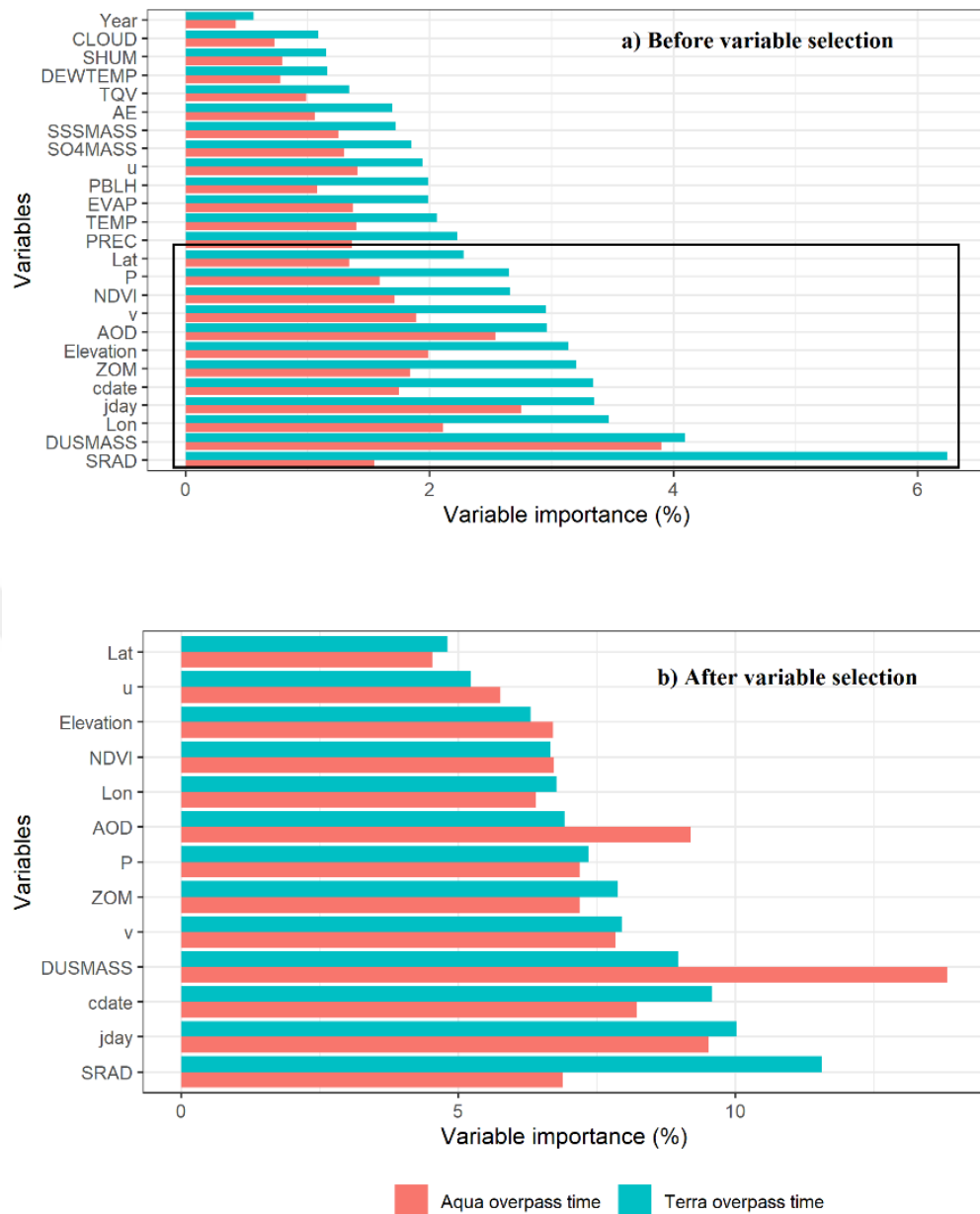


Figure 4.22 Variable importance for TAOD and AAOD model before (a) and after (b) variable selection

The most significant variables in the model, such as Lat, Lon, elevation, and NDVI, are constant in time. However, they were used to consider the spatial variation of PM_{10} concentrations over Turkey. AOD had not significant importance in the TAOD model. However, AOD was determined as the second most significant variable in the AAOD model. The low and high importance of AOD has been reported by Park et al. (2019) and Wei et al. (2020a), respectively. Variable importance results indicated that the relationship between PM_{10} and AOD could notably be affected by aerosol types in

PM₁₀ and solar radiation for both satellite overpass times. The inconstant results showed that the importance of AOD in the TAOD and AAOD models is mainly related to aerosol composition and satellite overpass time in Turkey. The higher importance scores of DUSMASS indicated the importance of the different sources of PM₁₀ over Turkey and the strong effect on PM₁₀ estimations. Results indicated that the sources of PM₁₀ are particularly natural in Turkey. The importance of wind components on PM₁₀ estimations was consistent with previous studies (Stafoggia et al., 2019). It is known that low wind speed conditions cause an increase in PM₁₀ concentrations. P is another significant meteorological variable since higher pressure, particularly with a low PBLH, blocks the vertical dispersion (Su et al., 2018).

Despite a relatively high correlation with PM₁₀ concentrations, lower importance scores were found for some independent variables. It can be explained that PCC indicates the linear relationship, while the importance score in RF represents the non-linear relationship. It mainly shows the effect of each variable on the final values of the RF model for PM₁₀ concentrations. It is expected that precipitation and humidity can highly affect PM₁₀ concentrations (Wei et al., 2020a). Their importance is not reasonable in the RF model. The main reason could be that AOD is retrieved only at cloud-free pixels where precipitation is not customarily characterized, and RH is low. This result is similar to the previous study carried out over Texas (Ghahremanloo et al., 2021) and China (Wei et al., 2020a). This finding can also be explained that the importance scores in the RF model represent the importance of independent variables in the splitting process during the construction of trees and do not indicate the physical contribution of the parameters to PM₁₀ concentrations (Ghahremanloo et al., 2021). Chen et al. (2018) and Nabavi et al. (2019) indicated that the integration of temporal variables into the estimation model (e.g., day of the year) could regulate daily estimated PM concentrations replacing remaining unknown parameters. Hence, jday and cdate are used as input to resolve the significant temporal variation of PM₁₀ in Turkey. Significant improvement was observed in the performance of the PM₁₀ estimation model with the inclusion of temporal information.

Some meteorological variables (SRAD, P, u, and v) indicated that the atmospheric conditions could affect the accurate estimation of ground-level PM₁₀ concentrations in the final TAOD and AAOD model. In addition to meteorological variables, land-related parameters such as NDVI, Elevation, and ZOM also showed the impact of land parameters on PM₁₀ estimation models. The most important predictors were spatiotemporal variables (SRAD, DUSMASS, AOD, P, ZOM, AOD, wind components, and Julian day) as they described PM variability both in time and in space. As the spatial terms, spatial coordinates, elevation, and NDVI showed the highest importance.

Variable importance from the RF models for each region is also determined to find out which variables were more important in the regional models. Figure 4.23 indicates that the variable importance of each satellite overpass time and region was different. Overall, DUSMASS, SRAD, and AOD were the three most significant variables for the Aqua overpass time, while SRAD, DUSMASS, and jday were most important for the Terra overpass time. The cdate information relatively provides long-term effects contributed to the TAOD model performance.

The importance of AOD in the RF models indicated a regional variation. TAOD was more likely to be the most critical variable in the southeastern Anatolia region, whereas AAOD was the most significant variable in the Mediterranean region. Elevation did not significantly contribute to PM₁₀ estimations except for the TAOD model over the Marmara region. According to regional variable importance results, DUSMASS, the most critical parameter in the AAOD model, was a highly significant contributor in the Mediterranean, eastern and southeastern Anatolia regions. It is known that desert dust contributes the most to these regions with higher mass concentrations compared to other regions. SRAD is the highest contributor to the TAOD model and mainly contributes to PM₁₀ estimates in the Aegean, Central Anatolia, Mediterranean and southeastern regions. It concluded that there was no stable pattern in regional variable importance. Regional inconsistent importance scores are shown in Figure 4.23.

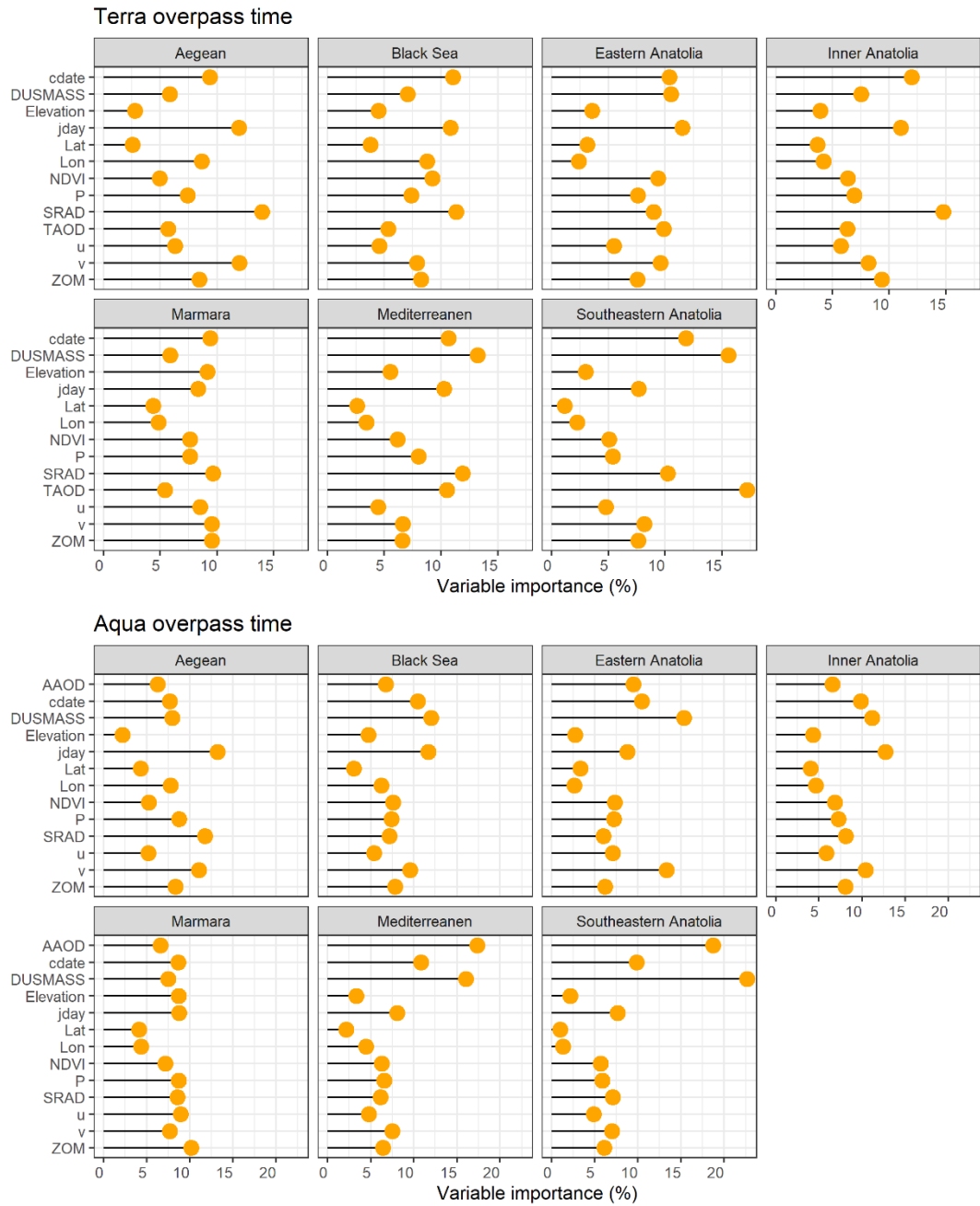


Figure 4.23 Variable importance scores of for regional RF models

4.5 Performance of the RF Models

4.5.1 National Scale Model

The estimated ground-level PM_{10} concentrations from two different RF models were compared based on MODIS-Terra (TAOD) and MODIS-Aqua (AAOD). Figure 4.24 indicates the scatterplots, including the model training and CV results of the RF

models with TAOD and AAODs. PM₁₀ estimation models based on satellite AOD and statistical relationships of AOD and PM₁₀ could not define the physical conditions of the relationship. This causes satellite and ground-level aerosol levels not to be well correlated in Turkey. Thus, a total of 13 variables including PM₁₀, AOD representing the Terra and Aqua satellites overpass time, SRAD, P, DUSMASS, u, v, NDVI, ZOM, Elevation, Lat, Lon, jday, and cdate were used for model training of national TAOD and AAOD model in this study.

A linear regression between PM₁₀ and AOD showed that only AOD is not a strong predictor over Turkey. With the random forest model, results indicated that the observed and estimated PM₁₀ concentrations were moderately correlated as the CV R reached 0.71 for the Terra overpass time, which was higher (61%) than the result of the linear regression model. The RMSE and MAE were 36.43 and 21.75 µg/m³, respectively (Figure 4.25). However, the increasing was smaller for the Aqua overpass time as the CV R reached 0.67 (~49%). The RMSE and MAE were 32.75 and 18.35 µg/m³, respectively. In the training dataset, R values reached the values higher than 0.90 for each satellite overpass time. It means that overfitting exists in the TAOD and AAOD models. Although a statistically significant difference was not found between the TAOD and AAOD models, TAOD model produced a slightly better R-value than the based on AAOD model. At the same time, RMSE and MAE were observed smaller in the AAOD-based RF model. Figure 4.25 shows that the estimated and observed PM₁₀ concentrations are densely distributed on both sides of the 1:1 line for each product.

A more careful data configuration between training and test samples with a larger sample size may mitigate such an overestimation. Further analysis of the new independent variables can increase the PM₁₀ estimation model performance since our national model highly underestimated the higher concentrations over the Turkey (Figure 4.25).

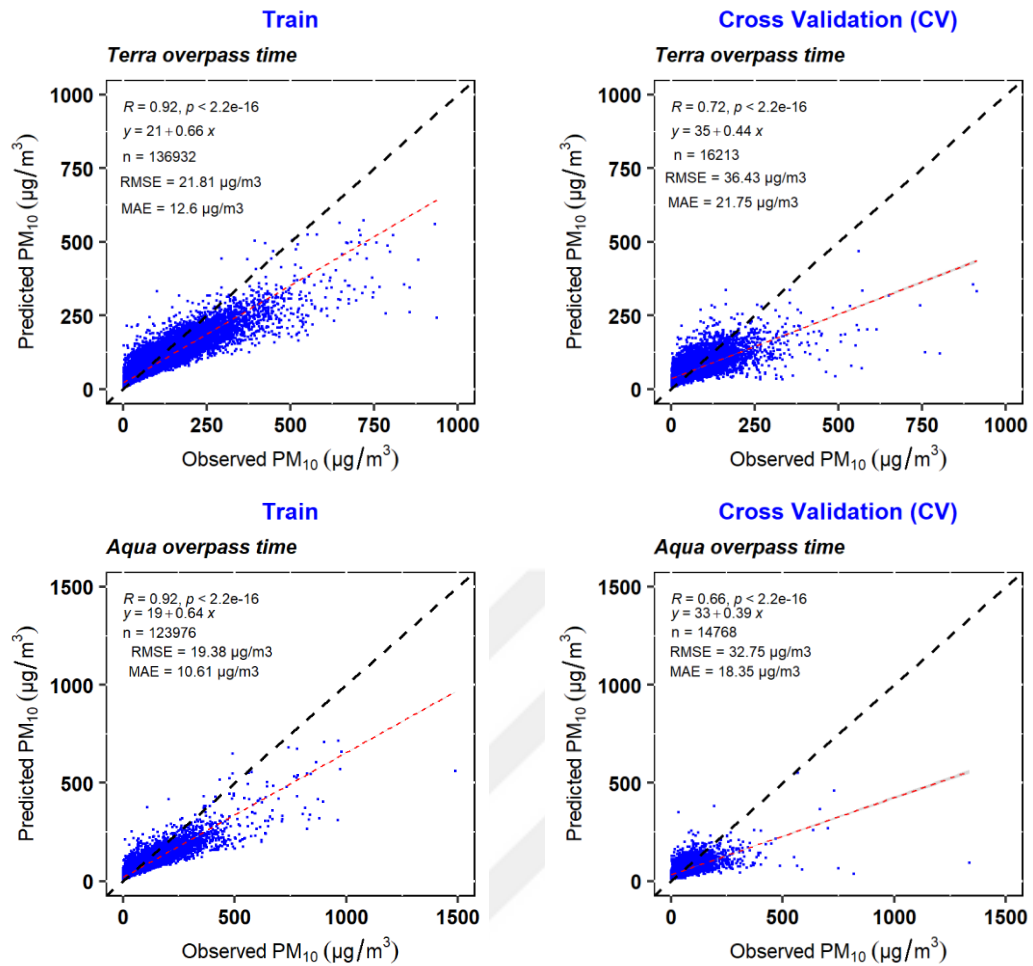


Figure 4.24 National RF model results for each satellite overpass time (black and red dotted lines represent 1:1 line and regression line, respectively)

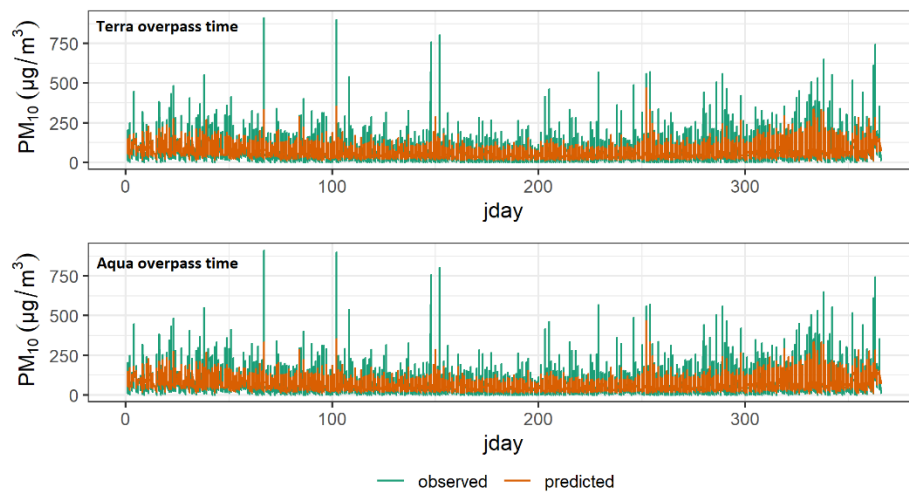


Figure 4.25 Variation of the observed and predicted daily PM_{10} concentrations by national TAOD and AAOD RF models

A random CV cannot indicate the ability of the model to make predictions beyond the location of the training data. Since the study aims to map PM₁₀ distribution over the region, it is needed to perform a spatial CV to validate the model in view of spatial mapping. Spatial CV was carried out to know how well the model can make predictions for areas without PM₁₀ monitoring stations. To find out this, it was repeatedly leaving the entire geographical region of one out and use them as test data during the CV. The index defines which regions are used for model training during each model run and reversely defines which regions are held back. Temporal CV had a similar procedure. Since the temporal variation of AOD is more significant than spatial variation over Turkey, temporal CV results were higher than spatial CV results in Table 4.11. For CV on a spatial predictive power, all models based on geographic regions have relatively poor accuracies ($R \sim 0.37-0.45$) since AOD and PM₁₀ concentrations are noticeably spatially uniform over the different regions. The spatial predictive power of the RF models showed higher performance compared to other regions in the Inner and Southeastern Anatolia regions ($R \sim 0.47$). The spatial predictive power of the national RF model was lower in the Marmara region ($R \sim 0.37$). Both satellites showed similar spatial predictive powers over the regions.

In this study two different temporal model validation method was selected. One of them randomly selected 10% of jday as testing set (Temporal CV-1), other was randomly selected days from January, April, July, and October as a testing set (Temporal CV-2). The temporal model performances were much better than those on the spatial model ($R \sim 0.60-0.69$). Temporal CV-1 performances for both satellites were higher than temporal CV-2. In general, the temporal variation enhances the national RF models performance. The CV results showed that more temporal variability was explained than spatial variability in the national scale model over Turkey. However, the space-time CV performs better. Models with fine resolution AOD products and more PM₁₀ monitoring stations over Turkey can provide more robust R to represent the spatial variation of estimated PM₁₀ over Turkey. Spatial and temporal predictive power of national TAOD and AAOD RF model is given in Table 4.11.

However, national RF models with Terra and Aqua AOD underestimated PM₁₀ concentrations (slope = 0.39-0.44, intercept = 33-36 µg/m³), indicating weak estimations on heavily polluted days. Other studies in the literature (Ma et al., 2016; Wei et al., 2019a) were also found such underestimation. It was also determined by the quadrant method that much lower correlation between AOD and low-level PM₁₀ concentrations, indicating that lower levels of PM₁₀ concentrations can potentially decrease the performance of the model. Di et al. (2016) also indicated the adverse effects of low PM concentrations on the model performance.

Table 4.11 Spatiotemporal predictive power of national TAOD and AAOD RF models

Predictive scale		n	r	RMSE (µg/m ³)	MAE (µg/m ³)	mean observed PM ₁₀ (µg/m ³)	mean predicted PM ₁₀ (µg/m ³)
Spatial (TAOD)	Aegean	19998	0.42	44.4	28.3	66.8	66.2
	Black Sea	24935	0.39	43.5	30.0	56.1	63.5
	Central Anatolia	21561	0.47	54.6	38.4	64.3	83.9
	Marmara	38126	0.37	43.9	29.7	63.7	62.9
	Mediterranean	17750	0.41	50.4	35.1	61.6	70.3
	South Anatolia	18420	0.47	42	29.9	50.9	63.9
	Southeastern Anatolia	12355	0.45	55	38.8	63.8	81.9
Temporal (TAOD)	Temporal CV-1	15474	0.69	37.7	21.9	61.4	62.2
	Temporal CV-2	5172	0.67	37.4	21	59.1	60.0
Spatial (AAOD)	Aegean	17747	0.36	37.5	24.4	57.4	58.2
	Black Sea	22797	0.36	35.1	23.9	46.3	54.6
	Central Anatolia	17192	0.45	41.3	29.3	52.5	64
	Marmara	38464	0.34	36.2	24.2	52.2	54
	Mediterranean	17193	0.34	49.3	32.2	52.8	65
	South Anatolia	14794	0.43	41.5	27.1	47	57.2
	Southeastern Anatolia	10557	0.47	50.2	34.0	55.9	72.0
Temporal (AAOD)	Temporal CV-1	14043	0.68	30.1	17.9	51.8	52.8
	Temporal CV-2	4719	0.60	35.4	18.2	50.3	50.8

Therefore, it is significant to evaluate model performance for different PM₁₀ levels. Quantile regression was performed for each national RF model for the following

quantiles: 0.1, 0.25, 0.5, 0.75, 0.90 and 0.95. The performance of QR models was then compared to the national models to determine the representativeness of different quantile ranges. The results of QR are shown in Figures 4.26 and 4.27. QR of the 50th and 75th percentile has the highest R (0.71-0.70), while those of the 10th and 95th percentile have the lowest R values (0.63-0.64). The QR with the R differs between 0.63-0.71 on the 10th, 25th, 50th, 75th, 90th, and 95th percentiles (Figures 4.26-4.27). The QR indicated that national STRF modeling was not capable of all quantiles. Regression results for lower quantiles (10th) were notably different from upper quantiles. Di et al. (2016) also found high uncertainty for low aerosol loadings.

However, the QR performance increased with quantile (Figures 4.26-4.27). QR also illustrated that the national model might not be adequate for upper quantiles (90th and 95th). Moreover, QR results of the 25th and 75th percentiles were notably different. Based on the national STRF and QR results, alternative ML models or variables must be integrated into the modeling step to represent the data adequately. The evident correlation was found between PM₁₀ and MODIS AOD at higher PM₁₀ concentrations in quadrant analysis. However, prediction ability was low, and it is not adequate qualitatively to identify areas of high PM₁₀ concentrations. Both RF models cannot be used to estimate high level PM₁₀ concentrations over Turkey. In this study, randomly selected 90% percent of the data points was used for national RF model's development as training data and remaining 10% was used as test data to determine model accuracy. Recently oversampling and subsampling methods were developed for the training dataset due to skewed distribution of PM₁₀ including many low-concentration and a few high-concentration data samples (Park et al., 2019). This can solve the overfitting problem of the national RF model. Many of the PM₁₀ concentrations used in this study consisted of low concentrations, whereas a small number of high PM₁₀ concentrations were available. This imbalance could cause the biased estimation with a significant underestimation of high-concentration data. Thus, oversampling, or subsampling methods have been recently used to handle unbalanced sample problem in training dataset.

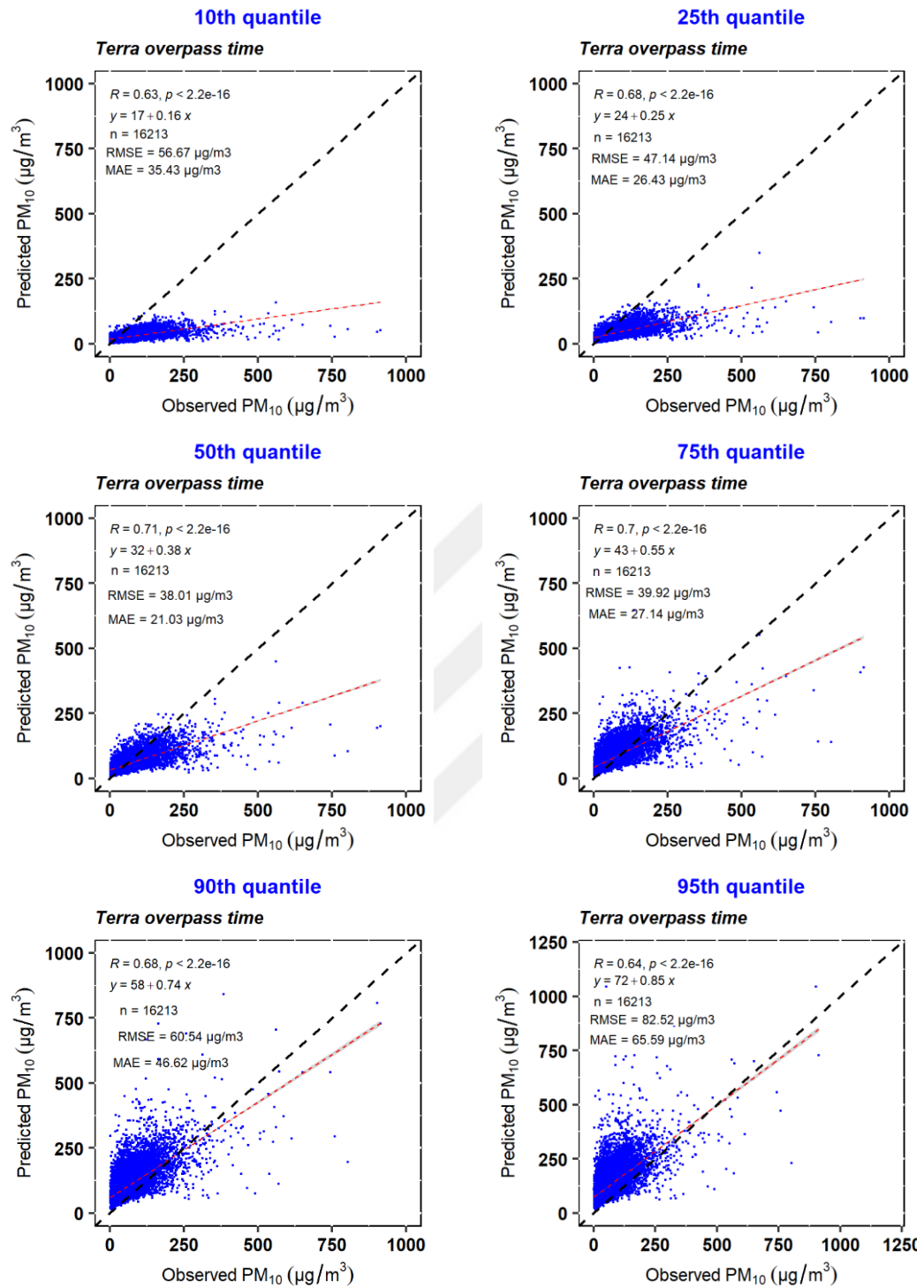


Figure 4.26 Quantile regression results for the Terra overpass time (black and red dotted lines represent 1:1 line and regression line, respectively)

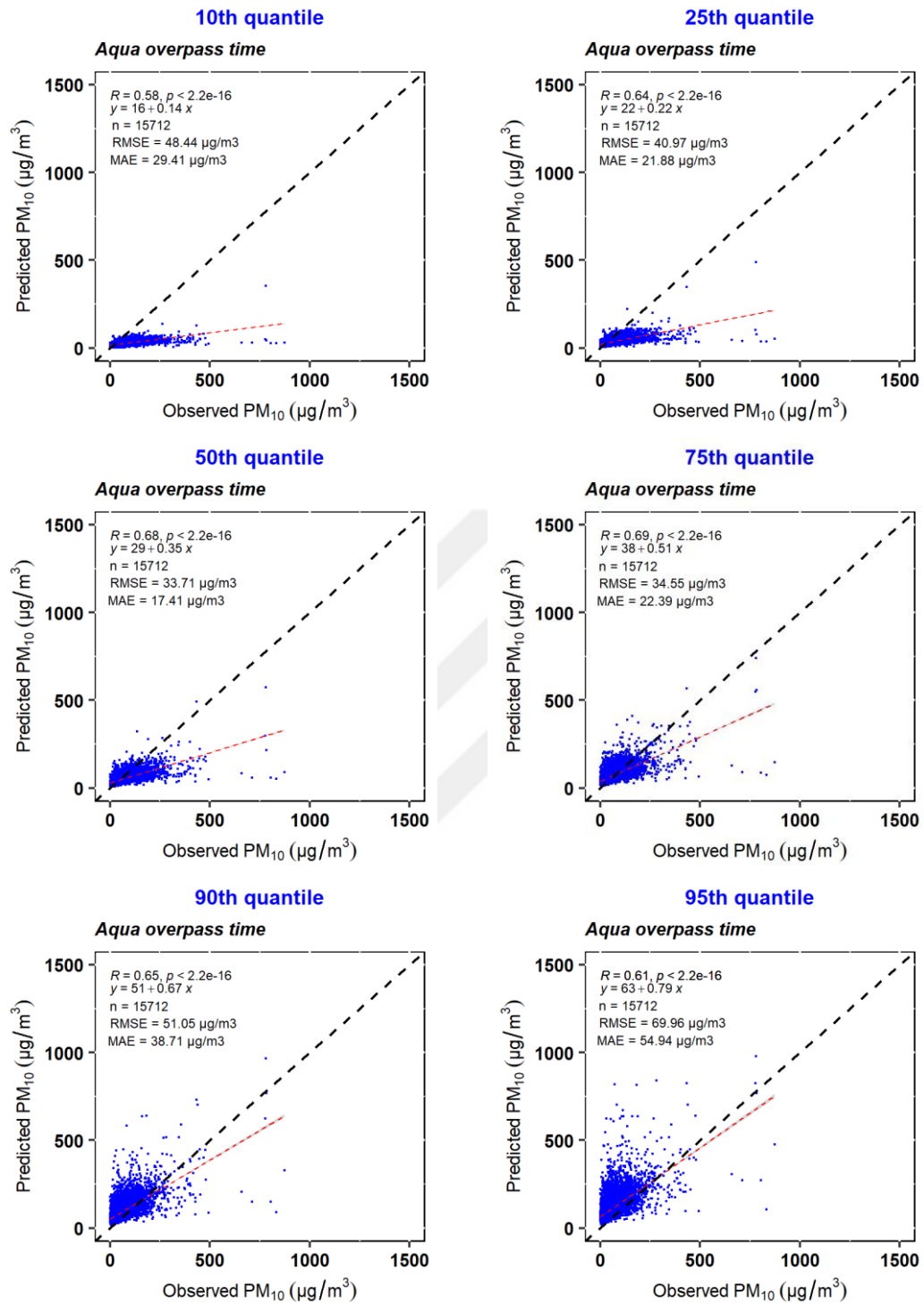


Figure 4.27 Quantile regression results for the Aqua overpass time (black and red dotted lines represent 1:1 line and regression line, respectively)

4.5.2 Regional and Site-Based Models

The region-based RF models were also developed in this study. The CV-R of the models for the regions ranged from 0.65-0.80. CV-RMSE ranged from 25.5 to 49.0 $\mu\text{g}/\text{m}^3$, where the lower performances were found in the model with Aqua AOD. Model performances are illustrated in Figures 4.28 and 4.29 for each satellite RF model. The highest performance was achieved in the Mediterranean region with an R-value of 0.80 with Terra AOD. The RMSE and MAE values were 34.5 $\mu\text{g}/\text{m}^3$ and 20.6 $\mu\text{g}/\text{m}^3$, respectively. The lowest model performance was found in the Southeastern Anatolia region ($R \sim 0.68$). However, with Aqua AOD the highest model performance was observed in the Southeastern region ($R \sim 0.75$). Also, the model performance over the Mediterranean region was poor with Aqua AOD. Satellite is key parameter to achieve higher performance of the regional model over Turkey.

With the lower performance of the regional models than the national model over the Inner and Southeastern Anatolia regions with Terra AOD, this AOD product is not applicable in these regions to estimate PM_{10} concentrations. More inadequate predictions in Southeastern Anatolia could be due to low data availability over the Terra overpass time and less ability of the available variables and AOD to capture the specific PM_{10} profile in the region characterized by significant contributions from desert regions.

Aqua AOD is not applicable for the Black Sea, Mediterranean, and Black Sea regions. In general, low R values (<0.70) show that model fitting is not adequate to an accurate estimate of ground-level PM_{10} . Regional models indicate that the models particularly underpredict high-level ground-level PM_{10} concentrations. Note that the Eastern Anatolia, Black Sea, and Mediterranean regions can genuinely encompass the national models with Terra AOD.

Figures 4.30-4.31 show the spatial distribution of site-based RF model performance over the 213 PM_{10} monitoring stations. Station-based CV results of daily PM_{10} estimates in Turkey using the STRF model offer similar performances with both

satellite data. The STRF-model estimated daily PM₁₀ concentrations wide correlation range with ground-level PM₁₀ concentrations ($R=0.25-0.90$) and with various estimation uncertainties ($RMSE = 7.7-75 \mu g/m^3$, $MAE = 6.2-49.6 \mu g/m^3$). Overall, the site-based R over the stations is 0.62, and the RMSE and MAE are $31.5 \mu g/m^3$ and $20.7 \mu g/m^3$, respectively. The station-based performance showed various performances, with high differences in evaluation metrics. Findings were consistent with the regional model results. With Aqua AOD the highest model performance was found in the stations located in the Southern Anatolia region ($R\sim 0.85$), lowest performance ($R\sim 0.1$) was observed in a station located in Central Anatolia with the low data number. Data numbers over stations shown in Figures 4.30-4.31 varied significantly and affected the site-based model performance. Best performance with Terra AOD was found in the stations over the Black Sea region ($R\sim 0.90$), whereas the lowest performance ($R\sim 0.25$) was observed in the station located in the Marmara region. Sixty-eight stations with Terra AOD and fifty-five stations with Aqua AOD estimate PM₁₀ concentrations at a reasonable level ($R>0.70$).

The regional patterns of the multiannual averaged PM₁₀ concentrations showed similar PM₁₀ levels (Figure 4.32). The multiannual average observed PM₁₀ concentrations were above the national limit value of $40 \mu g/m^3$. The Black Sea and eastern Anatolia regions showed the lowest average PM₁₀ concentrations compared to other geographical regions.

The highest average observed and estimated PM₁₀ concentrations were obtained in the Aegean and southeastern Anatolia regions with satellite overpass times. The observed and estimated PM₁₀ concentrations showed differences ranging from -2.1 to $0.2 \mu g/m^3$, indicating that they overestimated regional PM₁₀ concentrations by the regional models. In most cases, slightly higher estimated concentrations were found during the overpass of Aqua. The lowest deviation was found in the Marmara region with the TAOD model (~ 0.2), while the Mediterranean region had the lowest estimation error (~ 0.2) with the AAOD model. The highest deviation was found between the observed and estimated PM₁₀ in the Black Sea and Aegean regions with the TAOD and AAOD models, respectively.

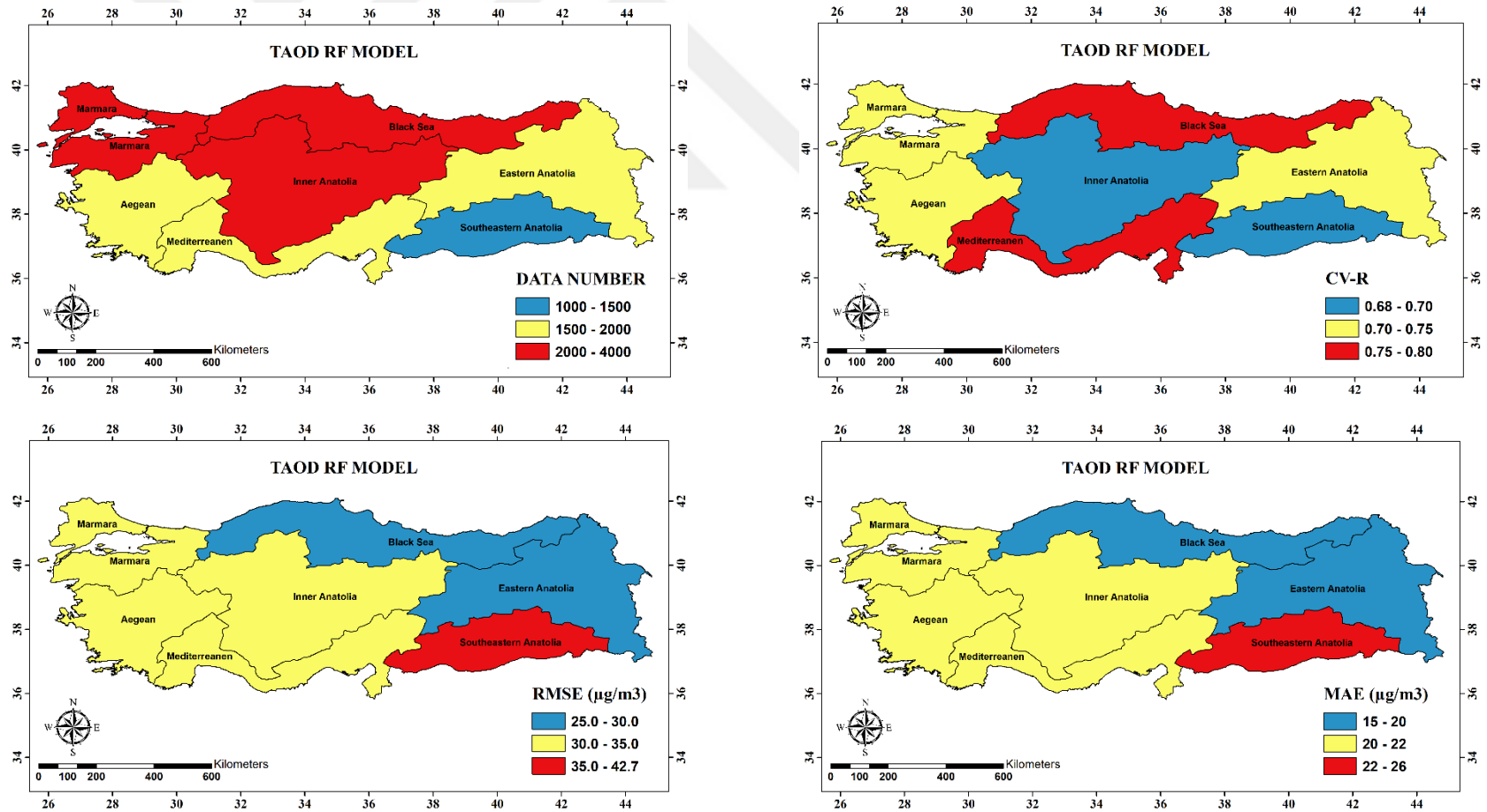


Figure 4.28 Regional TAOD RF model performance

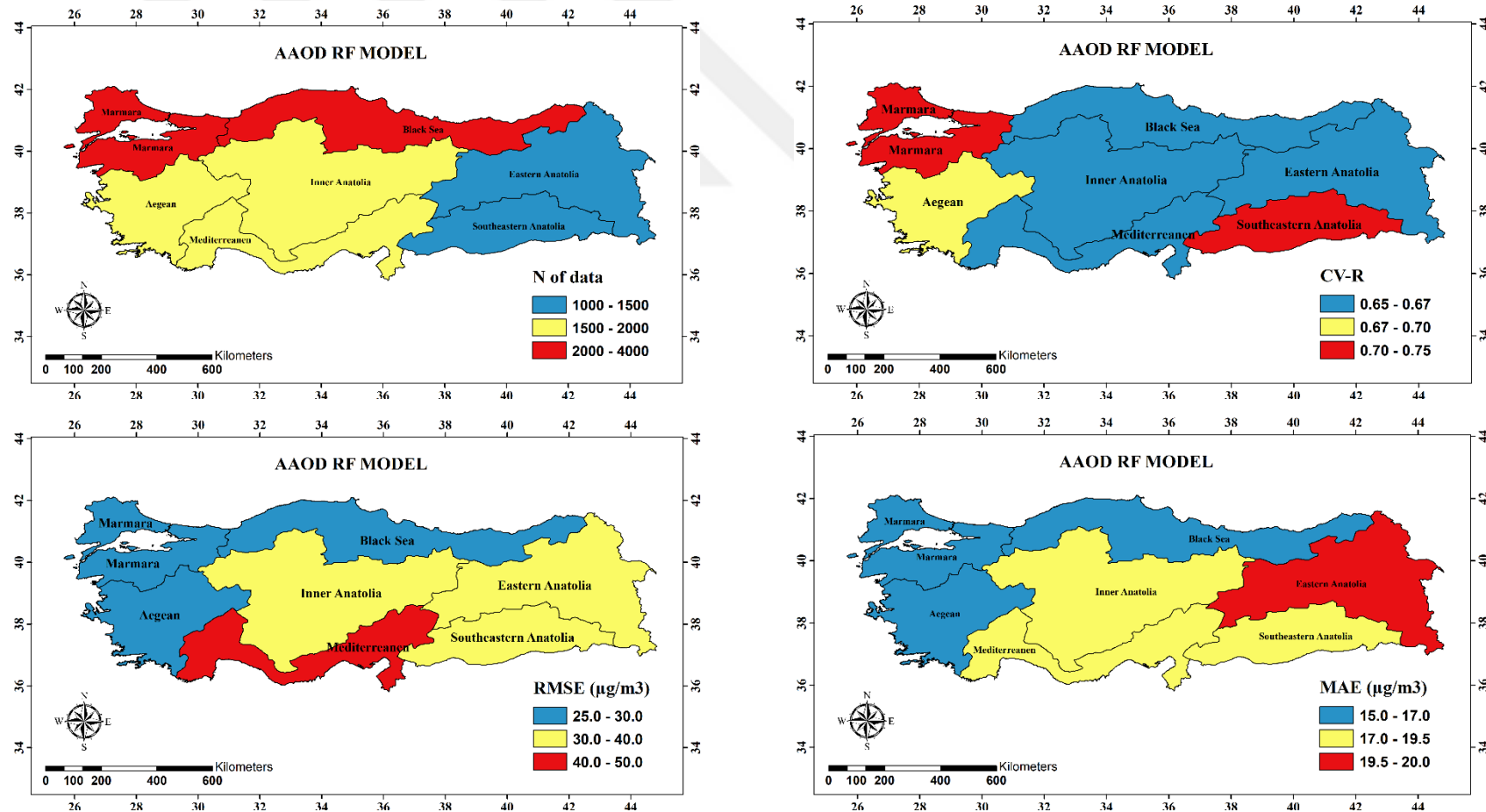


Figure 4.29 Regional AAOD RF model performance

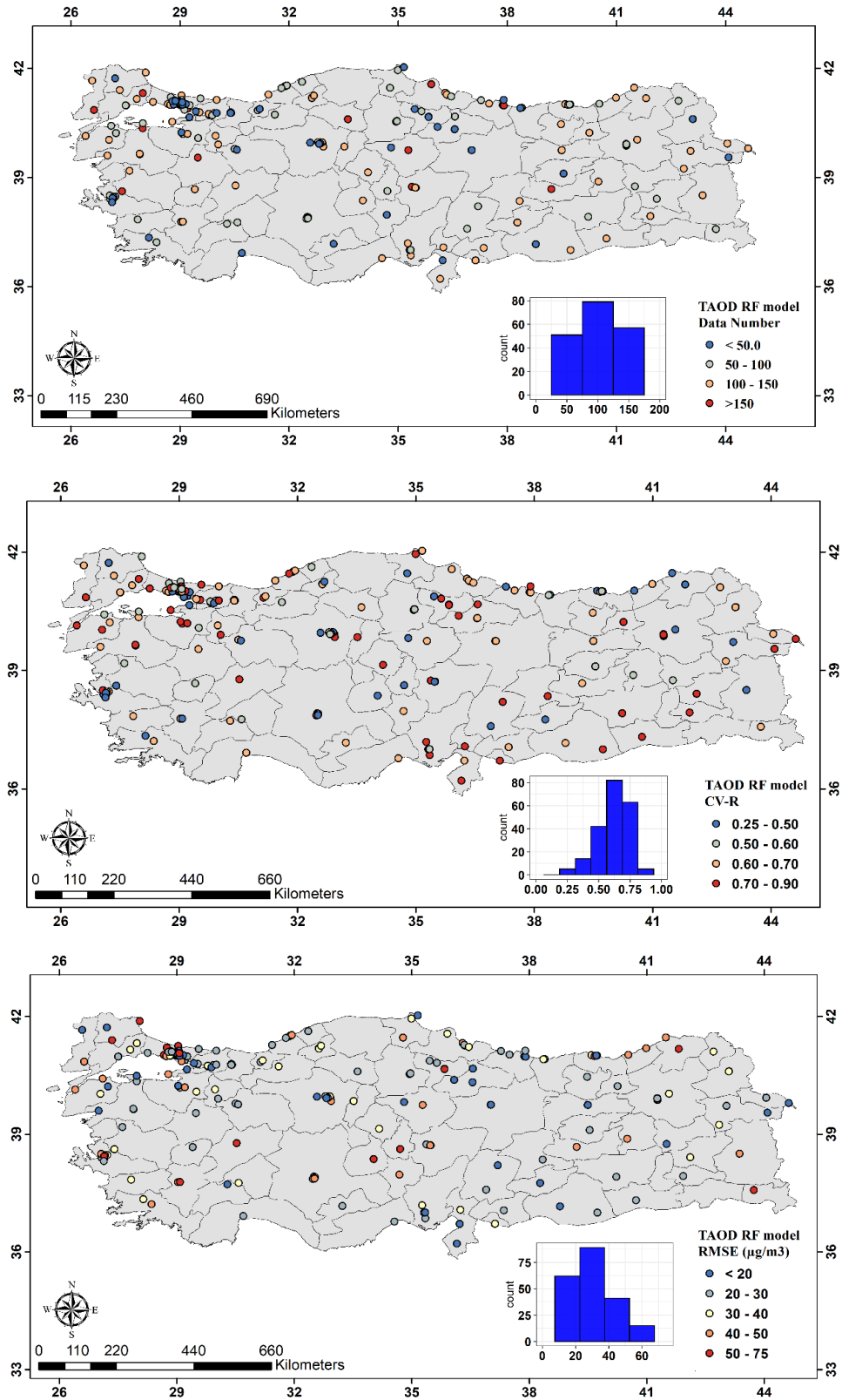


Figure 4.30 Site scale TAOD RF model performance

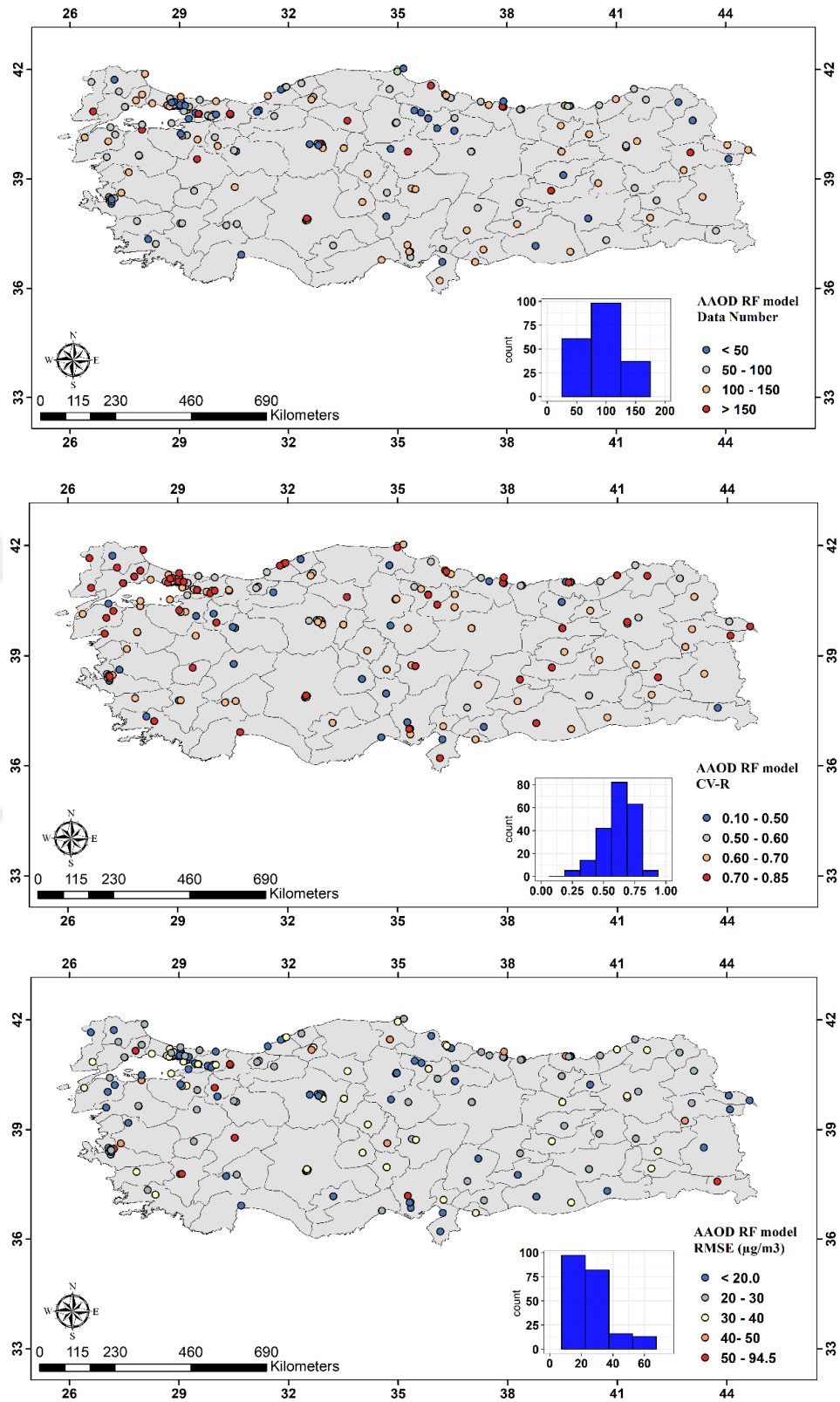


Figure 4.31 Site scale RF model performance for Aqua overpass time

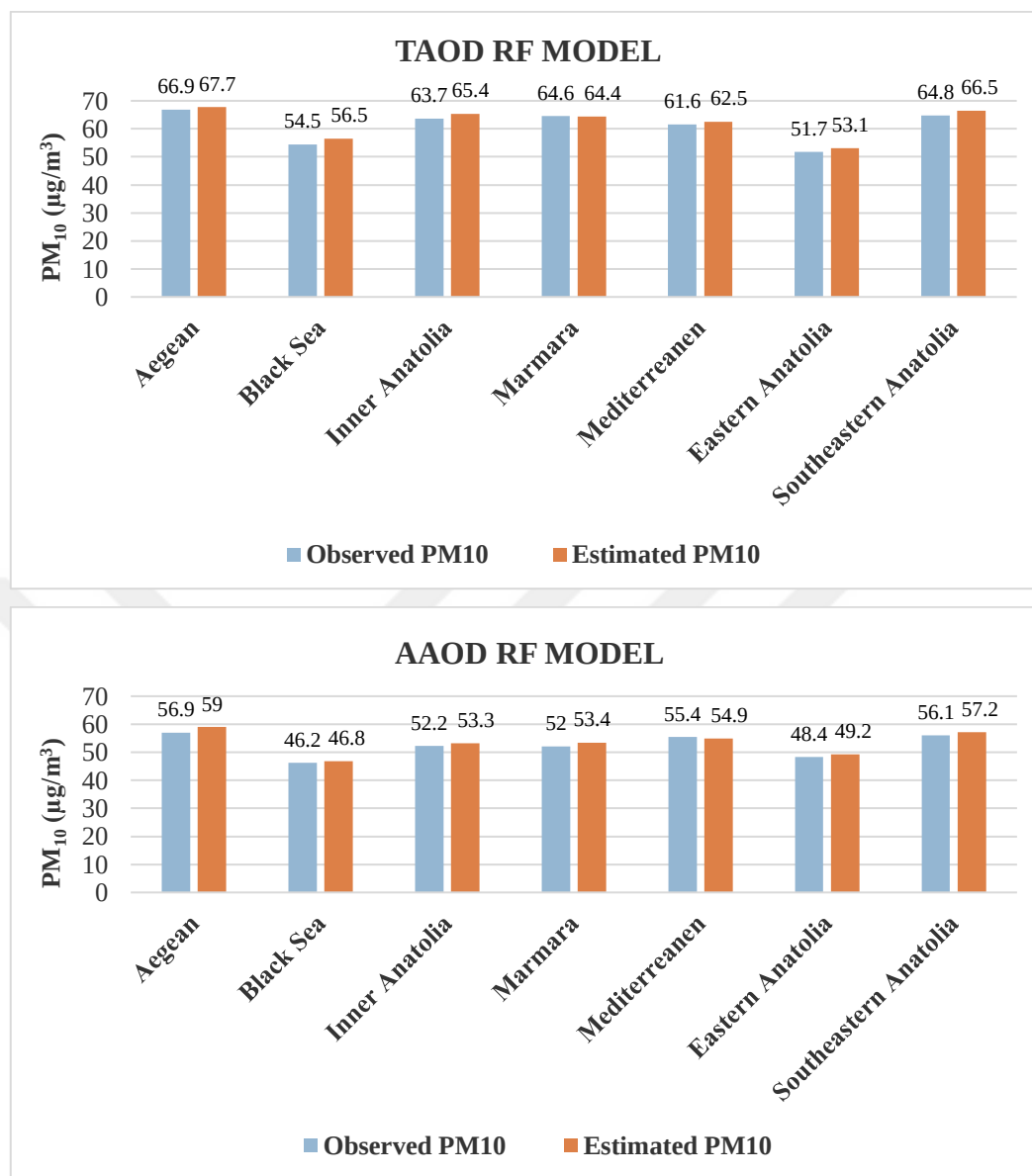


Figure 4.32 Illustration of regional estimated and observed PM₁₀

4.5.3 Seasonal Performance of RF Models

In general, the national RF model was not sufficiently estimated daily PM₁₀ concentrations at the seasonal scale, even though slightly varied model performances were found in different seasons (Table 4.12). Autumn ($R \sim 0.69$) and winter ($R \sim 0.67$) showed the highest accuracies (RMSE=39.9 $\mu\text{g}/\text{m}^3$, MAE=26 $\mu\text{g}/\text{m}^3$ and RMSE=44.6 $\mu\text{g}/\text{m}^3$, MAE=29.1 $\mu\text{g}/\text{m}^3$) for the Terra and Aqua satellite overpass times, respectively. However, summer showed the lowest accuracies ($R \sim 0.60$) with both satellite products. Summer and winter demonstrated similar performance with both

products due to lower concentrations of summer and higher concentrations in winter and autumn. Also, DB algorithm particularly underestimate AOD over in summer. In practice, MODIS AOD data often appear patchy; they can hardly retrieve AOD winter due to the minimal amount of valid data. Moreover, the MODIS AOD-PM₁₀ correlation is significantly stronger and positive in the summer months than in the winter months. The model was more stable in polluted seasons, particularly in autumn and winter. Park et al. (2019) and Staggofia et al. (2019) also found a high correlation in winter. The results were determined by seasonal differences in the correlations mainly depend on meteorological factors and PM₁₀ levels. Staggofia et al. (2019) also had the same limitations in their approach. Model performance was more unsatisfactory during the summer season in southern Italy. The lower performance of the calibration model in the summer months was also found in previous studies (Stafoggia et al., 2017; Park et al., 2019; Wei et al., 2019a) and deserves further investigation over Turkey.

Table 4.12 Seasonal performance of STRF models

Season		n	r	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	mean observed PM ₁₀ ($\mu\text{g}/\text{m}^3$)	mean predicted PM ₁₀ ($\mu\text{g}/\text{m}^3$)
TAOD	Spring	13985	0.65	35.5	23.1	61.1	63.3
	Summer	24664	0.60	25.7	15.6	48.3	49.3
	Autumn	19058	0.69	39.9	26.0	69.4	71.1
	Winter	4426	0.64	62.6	41.7	95.0	98.4
AAOD	Spring	11924	0.60	33.2	19.6	51.5	53.1
	Summer	22841	0.59	24.8	14.7	44.8	45.5
	Autumn	17336	0.65	36.1	21.1	56.0	57.4
	Winter	4226	0.67	44.6	29.1	72.2	74.2

CHAPTER FIVE

CONCLUSIONS & RECOMMENDATIONS

5.1 Comparison with other studies

Considering national calibration of the PM_{10} -AOD relationship, the best STRF model developed for entire Turkey can explain the 51% of the daily variability in the 10-fold CV with Terra AOD, surpassed previously published study (Zeydan & Wang, 2019) for estimation of the $PM_{2.5}$ concentrations over Turkey. However, regional validation results showed that model prediction ability could achieve higher values ($R^2 \sim 0.64$). Higher accuracy can be obtained by individual site scale ($R^2 \sim 0.81$).

It is not feasible to thoroughly compare the results of this study with the results from other studies. However, the results from this study have mainly lower estimation accuracy than studies in recent despite the machine learning approach. The developed STRF model in this study is mainly helpful at predicting temporal variations in PM_{10} concentrations. However, its capability to predict spatial variation in PM_{10} was weaker ($R \sim 0.36-0.47$). Since spatially, a slight variation of AOD and PM_{10} was observed over Turkey. Findings are similar to Yazdi et al. (2020). The ability of their $PM_{2.5}$ estimation model in the Greater London area to predict spatial variability was also weaker ($R^2 \sim 0.396$). Stafoggia et al. (2019) also found the lower spatial predictive ability in the RF model over Italy. However, estimation models in both studies surpassed our national model with the high spatiotemporal CV- R^2 results reported in Table 5.1.

Also, the performance of the RF model developed in the present study was comparable to that developed in Texas using the 1-km MAIAC AOD retrievals (CV $R=0.83-0.90$). Park et al. (2019) also found lower PM_{10} predictive performance ($R^2 \sim 0.58$) in South Korea before sampling adjustments in training data. Kamarul Zaman et al. (2017) developed an ANN model over Malaysia, and PM_{10} predictive power is similar to our results ($R^2=0.59$). This indicates that estimation models cannot capture random events over large areas. However, regional models with appropriate

AOD products and local auxiliary variables can benefit regions with variable PM₁₀ levels.

Unlike this study, the most recent research used the AOD from the MAIAC algorithm, which showed much more accurate validation results than the DT and DB algorithms, notably more minor estimation uncertainties, mainly over complex urban areas (Wei et al., 2019b). MAIAC AOD also has the 1-km spatial resolution higher than commonly used MODIS DT (3-10 km) and DB (10 km) products. With the help of this finer resolution, for example, Wei et al. (2021b) developed 1-km PM₁₀ estimates to provide more useful PM₁₀ distribution over China, particularly over at small and medium scales.

Table 5.1 Previous studies in the literature

Reference	Period	AOD Data	Spatial scale (approximately)	Temporal scale	Output	Method	R/R ²
Seo et al. (2015)	Dec 2011-Feb 2012	AERONE T	-	Hourly	PM ₁₀	Simple empirical model	R=0.81
Meng et al. (2016)	2004–2013	MODIS	10 km	Daily	PM ₁₀	LUR	R ² =0.79
Kamarul Zaman et al. (2017)	2007–2011	MODIS	10 km	Daily	PM ₁₀	ANN	R ² =0.59
Stafoggia et al. (2017)	2013–2015	MAIAC	1 km	Daily	PM ₁₀	Four-stage mixed model	R ² =0.58–0.72
Chen et al. (2018)	2014–2016	MODIS	10 km	Daily	PM ₁₀	RF GAM Non-linear exposure-Lag response model	R ² =0.78 R ² =0.50 R ² =0.47
Shtein et al. (2018)	2005–2015	MAIAC	1 km	Hourly	PM ₁₀	LME	R ² =0.91 (Terra) R ² =0.90 (Aqua)
Zang et al. (2018)	July 2015 - Dec 2016	AHI	5 km	Hourly	PM _{2.5}	PCA-GRNN	R ² =0.63
Ahmad et al. (2019)	2015–2017	MODIS	10 km	Daily	PM _{2.5}	MLR ANN	R=0.80

Table 5.1 continues

Park et al. (2019)	2015-2016	MODIS, GOCI	6 km	Daily	PM ₁₀	RF	R ² =0.58 (with original training samples) R ² =0.78 (with balanced training samples)
Stafoggia et al. (2019)	2013-2015	MAIAC, CAMS	1 km	Daily	PM ₁₀	RF	R ² = 0.73-0.75
Wei et al. (2019)	2016	MAIAC	1 km	Daily	PM _{2.5}	STRF	R ² =0.55
Wei et al. (2020a)	2018	MAIAC	1 km	Daily	PM _{2.5}	STET	R ² =0.89
Yazdi et al. (2020)	2005-2013	MAIAC	1 km	Daily	PM _{2.5}	Ensemble model with RF, KNN, GBM	R ² =0.82
Jiang et al. (2021)	Mar 2018- Feb 2019	AHI, MAIAC, CAMS	1 km	Hourly	PM _{2.5}	Two-stage RF model (Gap- filling: RF)	R ² =0.85
Wei et al. (2021a)	2015-2019	MAIAC	1 km	Daily	PM ₁₀	STET	R ² =0.86
This study	Jan 2008- Dec 2019	MODIS	10 km	Daily	PM₁₀	RF	R=0.72 (Terra) R=0.66 (Aqua)

5.2 Conclusions and Recommendations

Satellite data offers an excellent opportunity to support air quality management; thus, supporting ground-level measurements using remote sensing has been a progressive research area. Satellite AOD has been used by numerous studies to estimate PM_{2.5} or PM₁₀ due to the increasing air pollution over recent years (Chen et al., 2018; Huang et al., 2018; Wei et al., 2019a; Yan et al., 2020; Yazdi et al., 2020). However, most of the estimates are mainly focused on PM_{2.5} and highly polluted regions like China. According to a previous study in Turkey (Zeydan & Wang, 2019), traditional regression models restrict the PM estimation accuracy. In the present study, the improved RF models were constructed to decrease the uncertainty from

spatiotemporal variations by increasing the overall estimation performance of the ground-level PM concentrations. The national STRF model was firstly proposed in this study to estimate long-term PM₁₀ concentrations over 11 years based on AOD from different satellites in Turkey.

The PM₁₀-AOD correlations are low in Turkey. AAOD representing the early afternoon observations seems to explain better the condition of PM₁₀ than the late morning observations with TAOD. However, AOD can be an effective parameter with the help of strong local parameters for PM₁₀ estimations. A long-term dataset was collected, including ground-level PM₁₀ concentrations, AOD from two satellites, NDVI, elevation, meteorological and aerosol fields from January 2008 to December 2019 over Turkey. Then all collocated parameters were used for determining potential influencing variables from 25 independent variables based on the variable selection process. The number of independent variables was reduced, and the 13 effective parameters were selected for the final national RF model. Results indicated that the variables of surface radiation (SRAD), dust mass surface concentrations (DUSSMASS), Julian day (jday), cumulative date (cdate), surface roughness (ZOM), aerosol optical depth (AOD), pressure (P), and the normalized difference vegetation index (NDVI), latitude (Lat) and longitude (Lon) were the critical variables for PM₁₀ estimation over Turkey. Also, the variables of SRAD and DUSSMASS had the highest contribution to PM₁₀ estimation for the Terra and Aqua overpass time, respectively. Since solar radiation is affected by the concentration of air pollution and the path length of the solar beam through the urban atmosphere, higher concentrations observed in the morning hours can be more effective on the relationship between PM₁₀ and AOD. According to validation results, the Aqua satellite can capture high AOD values, and the DTB AOD from the Aqua satellite showed relatively higher performance in dusty season (spring) over coastal rural site. It means that Aqua overpass time can better represent dusty conditions over Turkey.

Then, national random forest model was applied over two different satellite overpass times to estimate daily PM₁₀ concentrations. The results indicated that the RF model, as the advanced ML method based on AOD, had a moderate capability to

estimate PM₁₀ concentrations over Turkey. However, analysis of PM₁₀ concentrations showed that both natural and anthropogenic factors could affect the spatiotemporal variation of PM₁₀ concentrations. The model performance also showed regionally different results. It was shown that the random forest model performance showed differences based on satellite overpass time.

However, the RF performance at the Terra overpass time was higher due to more significant effect on importance of auxiliary variables. The predictive performance of the models was also tested. Results indicated that the spatial predictive power of national RF models is low due to the low spatial variability of AOD and other auxiliary variables. In contrast, the final estimation model was dominated by temporal predictive power. However, none of the models was capable to fully estimate the PM₁₀ concentrations at both spatial and temporal scales. The RF models with TAOD and AAOD had similar performance and higher than the MLR model. High aerosol loadings detected by the MODIS AOD data are likely to have a higher ground-level PM₁₀ concentration. Therefore, they deserve more attention in air quality management. It was found that AOD from MODIS has significant potential in assessing moderate PM₁₀ concentrations (50th and 75th percentiles). Quadrant analysis showed that PM₁₀ and AOD agreed well over both high AOD and PM₁₀ conditions. The model was also more stable in polluted seasons, particularly in autumn and winter. However, the spatiotemporal model did not perform well in Turkey, especially over areas far from monitoring stations in the southern and eastern regions of Turkey. Although the overall performance of the model showed a moderate agreement between the estimated and observed PM₁₀ concentrations, a little higher overfitting existing in the model is mainly caused by the areas having small population and the number of monitoring stations. Under this circumstance, it is challenging to estimate high-resolution PM₁₀ concentrations at high accuracy without the overfitting problem.

The complex relationship between AOD and PM₁₀ and the accuracy of all the input parameters can cause inevitable uncertainties in satellite PM₁₀ estimates, as discussed in Liu (2015). Mainly, satellite AOD products have some limitations and disadvantages. Multiple approximations and assumptions in the AOD retrieval

algorithms involving several atmospheric parameters can cause significant uncertainties and errors in the PM estimation models. Detailed validation was proposed to compare the MODIS AOD with AERONET AOD to ensure MODIS AOD for PM₁₀ estimations. Results showed that AOD quality is low in the urban station than in the coastal rural station. However, there is not an adequate AERONET station to evaluate AOD quality over Turkey. Significant uncertainties might still exist in different regions. AOD quality is still unclear due to diverse meteorological conditions and complex aerosol types. More AERONET stations are required to improve data quality and more detailed analysis of AOD over Turkey.

AOD products with coarse spatial resolutions are an excellent candidate to monitor aerosol distributions in comprehensive coverage and large scales. However, AOD with a coarse spatial resolution can cause a decrease in the model accuracy since a large area is represented by each pixel over land surfaces. Most studies have recently preferred to use MAIAC AOD with the 1-km spatial resolution to estimate PM concentrations (Wei et al., 2019a; Stafoggia et al., 2019). Another limitation of satellite products is the non-random missingness. AOD data is mainly affected by cloud cover, bright surfaces, and high aerosol loadings incorrectly regarded as clouds (Xiao et al., 2017). High temporal resolution PM₁₀ estimation cannot be directly provided with non-randomly missing satellite data. Some studies conducted in China (Xie et al., 2015; Zheng et al., 2016) reported that missing AOD values introduce negative or positive biases in the prediction model. Since inconsistent and limited satellite measurements restrict the short-term use of AOD, a more comprehensive range of historical data may be needed in the future for detailed PM₁₀ estimation. Some studies in the literature used the ML models to fill AOD data combining collocated complementary satellite data which can increase the prediction power of the estimation model (Stafoggia et al., 2019; Jiang et al., 2021; Tuna Tuygun et al., 2021). In this study, MODIS DTB merged the AOD dataset used to improve the sampling frequency of AOD and consider the varying surface and meteorological conditions over Turkey. Secondly, used meteorological parameters and land-related data to solve the relationship between ground-level PM₁₀ and satellite AOD had different spatial

resolutions and temporal intervals. Thus, the collocation of data products can cause uncertainties in the PM_{10} estimation.

Further investigation of potential more robust model input parameters is required to increase model performances, especially in low vegetation cover and highly dust transported regions (e.g., the southern part of Turkey). Firstly, additional variables (e.g., visibility, emission inventory, population) with high spatial and temporal resolutions to overcome the complex structure of PM_{10} are strongly recommended to be integrated into the RF model. In a future study, the fractional AOD of the surface layer in the total AOD could be used as independent variables in the estimation model. More effective methods should be investigated to remove redundant variables to reduce the over-fitting and minimize noise caused by several input variables. This may increase both operation efficiency and model accuracy. Finally, both spatial and temporal predictors are crucial. Careful investigation is required before the development of the model; thus, advanced methods must be developed to improve the accuracy of PM_{10} estimates over Turkey. New approaches can also be developed in the evaluation of model performance. The performance of the prediction models is evaluated on training data. Models are evaluated on random subsets of the data in the cross-validation approach.

The dataset is randomly divided into folds without regard to which monitor, or region causes inappropriate estimation performance due to spatiotemporal autocorrelation in the dataset. Also, in remote locations far from the dense network of monitors located in urban areas, prediction accuracy will be low according to our spatial validation results. Thus, the spatial distribution of PM_{10} maps was not developed over Turkey. The basics of ML have been in progress. For better determination of the uncertainty sources, more studies should be developed over various regions. The PM_{10} estimation model developed in this study is the first step to estimate long-term PM_{10} concentrations based on AOD over Turkey. The STRF model performance in terms of estimation and predictive power is superior to the commonly preferred traditional MLR model developed in a previous study developed for $PM_{2.5}$ in Turkey (Zeydan & Wang, 2019).

In the future, the predictive power of the proposed model can be improved, and the model will also generate national or regional PM₁₀ maps instead of the site-based estimation results. Improvements can be made with the help of continuous enhancements in new satellites and satellite platforms and more advanced ML methods. According to the model evaluation results with stacking two ML method, proposed stacking models outperformed individual models (Chen et al., 2019b; Feng et al., 2020). In the future study, the stacking model may be applied to improve model performance integration of multiple ML approaches to the estimation model for Turkey. This method can affect the models in two ways by enhancing accuracy with an increased R^2 value and the robustness of the model. The spatial and temporal heterogeneity of factors for sub regional estimation can be studied to improve further accuracy and explanatory power of the model over Turkey.

REFERENCES

- Agacayak, T., Kindap, T., Unal, A., Pozzoli, L., Mallet, M., & Solmon, F. (2015). A case study for Saharan dust transport over Turkey via RegCM4.1 model. *Atmospheric Research*, 153, 392–403.
- Ahmad, M., Alam, K., Tariq, S., Anwar, S., Nasir, J., & Mansha, M. (2019). Estimating fine particulate concentration using a combined approach of linear regression and artificial neural network. *Atmospheric Environment*, 219 (2019), 117050.
- Aldabash, M., Balcik, F. B., & Glantz, P. (2020). Validation of MODIS C6.1 and MERRA-2 AOD using AERONET observations: A comparative study over Turkey. *Atmosphere*, 11(9), 905.
- Almazroui, M. (2019). A comparison study between AOD data from MODIS deep blue collections 51 and 06 and from AERONET over Saudi Arabia. *Atmospheric Research*, 225 (2019), 88–95.
- Al-Saadi, J., Szykman, J., Pierce, R. B., Kittaka, C., Neil, D., Chu, D. A., Remer, L., Gumley, L., Prins, E., Weinstock, L., MacDonald, C., Wayland, R., Dimmick, F., & Fishman, J. (2005). Improving national air quality forecasts with satellite aerosol observations. *Bulletin of the American Meteorological Society*, 86(9), 1249–1261.
- Bai, Y., Wu, L., Qin, K., Zhang, Y., Shen, Y., & Zhou, Y. (2016). A geographically and temporally weighted regression model for ground-level PM_{2.5} estimation from satellite-derived 500 m resolution AOD. *Remote Sensing*, 8(3), 8030262.
- Banerjee, T., Kumar, M., Mall, R. K., & Singh, R. S. (2017). Airing ‘clean air’ in Clean India Mission. *Environmental Science and Pollution Research*, 24(7), 6399–6413.

- Barnes, M. J., Brade, T. K., Mackenzie, A. R., Whyatt, J. D., Carruthers, D. J., Stocker, J., Cai, X., & Hewitt, C. N. (2014). Spatially varying surface roughness and ground-level air quality in an operational dispersion model. *Environmental Pollution*, 185, 44–51.
- Beloconi, A., Kamarianakis, Y., & Chrysoulakis, N. (2016). Estimating urban PM₁₀ and PM_{2.5} concentrations, based on synergistic MERIS/AATSR aerosol observations, land cover and morphology data. *Remote Sensing of Environment*, 172 (2016), 148–164.
- Benas, N., Beloconi, A., & Chrysoulakis, N. (2013). Estimation of urban PM₁₀ concentration, based on MODIS and MERIS/AATSR synergistic observations. *Atmospheric Environment*, 79(2013), 448–454.
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227.
- Bibi, H., Alam, K., Chishtie, F., Bibi, S., Shahid, I., & Blaschke, T. (2015). Intercomparison of MODIS, MISR, OMI, and CALIPSO aerosol optical depth retrievals for four locations on the Indo-Gangetic plains and validation against AERONET data. *Atmospheric Environment* 111(2015), 113–126.
- Bilal, M., Nazeer, M., Nichol, J., Qiu, Z., Wang, L., Bleiweiss, M. P., Shen, X., Campbell, J. R., & Lolli, S. (2019). Evaluation of Terra-MODIS C6 and C6.1 aerosol products against Beijing, Xiang He, and Xinglong AERONET sites in China during 2004-2014. *Remote Sensing*, 11(5), 486.
- Bilal, M., Nazeer, M., Qiu, Z., Ding, X., & Wei, J. (2018). Global validation of MODIS C6 and C6.1 merged aerosol products over diverse vegetated surfaces. *Remote Sensing*, 10(3), 475.

- Bilal, M., Nazeer, M., & Nichol, J. E. (2017). Validation of MODIS and VIIRS derived aerosol optical depth over complex coastal waters. *Atmospheric Research* 186 (2017), 43–50.
- Bilal, M., & Nichol, J. E. (2017). Evaluation of the NDVI-Based Pixel Selection Criteria of the MODIS C6 Dark Target and Deep Blue Combined Aerosol Product. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(8), 3448–3453.
- Bilal, M., Nichol, J. E., & Nazeer, M. (2016). Validation of Aqua-MODIS C051 and C006 Operational Aerosol Products Using AERONET Measurements over Pakistan. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), 2074–2080.
- Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199–215.
- Bright, J. M., & Gueymard, C. A. (2019). Climate-specific and global validation of MODIS Aqua and Terra aerosol optical depth at 452 AERONET stations. *Solar Energy*, 183(2019), 594–605.
- Brokamp, C., Jandarov, R., Hossain, M., & Ryan, P. (2018). Predicting daily urban fine particulate matter concentrations using a Random Forest model. *Environmental Science and Technology*, 52(7), 4173–4179.
- Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., Mittleman, M. A., Peters, A., Siscovick, D., Smith, S. C., Whitsel, L., & Kaufman, J. D. (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation*, 121(21), 2331–2378.

- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationary. *Geographical Analysis*, 28(4), 281–298.
- Butt, M. J., Assiri, M. E., & Ali, M. A. (2017). Assessment of AOD variability over Saudi Arabia using MODIS Deep Blue products. *Environmental Pollution*, 231, 143–153.
- Che, H., Yang, L., Liu, C., Xia, X., Wang, Y., Wang, H., Wang, H., Lu, X., & Zhang, X. (2019). Long-term validation of MODIS C6 and C6.1 Dark Target aerosol products over China using CARSNET and AERONET. *Chemosphere*, 236 (2019), 238.
- Chen, Z. Y., Zhang, T. H., Zhang, R., Zhu, Z. M., Yang, J., Chen, P. Y., Ou, C. Q., & Guo, Y. (2019a). Extreme gradient boosting model to estimate PM_{2.5} concentrations with missing-filled satellite data in China. *Atmospheric Environment*, 202(2019), 180–189.
- Chen, J., Yin, J., Zang, L., Zhang, T., & Zhao, M. (2019b). Stacking machine learning model for estimating hourly PM_{2.5} in China based on Himawari 8 aerosol optical depth data. *Science of the Total Environment*, 697(2019), 134021.
- Chen, G., Wang, Y., Li, S., Cao, W., Ren, H., Knibbs, L. D., Abramson, M. J., & Guo, Y. (2018). Spatiotemporal patterns of PM₁₀ concentrations over China during 2005–2016: A satellite-based estimation using the random forests approach. *Environmental Pollution*, 242(2018), 605–613.
- Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S., & Ghamisi, P. (2020). Spatial hazard assessment of the PM₁₀ using machine learning models in Barcelona, Spain. *Science of the Total Environment*, 701(2020), 134474.

- Chu, Y., Liu, Y., Li, X., Liu, Z., Lu, H., Lu, Y., Mao, Z., Chen, X., Li, N., Ren, M., Liu, F., Tian, L., Zhu, Z., & Xiang, H. (2016). A review on predicting ground PM_{2.5} concentration using satellite aerosol optical depth. *Atmosphere*, 7(10), 129.
- Chu, H. J., Huang, B., & Lin, C. Y. (2015). Modelling the spatio-temporal heterogeneity in the PM₁₀-PM_{2.5} relationships. *Atmospheric Environment*, 102(1), 176–182.
- Chu, D. A., Kaufman, Y. J., Zibordi, G., Chern, J. D., Mao, J., Li, C., & Holben, B. N. (2003). Global monitoring of air pollution over land from the Earth Observing System-Terra Moderate Resolution Imaging Spectroradiometer (MODIS). *Journal of Geophysical Research: Atmospheres*, 108(21), 3179.
- Chudnovsky, A. A., Kostinski, A., Lyapustin, A., & Koutrakis, P. (2013). Spatial scales of pollution from variable resolution satellite imaging. *Environmental Pollution*, 172(2013), 131–138.
- Cordero, L., Wu, Y., Gross, B. M., & Moshary, F. (2013). Assessing satellite AOD based and WRF/CMAQ output PM_{2.5} estimators. *Sensing Technologies for Global Health, Military Medicine, and Environmental Monitoring III*, 8723, 872319.
- Damascena, A. S., Yamasoe, M. A., Martins, V. S., Rosas, J., Benavente, N. R., Sánchez, M. P., Tanaka, N. I., & Saldiva, P. H. N. (2021). Exploring the relationship between high-resolution aerosol optical depth values and ground-level particulate matter concentrations in the Metropolitan Area of São Paulo. *Atmospheric Environment*, 244(2020), 117949.
- Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., & Schwartz, J. (2016). Assessing PM_{2.5} exposures with high spatiotemporal resolution across the continental United States. *Environmental Science and Technology*, 50(9), 4712–4721.

- Didan, K., Munoz, A.B., Solano, R. & Huete, A. (2015). *MODIS Vegetation Index User's Guide (MOD13 Series)*. University of Arizona: Vegetation Index and Phenology Lab. Retrieved December, 30 2020 from: https://vip.arizona.edu/MODIS_UsersGuide.php.
- Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L., & Samet, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of the American Medical Association*, 295(10), 1127–1134.
- Dubovik, O., & King, M. D. (2000). A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements. *Journal of Geophysical Research Atmospheres*, 105(D16), 20673–20696.
- Eck, T. F., Holben, B. N., Sinyuk, A., Pinker, R. T., Goloub, P., Chen, H., Chatenet, B., Li, Z., Singh, R. P., Tripathi, S. N., Reid, J. S., Giles, D. M., Dubovik, O., O'Neill, N. T., Smirnov, A., Wang, P., & Xia, X. (2010). Climatological aspects of the optical properties of fine/coarse mode aerosol mixtures. *Journal of Geophysical Research Atmospheres*, 115(19), 14002.
- Eck, T. F., Holben, B. N., Dubovik, O., Smirnov, A., Goloub, P., Chen, H. B., Chatenet, B., Gomes, L., Zhang, X. Y., Tsay, S. C., Ji, Q., Giles, D., & Slutsker, I. (2005). Columnar aerosol optical properties at AERONET sites in central eastern Asia and aerosol transport to the tropical mid-Pacific. *Journal of Geophysical Research D: Atmospheres*, 110(6), 1–18.
- Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., Slutsker, I., & Kinne, S. (1999). Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *Journal of Geophysical Research Atmospheres*, 104(24), 31333–31349.

- Elbir, T., Mangir, N., Kara, M., Simsir, S., Eren, T., Ozdemir, S. (2010). Development of a GIS-based decision support system for urban air quality management in the city of Istanbul. *Atmospheric Environment*, 44, 441-454.
- Elbir, T. & Muezzinoglu, A. (2004). Estimation of emission strengths of primary air pollutants in the city of Izmir, Turkey. *Atmospheric Environment*, 38, 1851-1857.
- Elbir, T. (2004). A GIS based decision support system for estimation, visualization and analysis of air pollution for large Turkish cities. *Atmospheric Environment*, 38, 4509-4517.
- Elbir, T., Muezzinoglu, A. & Bayram, A. (2000). Evaluation of some air pollution indicators in Turkey, *Environment International*, 26, 5-10.
- Emili, E., Popp, C., Petitta, M., Riffler, M., Wunderle, S., & Zebisch, M. (2010). PM₁₀ remote sensing from geostationary SEVIRI and polar-orbiting MODIS sensors over the complex terrain of the European Alpine region. *Remote Sensing of Environment*, 114(11), 2485–2499.
- Engel-Cox, J. A., Hoff, R. M., Rogers, R., Dimmick, F., Rush, A. C., Szykman, J. J., Al-Saadi, J., Chu, D. A., & Zell, E. R. (2006). Integrating lidar and satellite optical depth with ambient monitoring for 3-dimensional particulate characterization. *Atmospheric Environment*, 40(40), 8056–8067.
- Engel-Cox, J. A., Hoff, R. M., & Haymet, A. D. J. (2004). Recommendations on the use of satellite remote-sensing data for urban air quality. *Journal of the Air and Waste Management Association*, 54(11), 1360–1371.
- Erdun, H. Öztürk, A., Çapraz, Ö., Toros, H., Dursun, S. & Deniz, A. (2015). Spatial variation of PM₁₀ in Turkey. In 7th Atmospheric Science Symposium, Istanbul, Turkey, 311-323.

European Commission (EC) (2008). *EC Directive 2008/50/EC of the European Parliament and of the Council on Ambient Air Quality and Cleaner Air for Europe*. Brussels, Belgium.

European Commission (EC) (2007). *The Applicability of Remote Sensing in the Field of Air Pollution*. Institute for Environment and Sustainability, Italy.

European Commission (EC) (2004). *EC Directive 2004/107/EC of the European Parliament and of the Council relating to Arsenic, Cadmium, Mercury, Nickel and Polycyclic Aromatic Hydrocarbons in Ambient Air*. Brussels, Belgium.

European Commission (EC) (1999). *EC Directive 1999/30/EC of the European Parliament and of the Council Relating to Limit Values for Sulphur Dioxide, Nitrogen Dioxide and Oxides of Nitrogen, Particulate Matter and Lead in Ambient Air*. Brussels, Belgium.

European Commission (EC) (1996). *EC Directive 96/62/EC of the European Parliament and of the Council on Ambient Air Quality Assessment and Management*. Brussels, Belgium.

European Environment Agency (EEA) (2020). *Air quality in Europe — 2020 report*. EEA Report No 09/2020, Copenhagen, Denmark.

European Environment Agency (EEA) (2019). *Air quality in Europe — 2019 report*. EEA Report No 10/2019, Copenhagen, Denmark.

Fang, X., Zou, B., Liu, X., Sternberg, T., & Zhai, L. (2016). Satellite-based ground PM_{2.5} estimation using timely structure adaptive modeling. *Remote Sensing of Environment*, 186(2016), 152–163.

- Farahat, A. (2019). Comparative analysis of MODIS, MISR, and AERONET climatology over the Middle East and North Africa. *Annales Geophysicae*, 37(1), 49–64.
- Fernandes, A. P., Riffler, M., Ferreira, J., Wunderle, S., Borrego, C., & Tchepel, O. (2015). Comparisons of aerosol optical depth provided by SEVIRI satellite observations and CAMx air quality modelling. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 40, 187–193.
- Feng, L., Li, Y., Wang, Y., & Du, Q. (2020). Estimating hourly and continuous ground-level PM_{2.5} concentrations using an ensemble learning algorithm: The ST-stacking model. *Atmospheric Environment*, 223(2019), 117242.
- Feng, W., Li, H., Wang, S., Van Halm-Lutterodt, N., An, J., Liu, Y., Liu, M., Wang, X., & Guo, X. (2019). Short-term PM₁₀ and emergency department admissions for selective cardiovascular and respiratory diseases in Beijing, China. *Science of the Total Environment*, 657(2019), 213–221.
- Filonchyk, M., Hurynovich, V., & Yan, H. (2020). Trends in aerosol optical properties over Eastern Europe based on MODIS-Aqua. *Geoscience Frontiers*, 11(6), 2169–2181.
- Fine, P. M., Sioutas, C., & Solomon, P. A. (2008). Secondary particulate matter in the United States: Insights from the particulate matter supersites program and related studies. In *Journal of the Air and Waste Management Association*, 58(2), 234-253.
- Floutsi, A. A., Korras-Carraca, M. B., Matsoukas, C., Hatzianastassiou, N., & Biskos, G. (2016). Climatology and trends of aerosol optical depth over the Mediterranean basin during the last 12 years (2002-2014) based on Collection 006 MODIS-Aqua data. *Science of the Total Environment*, 551–552, 292–303.

- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., ... & Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454.
- Georgoulas, A. K., Alexandri, G., Kourtidis, K. A., Lelieveld, J., Zanis, P., Pöschl, U., Levy, R., Amiridis, V., Marinou, E., & Tsikerdekis, A. (2016). Spatiotemporal variability and contribution of different aerosol types to the aerosol optical depth over the Eastern Mediterranean. *Atmospheric Chemistry and Physics*, 16(21), 13853–13884.
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63, 3–42.
- Ghahremanloo, M., Choi, Y., Sayeed, A., Salman, A. K., Pan, S., & Amani, M. (2021). Estimating daily high-resolution PM_{2.5} concentrations over Texas: Machine Learning approach. *Atmospheric Environment*, 247(2021), 118209.
- Ghotbi, S., Sotoudeheian, S., & Arhami, M. (2016). Estimating urban ground-level PM₁₀ using MODIS 3km AOD product and meteorological parameters from WRF model. *Atmospheric Environment*, 141(2016), 333–346.
- Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., & Lyapustin, A. I. (2019). Advancements in the Aerosol Robotic Network (AERONET) Version 3 database - Automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmospheric Measurement Techniques*, 12(1), 169–209.

- Glantz, P., Freud, E., Johansson, C., Noone, K. J., & Tesche, M. (2019). Trends in MODIS and AERONET derived aerosol optical thickness over Northern Europe. *Tellus, Series B: Chemical and Physical Meteorology*, 71(1), 1–21.
- Guo, B., Wang, X., Pei, L., Su, Y., Zhang, D., & Wang, Y. (2021). Identifying the spatiotemporal dynamic of PM_{2.5} concentrations at multiple scales using geographically and temporally weighted regression model across China during 2015–2018. *Science of the Total Environment*, 751(2021), 141765.
- Guo, J., Xia, F., Zhang, Y., Liu, H., Li, J., Lou, M., He, J., Yan, Y., Wang, F., Min, M., & Zhai, P. (2017a). Impact of diurnal variability and meteorological factors on the PM_{2.5} - AOD relationship: Implications for PM_{2.5} remote sensing. *Environmental Pollution*, 221, 94–104.
- Guo, Y., Tang, Q., Gong, D. Y., & Zhang, Z. (2017b). Estimating ground-level PM_{2.5} concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sensing of Environment*, 198, 140–149.
- Guo, J., Miao, Y., Zhang, Y., Liu, H., Li, Z., Zhang, W., He, J., Lou, M., Yan, Y., Bian, L., & Zhai, P. (2016). The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data. *Atmospheric Chemistry and Physics*, 16(20), 13309–13319.
- Gupta, P., Remer, L. A., Levy, R. C., & Mattoo, S. (2018). Validation of MODIS 3km land aerosol optical depth from NASA's EOS Terra and Aqua missions. *Atmospheric Measurement Techniques*, 11(5), 3145–3159.
- Gupta, P., Levy, R. C., Mattoo, S., Remer, L. A., & Munchak, L. A. (2016). A surface reflectance scheme for retrieving aerosol optical depth over urban surfaces in MODIS Dark Target retrieval algorithm. *Atmospheric Measurement Techniques*, 9(7), 3293–3308.

- Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *Journal of Geophysical Research Atmospheres*, 114(14), D011496.
- Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, 40(30), 5880–5892.
- Hadjimitsis, D. G., & Clayton, C. (2009). Assessment of temporal variations of water quality in inland water bodies using atmospheric corrected satellite remotely sensed image data. *Environmental Monitoring and Assessment*, 159(1–4), 281–292.
- He, Q., & Huang, B. (2018). Satellite-based mapping of daily high-resolution ground PM_{2.5} in China via space-time regression modelling. *Remote Sensing of Environment*, 206(2018), 72–83.
- He, Q., Zhang, M., Huang, B., & Tong, X. (2017). MODIS 3 km and 10 km aerosol optical depth for China: Evaluation and comparison. *Atmospheric Environment*, 153(2017), 150–162.
- He, C., Liou, K.-N., Takano, Y., Zhang, R., Levy Zamora, M. & Yang, P. (2015). Variation of the radiative properties during black carbon aging: Theoretical and experimental intercomparison. *Atmospheric Chemistry and Physics*, 15(20), 11,967–11,980.
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modelling of spatial and spatio-temporal variables. *PeerJ*, 2018(8), e5518.

- Herman, J. R., Bhartia, P. K., Torres, O., Hsu, C., Seftor, C., & Celarier, E. (1997). Global distribution of UV-absorbing aerosols from Nimbus 7/TOMS data. *Journal of Geophysical Research Atmospheres*, 102(14), 16911–16922.
- Hoff, R. M., & Christopher, S. A. (2009). Remote sensing of particulate pollution from space: Have we reached the promised land? In *Journal of the Air and Waste Management Association*, 59(6), 645–675.
- Holben, B. N., Tanré, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., Newcomb, W. W., Schafer, J. S., Chatenet, B., Lavenu, F., Kaufman, Y. J., Vande Castle, J., Setzer, A., Markham, B., Clark, D., Frouin, R., Halthore, R., Karneli, A., O'Neill, N. T., ... & Zibordi, G. (2001). An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. *Journal of Geophysical Research Atmospheres*, 106(D11), 12067–12097.
- Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., & Smirnov, A. (1998). AERONET - A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, 66(1), 1–16.
- Horne, J. R., & Dabdub, D. (2017). Impact of global climate change on ozone, particulate matter, and secondary organic aerosol concentrations in California: A model perturbation analysis. *Atmospheric Environment*, 153, 1–17.
- Houthuijs, D., Breugelmans, O., Hoek, G., Vaskövi, É., Miháliková, E., Pastuszka, J. S., Jirik, V., Sachelarescu, S., Lolova, D., Meliefste, K., Uzunova, E., Marinescu, C., Volf, J., De Leeuw, F., Van De Wiel, H., Fletcher, T., Lebret, E., & Brunekreef, B. (2001). PM₁₀ and PM_{2.5} concentrations in Central and Eastern Europe: Results from the Cesar study. *Atmospheric Environment*, 35(15), 2757–2771.

- Hsu, N. C., Jeong, M. J., Bettenhausen, C., Sayer, A. M., Hansell, R., Seftor, C. S., Huang, J., & Tsay, S. C. (2013). Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research Atmospheres*, 118(16), 9296–9315.
- Hsu, N. C., Tsay, S. C., King, M. D., & Herman, J. R. (2004). Aerosol properties over bright-reflecting source regions. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3), 557–569.
- Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., & Liu, Y. (2017). Estimating PM_{2.5} Concentrations in the Conterminous United States Using the Random Forest Approach. *Environmental Science and Technology*, 51(12), 6936–6944.
- Hu, X. M., Ma, Z. Q., Lin, W., Zhang, H., Hu, J., Wang, Y., Xu, X., Fuentes, J. D., & Xue, M. (2014a). Impact of the Loess Plateau on the atmospheric boundary layer structure and air quality in the North China Plain: A case study. *Science of the Total Environment*, 499, 228–237.
- Hu, X., Waller, L., Lyapustin, A., Wang, Y., & Liu, Y. (2014b). 10-year spatial and temporal trends of PM_{2.5} concentrations in the southeastern US estimated using high-resolution satellite data. *Atmospheric Chemistry and Physics*, 14, 6301–6314.
- Huang, Y., Zhu, B., Zhou, X., Chen, D., Zhu, Z., Zhang, T., Gong, W., Ji, Y., Xia, X., & Wang, L. (2019). Evaluation and comparison of MODIS collection 6.1 and collection 6 dark target aerosol optical depth over mainland China under various conditions including spatiotemporal distribution, haze effects, and underlying surface. *Earth and Space Science*, 6(12), 2575–2592.

- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D., & Liu, Y. (2018). Predicting monthly high-resolution PM_{2.5} concentrations with random forest model in the North China Plain. *Environmental Pollution*, 242, 675–683.
- Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383–401.
- Ichoku, C., Allen Chu, D., Mattoo, S., Kaufman, Y. J., Remer, L. A., Tanré, D., Slutsker, I., & Holben, B. N. (2002). A spatio-temporal approach for global validation and analysis of MODIS aerosol products. *Geophysical Research Letters*, 29(12), 1-4.
- Ignatov, A., & Stowe, L. (2002). Aerosol retrievals from individual AVHRR channels. Part I: Retrieval algorithm and transition from Dave to 6S radiative transfer model. *Journal of the Atmospheric Sciences*, 59(3), 313–334.
- Jethva, H., Torres, O., & Yoshida, Y. (2019). Accuracy assessment of MODIS land aerosol optical thickness algorithms using AERONET measurements over North America. *Atmospheric Measurement Techniques*, 12(8), 4291–4307.
- Jiang, T., Chen, B., Nie, Z., Ren, Z., Xu, B., & Tang, S. (2021). Estimation of hourly full-coverage PM_{2.5} concentrations at 1-km resolution in China using a two-stage random forest model. *Atmospheric Research*, (2021), 248, 105146.
- Jin, Q., Crippa, P., & Pryor, S. C. (2020). Spatial characteristics and temporal evolution of the relationship between PM_{2.5} and aerosol optical depth over the eastern USA during 2003–2017. *Atmospheric Environment*, 239(2020), 117718.
- Just, A. C., Arfer, K. B., Rush, J., Dorman, M., Shtein, A., Lyapustin, A., & Kloog, I. (2020). Advancing methodologies for applying machine learning and

- evaluating spatiotemporal models of fine particulate matter (PM_{2.5}) using satellite data over large regions. *Atmospheric Environment*, 239 (2020), 117649.
- Kabatas, B., Unal, A., Pierce, R. B., Kindap, T., & Pozzoli, L. (2014). The contribution of Saharan dust in PM₁₀ concentration levels in Anatolian Peninsula of Turkey. *Science of the Total Environment*, 488–489(1), 413–421.
- Kacenelenbogen, M., Vaughan, M. A., Redemann, J., Hoff, R. M., Rogers, R. R., Ferrare, R. A., Russell, P. B., Hostetler, C. A., Hair, J. W., & Holben, B. N. (2011). An accuracy assessment of the CALIOP/CALIPSO version 2/version 3 daytime aerosol extinction product based on a detailed multi-sensor, multi-platform case study. *Atmospheric Chemistry and Physics*, 11(8), 3981–4000.
- Kamarul Zaman, N. A. F., Kanniah, K. D., & Kaskaoutis, D. G. (2017). Estimating particulate matter using satellite-based aerosol optical depth and meteorological variables in Malaysia. *Atmospheric Research*, 193, 142–162.
- Kanakidou, M., Mihalopoulos, N., Kindap, T., Im, U., Vrekoussis, M., Gerasopoulos, E., Dermitzaki, E., Unal, A., Koçak, M., Markakis, K., Melas, D., Kouvarakis, G., Youssef, A. F., Richter, A., Hatzianastassiou, N., Hilboll, A., Ebojie, F., Wittrock, F., von Savigny, C., ... & Moubasher, H. (2011). Megacities as hot spots of air pollution in the East Mediterranean. *Atmospheric Environment*, 45(6), 1223–1235.
- Kara, M., Dumanoglu, Y., Altiok, H., Elbir, T., Odabasi, M., Bayram, A. & Hopke, P.K. (2015). Characterization of PM using multiple site data in a heavily industrialized region of Turkey. *Aerosol and Air Quality Research*, 15, 11-27.
- Kara, M., Mangir, N., Bayram, A. & Elbir, T. (2014). A spatially high resolution and activity-based emissions inventory for the metropolitan area of Istanbul, Turkey. *Aerosol and Air Quality Research*, 14, 10-20.

- Karaca, F. (2012). Determination of air quality zones in Turkey. *Journal of the Air and Waste Management Association*, 62(4), 408–419.
- Kaufman, Y. J., Tanré, D., Remer, L. A., Vermote, E. F., Chu, A., & Holben, B. N. (1997). Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *Journal of Geophysical Research Atmospheres*, 102(14), 17051–17067.
- Kendall, M., Pala, K., Ucakli, S., & Gucer, S. (2011). Airborne particulate matter (PM_{2.5} and PM₁₀) and associated metals in urban Turkey. *Air Quality, Atmosphere and Health*, 4(3), 235–242.
- Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B., Wang, Y., Just, A. C., Schwartz, J., & Broday, D. M. (2015). Estimating daily PM_{2.5} and PM₁₀ across the complex geo-climate region of Israel using MAIAC satellite based AOD data. *Atmospheric Environment*, 122(2015), 409–416.
- Kloog, I., Melly, S. J., Ridgway, W. L., Coull, B. A., & Schwartz, J. (2012). Using new satellite-based exposure methods to study the association between pregnancy PM_{2.5} exposure, premature birth and birth weight in Massachusetts. *Environmental Health: A Global Access Science Source*, 11(1), 11-40.
- Koçak, M., Mihalopoulos, N., & Kubilay, N. (2007). Contributions of natural sources to high PM₁₀ and PM_{2.5} events in the eastern Mediterranean. *Atmospheric Environment*, 41(18), 3806–3818.
- Koelemeijer, R. B. A., Homan, C. D., & Matthijsen, J. (2006). Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmospheric Environment*, 40(27), 5304–5315.
- Kokhanovsky, A. A., Breon, F. M., Cacciari, A., Carboni, E., Diner, D., Di Nicolantonio, W., Grainger, R. G., Grey, W. M. F., Höller, R., Lee, K. H., Li, Z.,

- North, P. R. J., Sayer, A. M., Thomas, G. E., & von Hoyningen-Huene, W. (2007). Aerosol remote sensing over land: A comparison of satellite retrievals using different algorithms and instruments. *Atmospheric Research*, 85(3–4), 372–394.
- Kong, L., Xin, J., Zhang, W., & Wang, Y. (2016). The empirical correlations between PM_{2.5}, PM₁₀ and AOD in the Beijing metropolitan region and the PM_{2.5}, PM₁₀ distributions retrieved by MODIS. *Environmental Pollution*, 216, 350–360.
- Kubilay, N. & Koçak, M. (2009). *Doğu Akdeniz Bölgesindeki Aerosollerin İklimsel Önemi*. Tübitak Projesi Raporu, Proje No: 105Y368-GSRT, Mersin.
- Kutiel, H. & Türkeş, M. (2005). New evidence about the role of the North Sea – Caspian Pattern (NCP) on the temperature and precipitation regimes in continental central Turkey. *Geografiska Annaler: Series A, Physical Geography*, 87, 501-513.
- Lanzaco, B. L., Olcese, L. E., Palancar, G. G., & Toselli, B. M. (2016). A method to improve MODIS AOD values: Application to South America. *Aerosol and Air Quality Research*, 16(6), 1509–1522.
- Lee, H. J., Chatfield, R. B., & Strawa, A. W. (2016a). Enhancing the Applicability of Satellite Remote Sensing for PM_{2.5} Estimation Using MODIS Deep Blue AOD and Land Use Regression in California, United States. *Environmental Science and Technology*, 50(12), 6546–6555.
- Lee, M., Kloog, I., Chudnovsky, A., Lyapustin, A., Wang, Y., Melly, S., Coull, B., Koutrakis, P., & Schwartz, J. (2016b). Spatiotemporal prediction of fine particulate matter using high-resolution satellite images in the South-eastern US 2003-2011. In *Journal of Exposure Science and Environmental Epidemiology*, 26(4), 377–384.

- Lee, H. J., Kang, C. M., Coull, B. A., Bell, M. L., & Koutrakis, P. (2014). Assessment of primary and secondary ambient particle trends using satellite aerosol optical depth and ground speciation data in the New England region, United States. *Environmental Research*, 133, 103–110.
- Lee, H. J., Liu, Y., Coull, B. A., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict PM_{2.5} concentrations. *Atmospheric Chemistry and Physics*, 11(15), 7991–8002.
- Lelieveld, J., Berresheim, H., Borrmann, S., Crutzen, P. J., Dentener, F. J., Fischer, H., Feichter, J., Flatau, P. J., Heland, J., Holzinger, R., Kormann, R., Lawrence, M. G., Levin, Z., Markowicz, K. M., Mihalopoulos, N., Minikin, A., Ramanathan, V., De Reus, M., Roelofs, G. J., ... & Ziereis, H. (2002). Global air pollution crossroads over the Mediterranean. *Science*, 298(5594), 794 – 799.
- Levy, R. C., Mattoo, S., Sawyer, V., Shi, Y., Colarco, P. R., Lyapustin, A. I., Wang, Y., & Remer, L. A. (2018). Exploring systematic offsets between aerosol products from the two MODIS sensors. *Atmospheric Measurement Techniques*, 11(7), 4073–4092.
- Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., & Hsu, N. C. (2013). The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, 6(11), 2989–3034.
- Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., & Eck, T. F. (2010). Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *Atmospheric Chemistry and Physics*, 10(21), 10399–10420.
- Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., & Kaufman, Y. J. (2007). Second-generation operational algorithm: Retrieval of aerosol properties over

land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *Journal of Geophysical Research Atmospheres*, 112, D00781.

Li, L. (2020). A robust deep learning approach for spatiotemporal estimation of Satellite AOD and PM_{2.5}. *Remote Sensing*, 12(2), 264.

Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating Ground-Level PM_{2.5} by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach. *Geophysical Research Letters*, 44(23), 11,985-11,993.

Li, Z., Zhao, X., Kahn, R., Mishchenko, M., Remer, L., Lee, K. H., Wang, M., Laszlo, I., Nakajima, T., & Maring, H. (2009). Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: A review and perspective. *Annales Geophysicae*, 27(7), 755–2770.

Li, R. R., Remer, L., Kaufman, Y. J., Mattoo, S., Gao, B. C., & Vermote, E. (2005). Snow and ice mask for the MODIS aerosol products. *IEEE Geoscience and Remote Sensing Letters*, 2(3), 306–310.

Liaw A, & Wiener M. (2002). Classification and Regression by Random Forest. *R News*, 2/3, 18–22.

Lin, C. A., Chen, Y. C., Liu, C. Y., Chen, W. T., Seinfeld, J. H., & Chou, C. C. K. (2019). Satellite-derived correlation of SO₂, NO₂, and aerosol optical depth with meteorological conditions over East Asia from 2005 to 2015. *Remote Sensing*, 11(15), 1738.

Liu, J., Weng, F., & Li, Z. (2019). Satellite-based PM_{2.5} estimations directly from reflectance at the top of the atmosphere using a machine learning algorithm. *Atmospheric Environment*, 208(2019), 113–122.

- Liu, Y. (2015). *Particulate Matter Air Quality from Space – Advanced Statistical Modeling*. Retrieved January 1, 2020 from ParticulateMatter_Part3.pdf (nasa.gov).
- Liu, Y., Franklin, M., Kahn, R., & Koutrakis, P. (2007). Using aerosol optical thickness to predict ground-level PM_{2.5} concentrations in the St. Louis area: A comparison between MISR and MODIS. *Remote Sensing of Environment*, 107, 33-44.
- Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J., & Koutrakis, P. (2005). Estimating ground-level PM_{2.5} in the eastern United States using satellite remote sensing. *Environmental Science and Technology*, 39(9), 3269–3278.
- Liu, Y., Park, R. J., Jacob, D. J., Li, Q., Kilaru, V., & Sarnat, J. A. (2004). Mapping annual mean ground-level PM_{2.5} concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. *Journal of Geophysical Research D: Atmospheres*, 109(22), 1–10.
- Loria-Salazar, S. M. (2018). *Determining the effectiveness of satellite remote sensing products to quantify aerosol pollution in the semi-arid western U.S.A.* PhD Thesis, The Graduate School, USA.
- Lu, J., Zhang, Y., Chen, M., Wang, L., Zhao, S., Pu, X., & Chen, X. (2021). Estimation of monthly 1 km resolution PM_{2.5} concentrations using a random forest model over “2 + 26” cities, China. *Urban Climate* (2021), 100734.
- Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R., & Reid, J. S. (2011). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research Atmospheres*, 116(3), 14986.

- Ma, J., Yu, Z., Qu, Y., Xu, J., & Cao, Y. (2020). Application of the XGBoost machine learning method in PM_{2.5} prediction: A case study of shanghai. *Aerosol and Air Quality Research*, 20(1), 128–138.
- Ma, Z., Hu, X., Sayer, A. M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L., & Liu, Y. (2016). Satellite-based spatiotemporal trends in PM_{2.5} concentrations: China, 2004-2013. *Environmental Health Perspectives*, 124(2), 184–192.
- Marchetti, S., Longhin, E., Bengalli, R., Avino, P., Stabile, L., Buonanno, G., Colombo, A., Camatini, M., & Mantecca, P. (2019). In vitro lung toxicity of indoor PM₁₀ from a stove fueled with different biomasses. *Science of the Total Environment*, 649(2019), 1422-1433.
- Martin, R. V. (2008). Satellite remote sensing of surface air quality. *Atmospheric Environment*, 42(34), 7823–7843.
- Martins, V. S., Lyapustin, A., De Carvalho, L. A. S., Barbosa, C. C. F., & Novo, E. M. L. M. (2017). Validation of high-resolution MAIAC aerosol product over South America. *Journal of Geophysical Research*, 122(14), 7537–7559.
- Mathur, R. (2008). Estimating the impact of the 2004 Alaskan forest fires on episodic particular matter pollution over the eastern United States through assimilation of satellite-derived aerosol optical depths in a regional air quality model. *Journal of Geophysical Research Atmospheres*, 113(17), D17302.
- Mattoo, S. (2017). Aerosol dark target (10km & 3km) collection 6.1 changes. Retrieved February 1, 2021, from https://modisatmosphere.gsfc.nasa.gov/sites/default/files/ModAtmo/C061_Aerosol_Dark_Target_v2.pdf.

- Mehdipour, V., Stevenson, D. S., Memarianfard, M., & Sihag, P. (2018). Comparing different methods for statistical modeling of particulate matter in Tehran, Iran. *Air Quality, Atmosphere and Health*, 11, (10), 1155–1165.
- Meng, X., Liu, C., Zhang, L., Wang, W., Stowell, J., Kan, H., & Liu, Y. (2021). Estimating PM_{2.5} concentrations in North-eastern China with full spatiotemporal coverage, 2005–2016. *Remote Sensing of Environment*, 253(2021), 112203.
- Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., Xue, W., Wang, J., Wang, D., Kan, H., & Liu, Y. (2016). Estimating ground-level PM₁₀ in a Chinese city by combining satellite data, meteorological information and a land use regression model. *Environmental Pollution*, 208(2016), 177–184.
- Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling and Software*, 101(2018), 1–9.
- Mhawish, A., Banerjee, T., Broday, D. M., Misra, A., & Tripathi, S. N. (2017). Evaluation of MODIS Collection 6 aerosol retrieval algorithms over Indo-Gangetic Plain: Implications of aerosol types and mass loading. *Remote Sensing of Environment*, 201(2017), 297–313.
- Ministry of Environment & Urbanization (MEU) (2020). *Permanent Monitoring Center*. Retrieved December 30, 2020, from <https://www.havaizleme.gov.tr/>.
- Ministry of Environment & Urbanization (MEU) (2019). *National Air Quality Network*. Retrieved December 30, 2020 from <https://sim.csb.gov.tr/Intro/Uhkia>.
- Mirzaei, M., Bertazzon, S., & Couloigner, I. (2018). Modelling wildfire smoke pollution by integrating land use regression and remote sensing data: Regional

multi-temporal estimates for public health and exposure models. *Atmosphere*, 9(9), 9090335.

Muezzinoglu, A., Elbir, T. & Bayram, A. (2003). Air quality management in Izmir region of Turkey as required by clean air plans. *Water, Air and Soil Pollution: Focus*, 3, 307-323.

Munchak, L. A., Levy, R. C., Mattoo, S., Remer, L. A., Holben, B. N., Schafer, J. S., Hostetler, C. A., & Ferrare, R. A. (2013). MODIS 3 km aerosol product: Applications over land in an urban/suburban region. *Atmospheric Measurement Techniques*, 6(7), 1747–1759.

Nabavi, S.O, Haimbergera, L., & Abbasib, E. (2019). Assessing PM_{2.5} concentrations in Tehran, Iran, from space using MAIAC, deep blue, and dark target AOD and machine learning algorithms. *Atmospheric Pollution Research*, 10(2019), 889-903.

National Aeronautics and Space Administration (NASA) (2021a). *What is Remote Sensing*. Retrieved April 1, 2021 from <https://earthdata.nasa.gov/learn/backgrounders/remote-sensing#:~:text=Remote%20sensing%20is%20the%20acquiring,record%20reflected%20or%20emitted%20energy>.

National Aeronautics and Space Administration (NASA) (2021b). *MODIS standard collection 6.1 update*. Retrieved February 1, 2021, from <https://atmosphere-imager.gsfc.nasa.gov/documentation/collection-61>.

National Oceanic and Atmospheric Administration (NOAA) (2021). *SURFRAD Aerosol Optical Depth*. Retrieved March 1, 2021, from <https://www.esrl.noaa.gov/gmd/grad/surfrad/aod/index.html>.

National Aeronautics and Space Administration (NASA) (2020). *Aerosol Robotic Network*, Retrieved December 1, 2020, from <https://aeronet.gsfc.nasa.gov/>.

National Aeronautics and Space Administration (NASA) (2018a). *New and Upcoming Satellite Capabilities for Air Quality Monitoring. Data Analysis Tools for High Resolution Air Quality Satellite Datasets*. Retrieved February 1, 2021, https://appliedsciences.nasa.gov/sites/default/files/2020-11/HighResAQ_Part3b.pdf.

National Aeronautics and Space Administration (NASA) (2018b). *Theoretical Basis for Converting Satellite Observations to Ground-Level PM_{2.5} Concentrations*. Retrieved December 30, 2020, from D2P3_aod_pm25 (nasa.gov).

National Aeronautics and Space Administration (NASA) (2016). *Introduction to Satellite Remote Sensing for Air Quality Applications. Applied Remote Sensing Training Webinar Notes*. Retrieved September 1, 2020 from <https://arset.gsfc.nasa.gov/>.

National Aeronautics and Space Administration (NASA) (2009). *Catalog of Earth Satellite Orbits*, Retrieved September 1, 2018, from <https://earthobservatory.nasa.gov/Features/OrbitsCatalog/>.

Natural Resources of Canada (NRC) (2015). *Fundamentals of Remote Sensing*. Canada Center for Remote Sensing, Canada.

Nichol, J. E., & Bilal, M. (2016). Validation of MODIS 3 km resolution aerosol optical depth retrievals over Asia. *Remote Sensing*, 8(4), 8040328.

Nordio, F., Kloog, I., Coull, B. A., Chudnovsky, A., Grillo, P., Bertazzi, P. A., Baccarelli, A. A., & Schwartz, J. (2013). Estimating spatio-temporal resolved PM₁₀ aerosol mass concentrations using MODIS satellite data and land use regression over Lombardy, Italy. *Atmospheric Environment*, 74(2013), 227–236.

- O'Neill, N., Eck, T., Smirnov, A., Holben, B. & Thulasiraman, S. (2005). *Spectral Deconvolution Algorithm Technical Memo*. Retrieved January 1, 2020 from https://aeronet.gsfc.nasa.gov/new_web/PDF/tauf_tauc_technical_memo1.pdf.
- Ozdemir, E., Tuna Tuygun, G., & Elbir, T. (2020). Application of aerosol classification methods based on AERONET version 3 product over eastern Mediterranean and Black Sea. *Atmospheric Pollution Research*, 11(12), 2226–2243.
- Paciorek, C. J., & Liu, Y. (2009). Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter. *Environmental Health Perspectives*, 117(6), 904–909.
- Paciorek, C. J., Liu, Y., Moreno-Macias, H., & Kondragunta, S. (2008). Spatio-temporal associations between GOES aerosol optical depth retrievals and ground-level PM_{2.5}. *Environmental Science & Technology*, 42(15), 5800–5806.
- Park, S., Lee, J., Im, J., Song, C. K., Choi, M., Kim, J., Lee, S., Park, R., Kim, S. M., Yoon, J., Lee, D. W., & Quackenbush, L. J. (2020). Estimation of spatially continuous daytime particulate matter concentrations under all sky conditions through the synergistic use of satellite based AOD and numerical models. *Science of the Total Environment*, 713(2020), 136516.
- Park, S., Shin, M., Im, J., Song, C. K., Choi, M., Kim, J., Lee, S., Park, R., Kim, J., Lee, D. W., & Kim, S. K. (2019). Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea. *Atmospheric Chemistry and Physics*, 19(2), 1097–1113.
- Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term

- exposure to fine particulate air pollution. *Journal of the American Medical Association*, 287(9), 1132–1141.
- Préndez, M. M., Egido, M., Tomas, C., Seco, J., Calvo, A., & Romero, H. (1995). Correlation between solar radiation and total suspended particulate matter in Santiago, Chile-Preliminary results. *Atmospheric Environment*, 29(13), 1543–1551.
- Price, D. J., Kacarab, M., Cocker, D. R., Purvis-Roberts, K. L., & Silva, P. J. (2016). Effects of temperature on the formation of secondary organic aerosol from amine precursors. *Aerosol Science and Technology*, 50(11), 1216–1226.
- Probst, P., & Boulesteix, A. L. (2018). To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research* (2018), 18, 1–8.
- R Development Core Team (2020). *A Language and Environment for Statistical Computing*. In *R Foundation for Statistical Computing*. Retrieved January 1, 2019, from <https://www.R-project.org>. <http://www.r-project.org>.
- Randles, C. A., da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinozuka, Y., & Flynn, C. J. (2017). The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *Journal of Climate*, 30(17), 6823–6850.
- Ranjan, A. K., Patra, A. K., & Gorai, A. K. (2021). A review on estimation of particulate matter from satellite-based aerosol optical depth: Data, methods, and challenges. *Asia-Pacific Journal of Atmospheric Sciences*, 57(3), 679–699.
- Raptis, I. P., Kazadzis, S., Amiridis, V., Gkikas, A., Gerasopoulos, E., & Mihalopoulos, N. (2020). A decade of aerosol optical properties measurements over Athens, Greece. *Atmosphere*, 11(2), 11020154.

- Remer, L. A., Mattoo, S., Levy, R. C., & Munchak, L. A. (2013). MODIS 3 km aerosol product: Algorithm and global perspective. *Atmospheric Measurement Techniques*, 6(7), 1829–1844.
- Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R. R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., & Holben, B. N. (2005). The MODIS aerosol algorithm, products, and validation. *Journal of the Atmospheric Sciences*, 62(4), 947–973.
- Rodríguez, J. D., Pérez, A., & Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 569–575.
- Rupakheti, D., Kang, S., Bilal, M., Gong, J., Xia, X., & Cong, Z. (2019). Aerosol optical depth climatology over Central Asian countries based on Aqua-MODIS Collection 6.1 data: Aerosol variations and sources. *Atmospheric Environment*, 207(2019), 205–214.
- Sayer, A. M., Munchak, L. A., Hsu, N. C., Levy, R. C., Bettenhausen, C., & Jeong, M. J. (2014). MODIS Collection 6 aerosol products: Comparison between Aqua's Deep Blue, Dark Target, and “merged” data sets, and usage recommendations. *Journal of Geophysical Research*, 119(22), 13,965–13,989.
- Sayer, A. M., Hsu, N. C., Bettenhausen, C., & Jeong, M. J. (2013). Validation and uncertainty estimate for MODIS Collection 6 “Deep Blue” aerosol data. *Journal of Geophysical Research Atmospheres*, 118(14), 7864–7872.
- Schaap, M., Timmermans, R. M. A., Koelemeijer, R. B. A., de Leeuw, G., & Builtjes, P. J. H. (2008). Evaluation of MODIS aerosol optical thickness over Europe using sun photometer observations. *Atmospheric Environment*, 42(9), 2187–2197.

- Schäfer, K., Harbusch, A., Emeis, S., Koepke, P., & Wiegner, M. (2008). Correlation of aerosol mass near the ground with aerosol optical depth during two seasons in Munich. *Atmospheric Environment*, 42(18), 4036–4046.
- Segura, S., Estellés, V., Utrillas, M. P., & Martínez-Lozano, J. A. (2017). Long term analysis of the columnar and surface aerosol relationship at an urban European coastal site. *Atmospheric Environment*, 167(2017), 309–322.
- Seo, S., Kim, J., Lee, H., Jeong, U., Kim, W., Holben, B. N., Kim, S. W., Song, C. H., & Lim, J. H. (2015). Estimation of PM₁₀ concentrations over Seoul using multiple empirical models with AERONET and MODIS data collected during the DRAGON-Asia campaign. *Atmospheric Chemistry and Physics*, 15(1), 319–334.
- Sever, L., Alpert, P., Lyapustin, A., Wang, Y., & Chudnovsky, A. (2017). An example of aerosol pattern variability over bright surface using high resolution MODIS MAIAC: The eastern and western areas of the Dead Sea and environs. *Atmospheric Environment*, 165(2017), 359–369.
- Shtein, A., Karnieli, A., Katra, I., Raz, R., Levy, I., Lyapustin, A., Dorman, M., Broday, D. M., & Kloog, I. (2018). Estimating daily and intra-daily PM₁₀ and PM_{2.5} in Israel using a spatio-temporal hybrid modeling approach. *Atmospheric Environment*, 191(2018), 142–152.
- Shtein, A., Kloog, I., Schwartz, J., Silibello, C., Michelozzi, P., Gariazzo, C., Viegi, G., Forastiere, F., Karnieli, A., Just, A. C., & Stafoggia, M. (2019). Estimating Daily PM_{2.5} and PM₁₀ over Italy Using an Ensemble Model. *Environmental Science and Technology*, 54(1), 120–128.
- Sorek-Hamer, M., Just, A. C., & Kloog, I. (2016). Satellite remote sensing in epidemiological studies. *Current Opinion in Pediatrics*, 28(2), 228–234.

- Sorek-Hamer, M., Strawa, A. W., Chatfield, R. B., Esswein, R., Cohen, A., & Broday, D. M. (2013). Improved retrieval of PM_{2.5} from satellite data products using non-linear methods. *Environmental Pollution*, 182(2013), 417–423.
- Sotoudeheian, S. & Arhami, M. (2014). Estimating ground-level PM₁₀ using satellite remote sensing and ground-based meteorological measurements over Tehran. *Iranian Journal of Environmental Health Science and Engineering* 12(1), 122.
- Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de' Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Vieg, G., Kloog, I., & Schwartz, J. (2019). Estimation of daily PM₁₀ and PM_{2.5} concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environment International*, 124(2019), 170–179.
- Stafoggia, M., Schwartz, J., Badaloni, C., Bellander, T., Alessandrini, E., Cattani, G., de' Donato, F., Gaeta, A., Leone, G., Lyapustin, A., Sorek-Hamer, M., de Hoogh, K., Di, Q., Forastiere, F., & Kloog, I. (2017). Estimation of daily PM₁₀ concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. *Environment International*, 99(2017), 234–244.
- Stafoggia, M., Samoli, E., Alessandrini, E., Cadum, E., Ostro, B., Berti, G., Faustini, A., Jacquemin, B., Linares, C., Pascal, M., Randi, G., Ranzi, A., Stivanello, E., Forastiere, F., Angelini, P., Bisanti, L., Catrambone, M., Chiusolo, M., Davoli, M., ... & Pascal, L. (2013). Short-term associations between fine and coarse particulate matter and hospitalizations in Southern Europe: Results from the MED-PARTICLES project. *Environmental Health Perspectives*, 121(9), 1026–1033.
- Stirnberg, R., Cermak, J., & Andersen, H. (2018). An analysis of factors influencing the relationship between satellite-derived AOD and ground-level PM₁₀. *Remote Sensing*, 10(9), 10091353.

- Su, T., Li, Z., & Kahn, R. (2018). Relationships between the planetary boundary layer height and surface pollutants derived from lidar observations over China: Regional pattern and influencing factors. *Atmospheric Chemistry and Physics*, 18(21), 15921–15935.
- Sun, J., Gong, J., & Zhou, J. (2021). Estimating hourly PM_{2.5} concentrations in Beijing with satellite aerosol optical depth and a random forest approach. *Science of the Total Environment*, 762(2020),144502.
- Tao, M., Chen, L., Wang, Z., Tao, J., Che, H., Wang, X., & Wang, Y. (2015). Comparison and evaluation of the MODIS Collection 6 aerosol data in China. *Journal of Geophysical Research*, 120(14), 6992–7005.
- Tian, X., Liu, Q., Li, X., & Wei, J. (2018). Validation and comparison of MODIS C6.1 and C6 aerosol products over Beijing, China. *Remote Sensing*, 10(12), 1–18.
- Tian, J., & Chen, D. (2010). A semi-empirical model for predicting hourly ground-level fine particulate matter (PM_{2.5}) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sensing of Environment*, 114(2), 221–229.
- Toros, H., Erdun, H., Çapraz, Ö., Özer, B., Bozyazı Daylan, E., & Öztürk, A. İ. (2013). Air pollution and quality level in metropolitan Turkey for sustainable life. *European Journal of Science and Technology*,1(1),12-18.
- Tuna Tuygun, G. Gündoğdu, S., & Elbir, T. (2021). Estimation of ground-level particulate matter concentrations based on synergistic use of MODIS, MERRA-2 and AERONET AODs over a coastal site in the Eastern Mediterranean. *Atmospheric Environment*, 261, 118562.

- Tuna Tuygun, G., & Elbir, T. (2020). Long-term temporal analysis of the columnar and surface aerosol relationship with planetary boundary layer height at a southern coastal site of Turkey. *Atmospheric Pollution Research*, 11(12), 2259–2269.
- Tuna Tuygun, G., Elbir, T., Altug, H. & Gaga, E.E. (2017). Region-based emission inventory and air quality modeling study for multiple air pollutants in an interest priority site of Turkey. *Environmental Science and Pollution Research*, 24(9), 8242.
- Tuna Tuygun, G., Ozdemir, E., & Elbir, T. (2020). Evaluation of MODIS C6 and C6.1 Dark Target AOD products over Turkey based on NDVI and aerosol type. *Atmospheric Pollution Research*, 11(12), 2335-2349.
- Turkish Statistical Institute (TSI) (2021). *Population*. Retrieved April 1, 2021, from <https://www.tuik.gov.tr/Home/Index>.
- Tutsak, E., & Koçak, M. (2020). Optical and microphysical properties of the columnar Aerosol burden over the Eastern Mediterranean: Discrimination of Aerosol types. *Atmospheric Environment*, 229(2020),117463.
- Tutsak, E., & Koçak, M. (2019). Long-term measurements of aerosol optical and physical properties over the Eastern Mediterranean: Hygroscopic nature and source regions. *Atmospheric Environment*, 207(2019), 1–15.
- Türkeş, M. (1998). Influence of geopotential heights, cyclone frequency and Southern Oscillation on rainfall variations in Turkey. *International Journal of Climatology*,18, 649-680.
- Ustaoglu, E., & Aydinoglu, A. C. (2019). Regional variations of land-use development and land-use/cover change dynamics: A case study of Turkey. *Remote Sensing*, 11(7), 885.

- Van Donkelaar, A., Martin, R. V., Brauer, M., & Boys, B. L. (2015). Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environmental Health Perspectives*, 123(2), 135–143.
- Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, 118(6), 847–855.
- Van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM_{2.5} using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research Atmospheres*, 111(21), 1-10.
- Wang, Y., Yuan, Q., Li, T., Shen, H., Zheng, L., & Zhang, L. (2019). Evaluation and comparison of MODIS Collection 6.1 aerosol optical depth against AERONET over regions in China with multifarious underlying surfaces. *Atmospheric Environment*, 200(2019), 280–301.
- Wang, C., Liu, Q., Ying, N., Wang, X., & Ma, J. (2013). Air quality evaluation on an urban scale based on MODIS satellite images. *Atmospheric Research*, 132–133, 22–34.
- Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies. *Geophysical Research Letters*, 30(21), 2095.
- Wei, J., Li, Z., Sun, L., Xue, W., Ma, Z., Liu, L., Fan, T., & Cribb, M. (2021a). Extending the EOS Long-Term PM_{2.5} Data Records Since 2013 in China: Application to the VIIRS Deep Blue Aerosol Products. *IEEE Transactions on Geoscience and Remote Sensing*, 1-12.

- Wei, J., Li, Z., Xue, W., Sun, L., Fan, T., Liu, L., Su, T., & Cribb, M. (2021b). The ChinaHighPM10 dataset: generation, validation, and spatiotemporal variations from 2015 to 2019 across China. *Environment International*, 146(2020), 106290.
- Wei, J., Li, Z., Cribb, M., Huang, W., Xue, W., Sun, L., Guo, J., Peng, Y., Li, J., Lyapustin, A., Liu, L., Wu, H., & Song, Y. (2020a). Improved 1-km resolution PM_{2.5} estimates across China using enhanced space-time extremely randomized trees. *Atmospheric Chemistry and Physics*, 20(6), 3273–3289.
- Wei, J., Li, Z., Sun, L., Peng, Y., Liu, L., He, L., Qin, W., & Cribb, M. (2020b). MODIS Collection 6.1 3 km resolution aerosol optical depth product: global evaluation and uncertainty analysis. *Atmospheric Environment*, 240(2020), 117768.
- Wei, X., Chang, N. Bin, Bai, K., & Gao, W. (2020c). Satellite remote sensing of aerosol optical depth: advances, challenges, and perspectives. *Critical Reviews in Environmental Science and Technology*, 50(16), 1640–1725.
- Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., & Cribb, M. (2019a). Estimating 1-km-resolution PM_{2.5} concentrations across China using the space-time random forest approach. *Remote Sensing of Environment*, 231, 111221.
- Wei, J., Li, Z., Peng, Y., & Sun, L. (2019b). MODIS Collection 6.1 aerosol optical depth products over land and ocean: validation and comparison. *Atmospheric Environment*, 201(2018), 428–440.
- Wei, J., Sun, L., Huang, B., Bilal, M., Zhang, Z., & Wang, L. (2018). Verification, improvement, and application of aerosol optical depths in China Part 1: Inter-comparison of NPP-VIIRS and Aqua-MODIS. *Atmospheric Environment*, 175, 221–233.

- Weizhen, H., Zhengqiang, L., Yuhuan, Z., Hua, X., Ying, Z., Kaitao, L., Donghui, L., Peng, W., & Yan, M. (2014). Using support vector regression to predict PM₁₀ and PM_{2.5}. *IOP Conference Series: Earth and Environmental Science*, 17(1), 012268.
- World Health Organization (WHO) (2020). *Air Pollution*. Retrieved January 7, 2020 from <https://www.who.int/airpollution/en/>.
- Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 20035.
- Wu, R. (2014). Seasonal dependence of factors of year-to-year variations in South China AOD and Hong Kong air quality. *International Journal of Climatology*, 34(11), 3204–3220.
- Xia, X., Chen, H., Goloub, P., Zong, X., Zhang, W., & Wang, P. (2013). Climatological aspects of aerosol optical properties in North China Plain based on ground and satellite remote-sensing data. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 127(2013), 12–23.
- Xiao, Q., Wang, Y., Chang, H. H., Meng, X., Geng, G., Lyapustin, A., & Liu, Y. (2017). Full-coverage high-resolution daily PM_{2.5} estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sensing of Environment*, 199(2017), 437–446.
- Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B., & Bai, Y. (2015). Daily Estimation of Ground-Level PM_{2.5} Concentrations over Beijing Using 3 km Resolution MODIS AOD. *Environmental Science and Technology*, 49(20), 12280–12288.
- Yan, X., Zang, Z., Jiang, Y., Shi, W., Guo, Y., Li, D., Zhao, C., & Husi, L. (2021). A Spatial-Temporal Interpretable Deep Learning Model for improving

- interpretability and predictive accuracy of satellite-based PM_{2.5}. *Environmental Pollution*, 273(2021), 116459.
- Yan, X., Zang, Z., Luo, N., Jiang, Y., & Li, Z. (2020). New interpretable deep learning model to monitor real-time PM_{2.5} concentrations from satellite data. *Environment International*, 144(2020), 106060.
- Yan, X., Shi, W., Li, Z., Li, Z., Luo, N., Zhao, W., Wang, H., & Yu, X. (2017). Satellite-based PM_{2.5} estimation using fine-mode aerosol optical thickness over China. *Atmospheric Environment*, 170(2017), 290–302.
- Yang, L., Xu, H., & Yu, S. (2020). Estimating PM_{2.5} concentrations in Yangtze River Delta region of China using random forest model and the Top-of-Atmosphere reflectance. *Journal of Environmental Management*, 272(2020), 111061.
- Yang, Q., Yuan, Q., Yue, L., Li, T., Shen, H., & Zhang, L. (2019). The relationships between PM_{2.5} and aerosol optical depth (AOD) in mainland China: About and behind the spatio-temporal variations. *Environmental Pollution*, 248(2019), 526–535.
- Yang, W., Wang, G., & Bi, C. (2017). Analysis of long-range transport effects on PM_{2.5} during a short severe haze in Beijing, China. *Aerosol and Air Quality Research*, 17(6), 1510–1522.
- Yao, F., & Palmer, P. I. (2021). A model framework to reduce bias in ground-level PM_{2.5} concentrations inferred from satellite-retrieved AOD. *Atmospheric Environment*, 248(2021), 118217.
- Yazdi, M. D., Kuang, Z., Dimakopoulou, K., Barratt, B., Suel, E., Amini, H., Lyapustin, A., Katsouyanni, K., & Schwartz, J. (2020). Predicting fine particulate matter (PM_{2.5}) in the greater London area: An ensemble approach using machine learning methods. *Remote Sensing*, 12(6), 914.

- You, W., Zang, Z., Zhang, L., Zhang, M., Pan, X., & Li, Y. (2016). A nonlinear model for estimating ground-level PM₁₀ concentration in Xi'an using MODIS aerosol optical depth retrieval. *Atmospheric Research*, 168(2016), 169–179.
- Zeydan, Ö., & Wang, Y. (2019). Using MODIS derived aerosol optical depth to estimate ground-level PM_{2.5} concentrations over Turkey. *Atmospheric Pollution Research*, 10(5), 1565–1576.
- Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M. L., Shen, X., Zhu, L., & Zhang, M. (2017). Spatiotemporal prediction of continuous daily PM_{2.5} concentrations across China using a spatially explicit machine learning algorithm. *Atmospheric Environment*, 155(2017), 129–139.
- Zhang, T., He, W., Zheng, H., Cui, Y., Song, H., & Fu, S. (2021). Satellite-based ground PM_{2.5} estimation using a gradient boosting decision tree. *Chemosphere*, 268, 128801.
- Zhang, T., Zang, L., Wan, Y., Wang, W., & Zhang, Y. (2019). Ground-level PM_{2.5} estimation over urban agglomerations in China with high spatiotemporal resolution based on Himawari-8. *Science of the Total Environment*, 676, 535–544.
- Zhang, R., Di, B., Luo, Y., Deng, X., Grieneisen, M. L., Wang, Z., Yao, G., & Zhan, Y. (2018). A nonparametric approach to filling gaps in satellite-retrieved aerosol optical depth for estimating ambient PM_{2.5} levels. *Environmental Pollution*, 243, 998–1007.
- Zhang, T., Gong, W., Wang, W., Ji, Y., Zhu, Z., & Huang, Y. (2016). Ground level PM_{2.5} estimates over China using satellite-based geographically weighted regression (GWR) models are improved by including NO₂ and enhanced

vegetation index (EVI). *International Journal of Environmental Research and Public Health*, 13(12), 1215.

Zhang, Y., & Li, Z. (2015). Remote sensing of atmospheric fine particulate matter (PM_{2.5}) mass concentration near the ground from satellite observation. *Remote Sensing of Environment*, 160, 252–262.

Zhang, H., Lyapustin, A., Wang, Y., Kondragunta, S., Laszlo, I., Ciren, P., & Hoff, R. M. (2011). A multi-angle aerosol optical depth retrieval algorithm for geostationary satellite data over the United States. *Atmospheric Chemistry and Physics*, 11(23), 11977–11991.

Zhang, H., Hoff, R. M., & Engel-Cox, J. A. (2009). 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 and Waste Management Association*, 59(11), 1358–1369.

Zhao, C., Liu, Z., Wang, Q., Ban, J., Chen, N. X., & Li, T. (2019). High-resolution daily AOD estimated to full coverage using the random forest model approach in the Beijing-Tianjin-Hebei region. *Atmospheric Environment*, 203, 70–78.

Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., & Qiu, Y. (2017). Analysis of influential factors for the relationship between PM_{2.5} and AOD in Beijing. *Atmospheric Chemistry and Physics*, 17(21), 13473–13489.

APPENDICES

ABBREVIATION LIST

AAI	: Absorbing aerosol index
AERONET	: Aerosol Robotic Network
AHI	: Advanced Himawari Imager
AI	: Artificial intelligence
AIRS	: Advanced Infrared Radiation Sounder
ANN	: Artificial neural network
AOD	: Aerosol Optical Depth
BRNN	: Back-propagation regression neural networks
CACs	: Clean Air Centers
CALIPSO	: Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations,
PARASOL	: Polarization and Anisotropy of Reflectance's for Atmospheric Sciences coupled with Observations from a Lidar
CAMS	: Copernicus Atmosphere Monitoring Service
CLOUD	: Total cloud area fraction
CMAQ	: The Community Multi-Scale Air Quality
CTM	: Chemical transport model
CV	: Cross Validation
DB	: Deep Blue
DEWP	: Dew point temperature at 2 m
DT	: Dark Target
DTB	: Dark Target & Deep Blue combined
EC	: European Commission
EE	: Expected Error
EEA	: European Environmental Agency
EPA	: Environmental Protection Agency
ERT	: Extremely Randomized Trees
EVAP	: Evaporation
GAM	: Generalized additive models
GEOS-Chem	: Goddard Earth Observing System-Chem

GOCART	: Georgia Tech/Goddard Global Ozone Chemistry Aerosol Radiation and Transport
GOES	: Geostationary Operational Environmental Satellites
GRNN	: Generalized Regression Neural Networks
GWR	: Geographically Weighted Regression
KNN	: K-Nearest Neighbor
LAADS	: Level 1 Atmosphere Archive and Distribution System
LME	: Linear Mixed-Effect
LR	: Linear Regression
LUTs	: Look up Tables
LUR	: The Land-Use Regression
MAE	: Mean Absolute Error
MAIAC	: Multi-Angle Implementation of Atmospheric Correction
MEM	: Mixed-Effects Model
MERRA-2	: Modern-Era Retrospective Analysis for Research and Applications, Version 2
MEU	: Ministry of Environment and Urbanization
MISR	: Multi-angle imaging spectroradiometer
ML	: Machine Learning
MLR	: Multiple Linear Regression
MODIS	: Moderate Resolution Imaging Spectroradiometer
MSE	: Mean Squared Error
NASA	: National Aeronautics and Space Administration
NDVI	: Normalized Difference Vegetation Index
OBPG	: Ocean Biology Processing Group
OMI	: Ozone Monitoring Instrument
OOB	: Out-of-Bag
P	: Pressure
PBLH	: Planetary Boundary Layer Height
PCC	: Pearson's correlation coefficient
PM	: Particulate Matter
PREC	: Total Precipitation

QAC	: Quality Assurance Confidence
QR	: Quantile regression
R	: Correlation Coefficient
RF	: Random Forest
RH	: Relative Humidity
RMB	: Relative Mean Bias
RMSE	: Root Mean Square Error
SHUM	: Specific Humidity
SOA	: Secondary Organic Aerosols
SRAD	: Solar Radiation
STET	: Space-Time Extremely Randomized Trees
STRF	: Space Time Random Forest
SVM	: Support Vector Regression
T	: Temperature at 2m
TOA	: Top of Atmosphere
TQV	: Total Precipitable
u	: u-component of the wind
v	: v-component of the wind
VIF	: Variance Inflation Factor
VIIRS	: Visible Infrared Imaging Radiometer Suite
WHO	: World Health Organization
XGBoost:	: Extreme Gradient Boosting Model
ZOM	: Surface Roughness