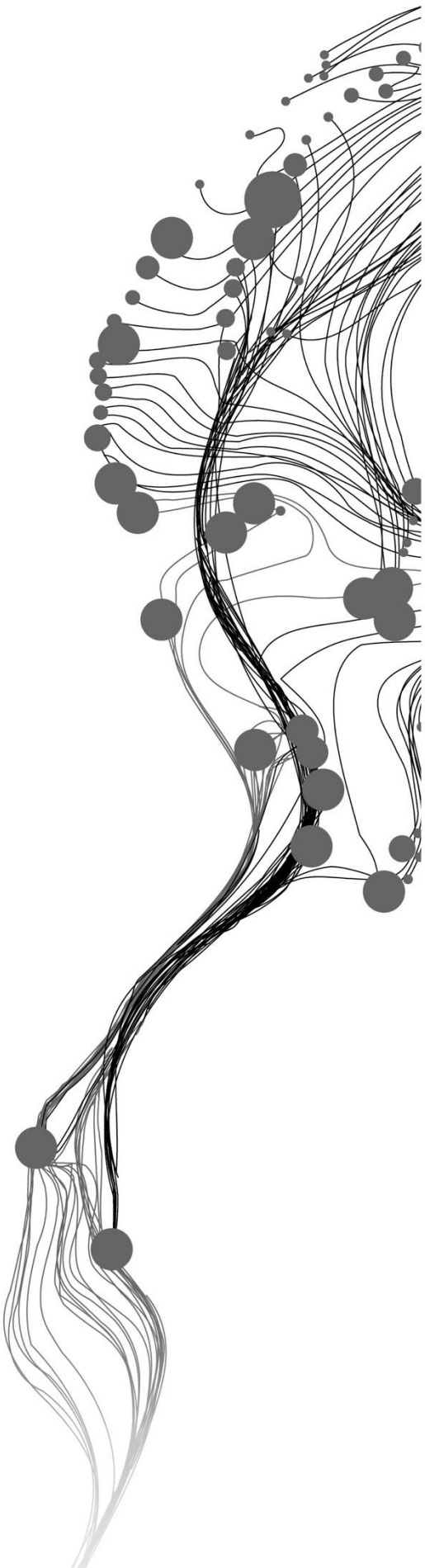
SEMI-EMPIRICAL MODELLING OF $PM_{2.5}$ USING SURFACE WEATHER, UPPER-AIR AND SATELLITE-RETRIEVED ATMOSPHERIC OBSERVATIONS

NOSHAN BHATTARAI

May, 2016

SUPERVISORS:
Ing. Valentijn Venus, MSc
Dr. Nicholas Hamm



NOSHAN BHATTARAI
Enschede, The Netherlands, May, 2016

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geo-information Science and Earth Observation For Environmental Modelling and Management

SUPERVISORS:
Drs. Valentijn Venus
Dr. Nicholas Hamm

THESIS ASSESSMENT BOARD:
Dr. Ir. C.A.J.M. De Bie (Chair)
Prof. Dr. A.A. Voinov (External Examiner, ITC-GIP)

ABSTRACT

Aerosol Optical Depth (AOD) is a measure of extinction of light by particles in the atmosphere. It is the integral of the atmospheric extinction by aerosol in the vertical column from surface to the top of the atmosphere. AOD is extracted using weighted average temporal interpolation method during resampling that are made time commensurate with $PM_{2.5,}$ measured at the surface air quality monitoring station. MODIS collection 5 aerosol and atmospheric profile product is used to retrieve AOD and other predictor variables from satellite retrieved atmospheric observations respectively. The surface weather observations such as wind speed, wind direction, temperature, relative humidity and upper-air observations (boundary layer height) are considered as a predictor variable to estimate $PM_{2.5}$. The overall goal of the study is to develop a semi-empirical model to explain the variability in surface air quality ($PM_{2.5}$) using time commensurate predictor variables. The closest AOD values observed, when satellite overpasses the station is extracted during resampling that are not made temporally commensurate, when $PM_{2.5}$ is observed at the surface air quality monitoring station. The data pairs of AOD, upper-air observation, surface weather observation and satellite retrieved atmospheric observations are created for with time commensurate and without time commensurate AOD. The first objective is to assess whether a semi-empirical model improves if the predictor variables are made temporally commensurate with the surface air quality ($PM_{2.5}$) observations that it tries to explain. The final selected variable for regression analysis are AOD, temperature and boundary layer height. The result shows that temporally commensurate predictor variables explain 29.6% variability in estimated $PM_{2.5}$. The $PM_{2.5}$ concentrations is a measurement of dry aerosol mass. The instantaneous AOD is measured in an ambient environment, thus it is affected by relative humidity. The AOD is corrected for relative humidity using hygroscopic growth factor, $f(RH)$ and is normalized by boundary layer height which are "meteo-scaled" AOD. The second objective of the study is to assess whether a semi-empirical model improves explained variability in $PM_{2.5}$ with meteo-scaled AOD that it tries to explain. The AOD is corrected using simple method when relative humidity is available. The AOD is corrected using advanced method (4.3.2) when experimental fitting curves of $f(RH)$ for different air mass type are available. The result shows that variability in $PM_{2.5}$ is improved by 4%, when AOD is corrected for time discrepancies (with the other predictor variables).

Key words: AOD, $PM_{2.5}$, temporally commensurate, $f(RH)$, meteo-scaled AOD, explained variability

ACKNOWLEDGEMENTS

I would express my sincere gratitude to my Employer Sonapur Cements Pvt. Ltd, who provided the great opportunity to work and contribute on "air quality" in Nepal. The communities around the factory, whose immense attention on air quality issues had led me to apply for GEM masters. So, I would like to deeply thank all the communities' members for their contribution in my decision to choose "air quality" as a subject to work on. I would like to dedicate this research to my family who are at home and abroad. Without their unconditional love, support and deep sacrifices, I would not have been studying here in the Netherlands.

I would like to deeply thank my supervisors Drs. Valentijn Venus and Dr. Nicholas Hamm for their continuous motivation, constructive feedbacks and phenomenal mentoring throughout the process. This research would not have been possible without their untiring guidance.

I would like to thank Ir. Henk Klein Baltink from KNMI, Mr. Hans Berkhout, Mr. Guus Stefess from RIVM for their help in obtaining essential information and dataset needed for the research. I owe special thanks to Mr. Willem Nieuwenhuis and Mr. Ali Bagislayici, who helped me to carry out parallel processing during my research, which saved lot of my time.

My special thanks goes to Dr. Pawan Gupta from NASA Goddard Space Flight Center who shaped my thoughts about the task in the beginning of the research.

I would like to thank all the teachers from Lund University and ITC, who in one way or another has inspired me to think big yet simple. I would like to thank Nuffic for providing me scholarship to experience once in a life time opportunity.

The time that I spent in Lund and ITC wouldn't have been as great as it is, without support, love and care from dynamic, awesome and great friends that one could have in their life, which is another achievement together with this thesis. I highly appreciate the role played by Phanintra, Laura, Emile, Mirza, Ipsit, Rushikesh, Rahul, John, Vera during this period, which was phenomenal.

Finally, I would like to thank ITC for providing me the great space to grow.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AERONET	Aerosol Robotic Network
AMF	Air Mass Factor
ANN	Artificial Neural Network
AOD	Aerosol Optical Depth
AOT	Aerosol Optical Thickness
BAM	Beta Attenuation Monitoring
BLH	Boundary Layer Height
BTD	Brightness Temperature Difference
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
CESAR	Cabauw Experimental Site for Atmospheric Research
CMAQ	Community Multiscale Air Quality
CPCB	Central Pollution Control Board
DAAC	Distributed Active Archive Centre
ECMWF	European Centre for Medium-Range Weather Forecasts
EPA	Environmental Protection Agency
FDMS	Filter Dynamics Measurement System
FRM	Federal Reference Method
GAM	Generalized Additive Model
GAW	Global Atmosphere Watch
GEOS	Geostationary Operational Environmental Satellite
GLM	Generalized Linear Regression
GWR	Geographically Weighted Regression
ISD	Integrated Surface Database
KNMI	Royal Netherlands Meteorological Institute
LiDAR	Light Detection and Ranging
McIDAS	Man Computer Interactive Data Access System
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	Moderate-Resolution Imaging Spectroradiometer
MSRFR	Multifilter Rotating Shadowband and Radiometer
NCDC	National Climatic Data Centre
NCEP	National Centre for Environmental Prediction
NLDAS	National American Land Data Assimilation System
NOAA	National Oceanic and Atmospheric Administration
OLS	Ordinary Least Squares
PHOTONS	Photometrie pour le Traitement Operationel de Normalization Satellitaire
PM	Particulate Matter
RH	Relative Humidity
RIVM	National Institute for Public Health and the Environment
RS	Remote Sensing
RUC	Rapid Update Cycle
TEOM	Tapered Element Oscillating Microbalance
TOA	Top of Atmosphere
TOMS	Total Ozone Mapping Spectrometer
WHO	World Health Organization
WMO	World Meteorological Organization

1. INTRODUCTION

1.1. Background and Motivation

Particulate matter (PM) is a group of major pollutants that affect the rural and urban environment. It is a complex mixture of solid and liquid particles that vary in size, composition and remains suspended in air. The sources of emission of particulate matters are due to anthropogenic activities and natural sources, which is directly emitted or reacted with other pollutants in the atmosphere to form particulate matter (United States Environmental Protection Agency, 2015). Anthropogenic sources includes aerosol from biomass burning, combustion from automobiles and emission from power plants and natural sources includes wind-blown dust, sea salt from oceans and volcanic eruptions (Kaufman, Tanré, & Boucher, 2002). The particulate matter which has aerodynamic diameter of 2.5 μm or smaller is called fine particulate matter ($PM_{2.5}$). $PM_{2.5}$ are approximately $1/30^{th}$ the diameter of the human hair (Figure 1-1). The toxicity of fine particles (particle size less than 2.5 μm in the aerodynamic diameter) is considered to be stronger as they can penetrate into the gas-exchange region of the lung than the coarse particles (particle size between 2.5 μm and 10 μm in the aerodynamic diameter) that can penetrate into the lower respiratory system.

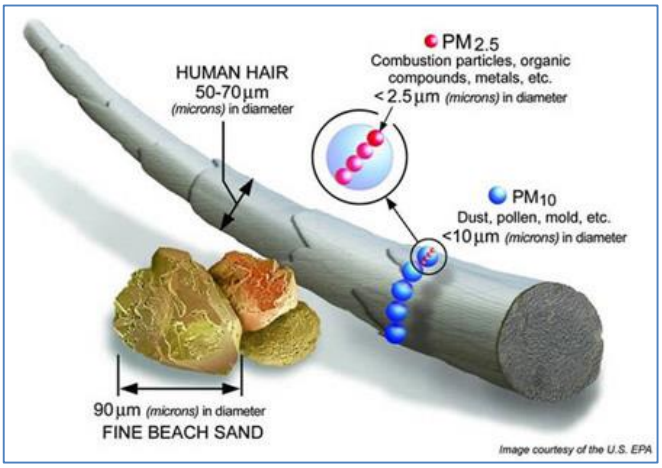


Figure 1-1 Size comparison of $PM_{2.5}$ (US EPA, 2015)

The smaller the size of pollutant, the longer it remains suspended in the atmosphere. The short term exposure to particulate matter could reduce life expectancy at 1-2 years (Brunekreef & Holgate, 2002) while the long-term exposure to $PM_{2.5}$ are more uncertain but are believed to have adverse effect on human health, mainly related to the respiratory and cardiovascular systems (Pope & Dockery, 2006). For example, in 2012, it was estimated that 3.7 million premature deaths was due to outdoor air pollution in both cities and rural areas worldwide, where 88% of occurred in low and middle income countries (WHO, 2014). Thus, to understand the effect of particulate matters on human health, it is necessary to monitor particulate matters on regular basis. The World Health Organization (WHO) has set a standard for ambient annual averaged, daily averaged $PM_{2.5}$ concentrations, which is 10 $\mu g/m^3$ and 25 $\mu g/m^3$ respectively (WHO, 2006) whereas the annual average concentration and daily average concentrations according to US EPA is 12 $\mu g/m^3$ and 35 $\mu g/m^3$. Likewise, the European Commission has set the annual average target concentration of 25 $\mu g/m^3$ for $PM_{2.5}$ exposure, which must be achieved by January 2015 (European Parliament & Council of the European Union, 2008). In addition, an interim reduction of 20 % has been proposed for annual average background $PM_{2.5}$ concentrations, which should be realized up to January 1, 2020.

Air pollution, in some of the Asian countries like China, India and Nepal, it was reported that the annual average concentration of $PM_{2.5}$ over Beijing from 2000 to 2010 were more than seven folds compared to WHO standards which could have significant impacts on atmospheric visibility and human health during those periods (Lang et al., 2013). Public concerns on air quality grown largely after severe haze event in December 2011 in Beijing. According to Zhang et al. (2013) after haze event, Beijing Environmental Protection Bureau started releasing for the first time, the set of real-time monitoring $PM_{2.5}$ data of 31 stations, which aided to carry out research on population exposure to $PM_{2.5}$ in the urban area of Beijing. Likewise, due to growing concern in air pollution in India, the Central Pollution Control Board (CPCB) initiated air quality monitoring of $PM_{2.5}$, forecasting during Commonwealth Games 2010 in six major cities including New Delhi (Sahu, Beig, & Parkhi, 2011).

In Nepal, recent studies (Giri, 2006; Panday, Prinn, & Schar, 2009; Aryal et al., 2009; Putero et al., 2015) suggests that the poor air pollution in Kathmandu has harmful effects on human health. However, none of them presented observations of $PM_{2.5}$ across valley. Review by Gurung & Bell (2013) on the state of scientific evidence on air pollution and human health in Nepal shows that very less assessment has been carried out on air quality and population exposure to $PM_{2.5}$. Thus, people are unaware of possible health consequences resulting from poor air quality. Chow et al. (2002) has mentioned that the health exposure assessment needs of monitoring aerosol by air quality agencies should focus on spatial scales ranging from 1 to 100 m and time scales of minutes to months. Kathmandu valley initially had six air quality monitoring stations at some urban and rural areas since 2000 operated by Ministry of Population and Environment back that time but it became dysfunctional in 2007 due to poor management. The recent five year plan has documented government plan to install 11 stations for data collection and monitoring of air quality of valley only to start with fiscal year 2015/2016 (Shrestha, 2015).

The *in situ* air quality monitoring stations at any urban and rural location can only estimate concentration of particulate matters ($PM_{2.5}$, PM_{10}) at particular location and time only. They are recorded at sparsely distributed locations on the ground at different time intervals (example, every hour, 8 h or 24 h). The satellite observation, which has large spatial coverage and reliable repeated measurements, could provide information on aerosol loading over an area. Thus, aerosol remote sensing could serve as a proxy for monitoring fine particulate matter air quality (Gupta et al., 2006) at particular area of interest. **Aerosol optical depth** (AOD), which measures the extinction of light and is the integral of the atmospheric extinction by aerosol in the vertical column from surface to the top of atmosphere (Equation 2-1) could be retrieved from satellite data (Holben et al., 1998). Several studies (Engel-Cox, Holloman, Coutant, & Hoff, 2004; van Donkelaar, Martin, & Park, 2006; Schaap et al., 2009) has established the empirical relationship between $PM_{2.5}$ and AOD to estimate the surface level concentration, which requires some in situ measurements to train empirical models.

The review of data in literature review (Table 2-1) shows that MODIS retrieved AOD can be used to study linear relationship between satellite retrieved AOD and PM2.5. MODIS is an instrument which operate on the Tera and Aqua satellite, which observes the entire earth on every other day. The Terra and Aqua satellite overpass any place on earth at approximately 10:30 and 13:30 local time. MODIS provides AOD data in collections, which new and improved science algorithms are developed. Till now, MODIS has Collection 001, 003, 004, 005, 051 and 006. AOD is available in level 2 MODIS **"aerosol product data"** as hierarchical data format, whereas level 1B MODIS data consist of calibrated and geo-located radiances for its 36 bands. MODIS also has "**atmospheric profile product"** which monitors profiles of atmospheric temperature, moisture and atmospheric stability.

For Netherlands, Schaap et al. (2009) found a relationship of Aerosol Robotic Network (AERONET) retrieved AOD with ground measurement of $PM_{2.5}$ concentration at Cabauw station. AERONET is a globally distributed observations of aerosol optical depth (Holben et al.,1998). The validation of AERONET AOD and MODIS AOD with PM2.5 explained 52% of the variability in $PM_{2.5}$. Estimating surface PM in particular space and time through the relationship of $PM_{2.5}$-AOD is not straight forward because AOD is a columnar, optical measurement while $PM_{2.5}$ measurement at ground is a surface and point based, dry mass measurement of $PM_{2.5}$ and it is assumed for quantitative analysis, when most particles are concentrated and well mixed in the boundary layer, satellite AOD contains a strong signal of ground level particle concentration (Gupta, 2008). The global validation study (Levy et al., 2010) shows the retrieval accuracy of AOD from **MODIS Collection 005** is 0.05 ± 0.15% for spatiotemporally matched AOD at 550 and the prediction of $PM_{2.5}$ from AOD is ±30% in the most careful studies (Hoff & Christopher, 2009).The nature and sources of aerosols and meteorological and climatic conditions vary regionally and could play important role in AOD retrieval and its association with PM. So, examining meteorological factors could improve empirical relationship of AOD-PM (Kumar, Chu, Foster, Peters, & Willis, 2011).

1.2. Research Problem

The global validation study of AOD for MODIS collection 005 shows that accuracy of MODIS is of 0.05 ± 15%×AOD over land (Levy et al., 2010). Other validation study (Koelemeijer, Homan, & Matthijsen, 2006) for time-series of MODIS and AERONET AOD show a correlation of 0.72, averaged over 36 AERONET stations of Europe for 2003. It is found that the retrieved AOD has minimum value of -0.049 and maximum of 0.68 with weighted average temporal interpolation method and minimum value of -0.05 and maximum of 0.72 with nearest neighbourhood method, which is well within the typical valid range of AOD retrieval of -0.05 to 5 for MODIS aerosol product (MOD04, MYD04) in current study. However, it is generally classified as heavy haze if AOD values are over unity (Engel-Cox et al., 2004;.Engel-Cox et al., 2006). Thus, it is assumed (research assumption 1 of 1.3.5) that the MODIS AOD retrieved for the study period of 2013-2014 for the Cabauw air quality monitoring station is valid for further study.

Veefkind, Hage, & Brink (1996) measured AOD of the atmospheric boundary layer using Nephelometer for a clear sky day, when there was a single aerosol layer as found in Light Detection and Ranging (LiDAR) probing thus, assuming that BLH from LiDAR is representative for the aerosol height within the first layer. Liu, Sarnat, Kilaru, Jacob, & Koutrakis (2005) assumed that the aerosol particles are correlated to surface particulate concentrations at different altitudes and the aerosol particles vertical profile is smooth to establish the relationship between AOD and $PM_{2.5}$. Similarly, Koelemeijer et al. (2006) found that the correction of AOD and $PM_{2.5}$ improved when the AOD is divided by BLH assuming that there is a single homogenous layer and no overlapping of aerosols in the atmosphere. Wang & Christopher (2003) also assumed that the vertical mixing produces well mixed BLH in cloud-free conditions thus resulting in favourable positive correlations between MODIS AOD and $PM_{2.5}$ mass. Thus, in our research, first significant height of atmospheric boundary layer which is called **Boundary layer height** (BLH) retrieved from backscattering profile of **Vaisala LD-40 Ceilometer** is used. Atmospheric boundary layer is the lowest layer of the atmosphere that is in direct contact with Earth's surface. LD-40 Ceilometer is a commercial **LiDAR** system operated by Royal Netherlands Meteorological Institute (KNMI) in The Netherlands. It is used to establish relationship between AOD and $PM_{2.5}$, assuming (research assumption 2 of 1.3.5) that the aerosol particles are correlated within the first significant height of boundary layer detected by LiDAR with surface level $PM_{2.5}$ concentrations.

AOD is measured under ambient condition which is influenced by relative humidity (RH) whereas $PM_{2.5}$ is measured in dry condition, which represents the dry mass concentration measured at RH below 40-50 % (Gupta et al., 2006; Liu et al., 2005). Veefkind et al. (1996) stated that ammonium sulfate and ammonium

nitrate are major aerosol which are dominated by Continental and maritime air masses in The Netherlands. When RH is high (RH>70%), which is frequently found in the current study period (2013-2014), the hygroscopic particles such as ammonium sulfate and ammonium nitrate can grow 2-10 times in size, increasing particle light scattering (Malm, Day, & Kreidenweis, 2000).

Review of literature (Table 2-1) and specifically Koelemeijer et al.(2006) shows that the correlation between AOD and PM2.5 would improve if AOD is corrected for BLH and hygroscopic growth of aerosol particles represented as $f(RH)$ (Wang & Christopher, 2003) for surface RH (Hutchison et al., 2008; Tian & Chen, 2010). Thus, in current study the satellite retrieved AOD is corrected for BLH and hygroscopic growth of aerosol particles. The method when only RH is available for correcting AOD is referred as **"Simple correction method/reference method"** hereafter (Equation 4-4). It is referred as **"Advanced Method",** when there is availability of experimental fitting curve of $f(RH)$ for different air mass type for correction of AOD with $f(RH)$. Koelemeijer et al. (2006) has defined the term **"meteo-scaled"** for AOD normalized by BLH and corrected with $f(RH)$ as **AOD* (**Equation 4-3):

Studies (Engel-Cox et al., 2006; Koelemeijer et al., 2006; Gupta & Christopher, 2009) used PM2.5 measurment closest in time to instantaneous MODIS Terra and Aqua AOD measured value during the satellite overpass time to create data pairs to established relationship between AOD and PM2.5. Each MODIS level 2 product file covers a five-minute time granule. The same location can have more than one AOD observation in a day. Thus, estimated AOD that is closer to the time when PM2.5 is measured at the station is temporally matched during resampling, when satellite overpass the air quality monitoring station. Hence, the data pair of PM2.5 and AOD is obtained for 10:30 and 13:30. But, PM2.5 is an hourly measurement at air quality monitoring station and AOD is an integral of instantaneous observation from surface to the top of atmosphere in a vertical column during the satellite overpass time. To elaborate further, if the satellite overpasses the station at supposedly 10:15 and 13:00 in a day, the measurement at air quality monitoring station at 10:30 is paired with AOD observed at 10:15 during temporal matching. Similarly, the measurement at air quality monitoring station at 13:30 is paired with AOD observed at 13:00. And, supposedly if there is only one observation in a day at 10:15, each PM2.5 measurements at air quality monitoring station at 10:30 and 13:30 are paired with that one AOD observation providing two data pairs during temporal matching. This process of pairing which has **"without time commensurate"** AOD, which may ignore the change in atmosphere (e.g. wind speed, wind direction, relative humidity, temperature, boundary layer height) within air quality measurement durations between 10:30 and 13:30.

The novelty of this research is to extract and prepare a matched data pairs of satellite retrieved AOD from Terra and Aqua satellites level 2 MODIS aerosol product (MOD04 and MYD04 respectively) that are made "**time commensurate"** during resampling with ground level measured $PM_{2.5}$ concentrations at air quality monitoring station. The satellite retrieved AOD is temporally matched with the ground level $PM_{2.5}$ measurement time exactly at 10:30 and 13:30 (Terra and Aqua overpass time respectively), when the satellite overpasses the air quality monitoring station. As NASA, (2016) highlights that the observation area closer to the poles will have increased number of overpasses, thus, daily MODIS observations increases due to overlapping orbits. Hence, the same location can have more than one observation in a day. To elaborate further, if the satellite overpasses the station at supposedly 09:45 and 10:15 in a day, the AOD value at air quality monitoring station at 10:30 is obtained with weighted average temporal interpolation during resampling of estimated AOD values observed at 09:45 and 10:15 by the satellite. And supposedly, if there is only one satellite observation in a day at 10:15, only PM2.5 measured at air quality monitoring station at 10:30 is paired with AOD estimated at 10:15 without interpolation. Thus, there is only one data pair in a day. This process of pairing has **"with time commensurate"** AOD, which may consider the change in

atmosphere (e.g. wind speed, wind direction, relative humidity, temperature, boundary layer height) within air quality measurements between 10:30 and 13:30.

AOD is a measure of the integral of the amount of light attenuated in a vertical column from surface to the top of atmosphere, when it overpasses the station. It is assumed (research assumption 3 of 1.3.5) that the surface level $PM_{2.5}$ concentration measured at 10:30, 11:30, 12:30 and 13:30 hours are representative of the ambient concentration of particulate matters at the station at that hour. **McIDAS-V** (Space Science and Engineering Center, 2000), is used for resampling, which has an option to interpolate AOD values using weighted average method. McIDAS-V is a free, open source, visualization and data analysis software package (*McIDAS-V User's Guide*, 2015).

With the available data pairs, multiple linear regression modelling (Equation 4-2) is carried out with AOD and surface level PM2.5 (dependent variable). AOD is considered as a predictor variable in this thesis. When the satellite overpasses the air quality monitoring station, the potential **"predictor variables"** (Table 3-5) is extracted from Terra/Aqua level 2 MODIS atmospheric profile product (MOD07 and MYD07 respectively) that are made commensurate with the time AOD was observed (referred hereafter as "**satellite retrieved atmospheric observations"**). The surface weather variables such as relative humidity, wind speed, wind direction and temperature (hereafter referred as **"surface weather observations"**) and boundary layer height referred as **"upper-air observations"** are used as additional predictor variable that could improve variability in estimated $PM_{2.5}$ concentrations at surface. It is however assumed (research assumption 4 of 1.3.5) that the surface weather observations at 10:30 and 13:30 are representative of the surface air weather condition when the instantaneous AOD was observed at the station.

Thereafter it is assessed whether these predictor variables such as upper air observations, satellite retrieved atmospheric observations and surface weather observations that are made temporally commensurate could improve the variability in PM2.5 that it tries to explain. The **semi-empirical modelling** approach is applied as the modelling is carried out on the basis of the knowledge of processes involved to estimate PM2.5 with all predictor variables. Thus, a semi-empirical model is assessed with time commensurate AOD and without time commensurate AOD. The semi-empirical model obtained with both time commensurate and without time commensurate AOD with PM2.5 is assessed with meto-scaled AOD separately if it could improve the variability in PM2.5

1.3. Research Objectives

1.3.1. Overall Research Objective

The overall objective of the study is to assess how much the predictor variables which are extracted as point based time series data pairs of satellite retrieved atmospheric observations that are made temporally commensurate, upper air and surface weather conditions explains the variability in surface level $PM_{2.5}$.

1.3.2. Specific Research Objectives

The overall goal of this study is to develop a semi-empirical model to explain variability in surface air quality ($PM_{2.5}$) using time commensurate AOD, upper air, surface weather, and satellite-retrieved atmospheric observations:

1. To assess whether a semi-empirical model improves if the predictor variables are made temporally commensurate with the surface air quality ($PM_{2.5}$) observations it tries to explain.
2. To assess whether a semi-empirical model improves with meteo-scaled AOD it tries to explain.

1.3.3. Research Questions

1. Does the explained variability in $PM_{2.5}$ increase (assessed from the model's coefficient of determination, $R^2_{adj.}$) if all the predictor variables retrieved from the MODIS aerosol and atmospheric products are temporally commensurate?

2. Does the explained variability in $PM_{2.5}$ increase (assessed from the model's coefficient of determination, $R^2_{adj.}$) in a semi-empirical model with meteo-scaled AOD?

1.3.4. Hypotheses

Hypothesis 1

H0: There is no significant difference between semi-empirical model with time commensurate predictor variables and without time commensurate predictor variables at 90% confidence interval (i.e. all regression coefficients are zero, *i.e.,* none of the regressors contribute significantly to the prediction of PM2.5).

H1: There is significant difference between semi-empirical model with time commensurate predictor at 90% confidence interval (at least one regression coefficient is significantly different from zero i.e., at least one regressor contributes significantly to the prediction of PM2.5).

Hypothesis 2

H0: There is no significant difference between semi-empirical model with time commensurate meteo-scaled AOD and without time commensurate meteo-scaled AOD at 90% confidence interval (i.e. all regression coefficients are zero, *i.e.,* none of the regressors contribute significantly to the prediction of PM2.5).

H1: There is significant difference between semi-empirical model with time commensurate meteo-scaled AOD at 90% confidence interval (at least one regression coefficient is significantly different from zero i.e., at least one regressor contributes significantly to the prediction of PM2.5).

1.3.5. Research assumptions

1. It is assumed that the Terra, Aqua MODIS AOD retrieved for the study period of 2013-2014 at Cabauw air quality monitoring station is within the accuracy of the global validation studies of MODIS collection 051 product compared to AERONET AOD.

2. It is assumed that when only first significant height of boundary layer was detected in LiDAR, it is representative of the aerosol height within the boundary layer assuming there is a single well-mixed layer during satellite overpass time and the vertical distribution of particles above the boundary layer is relatively smooth.

3. Since, AOD is an instantaneous measure of the integral of the amount of light attenuated in a vertical column from surface to the top of atmosphere, when it overpasses the station, it is assumed that the surface level $PM_{2.5}$ concentration measured at 10:30 and 13:30 hours are representative of the ambient concentration of particulate matters at the station.

4. It is assumed that the surface weather measured at nearest weather station and upper air observations at 10:30 and 13:30 hours measured at air quality monitoring station are representative of the surface weather and upper air observations at Cabauw, when the instantaneous AOD is observed at the air quality monitoring station, when the Terra, Aqua MODIS overpass the station at 10:30 and 13:30 hours respectively.

5. It is assumed that the experimental $f(RH = 72\%, \lambda = 474 \text{ nm})$ found in Petten in the Netherlands for November 17, 1993 by Veefkind et al., (1996) could be adapted in Cabauw with modification on offset values based on experimental study of $f(RH = 85\%, \lambda = 550 \text{ nm})$ values during June to October 2009 for different air mass types found at Cabauw in Zieger et al., (2013).

6. It is assumed that the $f(RH = 85\%, \lambda = 550 \text{ nm})$ found in experimental study at Cabauw by Zieger et al., (2011) for different air mass types such as maritime, continental south, maritime heavily polluted, maritime slightly polluted, continental east and all air mass together could be grouped into two types of air mass as "maritime" by grouping maritime and maritime, slightly polluted and "others" by grouping continental south, continental east and maritime, heavily polluted. It is also assumed that the coefficient of fitting curve can be averaged to find the curve fitting for grouped air mass types.

7. It is assumed that having wind direction at surface weather observation station at Cabauw is enough representative of where the air mass is coming from.

1.3.6. Thesis structure

The thesis is divided into six chapters. Chapter 1 consist of Background, Research Motivation with research problem, objectives, questions, assumptions and hypotheses. Chapter 2 provides literature review on previous studies related to estimating PM2.5 using satellite retrieved AOD. Chapter 3 provides description of study area, data's used for study, pre-processing and processing of all available data. Methods parts is provided in Chapter 4. Chapter 5 consists of result and discussion. And finally, Chapter 6 provides the Conclusion, Limitations and Recommendation for future research.

2. LITERATURE REVIEW

2.1. Aerosol Optical Depth

The AOD is an integral measure of aerosol extinction coefficient in a vertical column due to scattering and absorption by aerosol properties from the surface to the top of the atmosphere as presented in Equation 2-1 (Levy, 2009).

$$\tau(\lambda) = \int_{0}^{TOA} \beta_{ext,p}(\lambda, z)\, dz \qquad \text{Equation 2-1}$$

where $\tau(\lambda)$ represents AOD at a wavelength (λ), subscript p represents the contribution from the particles (to be separated from molecular or Rayleigh optical depth), $\beta_{ext}(\lambda, z)$ is the aerosol extinction coefficient at height z above the ground, for wavelength λ. Aerosol extinction coefficient represents the area extinction for a unit mass of the aerosol (units: m^2/g) and depends on extinction efficiencies (Q_{ext}) and the average particle density (gm^{-3}), ρ. The extinction efficiencies, Q_{ext} for aerosol is obtained from Mie code.

The satellite AOD is related to PM with the relationship derived for a single homogeneous atmospheric layer in cloud-free skies and with no overlapping of aerosols through Equation 2-2 (Koelemeijer et al., 2006).

$$AOD = (PM) \times (BLH) \times f(RH) \times \frac{3Q_{ext,dry}}{4\rho r_{eff}} \qquad \text{Equation 2-2}$$

where, BLH is the planetary boundary layer height; $f(RH)$ is aerosol hygroscopic growth factor; ρ is the aerosol mass density (gm^{-3}); $Q_{ext,dry}$ is the extinction efficiency under dry conditions; and r_{eff} is the particle effective radius.

However, using Equation 2-2 for converting AOD data to surface level $PM_{2.5}$ is not straight forward as it requires knowledge of factors that influences AOD-$PM_{2.5}$ relationship (more details in Chapter 2.4.1) such as relative humidity, altitude of aerosol layer and aerosol composition (Wang & Christopher, 2003; Engel-Cox et al., 2004). However, the study found other approaches to relate satellite retrieve AOD with surface level $PM_{2.5}$. The first study on relating AOD with surface level $PM_{2.5}$ is found by Chu et al., (2003), where the correlation between satellite AEROENT AOD and surface level $PM_{2.5}$ is studied. The details about the similar studies and previous related work is presented in Chapter 2.5.

NASA Level 1 and Atmosphere Archive and Distribution System ((NASA, 2016b), provides level 2 aerosol product files that are stored in Hierarchical Data Format (HDF), which contains AOD as a scientific data set. The level 2 aerosol product was derived using level 1 spectral radiance data. However, it is not discussed in this thesis how the spectral radiance is converted into AOD. The details about AOD retrieval algorithm over land to obtain AOD from level 1 radiances data can be found in Levy et al., (2009).

The AOD is commonly retrieved at 550 nm from MODIS aerosol product. It is also termed as Aerosol Optical Thickness (AOT). The AOD at land (mainly vegetated surfaces) is retrieved using dark target

algorithm. The algorithms are only applied over daytime, cloud free and snow/ice-free pixels to retrieve AOD at 550 nm. The details of the AOD retrieval over Land and Ocean is provided in MODIS algorithm version 5.2. (Levy et al., 2007; Levy et al., 2007). Dark target algorithm employs three spectral reflectance at 0.47 μm, 0.66 μm and 2.1 μm to retrieve AOD at 550 nm. As the vegetated surface in not "dark" at 550 in the "green" wavelength channel of MODIS, AOD cannot be retrieved at 550 nm directly. It is derived at 0.47 μm and 0.66 μm and is interpolated to 550 nm (Levy et al., 2010).

2.2. AERONET and Available Sun Photometer Networks

The aerosol optical depth is routinely monitored using the CIMEL sun photometer, which is a solar-powered weather hardy robotically pointed sun and sky spectral radiometer. The radiometer makes aerosol optical depth measurements in eight spectral bands between 340 nm and 1020 nm but basically, aerosol optical depth measured at four wavelengths 440 nm, 670 nm, 870 nm and 1020 nm are the standard measurements. These data are used in the AERONET standard procedures to retrieve information on columnar aerosol characteristics such as the aerosol optical depth, size distribution and phase function (Holben et al., 1998). AEROENT AOD is provided in different levels such as Level 1.0 (unscreened), Level 1.5 (cloud-screened) and Level 2.0. (quality-assured).

There are many large networks available globally that use sun photometry to measure Aerosol optical depth namely AERONET (Aerosol Robotic Network), PHOTONS (Photometrie pour le Traitement Operationel de Normalization Satellitaire), SKYNET, MSRFR (Multifilter Rotating Shadowband and Radiometer), AEROCAN, Polar AOD, German AOD Network, SibRad and world meteorological organization global atmosphere watch (WMO/GAW). These instruments are stable, well-calibrated, well-characterized which serves as a ground truth for satellite remote sensing (Hoff & Christopher, 2009).

2.3. Uncertainties on AOD Retrieval

In order to use the aerosol optical depth (AOD at 550 nm) for the further study, the literature review on uncertainties in the retrieval of AOD from MODIS is carried out. The global validation study shows that the accuracy of MODIS collection 005 retrieved AOD values is 0.05 ±0.15% compared to AERONET AOD (Levy et al., 2010).

The study found possible factors such as radiometric calibration (Levy et al., 2007), Atmospheric Correction for Surface Reflectance (Levy et al., 2007; Levy et al., 2010), Selection of aerosol model (Levy et al., 2007), Cloud Screening (Remer et al., 2005), Systematic Error such as Angstrom Exponent (Levy et al., 2010), Surface effect (Luo et al., 2005; Lyapustin, 2001), Observation Geometry (Gatebe et al., 2001; Remer et al., 2001) and Cloud Fraction (Levy et al., 2010) that could add uncertainties in the retrieval of AOD from MODIS.

2.4. Empirical relationship of AOD and $PM_{2.5}$

Irrespective of availability of data, the understanding of PM concentration and monitoring would greatly benefit from consistent and accurate representation in the form of PM maps over a region or air quality index would be simpler for public to visualize the quality of air we breathe. Although, the ground based measurement may be precise but they are only point measurement and doesn't necessarily represent large area because aerosol sources vary over small spatial scales and their lifetime in atmosphere varies depending on particle size, chemical composition, atmospheric humidity and precipitation (Gupta et al., 2006). Thus, in case of particulate matter, in situ observations approach is hampered by the difficulties in the sampling techniques and variation of particulate matter concentration in space and time. Therefore, cost effective

satellite remote sensing techniques is utilise to monitor highly variable aerosol fields on regional scales, where ground based observations are scarce or lacking altogether and it also has benefit of large spatial coverage. Recent satellite data from Moderate-resolution Imaging Spectroradiometer (Chu et al., 2003) or Multiangle Imaging Spectro-Radiometer (Hu et al., 2014; Liu et. al., 2007) can be used to derive indirect estimates of surface $PM_{2.5}$ using statistical method.

Lee et al., (2011) has proposed a statistical model for daily calibration of MODIS AOD to investigate the spatial patterns of ground $PM_{2.5}$. Kloog et al., (2012) extended temporal aspect by incorporating fine scale land use regression to better predict $PM_{2.5}$ concentrations for days with or without satellite AOD measures. The satellite retrieved AOD is dependent on aerosol mass concentration, mass extinction efficiency, hygroscopic growth factor, and effective aerosol scale height (Gupta et al., 2006). Satellite retrieved AOD is processed for cloud contamination before feeding into statistical model to better increase the accuracy of prediction and considering several factors affecting AOD such as relative humidity (Justice et al., 2009), temperature, height of planetary boundary layer in PM-AOD association could explained the variability in atmosphere in a well-mixed condition (Gupta & Christopher, 2009).

van Donkelaar et al., (2010) presented methods to estimate $PM_{2.5}$ globally by establishing a relationship between AOD retrieved from satellite, which requires modelled value for η (depends upon aerosol size, aerosol type, diurnal variation, relative humidity, and the vertical structure of aerosol extinction) for a grid, which is computed from GEOS-CHEM, a global chemical transport model, which is data driven methodology and the robustness of such output depends input emission and meteorological data (Kumar, 2010). There is another approach to predict surface $PM_{2.5}$ using statistical method to established an empirical relationship between AOD retrieved from satellite and ground measurement $PM_{2.5}$ and predicting $PM_{2.5}$ to identify time and space resolve estimate is by using space time Kriging (Kumar et al., 2011) which provides the way to predict $PM_{2.5}$ in different time and space at a spatial location. However, Space time geo-statistical Kriging model is potentially accurate than the remote sensing (RS) for deriving estimates over shorter time scales (e.g., yearly or monthly) and is more accurate for locations that are about 100 km from a ground based monitoring station (Lee et al., 2012). Moreover, RS data provides a useful estimate of pollution levels in the absence of extensive local ground based monitor networks, particularly when the nearest monitor is located more than 100 kilometres away using an integrated remote sensing-chemical transport model approach (van Donkelaar et al., 2010).

2.4.1. Surface Weather Observations Factors Affecting AOD and $PM_{2.5}$ Relationship

Study (Seinfeld & Pandis, 2006) shows that the surface weather conditions that strongly influence the concentration of fine particles in the atmosphere consist of temperature, relative humidity and height of planetary boundary layer. The increase in temperature at the surface accelerates the generation of secondary particles by enhancing the photochemical reactions during day time and thus, increases the amount of concentration of fine particles within the mixing height. Secondary particles are those that are not directly emitted and are form in the atmosphere from other gaseous pollutants, particularly sulphur dioxide, nitrogen oxides, ammonia and volatile organic compounds. By nature, the temperature decreases with increase in height (adiabatic lapse rate) but during strong temperature inversion, it could reduce the vertical mixing, when the boundary layer height is lower as shown in lower part of Figure 2-1, and thus increases chemical concentration of precursors. The higher the concentration of precursors, the faster will be the chemical process to convert the gaseous into fine particles. Thus, the concentration will increase during temperature inversion as well (Wang & Martin, 2007). The sulphate particles tend to have larger extinction coefficients in the atmosphere (Chin et al., 2002). Thus, AOD would correspond to less PM2.5 concentration, when there is more sulfate particles in the air.

Study (Koelemeijer et al., 2006; Liu et al., 2005) have suggested that relative humidity accounts for optical extinction of aerosol in the atmosphere. In case of hygroscopic particles such as ammonium sulphate and ammonium nitrate, the increase in relative humidity can increases the growth up to 2-10 times in particle size, resulting in the significant scattering efficiencies (Malm et al., 2000) and overestimation of aerosol mass. $PM_{2.5}$ measurements at the surface are the measurement of dry mass, where the filter storage and weighing are conducted under controlled RH conditions (at 40-50% RH). Thus, the same AOD values at higher RH would give rise to small particles of dry mass as compared to those at low RH conditions (Liu et al., 2005). The increase in wind speed causes the greater turbulence in the atmosphere, which enhances the vertical mixing. Thus, the concentration of fine particles decrease with increase in wind speed and vice versa (Liu et al., 2007).

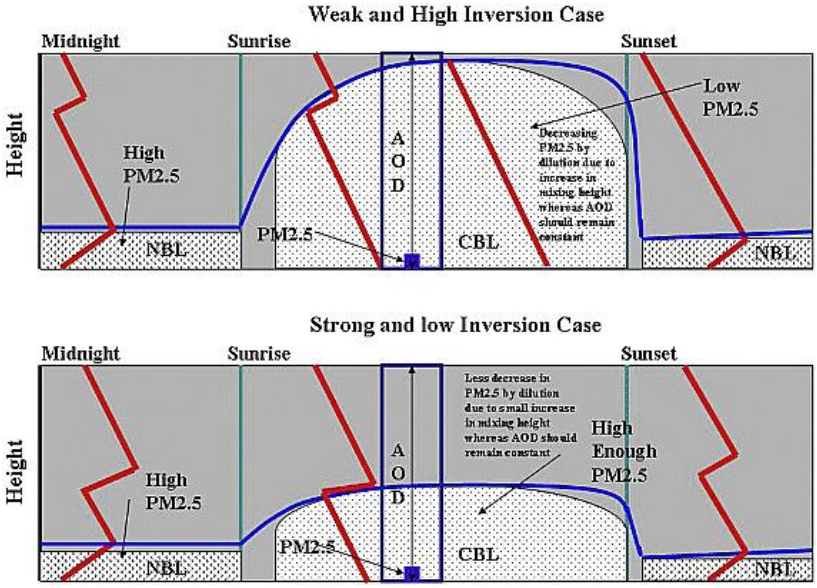


Figure 2-1 Schematic representation of boundary layer height and concentration of PM2.5 at a monitoring station (Gupta & Christopher, 2009)

The change in height of planetary boundary layer (BLH) has significant impact on the association of AOD and $PM_{2.5}$ concentrations. The same particle concentration trapped within a boundary layer with a low mixing height will have less dilution as shown in lower part of Figure 2-1 so the concentration will be higher at the surface measured at a station. In contrast, with increase in height of boundary layer, the dilution of particles increases in the mixing layer so, the concentration measured at a station will be decreasing which is shown in Figure 2-1 top part (Gupta & Christopher, 2009). However, AOD is a columnar measurement and so, it should remain constant within the boundary layer in both strong temperature inversion and weak inversion. Also, a thicker BLH usually corresponds to a lower aerosol density for a given AOD values. Thus, accounting boundary layer height is important in estimating PM2.5 using satellite retrieved AOD. Therefore, BLH is regarded as a denominator of AOD in relationship to AOD-PM2.5 association.

2.5. Previous Works

The relationships of satellite retrieved AOD and surface level $PM_{2.5}$ concentrations has been first investigated in 2003 by Chu et al., (2003) to estimate particulate matters on regional and urban scale. The study areas were chosen as eastern China and India as most populated region, eastern United States/Canada and western Europe as industrialized regions. The 24-hour PM_{10} concentration was temporally matched with daily averaged AEROENT 1.5 data from August to October 2000 and the linear correlation coefficient of 0.82 was found in northern Italy Thus, the result showed the potential use of MODIS aerosol retrieved AOD for air quality assessment.

Wang & Christopher, (2003) study in Jefferson county, Alabama compared MODIS collection 4 Terra and Aqua for daily average of $PM_{2.5}$ in 7 stations. For temporal matching, the hourly PM2.5 observations was averaged centred around the satellite overpass time of Terra and Aqua (±60 min) to compare MODIS AOD with PM2.5. The linear correlation coefficient for both MODIS Terra and Aqua was found to be 0.7, suggesting that surface level $PM_{2.5}$ mass is reflected in MODIS column AOD data. The study shows the diurnal changes of $PM_{2.5}$ mass under cloud free conditions is due to the diurnal changes of the height of the planetary boundary layer, which was found to be maximum in the afternoon around 1:00~2:00. The study infers that the positive correlation of AOD and $PM_{2.5}$ is favoured during the satellite overpass time due to well-mixed boundary layer. However, the study also underlines that the relationship between AOD and $PM_{2.5}$ are affected by hygroscopic growth of particles and extinction efficiency of particles (Q_{ext}).

Engel-Cox et al., (2004) examined the correlation between satellite retrieved AOD and surface level $PM_{2.5}$ over 1138 stations across United States across EPA air quality monitoring sites for April to September 2002. The spatial and temporal collocation of satellite measurements and ground measurements was carried out to pair satellite measurements with ground measurements. For spatial collection, the distance between MODIS pixel and PM2.5 stations using latitude and longitude information is calculated and the pixel within 0.5^o (40-50 km) were averaged over PM2.5 stations. For temporal matching of hourly ground-based measurements, only the measurements that are close to hour (±60 min) corresponding to satellite overpass time was considered. However, the average daily PM2.5 measurements was paired with satellite retrieved AOD measured on the same day. The study found that the overall linear correlation coefficient for daily average and hourly average to be 0.43 and 0.4 respectively across all sites. However, the correlations are stronger in the Eastern half of United States and weaker in the Western part of United States. Engel-Cox et al., (2004) provides one reason as the differences in the datasets may explain some variation as PM2.5 is a surface measurement while AOD is a measure of aerosol scattering in a vertical column from ground to top of the atmosphere. The study highlights that necessity of aerosol vertical distribution data such as from LiDAR systems to determine the effects aerosol height in the variation of correlation of AOD and PM2.5 from east to west or city to city.

With reference to Engel-Cox et al., (2004) study, Engel-Cox et al., (2006) examined the impact of elevated aerosols above the boundary layer on 4 stations at Old town Baltimore. The study found slight improvement in correlation of LiDAR retrieved AOD and PM2.5 from 0.56 to 0.65, when AOD is corrected for height of planetary boundary layer eliminating elevated aerosols in comparison to correlation of LiDAR retrieved total optical depth and PM2.5.

Later, Gupta et al., (2006) has applied the direct correlative model established using Terra and Aqua MODIS satellite retrieved AOD and surface level PM2.5 to estimate $PM_{2.5}$ concentrations over global cities (Hong Kong, Bern, Sydney, Delhi, New York). For temporal matching, all the 24-hr observations of PM2.5 is averaged within one hour (±60 min) of satellite overpass time. For spatial collection, the distance between MODIS pixel and PM2.5 stations using latitude and longitude information is calculated and the pixel within 0.2^o (20-25 km) were averaged over PM2.5 stations. The correlation coefficient averaged over all the global cities stations was reported to be 0.96 which concluded that the satellite retrieved AOD can be used for air quality studies over large spatial area. However, the linear correlation coefficient of hourly PM2.5 and satellite retrieved AOD for individual region is 0.14 (Switzerland), 0.35 (Sydney), 0.40 (Hong Kong), 0.41(Delhi) and 0.6 (New York). In the same study, the sensitivity study of height of planetary boundary layer effect on linear correlation is assessed over 5 stations in Texas. The study found that the linear correlation coefficient value of 0.94 for only those measurements which were between 100-200 m boundary layer height and where, RH less than 50 % were observed but the value of linear correlation coefficient

decreases to 0.36 when boundary layer height were close to 1 km (800-1300 m), which showed that the AOD-$PM_{2.5}$ relationship strongly depends upon the height of boundary layer in the atmosphere as such the concentration of $PM_{2.5}$ decreases with increasing boundary layer height. However, study also highlighted the necessity of aerosol vertical distribution profile and other factors such as hygroscopic growth of aerosol particles, as pointed out by Wang & Christopher, (2003) to further refine the analysis.

Thus, Koelemeijer et al., (2006) study in Europe, introduced the height of planetary boundary layer and relative humidity extracted from European Centre for Medium-Range Weather Forecasts (ECMWF) archive into the correlative model of AOD-$PM_{2.5}$ association to improve the correlation of AOD and PM2.5. For temporal matching of hourly ground-based measurements, only the measurements that are close to hour (±60 min) corresponding to satellite overpass time was considered. The average daily PM2.5 measurements was compared to AOD (instantaneous) measured on the same day. The average correlation coefficient of daily averaged $PM_{2.5}$ and hourly averaged PM2.5 increased from 0.27 to 0.48 and 0.38 to 0.59 respectively, when AOD is divided by PBL and $f(RH)$, which is the aerosol hygroscopic growth factor, which describes the increase of the aerosol extinction cross-section or scattering coefficient with RH for both daily and hourly averaged $PM_{2.5}$ for air quality monitoring station with dominant rural background concentration. However, the correlation of $PM_{2.5}$ with 1/BLH didn't improve significantly for both hourly and daily average $PM_{2.5}$ concentrations and there was no correlation found between PM2.5 with $1/f(RH)$.

Gupta & Christopher, (2008) study in North Birmingham, Alabama for MODIS collection 5 Terra product for a single site. The study assessed different criteria such as box sizes (5 × 5 pixels, 4 × 4 pixels and 3 × 3 pixels) and quality flags to obtain coincident MODIS Terra retrieved AOD over PM2.5 location to verify the assumption of using 5 × 5 pixels of Ichoku et al., (2002). The results showed that changing box size from 5 × 5 pixels to 3 × 3 pixels do not significantly affect AOD-PM2.5 relationship for daily averages. However, the box available data points were reduced by 8 % in 3 × 3 pixels. In case of quality flags, it is found critical for daily averages. The study shows that the correlation of $PM_{2.5}$ and AOD is higher when hourly averaged $PM_{2.5}$ is considered in the correlative model than daily averaged PM2.5 as the instantaneous MODIS AOD correlate better with hourly averaged PM2.5 compared to 24 hr averaged PM2.5 due to diurnal variations in PM2.5 mass measurements. The correlation was increased from 0.52 to 0.62 for the correlation between hourly averaged $PM_{2.5}$ and satellite retrieved AOD compared to daily average PM2.5 data. However, the study shows that the relationship between AOD and PM2.5 varies with location and time (Engel-Cox et al., 2004; Gupta et al., 2006; Koelemeijer et al., 2006). The study found that there was only 2 μg/m3 of difference in monthly, seasonal and annual means estimated using relationship of Terra MODIS (with 50 % of time sampled over the location) and surface PM2.5 observations, when compared to 100 % data availability of monitoring station. Thus, the study concluded that the satellite data can be used for evaluating air quality of a region in absence of ground measurements data.

Schaap et al., (2009) collocated PM2.5 data to established relationship with AERONET AOD at Cabauw, The Netherlands. The established AERONET AOD and PM2.5 relationship was investigated for the different time window for which the MODIS overpasses the station. The correlation was increasing when using PM2.5 measurements of mid-day thus, the relationship for central time window (11:00-15:00) was used for further analysis. The study found that the inclusion of mixing height doesn't improve the correlation between satellite retrieved AOD from MODIS collection 5 and $PM_{2.5}$ measured from a Tapered Element Oscillating Microbalance (TEOM) with Filter Dynamics Measurement System (TEOM-FDMS). The boundary layer height for study was considered from National Institute for Public Health and the Environment (RIVM) backscatter LIDAR station of de Bilt and meteorological variables such wind speed and wind direction were considered from Cabauw Experimental Site for Atmospheric Research (CESAR) data portal were used to assess if the air mass origin would effect on AEROENT AOD-PM2.5 relationship. It is mentioned that a relationship of AOD and $PM_{2.5}$ for the continental air mass originating from south,

south east and east were identical. However, the air mass originating from west do not show positive relation with AOD and PM2.5, where the results was not shown. Also, it is mentioned that the longer time series is necessary for statistical analysis of effect of air mass origin on AOD-PM2.5 relationship. So, it was not further considered in the study. The study compared $PM_{2.5}$ with AERONET and MODIS retrieved AOD. The slope was found to be 120 μg/m3 and bias of 5.1 μg/m3. The TEOM with FDMS accounts for loss of volatile components (e.g. Ammonium nitrate) from the TEOM at standard temperature of 40 degrees Celsius. It should be noted that the $PM_{2.5}$ concentration measured from TEOM-FDMS instrument at Cabauw and other studies with PM2.5 measured with instruments like TEOM, Beta attenuation monitoring (BAM) method results cannot be compared until correction factors were used (Hoff & Christopher, 2009).

Recent study (Guo et al., 2014) in Beijing area by showed the correlation analysis of MODIS AOD with ground based $PM_{2.5}$ for different seasons (spring, summer and autumn), where the correlation was higher for autumn (0.731) compared to summer (0.597). The study showed that the correlation coefficients between MODIS AOD and ground $PM_{2.5}$ is improved by about 11 % and 4 % in summer, when AOD is corrected for BLH only and with both BLH and f (RH) respectively, while the correlations in spring and autumn was barely changed with. However, the prediction error of the model was higher for summer with RMSE of 55.09±32.86. The BLH data for the study area was obtained from vertical sounding profiles and RH was obtained from Global Surface Hourly database, which is a part of Integrated Surface Database (ISD) provided by National Oceanic and Atmospheric Administration (NOAA's) National Climatic Data Centre (NCDC). The hourly PM2.5 observations was averaged centred around the satellite overpass time of Terra and Aqua to compare MODIS AOD with PM2.5, to minimize the temporal noises in the PM2.5 data. For temporal collocation, the time window of ±150 min was taken to include more data points of PM2.5 and AOD compared to averaging PM2.5 in ±60 min reported in other studies (Wang & Christopher, 2003). For validation of MODIS AOD with AERONET AOD, spatial collocation, 5 × 5-pixel box (50 km × 50 km) was chosen over the surface measurement location to average AOD_{550} pixels and to match with ±30 min average AOD_{550} from AERONET centred around the MODIS overpass time. The 50 km × 50 km window of MODIS AOD was taken assuming that it would take 50 $kmhr^{-1}$ for aerosol mass (air mass) to transport in mid-troposphere based on analysis of Total Ozone Mapping Spectrometer (TOMS) aerosol index images over the Atlantic Ocean, which would match a 15-min observations over one-hour time period of ground measurement at AERONET station to represent a similar air mass as observed by MODIS, which is the standard strategy to obtain coincident PM2.5 and MODIS AOD at a location (Ichoku et al., 2002).

Liu et al., (2005) study in eastern United States has used MISR retrieved AOD with 346 EPA sites and ground level PM2.5 measurements using a generalized linear regression (GLM) model. GEOS-3 (Geostationary Operational Environmental Satellite) meteorological data such as planetary boundary layer and RH was included in the model to account for the variation in particle vertical profiles, composition and optical properties. The mean and standard deviation of the AOD measurements for each 3 × 3 pixels (17.6 × 17.6 km^2) centred at a given EPA site was matched with the daily average PM2.5 measurement of the same day. The BLH and RH interpolated at 10:00-11:00 were also matched to EPA MISR (Multi-angle Imaging SpectroRadiometer) measurements. Since, the repetition of MISR AOD measurement varies 2-9 days and ground level PM2.5 was measured in various schedules in monitoring stations (daily, every third day, every sixth day), the matching process was considered as a random sampling of ground observations of PM2.5 over a given area. The overall empirical model explained 48 % of the variability in the $PM_{2.5}$ concentrations where, MISR AOD and BLH were able to explain 18 % and 15 % of the variation in $PM_{2.5}$ concentrations respectively. In regression model, AOD, BLH and RH were fit using power-law whereas the RH was expressed in an exponential form which is a simplified representation of the particle growth effect based on the regression statistics but in reality, the particle growth does not grow strictly exponentially and depends on the particle composition (Malm et al., 2000). The regression analysis of model estimated the

regression coefficients of power dependence of AOD, BLH and RH as 0.447±0.022, -0.361± 0.023 and exp. (-0.634±0.115) respectively. The estimated power of AOD was positive meaning the surface concentration of $PM_{2.5}$ varies directly with increase in MISR AOD measurements. The negative sign of BLH denotes that the BLH is indirectly proportional to the surface concentration of $PM_{2.5}$ and the negative parameter estimate of exponential of RH indicated that the same AOD values would correspond to lower $PM_{2.5}$ concentrations with increase in RH. The study concluded the inclusion of information on vertical distribution of particle mass would improve the model predictability.

Liu et al. (2007) study at St Louis, Missouri has compared both MISR and MODIS to predict ground-level $PM_{2.5}$ concentration using daily average $PM_{2.5}$ concentration of 22 EPA FRM (Federal Reference Method) monitoring stations. The meteorological data such as height of planetary boundary layer (PBL), surface wind speed, wind direction, relative humidity, and air temperature at surface were extracted from National Oceanic and Atmospheric Administration (NOAA)'s Rapid Update Cycle (RUC) model at the spatial resolution of 20 km. The RUC model is a high-frequency operational weather forecast and data assimilation system developed at NOAA. The predictor variables such as surface air temperature, surface wind speed, PBL, MODIS/MISR AOD and two categorical variables wind direction, and seasonal indicators are used to develop generalized linear regression model. The hourly values of RUC extracted meteorological parameters were averaged between 10:00-12:00 corresponding to MODIS overpass time and then, PM2.5 concentrations is matched with the averaged RUC meteorological parameters from the RUC grid cell, where EPA site falls. To match regional measurement of RUC data with point measurement of PM2.5, 30 km search radius around each EPA monitoring site was used for averaging MODIS AOD pixels. Finally, MODIS AOD is matched with PM2.5 concentrations and the time-averaged RUC20 meteorological parameters by date. The average temporal spacing was three days for consecutive temporally matched daily measured PM2.5 concentration and MODIS AOD.

MODIS AOD, BLH and wind were assumed to have power law functional forms to account for non-linear relationship with $PM_{2.5}$ concentration and temperature were assumed to have an exponential form to account temperature fluctuations with $PM_{2.5}$ concentrations. RH didn't improve model performance so, was not included in final model. The regression analysis of model estimated the regression coefficients of power dependence of AOD for Spring and non-spring, PBL, Wind, Temperature as 0.20, 0.45, -0.14, -0.17 and exp (-0.005) respectively. The negative sign of regression coefficient of wind speed shows that MODIS AOD predicts lower $PM_{2.5}$ concentrations at higher wind speed. The greater turbulence caused by wind speed enhances vertical mixing so, predicted $PM_{2.5}$ would be lower. In case of temperature, the increase in air temperature accelerates the generation of secondary particles near the surface, causing a higher proportion of particle mass in the mixed layer, thus increase in temperature would increase the $PM_{2.5}$ concentrations, whereas the PBL is not significant in MISR. The meteorological variables have accounted for 23 % of variability in $PM_{2.5}$ concentrations.

Pelletier, Santer, & Vidot (2007) study in Lille, France first used the linear approach with only AERONET AOD as a predictor variable to established relationship with PM_{10} and found that the linear model could only explain 27% of variability and the average uncertainty of model was 35% in estimating the PM_{10} concentration. Thus, the model failed to accurately explain the data. The study then used auxiliary meteorological variables such as wind vector, pressure, relative humidity, perceptible water and Julian date as a function that are added to form an additive varying coefficient model. The performance of the model is improved with an average uncertainty of less than 20% and found linear correlation of 0.87 between fitted and expected PM10, when such linear relationship of AOD and PM2.5 is conditioned on auxiliary parameters. The meteorological variables were used in the study were collected from National Centre for Environmental Prediction (NCEP) through the Distributed Active Archive Centre (DAAC). Penalized smoothing splines were used to avoid over-fitting of the coefficient of free parameters of model, which are

to be estimated from the data, which determines the shape of the functions. It is found that the PM_{10} concentration has negative relationship with RH and positive relationship with pressure and perceptible water.

In (Gupta & Christopher, 2009) study, first satellite retrieved AOD were spatially and temporally matched with ground level PM2.5 observations over station. Secondly, hourly RUC parameters averaged during the satellite over time of MODIS over PM2.5 station using similar data integration method as in Gupta & Christopher, (2008), which were later paired with satellite AOD and PM2.5 match-up. The two-variate regression method and multivariate regression method is applied in South eastern US to estimate PM2.5 in the data pair of spatially and temporally matched RUC20 parameters, AOD and PM2.5 observations. The study showed that the correlation increased from mean value of 0.60 over 22 stations with only MODIS satellite retrieved AOD (collection 5) to 0.71, when meteorological variables such as air temperature, surface relative humidity, wind speed and height of planetary boundary layer and cloud fraction was introduced into the multivariate regression model. The hourly analysis data of temperature, surface RH, wind speed and height of planetary boundary layer at 20 × 20 spatial resolution obtained from RUC20 model was used in the study. The correlation coefficient over each station varies and considering all stations together, about 13.1 % increase in correlation coefficient of hourly average surface level $PM_{2.5}$ concentrations is estimated. Stepwise multiple regression analysis showed the first order influence of temperature on $PM_{2.5}$-AOD relationship and second order impact is due to BLH.

Liu et al., (2009) study at Massachusetts (except for Cape Cod) and part of surrounding states has used AOD retrieved from Geostationary Operational Environmental (GOES) Satellite aerosol/smoke product and meteorological parameters such as mixing height, RH, air temperature and wind speed from RUC model. Land use information such as population information, road type to estimate regional spatial and temporal variability of $PM_{2.5}$ are also used as a covariate in a model. The GOES satellite allows AOD retrieval in 30-min frequencies between sunrise and sunset, which is reported at roughly 6.5 km * 2.4 km rectangular GOES pixels. However, GOES retrieved AOD is averaged between 10:00 to 15:00 hours local time to generate daily averaged AOD estimates to better match with daily averaged PM2.5 concentration. The AOD is interpolated using weighted average using Thiessen polygons intersecting with a 4 km grid which were matched with the U.S. EPA site, that falls in the same grid cell. The RUC parameters which centroid were also matched with the EPA site with the nearest neighbourhood method for developing a spatial model. The study provides the comparison on capability of prediction of $PM_{2.5}$ using non-AOD model, which was developed only using meteorological and land use information as predictor of $PM_{2.5}$ concentrations with AOD model developed using GOES retrieved AOD and metrological information's. The comparison is carried out in areas, where AOD is missing during the study period by using two-stage generalized additive models (GAM). A linear regression between fitted and observed $PM_{2.5}$ concentrations produced adjusted R-squared of 0.79 for AOD model (Correlation coefficient, R=0.89) and 0.48 for non-AOD model (Correlation coefficient, R=0.70), which showed that PM2.5 concentrations is predicted well when AOD values are present.

Hu et al., (2013) study at Atlanta metro area for year 2003 has developed a geographically weighted regression (GWR) model to examine the relationship among $PM_{2.5}$, aerosol optical depth, meteorological parameters such as wind speed, temperature, relative humidity, height of planetary boundary layer and land use information to examine the spatial variability and spatial non-stationarity of the $PM_{2.5}$ and AOD by producing local regression results. The meteorological data were collected from National American Land Data Assimilation System (NLDAS) and the Land use information such as forest cover were used. The study used Community Multiscale Air Quality (CMAQ) grid with spatial resolution of 12 km × 12 km as a based grid for prediction. MODIS AOD data is resampled using a nearest neighbourhood approach to the based grid of CMAQ. The hourly NLDAS meteorological measurements for 10:00 to 4:00 were averaged

to generate daily averaged meteorological value. The data pair of AOD and meteorological data sampled from CMAQ and NLDAS grid which are closest to the PM2.5 measurement site are paired to carry out geographically weighted regression. The overall mean local variance explained by the geographically weighted regression model was 0.61 with prediction accuracy of 83% compared to a GWR model developed with AOD as only predictor where overall mean R-squared obtained was only 0.38. Thus, the study found showed that the incorporation of NLDAS meteorological variables can significantly improve the model performance. The study also compared the mean adjusted R-squared with Ordinary Least Squares (OLS) model developed for the same study and found that it only explains 47 % of variability. Thus, the study suggests that the GWR (Geographically Weighted Regression) model has better performance than the OLS model.

The AOD and PMx relationship found in literature review is presented as an overview in Table **2-1**

Table 2-1 Overview of previous works on AOD-PM association

Date	Region	AOD	Station	$PM_{2.5}/PM_{10}$	Intercept	Slope	Correlation coefficient, R	R-squared	Predictor variables other than AOD					Prediction Error	Author
									BLH	RH	Temper-ature	Wind Speed	Wind Direction		
Aug-Oct, 2000	Italy	AERONET Level 1.5	1	PM_{10} (24-hr)	8	54.7	0.82	NA	NC	NC	NC	NC	NC	NA	Chu et al., (2003)
2002	Alabama	MODIS (Terra) 4	7	$PM_{2.5}$ (24-hr)	-0.23	77	0.67	NA	NC	NC	NC	NC	NC	NA	Wang & Christopher, (2003)
		MODIS (Aqua) 4			1.93	66.67	0.76	NA							
		MODIS (both Terra and Aqua)			0.85	71.43	0.7	NA							
April to September, 2002	United States	MODIS (Terra) 4	1338	$PM_{2.5}$ (24-hr)	7.54	18.66	0.43	NA		NC	NC	NC	NC	NA	Engel-Cox et al., (2004)
				$PM_{2.5}$ (hourly)	6.35	22.55	0.4	NA							
July 1 to August 30, 2004	Old Town Baltimore	MODIS 4 (both Terra and Aqua)	4	$PM_{2.5}$ (24-hr)	11.1	25.3	0.57	NA	Correlation of LiDAR AOD and hourly PM2.5 below boundary layer increases from 0.56 to 0.65 compared to when LiDAR total optical depth.	NC	NC	NC	NC	NA	Engel-Cox et al., (2006)
				$PM_{2.5}$ (hourly)	5.22	31.05	0.65	NA							
July 2002 to November 2003	Global cities	MODIS (both Terra and Aqua)	26	$PM_{2.5}$ (24-hr)	-24.83	166.7	0.96	NA	Sensitivity study showed that R value increased to 0.94 for 100-200 m mixing height, but R value decreases to 0.36 when mixing height of 800-1300 m.		NC	NC	Surface WS were used to calculate BLH.	NA	Gupta et al., (2006)
2003	Europe	MODIS (both Terra and Aqua) 4	53	$PM_{2.5}$ (24-hr)	NA	NA	0.27	NA	BLH and RH was extracted from ECMWF archive. R value increased to 0.48 for AOD/(BLH*f(RH))		NC	NC	NC	NA	Koelemeijer et al., (2006)

Date	Region	AOD	Station	$PM_{2.5}/PM_{10}$	Intercept	Slope	Correlation coefficient, R	R-squared	Predictor variables other than AOD					Prediction Error	Author
									BLH	RH	Temper-ature	Wind Speed	Wind Direction		
			9	$PM_{2.5}$ (hourly)	NA	NA	0.38	NA	R value increased to 0.59 for AOD/(BLH*f(RH))		NC	NC	NC	NA	
1999 to 2002	Lille	AERONET Level 1.5	5	PM_{10} (24-hr)			0.87		NC	Data collected from NCEP through DAAC. Negative relationship with PM10 concentration	NC	Data collected from NCEP through DAAC. Less effect on correlation	NC	Standard deviation of residuals to about 0.20	Pelletier et al., (2007)
February 200 to 30 June 2006	North Birmingham, Alabama	MODIS (Terra) 5	1	$PM_{2.5}$ (24-hr)	15.8	27.5	0.52	NA	NC	NC	NC	NC	NC	NA	Gupta & Christopher, (2008)
				$PM_{2.5}$ (hourly)	8.8	29.4	0.62	NA							
August 2, 2006 to 6 May 2007	Cabauw	AERONET, MODIS (both Terra and Aqua) 5	1	$PM_{2.5}$ (24-hr)	5	120	NA	0.51	Lidar data from de Bilt was considered. No improvement in correlation between PM2.5 and AOD/BLH, AOD/Maximum BLH (Results not shown)	NC	NC	Linked with Air mass origin. Not considered for further analysis as longer time series data is needed for statistical analysis (Results not shown).		NA	Schaap et al., (2009)
March to November 2012	Beijing	MODIS (Terra and Aqua) 5	1	$PM_{2.5}$ (spring-MAM)	-24.45	90.9	0.685	NA	BLH data is collected from vertical sounding profile of radiosonde observation. RH data was collected from Global surface hourly database provided by NOAA's National Climatic Data Center. No		NC	NC	NC	Mean RMSE$\pm\sigma$= 41.502$\pm$ 17.8	(Guo et al., 2014)

Date	Region	AOD	Station	$PM_{2.5}/PM_{10}$	Intercept	Slope	Correlation coefficient, R	R-squared	Predictor variables other than AOD					Prediction Error	Author
									BLH	RH	Temper-ature	Wind Speed	Wind Direction		
									significant improvement in the correlation when corrected for BLH and f(RH)						
				$PM_{2.5}$ (summer-JJA)	-32.55	111.1	0.597	NA	Correlation increased from 0.597 to 0.660 when AOD/BLH	Correlation improved from 0.597 to 0.621 when AOD/BLH*f(RH)	NC	NC	NC	Mean RMSE±σ= 55.094± 32.856	
				$PM_{2.5}$ (Autumn-SON)	-43.8	200	0.731	NA	No significant improvement in correlation for AOD/BLH and AOD/BLH*f(RH) respectively.		NC	NC	NC	Mean RMSE±σ= 53.173± 20.186	
2001	Eastern US	MISR (Level 2)	346	$PM_{2.5}$ (24-hr)	NA	NA	NA	0.48	It was retrieved from GEOS-3 satellite. Significant, increased 15% variability and estimated regression coefficients is -0.361± 0.023.	It was retrieved from GEOS-3 satellite. Significant, Estimated regression coefficient is -0.634±0.115.	NC	NC	NC	RMSE= 6.2 µg/m3, Relative Error= 45%	Liu et al., (2005)
2003	St. Louis	MODIS (Collection 4)	22	$PM_{2.5}$ (24-hr)	NA	NA	0.69	0.51	It was retrieved from NOAA's RUC20 model. Significant, Regression coefficient of -0.14.	It was retrieved from NOAA's RUC20 model. Not significant in the model.	It was retrieved from NOAA's RUC20 model. Regression coefficient of exp (0.005), Significant	It was retrieved from NOAA's RUC20 model. Regression coefficient of -0.17, Singificant	It was retrieved from NOAA's RUC20 model. Regression coefficient of 1.15 for E, 1.12 for N, 1.55	RMSE= 5.6 µg/m3, Relative Error= 39 %	Liu et al., (2007)

Date	Region	AOD	Station	$PM_{2.5}/PM_{10}$	Intercept	Slope	Correlation coefficient, R	R-squared	Predictor variables other than AOD					Prediction Error	Author
									BLH	RH	Temper-ature	Wind Speed	Wind Direction		
													for S, 1 for W, Significant		
2004-2006	South eastern US	MODIS Terra (collection 5)	85	$PM_{2.5}$ (hourly)	NA	NA	0.6	NA	It was extracted from RUC20 model. 1.3% increment in correlation	It was extracted from RUC20 model. 2.3% increment in correlation	It was extracted from RUC20 model. 7.3% increment in correlation	It was extracted from RUC20 model. 1.2% increment in correlation	NC	Average Uncertainty= 34 % (hourly) and 24 % (daily)	(Gupta & Christopher, 2009)
April 2003 to June 2005	Massachusetts	GOES	32	$PM_{2.5}$ (24-hr)	NA	NA	0.89	0.79	It was extracted from RUC20 model.					Mean Relative Error= 30%	Liu et al., (2009)
2003	Atlanta Metro Area	MODIS (Terra and Aqua) 5	119	$PM_{2.5}$ (24-hr)	NA	NA	0.94	o.61	It was collected from NLDAS. Median value of regression coefficient of -0.003	It was collected from NLDAS. Median value of regression coefficient of -0.08	It was collected from NLDAS. Regression coefficient of 0.36	It was collected from NLDAS. Regression coefficient of -0.33	NC	83 % in model fitting	Hu et al., (2013)

Note: NC: Not Considered, NA: Not Available

3. DATA

3.1. Data

The data collected for the study consist of secondary data such as satellite retrieved AOD data retrieved from MODIS Terra, Aqua satellite aerosol product, atmospheric profile product, surface observation of $PM_{2.5}$ data, meteorological data such relative humidity, temperature, wind speed, wind direction and height of Planetary boundary layer (BLH) which are briefly presented in Table 3-1.

Table 3-1 Description of data under study

S. N.	Data Type	Source	Data provider	Description
1	$PM_{2.5}$	RIVM	Drs. Valentijn Venus	Hourly $PM_{2.5}$ data (in μg/m3) for 2013-01-01 to 2014-12-31
2	Temperature	Royal Netherlands Meteorological Institute (KNMI) (http://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi)	Downloaded from KNMI web for weather data	Temperature (in degrees Celsius) at 1.50 m height during the observation
3	Relative Humidity	KNMI	Downloaded from KNMI web for weather data	Relative humidity (in percentage) at 1.50 m height during the observation
4	Wind Speed	KNMI	Downloaded from KNMI web for weather data	Hourly average wind speed (in m/s)
5	Wind Direction	KNMI	Downloaded from KNMI web for weather data	Wind Direction (degrees) averaged over the last 10 minutes of the last hour (360 = North, 90 = East, 180 = South, 270 = West, 0 & 990 = variable wind
6	Height of Planetary Boundary Layer	KNMI ftp-server (bbc.knmi.nl)	Henk Klein Baltink	Every 10 minutes' observation data collected LiDAR backscattering profile of Cabauw Ceilometer station
7	Aerosol Products (MXD04)	NASA LAADS WEB (https://ladsweb.nascom.nasa.gov/data/search.html)	NASA LAADS WEB	Aerosol Optical Depth (Optical_Depth_Land_And_Ocean at 550 nm extracted from MODIS Aqua, Terra Collection 051 level 2
8	Atmospheric Profile products (MXD07)	NASA LAADS WEB (https://ladsweb.nascom.nasa.gov/data/search.html)	NASA LAADS WEB	Retrieved Temperature-Level 18 (920 hPa), 19 (950 hPa), 20 (1000 hPa); Retrieved Moisture-Level 18 (920 hPa), 19 (950 hPa), 20 (1000 hPa); Retrieved Height-Level 18 (920 hPa), 19 (950 hPa), 20 (1000 hPa); Brightness Temperature-Band 10, 11, 12; Surface Skin Temperature, Surface Pressure, Surface Elevation, Tropopause Height, Lifted Index, K Index, Water Vapor, Water Vapor-Low, Water Vapor-High, Water Vapor-Direct extracted from MODIS Aqua, Terra Collection 051 level 2

3.2. Study Area

The study area (Figure 3-1) is Cabauw, which is located (51.974° N, 4.923° E) in western part of The Netherlands, 20 km southwest of the city of Utrecht and 45 km from the North Sea. The measurement site is classified as rural regional background concentration where, the dominant local sources of pollution at the site are anthropogenic and naturally occurring background concentrations. Study by Matthijsen & ten Brink, (2007) reported that the dominant contribution of main chemical components to average $PM_{2.5}$ background concentration at Cabauw are secondary inorganic aerosols (combination of sulphate, nitrate and ammonium), carbon, sea salt and mineral dust. Cabauw air quality monitoring station is situated inside the Cabauw Experimental Site for Atmospheric Research (CESAR), which is the experimental atmospheric research centre (http://www.cesar-observatory.nl/) in The Netherlands. Royal Netherlands Meteorological Institute (KNMI) operated weather station and LiDAR station measuring height of planetary boundary layer is also located within the CESAR Observatory. The surrounding nearby areas of CESAR observatory site mostly consist of agricultural lands. The surface elevation differs only few meters (<20 km radius around Cabauw). The ground surface consists of windbreaks such as Orchard trees and low houses over distant East. The landscape is mostly open consisting of short mowed grass-land on the West, which is the dominant wind direction at Cabauw. However, the distant North and South consist of mixed landscapes, mostly pasture and windbreaks such as Orchard trees (Monna & Bosveld, 2013).

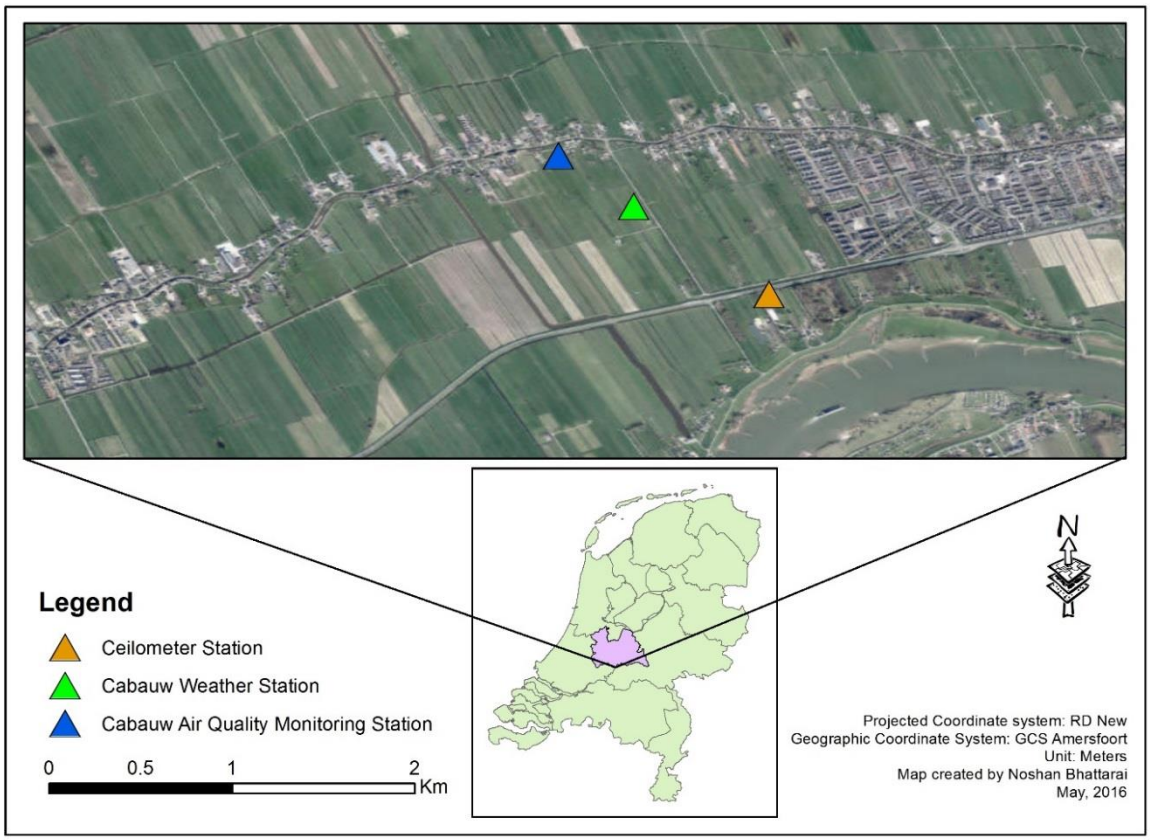


Figure 3-1 Location map of air quality monitoring station at Cabauw

3.2.1. MODIS Observations

3.2.1.1. MODIS Aerosol Product

MODIS is an instrument that operates on the Terra and Aqua Satellite. With a swath width of 2330 Km, and with two of the same sensors in orbit, MODIS observes the entire earth on every other day. The Terra and Aqua satellite overpass Netherlands at approximately 10:30 and 13:30 local time. It captures spectral

radiances in 36 spectral bands ranging in wavelength from 0.4 μm to 14.4 μm. The band spatial resolutions are grouped into 250 m (Band 1 and Band 2), 500 m (Band 3 to Band 7) and 1000 m (Band 8 to Band 36). The MODIS aerosol level 2 product over Netherlands were acquired from the Level 1 Atmosphere Archive and Distribution System (LAADS). The total size for the storage capacity of gridded swath data of MODIS level 2 aerosol product were about 5 gigabytes. The daily level 2 data were produced at a spatial resolution of 10 km × 10 km at nadir. The collection 051 Aerosol product is used to extract AOD Optical_Depth_Land_And_Ocean at 550 nm for both Terra (MOD04) and Aqua (MYD04) for the time period 2013 to 2014. MODIS data "collections" are basically a data version of MODIS. MODIS dataset which are reprocessed (from launch), when new and improved science algorithms are developed and then tagged as a new "Collection". There is six MODIS data Collection that has been processed since the launch of MODIS Terra in early 2000. The Collection available so far are Collection 001, 003, 004, 005, 051 and 006.

3.2.1.2. MODIS Atmospheric Profile product (Satellite retrieved atmospheric observations)

The MODIS atmospheric profile product over The Netherlands were acquired from the Level 1 Atmosphere Archive and Distribution System (LAADS). The daily level 2 data were produced at a spatial resolution of 5 km × 5 km at nadir. The similar MODIS collection 051 compared to MODIS aerosol profile product is used to extract scientific data sets (SDS) atmospheric profile product for the time period 2013 to 2014. The parameters related to atmospheric stability, temperature and moisture profiles and atmospheric water vapour is selected to retrieve from MODIS atmospheric profile product for both Terra (MOD07_L2) and Aqua (MYD07_L2) satellite as it would provide information on the characterization of the atmosphere (Seemann et al., 2006). The total size for the storage capacity of gridded swath data of MODIS Level 2 atmospheric product were about 45 gigabytes. The MODIS temperature and moisture profiles are produced at 20 vertical atmospheric levels of air pressure. The air pressure level from surface to upper air are at 1000, 950, 920, 850, 780, 700, 620, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 10 and 5 hPa. In our study, satellite retrieved temperature, pressure and height were extracted at lower air pressure level of 1000 (level-20), 950 (level-19), 920 hPa (level-18) because in most situation, the height of the boundary layer lies at an elevation below 600 hPa (Feng et al., 2015).

MODIS atmospheric profile product was used to extract potential predictor variables such as Retrieved Temperature profile-Level 18 (920 hPa), Retrieved Temperature profile-Level 19 (950 hPa), Retrieved Temperature profile-Level 20 (1000 hPa), Retrieved Moisture profile-Level 18 (920 hPa), Retrieved Moisture-Level (950 hPa), Retrieved Moisture profile-Level 20 (1000 hPa), Retrieved Height-Level 18 (920 hPa), Retrieved Height-Level 19 (950 hPa), Retrieved Height-Level 20 (1000 hPa), Brightness Temperature-Band 10, Brightness Temperature- Band 11, Brightness Temperature- Band 12, Cloud Mask, Surface Skin Temperature, Surface Pressure, Surface Elevation, Processing Flag, Tropopause Height, Lifted Index, K Index, Water Vapor, Water Vapor-Low, Water Vapor-High, Water Vapor-Direct.

3.2.2. Ground Observations

The ground observation of hourly $PM_{2.5}$ concentrations is obtained from Valentijn Venus for time period 2013-2014 for Cabauw Rural Air Quality Monitoring Network of the Netherlands operated by Netherlands National Institute for Public Health and the Environment (in Dutch abbreviated to RIVM). According to Guus Stefess (personal communication, September 1, 2015), the measured values were hourly average values which were obtained by automatic sampling every hour from 00:09 to 00:51 hr from January 2013 to July 2014. The measured hourly average PM2.5 concentrations were exactly at specified minutes from August to December 2014. According to Guus Stefess (personal communication, May 18, 2016), reference method (gravimetric) was used for hourly PM2.5 measurements in 2013 and since January 2014, equivalence method (BAM) was used. Rural Air Quality Monitoring Network is the translated name for *Landelijk Meetnet*

Luchtkwaliteit (LML) which is the national air quality measurement network of the Netherlands. The LML network is maintained by *Rijksinstituut voor Volksgezondheid en Milieu* translated as Netherlands National Institute for Public Health and the Environment (RIVM, 2016), which publishes observational data at http://www.lml.rivm.nl/gevalideerd/index.php in Central European Time (CET).

3.2.3. Surface Weather Observations

The hourly data from surface weather observation for Cabauw, in the Netherlands were obtained from Royal Netherlands Meteorological Institute (in Dutch abbreviated to KNMI). *Koninklijk Nederlands Meteorologisch Instituut*, which collects and publishes hourly data on surface weather condition (e.g. wind speed, wind direction, temperature and relative humidity). For the study period, 2013-2014, these were obtained from the KNMI for the nearest weather station to Cabauw air quality monitoring (see Table 3-2).

KNMI operates a LD-40 Ceilometer (LiDAR) land station at Cabauw for retrieving the height of planetary boundary (BLH) layer (Haij et al., 2007) which is referred as **"upper-air"** observations. KNMI provides BLH data every 10 minutes in Universal Coordinated Time (UTC). The data was collected (Henk Klein Baltink, personal communication, September 29, 2015) through KNMI ftp server (bbc.knmi.nl). For the study period 2013-2014, these were obtained from the KNMI for the nearest Ceilometer station to Cabauw air quality monitoring station (see Table 3-3).

Table 3-2 Nearest weather station from rural air quality monitoring network station

Air quality monitoring station	Latitude	Longitude	Nearest Weather station	Nearest Distance (Km)
Cabauw-Wielsekade	51.974	4.923	Cabauw	0.4

Table 3-3 Nearest Ceilometer station from rural air quality monitoring network station

Air quality monitoring station	Latitude	Longitude	Nearest Ceilometer station	Nearest Distance (Km)
Cabauw-Wielsekade	51.974	4.923	Cabauw	1.1

3.2.4. Data pre-processing and processing

3.2.4.1. Selection of AOD variable and MODIS collection

For the quantitative analysis, the variable considered was aerosol optical depth (the variable "Optical_Depth_Land_And_Ocean", which is the optical depth of aerosol from both the land and ocean models at 550 nm with for the study period (2013-2014). Similar studies carried out in Engel-Cox et al., (2004); Engel-Cox et al., (2006) and Gupta & Christopher, (2008) has also considered aerosol optical depth (the variable "Optical_Depth_Land_And_Ocean") from MODIS Level 2 aerosol product (collection 5) in their studies. The data values were stored as integer in the gridded swath data of MODIS aerosol gridded satellite image. However, the scaling factor is used to convert the extracted AOD values into meaningful physical values using the scaling factor of 0.001 (http://modis-atmos.gsfc.nasa.gov/_specs_c51/MOD04_L2.CDL.fs).

The aerosol data retrieved over land and ocean are processed as collection 5 or C005 for Terra and Collection 51 or C051 for Aqua (Levy et al., 2010). However, gridded swath data of both MODIS Terra and Aqua Collection 051 were used in the current study. It is mentioned in LAAD web (https://ladsweb.nascom.nasa.gov/data/search.html) that collection 051 consists of full set of Aqua products (MYD04_L2) and Terra products (MOD04_L2), which could no longer be found in Collection 005 resulted in choosing Collection 051. According to Lorraine Remer (personal communication, 29 March,

2016), there should be no difference between Optical_Depth_Land_And_Ocean (AOD) retrieved from Collection 051 and Collection 005 scientific data set. The summary document on Collection 005 MODIS aerosol product (04_L2) highlights that the original SDS for Land and Ocean product only contains land products with QA=3 for level 2 aerosol products (Remer et al., 2002).

3.2.4.2. Data pre-processing

In data pre-processing, total 45 gigabytes of data were processed to retrieve AOD and upper air observations (Table 3-5) at a location, where PM2.5 is measured using McIDAS-V. McIDAS-V is a free, open source, visualization and data analysis software package. Since, MODIS gridded swath data are only available for instantaneous observations and surface measured PM2.5 concentrations are available for hourly measurement, thus instead of having to manually select each gridded satellite data that matches the PM2.5 observation time to extract AOD and atmospheric observations variables at air quality monitoring station, McIDAS-V does it automatically. It has the time matching feature to accommodate data with different temporal frequencies which is an added advantage over other image processing software to process MODIS gridded swath data (*McIDAS-V User's Guide*, 2015).

As the upper-air, surface weather observations such as height of planetary boundary layer (BLH), temperature, relative humidity, wind speed, wind direction are recorded in UTC., the hourly $PM_{2.5}$ concentrations which was recorded in CET time zone format is converted to UTC format. The missing values and negative hourly PM2.5 concentrations values were replaced by NA. Since, wind speed and temperature were provided as an integer value by KNMI, it is converted to real measured value with scaling factor of 0.1. In case of BLH, first of all the fill value -999.0 is replaced with NA. The data were provided for every ten minutes so, they were hourly averaged. Since, first significant height retrieved by LiDAR at Cabauw is only considered as a boundary layer height, when there was no retrieval of second significant height by the LiDAR. It is carried out for analysis as to assure that there was only single boundary layer height of the aerosol in the atmosphere which represents the well mixed condition **(Research assumptions).** The BLH in meters was converted to kilometres. And, the temperature was converted to kelvin from degree Celsius for uniformity as some of the variables retrieved from upper air observations were in kelvin.

According to Henk Klein Baltink (personal communication, September 29, 2015), the LiDAR didn't perform well within the study period and a usual false hits were reported below 130 m above ground level and above 1300 m, so, the underestimation of BLH may occur between 130-1200 m above ground level so, it was not advised to use BLH data outside the range. The maximum and minimum retrieved boundary layer height were 140 m and 1300 m respectively, when the BLH of only first significant height was considered which agrees Henk's recommendations.

The surface weather observations that needs scaling as presented in Table 3-4 is converted to the real measured values. The upper-air and satellite retrieved atmospheric observations that needs scaling as presented in Table 3-5 is converted to the physical meaningful values using Equation 3-1 (http://modis-atmos.gsfc.nasa.gov/MOD07_L2/format.html).

$$Real\ value = Scale\ factor \times (Stored\ value - Add\ offset) \quad \text{Equation 3-1}$$

The matched values of hourly PM2.5, AOD, upper air observations variables and surface weather observations are prepared to form a dataset for data pairing as shown in Figure 3-2. The details on matching and data pairing is explained in Methods.

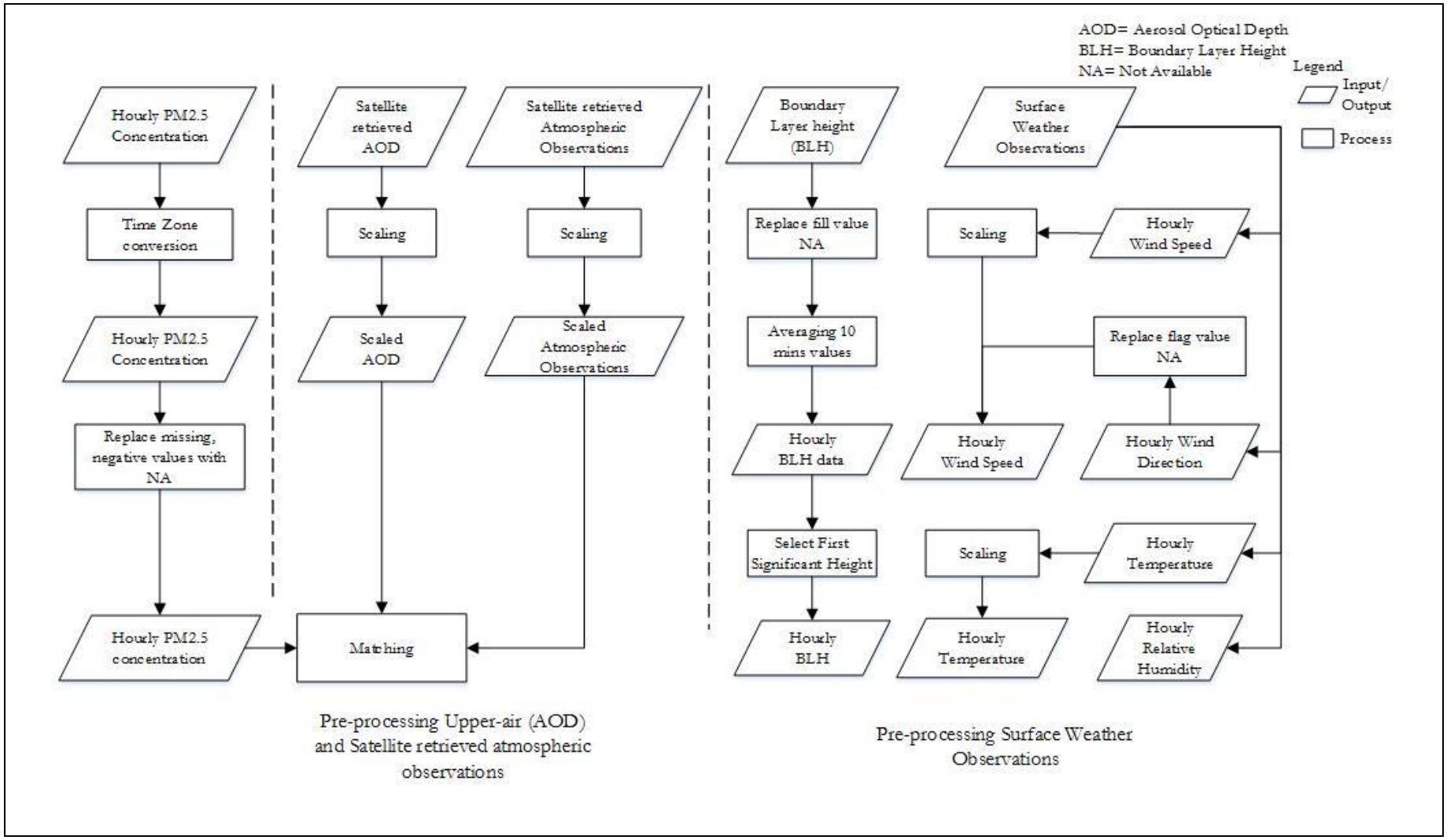


Figure 3-2 Processing flowchart of hourly PM2.5 concentrations, surface weather, upper-air (AOD), satellite retrieved atmospheric observations dataset for pairing

Table 3-4 List of Predictor variables

Predictor Variable	Unit	Scaling factor
Wind Speed	m/s	0.1
Wind Direction	Degrees	
Temperature	kelvin	0.1
Relative Humidity	%	-
Height of Planetary boundary layer	Km	-

Table 3-5 List of Predictor variable extracted from MODIS (Aqua, Terra) collection 051 level 2 products (MXD04, MXD07)

Predictor Variable	Unit	Scaling factor	Offset	Physical range	
0.55 μm Corrected Optical Depth Land	-	0.001	0	-0.05	5
Retrieved Temperature-Level 18	K	0.01	-15000	150	350
Retrieved Temperature-Level 19	K	0.01	-15000	150	350
Retrieved Temperature-Level 20	K	0.01	-15000	150	350
Retrieved Moisture-Level 18	K	0.01	-15000	150	350
Retrieved Moisture-Level 19	K	0.01	-15000	150	350
Retrieved Moisture-Level 20	K	0.01	-15000	150	350
Retrieved Height-Level 18	m	1	-32500	0	65000
Retrieved Height-Level 19	m	1	-32500	0	65000
Retrieved Height-Level 20	m	1	-32500	0	65000
Brightness Temperature- Band 10	K	0.01	-15000	150	350
Brightness Temperature- Band 11	K	0.01	-15000	150	350
Brightness Temperature- Band 12	K	0.01	-15000	150	350
Surface Skin Temperature	K	0.01	-15000	0	20000
Surface Elevation	m	1	0	-400	8840
Lifted Index	K	0.01	0	-20	40
Water Vapor	cm	0.001	0	0	20
Water Vapor-Low	cm	0.001	0	0	20
Water Vapor-High	cm	0.001	0	0	20
Water Vapor-Direct	cm	0.001	0	0	20

3.2.4.3. Data processing

The data availability for ground measurement data and satellite retrieved data after the processing is presented in Table 3-6 and Table 3-7 respectively.

Table 3-6 List of ground Measurement data available after processing for the study period

Variables	Data Available (N=2920)	Data Available (%)	Remarks
PM2.5 concentrations	2658	91.03	Hourly data of 10;30- 13;30 from January 2013 to December 2014
Wind Direction	2878	98.56	
Wind Speed	2920	100.00	
Temperature	2920	100.00	
Relative Humidity	2739	93.80	
Boundary Layer Height	1940	66.44	

Table 3-7 List of satellite-retrieved data available after processing for the study period

Predictor Variables	Data Available (N=2920)	Data Available (%)	Remarks	Data Available (N=2920)	Data Available (%)	Remarks
Optical Depth Land and Ocean	240	8.22	Hourly temporally interpolated data 10;30- 13;30, January 2013 to December 2014 using Weighted average method (with time commensurate data)	227	7.77	Hourly temporally interpolated data 10;30- 13;30, January 2013 to December 2014 using Nearest Neighbourhood method (without time commensurate data)
Brightness Temperature- Band 10	975	33.39		2046	70.07	
Brightness Temperature- Band 11	975	33.39		2046	70.07	
Brightness Temperature- Band 12	975	33.39		2046	70.07	
Surface Skin Temperature	975	33.39		2046	70.07	
Lifted Index	975	33.39		2046	70.07	
Water Vapor	975	33.39		2046	70.07	
Water Vapor-Low	863	29.55		2046	70.07	
Water Vapor-High	430	14.73		2046	70.07	
Water Vapor-Direct	975	33.39		1970	67.47	
Retrieved Temperature-Level 18	975	33.39		2046	70.07	
Retrieved Temperature-Level 19	975	33.39		2046	70.07	
Retrieved Temperature-Level 20	975	33.39		2046	70.07	
Retrieved Moisture-Level 18	975	33.39		2357	80.72	
Retrieved Moisture-Level 19	975	33.39		2557	87.57	
Retrieved Moisture-Level 20	975	33.39		976	33.42	
Retrieved Height-Level 18	975	33.39		2046	70.07	
Retrieved Height-Level 19	975	33.39		2046	70.07	
Retrieved Height-Level 20	975	33.39		2046	70.07	

4. METHODS

Data integration process is carried out to prepare data pairs (Figure 4-2). The predictor variables (Table 3-6 and Table 3-7) is assessed for their correlation, bivariate regression, multicollinearity, stepwise regression to select variables for multiple linear regression modelling. AOD is corrected for hygroscopic growth of particles, $f(RH)$ and is normalized by boundary layer height. The simple method of correction is used when only relative humidity is available. The advanced method of correction is used when there is availability of experimental value of $f(RH)$ and fitting curves for air mass types. The AOD corrected for relative humidity and is normalized by boundary layer height is termed as "meteo-scaled" AOD (Equation 4-3). The multiple linear regression modelling is carried out using meteo-scaled AOD obtained from simple method and advanced method together with final selected variables. However, correlation assessment, bivariate regression analysis, multicollinearity and stepwise regression is carried out again for the selected predictor variable including meteo-scaled AOD. The accuracy of the model is performed using cross-validation method. The output of the regression models is assessed using models coefficient of determination. The overview of the method is shown in Figure 4-1.

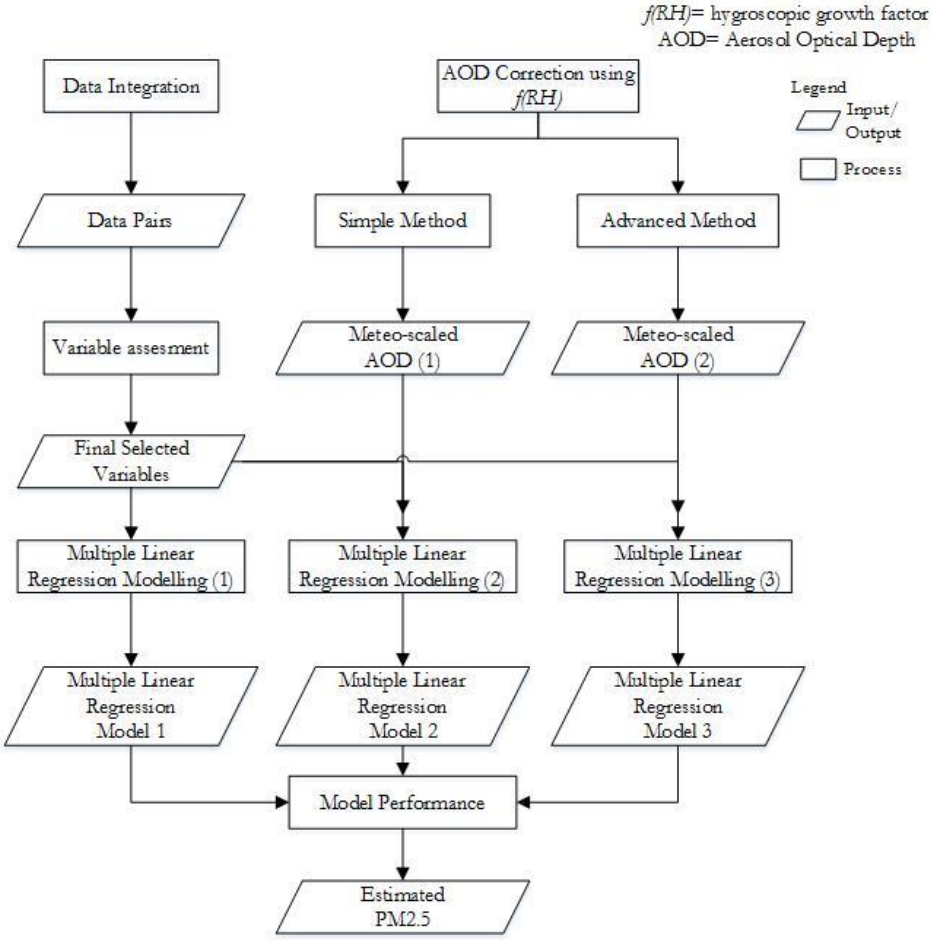


Figure 4-1 Overview of methods used in the current study

4.1. Data Integration

Ichoku et al., (2002) highlights that PM2.5 measurements is a point based measurement and usually it represents PM2.5 measured for an hour or averaged over daily time scales, whereas satellite measures instantaneous AOD at the station depending upon revisit time of satellite (for MODIS Terra, 10:30 and MODIS Aqua, 13:30).The first step in data integration is to temporally match satellite retrieved AOD, satellite retrieved atmospheric observations, upper-air and surface level PM2.5 concentration with the MODIS overpass time at the air quality monitoring station (Wang & Christopher, 2003).

MODIS is an instrument that operates on the Terra and Aqua Satellite. With a swath width of 2330 Km, and with two of the same sensors in orbit, MODIS observes the entire Earth every other day. MODIS instrument, with its Terra and Aqua satellite overpass Netherlands at approximately 10:30 and 13:30 p.m. local time. The MODIS satellites have a ± 55-degree scanning pattern and orbit at 705 km.

During resampling, weighted average interpolation method is available as an option to interpolate in McIDAS-V (Space Science and Engineering Center, 2000), the software used for temporal interpolation of instantaneous observations of AOD retrieved by MODIS Terra and Aqua Satellite.

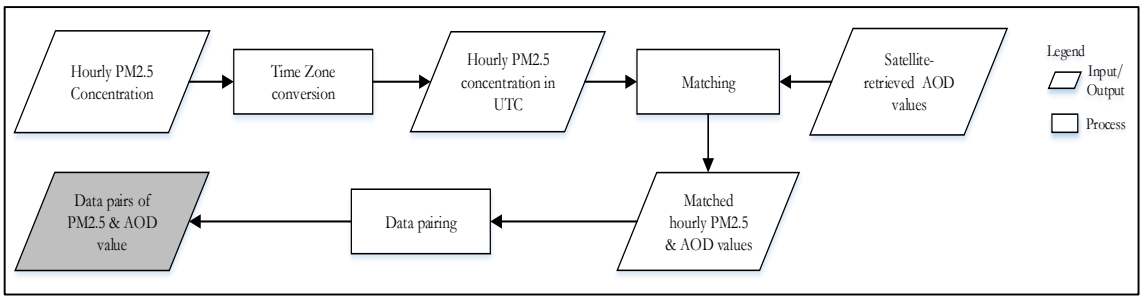


Figure 4-2 Process of Extracting Data Pairs: Point-based Time-series extracting from satellite-retrieved AOD (from TERRA/AQUA MODIS, MOD04 and MYD04 respectively) corresponding to hourly PM2.5-observations from surface air-quality monitoring stations

The point-based time-series (from TERRA/AQUA MODIS, MOD04 and MYD04 respectively) corresponding to hourly PM2.5-observations from surface air-quality monitoring stations (Figure 4-2) is extracted. Each MODIS level 2 product file covers a five-minute time granule. Thus, the same location can have more than one observation in a day. As NASA, (2016) highlights that the observation area closer to the poles will have increased number of overpasses thus, daily MODIS observations increases due to overlapping orbits.

Studies (Engel-Cox et al., 2006; Koelemeijer et al., 2006; Gupta & Christopher, 2009) used PM2.5 measurment closest in time to instantaneous MODIS Terra and Aqua AOD measured value during the satellite overpass time to create data pairs to established relationship between AOD and PM2.5. In the current study, the estimated AOD that is closer to the time when PM2.5 is measured at the station is temporally matched during resampling, when satellite overpass the air quality monitoring station. Hence, the data pair of PM2.5 and AOD is obtained for 10:30, 11:30, 12:30 and 13:30. But, PM2.5 is an hourly measurement at air quality monitoring station and AOD is an integral of instantaneous observation from surface to the top of atmosphere in a vertical column during the satellite overpass time. Thus, these data pairs that are created when estimated AOD that is closer to the time when PM2.5 is measured at the station is temporally matched, are referred as **"without time commensurate"** in the current study.

The satellite retrieved AOD values is temporally matched and extracted after resampling with the ground level $PM_{2.5}$ measurement time at 10:30, 11:30, 12:30 and 13:30 using "weighted average temporal

interpolation", when the satellite overpasses the air quality monitoring station. The data pairs are prepared with a matched satellite retrieved AOD from Terra and Aqua satellites level 2 MODIS aerosol product (MOD04 and MYD04 respectively) that are made "**time commensurate"** with ground level measured $PM_{2.5}$ concentrations at air quality monitoring station.

Similarly, the process is repeated to obtain point-based time series data pairs of satellite retrieved atmospheric observations and PM2.5 concentrations. Then, upper-air, surface weather observations that is measured at 10:30, 11:30, 12:30 and 13:30 are used to create data pairs with PM2.5 observations, AOD and satellite retrieved atmospheric observations at 10:30, 11:30, 12:30 and 13:30.

Thus, it is assessed how the variability in estimated PM2.5 differs from data pairs "with time commensurate" and "without time commensurate" AOD.

4.2. Assessment of predictors variables for Multiple Linear Regression Modelling

A review of the literature (Chu et al., 2003; Gupta et al., 2006; Wang & Christopher, 2003) on related work on data and method suggests that MODIS column AOD data is indicative of surface level PM2.5 concentrations. Studies (Gupta et al., 2006; Schaap et al., 2009) shows that the surface level PM2.5 can be estimated using satellite retrieved AOD using simple regression modelling approach. However, the review on previous studies (Gupta et al., 2006; Pelletier et al., 2007) shows that the performance of the simple regression model would improve if the surface weather predictor variables are incorporated using multiple linear regression modelling approach.

The linear relationship between surface measured PM2.5 and satellite retrieved AOD is assessed to estimate PM2.5. The surface weather (Table 3-4), upper air and satellite retrieved atmospheric observations (Table 3-5) are also assessed as potential predictor variables. The surface observations consist of PM2.5 measurements at the surface air quality monitoring stations. The surface weather observations consist of wind speed, wind direction, temperature, relative humidity measured at the surface weather meteorological station. The height of boundary layer measured at Ceilometer station is also considered as a upper-air observation. All upper air and satellite retrieved atmospheric observations, surface weather observations are considered as predictor variables. All predictor variables were temporally matched with the PM2.5 (dependent variable) observation time. It was assessed to find the selected the candidate predictor variable to carry out multiple linear regression modelling. So, the assessment allows to identify if the AOD, surface weather, upper air and atmospheric observations would improve the explained variability in PM2.5.

The following analysis were carried out to select final candidate predictor variables:

4.2.1. Correlation assessment

Pearson correlation of PM2.5 and all predictors variables is first assessed. It is carried out to understand the strength of linear relationship and direction of effect of PM2.5 and each predictor variables. It is mainly assessed if there is a statistically significant relationship (p-value <0.1) between PM2.5 and each predictor variables. If there was no statistically significant relationship between PM2.5 and predictor variables, they were not chosen.

4.2.2. Bivariate Regression Analysis

Also, each variable was regressed against the $PM_{2.5}$ concentration using bivariate regression. The variables which were statistically significant (p-value<0.1) were only chosen and their corresponding adjusted R-squared is recorded. The predictor variable which gave highest adjusted R-squared was selected as a first variable to start a modelling with.

4.2.3. Multicollinearity Check

The collinearity between the variables can increases the estimates of the parameters variance (O'Brien, 2007). The presence of severity of multicollinearity between the candidate predictor variables can be quantified using the variance inflation factor (VIF) (Quinn & Keough, 2002).

VIF is a measure of strength of relationship between each predictor variable and all others predictor variables. VIF performs separate regression analysis for each independent variable on the remaining independent variables. The adjusted R-squared yields of regression analysis which is used to calculate VIF (Equation 4-1). As adjusted R-squared increases, VIF increases. Thus, large values of VIF suggests that one predictor variable is highly correlated to other predictor variables. Hence, multicollinearity might be a problem.

$$VIF = \frac{1}{1 - R^2_{adj.}} \qquad \text{Equation 4-1}$$

Where, $R^2_{adj.}$ is the adjusted R-squared.

As a rule of thumb, VIF greater than 10 indicates that moderate collinearity might be a problem whereas values greater than 30 are indicative of severe multicollinearity (Dormann et al., 2013). Thus, only candidate variables with VIF<10 were included in the final model.

4.2.4. Stepwise Regression

The supervised forward selection stepwise regression procedure of Vienneau et al., (2010) was followed after correlation assessment, bivariate regression analysis and multicollinearity checking of predictor variables on the remaining predictor variables.

Stepwise forward regression starts with only dependent variable in the model. The predictor variable with highest adjust R-squared in bivariate regression is used as a first variable. The predictor variables were added individually in the model with first variable and the resulting adjusted R-squared is recorded. The selection of the candidate predictor variables added individually in turn was retain in the model on the basis of following criteria's:

1. The additional contribution to the adjusted R-squared together with first variable was at least increased by 1 %.
2. The p-value for all variables already in the model remained below 0.1.

The process of addition of predictor variables continued until there was no increase in total adjusted R-squared in a forward stepwise regression analysis. Hence, the followed procedure allowed logical selection of predictors variables that maximizes model explained variance.

4.2.5. Multiple Linear Regression Modelling

Prior to multiple linear regression modelling, the final selected candidate variables nature of the data is explored by using scatterplot. The 90 % confidence interval (CI) lines is drawn across the slope of linear regression line. It is carried out to understand linear relationship between PM2.5 and predictor variables.

Then after, the selected candidate predictor variables with dependent variable, PM2.5 is used to carry out multiple linear regression modelling.

The general form of multiple linear regression model is shown in Equation 4-2,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} + \epsilon$$ Equation 4-2

Where, p represents the total number of variables in the model, Y represents the dependent variables and X_1 to X_{p-1} represents the predictor variables, β_0 to β_{p-1} represents the regression coefficient of predictor variables and ϵ is the residuals.

4.2.6. Relative Importance of the variable

As the predictor variables are correlated with each other, the proportionate contribution of each predictor variable is calculated using R package **"Relaimpo"** (Grömping, 2010). It decomposes R-squared considering both direct effect and its effect when combined with others predictor variables in the regression equation (Grömping, 2006), which produces similar result compared to R package **"hier.part"** for "hierarchical partitioning" of R-squared, which has advantage over computation time. Grömping, (2007) considered some criteria as follows for decomposition of R-squared;

- The sum of the all shares of predictor variables decomposed variance is the model variance.
- The variance shared should be non-negative.
- Predictor with non-zero regression coefficient should receive a non-zero value.
- Predictor with zero regression coefficient should receive a zero share.

4.3. Correction of AOD for Relative Humidity using $f(RH)$

4.3.1. Simple correction method/reference method

Koelemeijer et al., (2006) has defined the corrected AOD for $f(RH)$ and BLH as "**meteo-scaled**" AOD (Equation 4-3). Meteo-scaled AOD value is replaced with AOD in regression model obtain in a form of Equation 4-2 to estimate PM2.5 concentrations in at the surface.

$$AOD^* = \frac{AOD}{BLH \times f(RH)}$$ Equation 4-3

Where, AOD is aerosol optical depth, BLH is the boundary layer height, $f(RH)$ is hygroscopic growth factor and AOD^* is the meteo-scaled AOD.

Koelemeijer et al. (2006) followed the method of Veefkind et al. (1996) to find a fitting curve of $f(RH)$. However, some studies (Li et al., 2005; Tsai et al., 2011) used simple method (Equation 4-4) as in to account for the effect of RH for hygroscopic growth of aerosol in correlating PM (PM10 and PM2.5) with AOD respectively. This method is carried out when only RH was available and without consideration of air mass types. Thus, it is followed as a "**simple method"** in current study.

$$f(RH) = \frac{1}{(1 - \frac{RH}{100})}$$ Equation 4-4

where, RH is relative humidity and $f(RH)$ is the hygroscopic growth factor.

4.3.2. Advanced Method

The advanced method of correction of AOD for $f(RH)$ is carried out when there was availability of $f(RH)$ and different air mass type for which $f(RH)$ was measured. During study, two sources of information is used to find $f(RH)$ fitting curve at Cabauw. The **"first"** is from Veefkind et al., (1996), which provided the experimental measurement value using Equation 4-5 for **Petten**, the north west of The Netherlands.. The

diurnal time series plot of hygroscopic growth factor curve, $f(RH)$ and relative humidity (RH) (which is called "**humidogram**") was found for Petten, which Koelemeijer et al. (2006) followed in his study.

$$f(RH) = \frac{B_{sp}(RH)}{B_{sp,dry}} \quad \text{Equation 4-5}$$

where, B_{sp} is the light scattering per unit aerosol mass (m^{-1}) and $B_{sp,dry}$ is the particle scattering coefficient (m^{-1}) at a low relative humidity (RH<40%).

The **"other one"** iss from Zieger et al. (2013), which provided (Table 4-1) median, 25th percentile and 75th percentile value of $f(RH = 85\%, \lambda = 550\ \text{nm})$ for different air mass type such as maritime; maritime, slightly polluted; maritime, slightly polluted, continental south; and continental east at **Cabauw**. Both available information was assessed to obtain the fitting curve of $f(RH)$ that can represent the dominant air mass type found in study area to correct AOD accordingly.

The detailed method for assessment on obtaining fitting curve with available information on $f(RH)$ and air mass type at Petten and Cabauw are described below;

4.3.2.1. Method to obtain Fitting curve using Petten data

At first, the fitting curve was obtained (Equation 4-6) from re-plotting of the data (relative humidity and corresponding $f(RH)$ extracted from humidogram for November 17, 1993 found at Veefkind et al. (1996). The fitting curve obtained after re-plotting of extracted Petten data is considered as a "**reference**", which is shown in Figure 4-3. The study was carried out when there were clear skies, strong inversions and persistent easterly winds associated with continental air masses in November 17, 1993.

$$f(RH) = 0.0004 \times RH^2 - 0.0309 \times RH + 1.93 \quad \text{Equation 4-6}$$

Zieger et al., (2011) showed the characteristics of fine and coarse (sea salt) aerosol mass for $f(RH)$ at Cabauw during the experimental study from June to October 2009 at RH=85 %. In the continuation study of Zieger et al. (2011), Zieger et al. (2013) presented $f(RH = 85\%, \lambda = 550\ \text{nm})$ instantaneous values for different air mass type (Table 4-1). It is also mentioned in the same study that there was no clear wavelength dependency found in the measurement period between the range of 450-700 nm. Thus, the assumptions are made to modify $f(RH)$ for Cabauw (see Research assumptions). The offset values for different air mass type is obtained by changing the offset of Equation 4-6. The offset is found out when Equation 4-6 offset value is changed in such a way that it would yield the exact reported value (Table 4-1) of median at $f(RH = 85\%, \lambda = 550\ \text{nm})$. The process is repeated to find the offset for 25th percentile and 75th percentile respectively and it is reported in Table 4-2.

As per the assumption (see Research assumptions) maritime and maritime slightly polluted air mass type is grouped into **"Maritime"** as shown in Figure 4-3. Similarly, continental south, continental east and maritime, heavily polluted air mass type is grouped as **"Others"** (Table 4-1) which is also shown in Figure 4-3. The corresponding $f(RH)$ values of the different air mass type were averaged to obtained $f(RH)$ value for grouped maritime and others air mass type respectively. Similarly, the offset value for different air mass type were averaged to obtained offset value for the grouped air mass type (Table 4-2).

Table 4-1 Percentile value of $f(RH = 85\%, \lambda = 550\ \text{nm})$ for air mass types found at Cabauw

Air mass type	Median	25th percentile	75th percentile
Maritime	3.38	3.16	3.6

Air mass type	Median	25th percentile	75th percentile
Continental South	1.86	1.76	2.02
Maritime, heavy polluted	2	1.82	2.04
Maritime, slightly polluted	2.96	2.81	3.11
Continental East	2.26	2.13	2.37
Maritime, Grouped	3.17	2.99	3.36
Others, Grouped	2.04	1.90	2.14

Table 4-2 Offset values for air mass typed found in Cabauw for $f(RH = 85\%, \lambda = 550\ \text{nm})$ to modify $f(RH = 72\%, \lambda = 475\ \text{nm})$ found in Petten

Air mass type	Median	25th Percentile	75th Percentile
Maritime	3.12	2.90	3.34
Continental South	1.60	1.50	1.76
Maritime, heavy polluted	1.74	1.56	1.78
Maritime, slightly polluted	2.70	2.55	2.85
Continental East	2.00	1.87	2.11
Maritime (grouped)	2.91	2.72	3.09
Others (grouped air mass type 2, 3 & 5)	1.78	1.64	1.88

The fitting curve equations of median values of $f(RH)$ for grouped maritime and others air mass is shown in Equation 4-7 and Equation 4-8 respectively. The polynomial curve for grouped maritime and other air mass type together with all other air mass type is shown in Figure 4-4.

Maritime (grouped) air mass,

$$f(RH) = 0.0004 \times RH^2 - 0.0309 \times RH + 2.91 \quad \text{Equation 4-7}$$

Others (grouped) air mass,

$$f(RH) = 0.0004 \times RH^2 - 0.0309 \times RH + 1.78 \quad \text{Equation 4-8}$$

4.3.2.2. Method to obtain Fitting curve using Cabauw data

The average humidograms was found for different air mass type in Zieger et al. (2013) as "**Figure 8**" for maritime (Equation 4-9), continental south (Equation 4-10), continental east (Equation 4-11), maritime heavily polluted (Equation 4-12) and maritime slightly polluted (Equation 4-13). First, the average fitting curve was obtained from re-plotting of the extracted data (relative humidity and $f(RH)$) for average humidograms (Figure 4-5) at Cabauw.

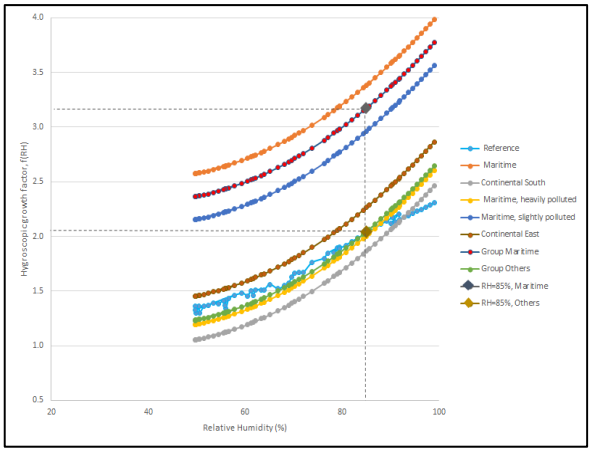


Figure 4-3 Fitting curve of humidogram for $f(RH = 72\%, \lambda = 475\ \text{nm})$ found at Petten (Veefkind et al., (1996), referred as reference) with modified fitting curve of humidograms for different air mass type found at Cabauw after the reference. For Petten, diurnal time series of $f(RH)$ was available. However, at Cabauw only the instantaneous $f(RH = 85\%, \lambda = 550\ \text{nm})$ was available.

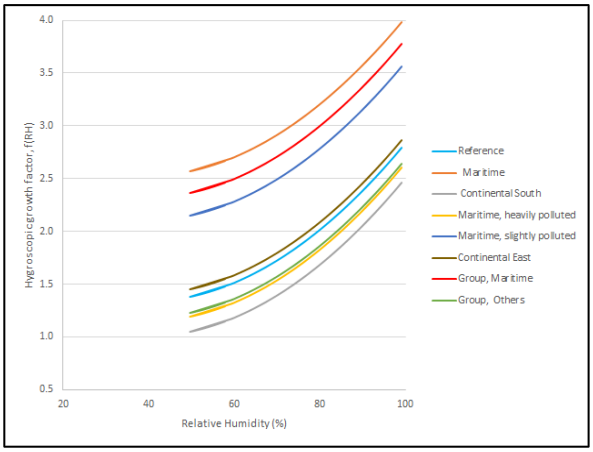


Figure 4-4 Polynomial curve of humidogram for $f(RH = 72\%, \lambda = 475\ \text{nm})$ found for modified fitting curve for different air mass type found at Cabauw after the reference.

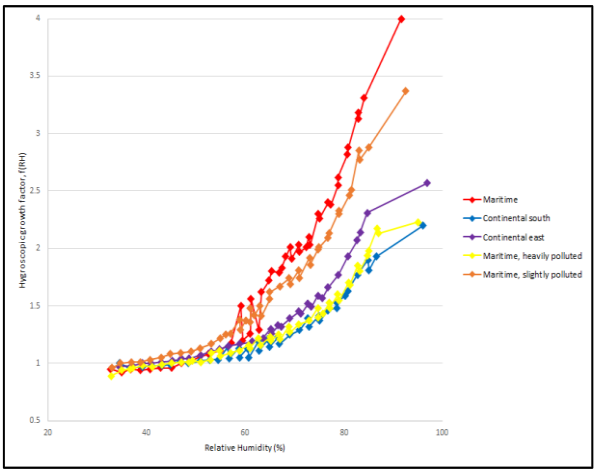


Figure 4-5 Average fitting curve of humidogram for $f(RH = 85\%, \lambda = 550\ \text{nm})$ for different air mass type found at Cabauw ((Zieger et al., 2013)

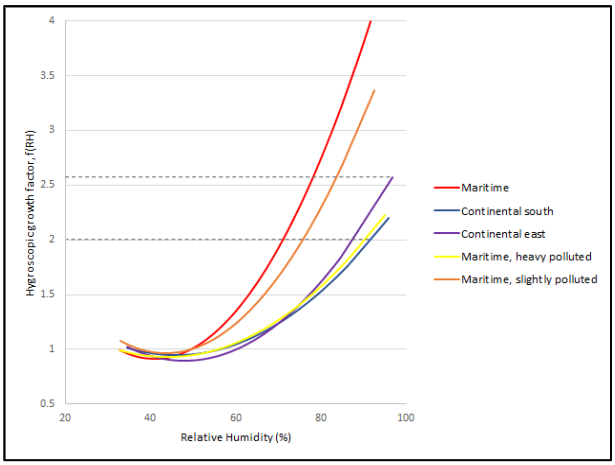


Figure 4-6 Polynomial curve of average humidogram for $f(RH = 85\%, \lambda = 550\ \text{nm})$ for different air mass type found at Cabauw

The polynomial curve of average humidogram plotted for all air mass type found at Cabauw is shown in Figure 4-6.

$$f(RH) = 0.0012 \times RH^2 - 0.0983 \times RH + 2.9303 \quad \text{Equation 4-9}$$

$$f(RH) = 0.0005 \times RH^2 - 0.046 \times RH + 2.0088 \quad \text{Equation 4-10}$$

$$f(RH) = 0.0007 \times RH^2 - 0.0671 \times RH + 2.5013 \quad \text{Equation 4-11}$$

$$f(RH) = 0.0005 \times RH^2 - 0.044 \times RH + 1.8951 \quad \text{Equation 4-12}$$

$$f(RH) = 0.001 \times RH^2 - 0.0874 \times RH + 2.8537 \quad \text{Equation 4-13}$$

Thus, as per the assumption (see Research assumptions) maritime and maritime slightly polluted air mass type is grouped into **"Maritime"**. Similarly, continental south, continental east and maritime, heavily

polluted air mass type is grouped as **"Others"**. The fitting curve for grouped air mass "Maritime" and "Other" were obtained by averaging the coefficient of polynomial for corresponding air mass type in grouping as Equation 4-14 and Equation 4-15 respectively.

$$f(RH) = 0.0011 \times RH^2 - 0.09265 \times RH + 2.892 \quad \text{Equation 4-14}$$

$$f(RH) = 0.0006 \times RH^2 - 0.0524 \times RH + 2.1351 \quad \text{Equation 4-15}$$

4.4. Selection of fitting curve of f(RH) between "Maritime" and "Others" air mass type

4.4.1. Conditional statements

To select the fitting curve of $f(RH)$ between "Maritime" and "Others" air mass type at Cabauw, some conditional statements were used. The conditional statements were formulated based on the 72 h air mass back trajectories measured at Cabauw presented in "**Figure 3"** of Zieger et al. (2013) is used. Thus, the wind direction (WDD) is used to formulate a conditional statement (see Research assumptions). The wind direction coming from the North Atlantic Sea and the North Sea are set as criteria for wind direction. Also, from the same figure, it is inferred that maritime air masses have $f(RH)$ greater than 2 and "others" air masses have $f(RH)$ less than 2. The lower dotted line is used to show $f(RH)$ value of 2 (see Figure 4-6). The maximum value of $f(RH)$ for other air mass was found to be 2.57 (for continental east) from same study which is plotted as an upper dotted line (see Figure 4-6). Thus, values of $f(RH)$ is also used as a conditional statement.

Specifically, the following conditional statements were as used;

a) IF(AND(OR(WDD<20, WDD>225), then maritime else others
b) IF(AND(WDD<357.5, WDD>320), then maritime else others.
c) IF $f(RH) >$2, then maritime else others.
d) IF $f(RH) > 2.57$, then maritime else others.

These conditional statement were used in the data pairs with time commensurate AOD (Table 5-1) and $f(RH)$ obtained using simple method (Equation 4-4). The Boolean 1 is used if the condition was **"TRUE"** for maritime and Boolean 0 is used it the conditions was **"FALSE"**. The f(RH) which obtains higher percentage of conditions is chosen, on the basis of the final score obtained in four conditions to correct AOD.

4.4.2. Sensitivity Analysis

At first, sensitivity analysis is carried out to assess the sensitivity and uncertainty in estimated PM2.5 with change in $f(RH)$ for grouped "maritime" and "others" air mass type for Petten. The modified value of median, 25th percentile and 75th percentile values for $f(RH = 72\%,\ \lambda = 475\ \text{nm})$ at Petten after Cabauw (Table 4-3) is used. The values of $f(RH)$ for different percentiles for grouped air mass types is considered keeping BLH and AOD variables same in Equation 4-17 for sensitivity and uncertainty analysis.

Table 4-3 Percentile value of $f(RH = 72\%,\ \lambda = 475\ \text{nm})$ for air mass types found at Petten after Cabauw

Air mass type	Median	25th percentile	75th percentile
Maritime, Grouped	3.17	2.99	3.36
Others, Grouped	2.04	1.9	2.14

In a similar way, the sensitivity and uncertainty in estimated PM2.5 with change in $f(RH)$ for grouped "maritime" and "others" air mass type for Cabauw is carried out. The median, 25th percentile and 75th

percentile values for $f(RH = 85\%,\ \lambda = 550\ \text{nm})$ as presented in Table 4-1 is used. The values of $f(RH)$ for different percentiles for grouped air mass types is considered keeping BLH and AOD values same in Equation 4-17 for sensitivity and uncertainty analysis. The procedure used for sensitivity analysis is described below;

Schaap et al., (2009) found the relationship of MODIS AOD and $PM_{2.5}$ for Cabauw, (Equation 4-16Equation 4-16), which was used in sensitivity study to quantify the effects on estimated PM2.5 with correction of AOD for hygroscopic growth factor, $f(RH)$ with different air mass type.

$$PM_{2.5} = 120.01 \times AOD + 5.1 \qquad \text{Equation 4-16}$$

"Meteo-scaled" AOD value as defined by Koelemeijer et al. (2006) (Equation 4-3) is inserted into Equation 4-16 to obtained Equation 4-17 to estimate $PM_{2.5}$ concentrations at the surface.

$$PM_{2.5} = 120.01 \times \frac{AOD}{BLH \times f(RH)} + 5.1 \qquad \text{Equation 4-17}$$

The $f(RH)$ of either grouped "maritime" or "others" air mass type is chosen on the basis of sensitivity of change in $f(RH)$ values of air mass to estimated PM2.5. The air mass type showing high sensitivity to estimated PM2.5 with change in $f(RH)$ values between 25th percentile and 75th percentile could support in finding dominant air mass type at Cabauw.

4.5. Multiple Linear Regression modelling with Meteo-scaled AOD

The AOD predictor variable is replaced with Meteo-scaled AOD obtained with simple method in Table 5-5. The bivariate regression is carried out for meteo-scaled AOD. If the predictor variable was statistically significant (p-value<0.1) then, it was chosen. The multicollinearity check for meto-scaled AOD is also carried out. If VIF<10, meteo-scaled AOD was selected for stepwise forward regression. The similar criteria for selection of variables in Stepwise Regression (4.2.4) is followed. The final selected candidate variables were used to obtain regression models using its general form as in Equation 4-2. The process is repeated for Meteo-scaled AOD obtained with advanced method. The regression modelling with meteo-scaled AOD is carried out for data pairs of with time commensurate and with time commensurate.

4.6. Model Performance

Cross validation method followed by similar study (Guo et al., 2014) is adopted to test the performance of the derived regression models for with time commensurate and without time commensurate data pairs in the current study. Initially, the data pair of the selected predictor variables is randomly split into 90% as model training and 10% as model testing samples along with the dependent PM2.5 variable. Then, a multiple linear regression model is established based on the training samples, which is used to estimate PM2.5 concentrations in the model testing samples. Thus, a linear regression equation between observed and estimated PM2.5 concentrations was fitted to obtain the slope and intercept values. Correlation coefficient, root mean square error (RMSE) of estimated and observed PM2.5 was calculated. The slope and intercept of the linear regression equation of estimated and observed PM2.5 is also calculated. The process is repeated for 1000 times with each time removing 10% of samples for model testing and fitting the model for other 90% training samples.

4.7. Used Software and package

The software and packages that are specifically used in the pre- processing, processing, analysing and visualization is presented in Table 4-4.

Table 4-4 Description of used software during the study

Software	Package/Extension	Description
McIDAS	McIDAS	To extract AOD value at a location from gridded swath data of MODIS Level 2 aerosol and atmospheric product For time matching of instantaneous satellite measurement with the PM2.5 measurement time at Cabauw air quality monitoring station For temporal interpolation of AOD using weighted average method during resampling
X-win 32		For parallel processing using McIDAS-V through University of Twente Linux server
R	ggplot2	For data visualization
		For correlation matrix calculation, bivariate regression, variance inflation factor calculation, stepwise regression, cross validation
	Relaimpo	To find the relative importance of a predictor variable in a multiple linear regression model
ArcGIS 10.2.2	Proximity toolset	To find the nearest distance from Cabauw air quality monitoring station to the surface weather stations and Ceilometer stations
MS Excel		To handle text files

5. RESULT AND DISCUSSION

5.1. Data Pairs

With the available data set, the final matched data set includes data pairs of AOD, PM2.5, surface weather, upper-air (BLH) and satellite-retrieved atmospheric observations (Table 5-1) for without time commensurate and with time commensurate data pairs, when the observation is available for each variable.

Table 5-1 Number of Data pairs available for dependent and predictor variables

Data pairs type	No. of data pairs
With time commensurate	34
Without time commensurate	86

If the satellite overpasses the station at supposedly 10:15 and 13:00 in a day, the measurement at air quality monitoring station at 10:30 is paired with AOD observed at 10:15 during temporal matching. Similarly, the measurement at air quality monitoring station at 13:30 is paired with AOD observed at 13:00. And, supposedly if there is only one observation in a day at 10:15, each PM2.5 measurements at air quality monitoring station at 10:30 and 13:30 are paired with that one AOD observation providing two data pairs during temporal matching. This process of pairing which has **"without time commensurate"** AOD, which may ignore the change in atmosphere (e.g. wind speed, wind direction, relative humidity, temperature, boundary layer height) within air quality measurement durations between 10:30 and 13:30.

If the satellite overpasses the station at supposedly 09:30 and 10:15 in a day, the AOD value at air quality monitoring station at 10:30 is obtained with weighted average temporal interpolation of estimated AOD values observed at 09:15 and 10:15 by the satellite. And supposedly, if there is only one satellite observation in a day at 10:15, only PM2.5 measured at air quality monitoring station at 10:30 is paired with AOD estimated at 10:15 without interpolation. Thus, there is only one data pair in a day. This process of pairing has **"with time commensurate"** AOD, which may consider the change in atmosphere (e.g. wind speed, wind direction, relative humidity, temperature, boundary layer height) within air quality measurements between 10:30 and 13:30.

As per the Research assumptions, it is considered that boundary layer is representative as well mixed layer when single boundary layer during the satellite overpass time is detected by LiDAR. The potential temperature, humidity, wind speed and wind direction are nearly constant with height over the bulk of the mixed layer (Stull, 1988). The potential temperature is the temperature that a mass of air would have if it is brought adiabatically to average pressure at mean sea level (1000 hPa). However, the convective time scale is about 10-20 minutes for air to circulate between the surface to top of the mixed layer thus, the inversion might occur. In such case, the instantaneous AOD may change during measurement durations at 10:30, 11:30, 12:30 and 12:30 within the vertical column of atmosphere. So, if AOD was not made time commensurate, it might not be representative.

The data pairs created consists of satellite retrieved temporally interpolated AOD which is matched with surface level PM2.5 concentrations, surface weather, upper-air and satellite-retrieved atmospheric observations. PM2.5 concentrations are measured at every 0:09 to 0:51 minute at Cabauw air quality

monitoring station for most of the study period. So, there is already 18 minutes of discrepancy in PM2.5 measurements in an hour. However, they were reported as hourly average values for 2013-2014. According to Guus Stefess (personal communication, May 18, 2016), the time discrepancies in automatic sampling instrument was adjusted for measurement since 2014 at Cabauw. The quick look in the final data pairs for both with time commensurate AOD (34) and data pairs without time commensurate AOD (86) shows that all the data pairs for measured PM2.5 observed consisted of measurement in 2013. As AOD is an integral of instantaneous measurement in a vertical column of atmosphere when MODIS Terra, Aqua overpasses at 10:30 and 13:30 local time respectively at any area on the surface of the earth, it is assumed (see **Research assumptions**) that PM2.5 concentrations at 10:30, 11:30, 12:30 and 13:30 is representative for 10:00, 11:00, 12:00, 13:00 at surface respectively. And similarly it is assumed that surface weather observation at 10:30, 11:30, 12:30 and 13:30 is representative for 10:00, 11:00, 12:00, 13:00 respectively. This is carried out as instantaneous AOD is observed.

5.2. Assessment of predictors variables for Multiple Linear Regression Modelling

5.2.1. Correlation Assessment

The Pearson correlation of PM2.5 and predictor variables for without time commensurate and with time commensurate data pairs is summarized in Table 5-2. The selected predictor variables whose correlation is statistically significant (p-value<0.1) were only presented. The table also presents the direction of effect (positive or negative) of correlation of predictor variables and dependent variable in correlation column.

Table 5-2 Summarized statistical characteristics of predictor variables with p-value < 0.1 for Pearson correlation

Dependent Variable	Predictor Variable	Symbol used	R*	Correlation	p-value	Data pairs type
PM2.5	Aerosol Optical Depth	AOD	0.34	Positive	0.05	With time commensurate
	Temperature	TEMP	-0.26	Negative	0.13	
	Wind Speed	WS	0.24	Positive	0.02	Without time commensurate
	Temperature	TEMP	-0.40	Negative	0.001	
	Relative Humidity	RH	0.33	Positive	0.00	
	Boundary Layer Height	BLH	-0.22	Negative	0.04	
	Aerosol Optical Depth	AOD	0.11	Positive	0.30	
	Surface Skin Temperature	STEMP	0.30	Positive	0.01	
	Lifted Index	LI	0.27	Positive	0.01	
	Water Vapor	WVAP	-0.23	Negative	0.03	
	Water Vapor Low	WVAPL	-0.25	Negative	0.02	
	Water Vapor Direct	WVAPD	-0.22	Negative	0.04	

Note: * R is the correlation coefficient between PM2.5 and corresponding predictor variable.

The correlation (Table 5-2) of AOD without time commensurate is much lower (0.11) than that has been found for hourly PM2.5 (0.39) and daily PM2.5 (0.27) in similar study in Koelemeijer et al. (2006) in Europe. However, in Koelemeijer et al., (2006) data pairs were created only within the satellite overpass time for hourly data pairs. The correlation is close to the current study results for data pairs with time commensurate AOD (0.34). Although, there is change in AOD retrieval algorithm between MODIS collection 4 and collection (Levy et al., 2009), it should be also noted that, in current study, MODIS Terra and Aqua collection 051 is used to extract AOD at air quality monitoring location whereas in Koelemeijer et al. (2006)

MODIS Terra, Aqua collection 4 was used. And, in current study, the hourly data pairs were created for 10:30, 11:30, 12:30 and 13:30 (i.e. four data pairs in between 3 hr) which is different to two data pairs for Terra and Aqua collection 4 in two hours in Koelemeijer et al., (2006) study.

Recent study (Guo et al., 2014) has used time window of 5 hr to include more data point of PM2.5 and AOD with respect to MODIS Terra and Aqua collection 5 overpass time Usually, the hourly PM2.5 concentrations is averaged centred around the satellite overpass time of Terra and Aqua to reduce the temporal noise of PM2.5 to correlate AOD and PM2.5. However, in current study, the correlation is assessed with data pairs of instantaneous AOD that is temporally interpolated with and without time commensurate AOD with the corresponding PM2.5 measurement time at the surface. Gupta & Christopher, (2008) found in North Brimingham, Alabama that the correlation of $PM_{2.5}$ and AOD retrieved from MODIS Terra collection 5 is increased from 0.52 to 0.62 when hourly averaged $PM_{2.5}$ is considered than daily averaged PM2.5 centred around satellite overpass time. It may be that the instantaneous MODIS AOD correlate better with hourly averaged PM2.5 compared to 24 hr averaged PM2.5 due to diurnal variations in PM2.5 mass measurements. The study shows that the relationship between AOD and PM2.5 varies with location and time (Engel-Cox et al., 2004; Gupta et al., 2006; Koelemeijer et al., 2006).

5.2.2. Bivariate Regression Analysis

The selected variables which were statistically significant (p-value<0.1) in bivariate regression analysis were shown in Table 5-3. The adjusted R-squared of those variables were also presented. No new predictor variable was added to the predictor variables (Table 5-2) which were found to be significant in correlation assessment.

It is found that AOD is not statistically significant (p-value>0.1) in bivariate regression with PM2.5.for without time commensurate data pairs. However, it might be statistically significant and add some variance in regression equation may be due to the relationship with other predictor variables, so it is included for stepwise regression analysis. However, due to limited knowledge about the response of meteorological predictor variables to PM2.5 at the surface, the interactions has not been addressed in this thesis. Similarly, temperature is not statistically significant (p-value>0.1). It is still considered as a predictor variable assuming that there might be relationship with other predictor variables. Thus, it is assessed if it could add some variance together with other variables in stepwise regression analysis.

Table 5-3 Summarized statistical characteristics of predictor variables with p-value < 0.1 in bivariate regression models

Variables	Adj R-squared	Intercept	Regression coefficient	p-value	Data pairs type
Aerosol Optical Depth	0.09	12.48	19.76	0.05	With time commensurate
Temperature*	0.04	121.13	-0.36	0.13	
Wind Speed	0.05	8.54	1.17	0.02	Without time commensurate
Temperature	0.15	161.46	-0.50	0.001	
Relative Humidity	0.10	-1.05	0.28	0.00	
Boundary Layer Height	0.04	20.51	-8.17	0.04	
Aerosol Optical Depth*	0.001	12.98	5.83	0.30	
Surface Skin Temperature	0.08	-233.76	1.64	0.01	
Lifted Index	0.06	12.27	69.88	0.01	
Water Vapor	0.04	18.93	-2405.82	0.03	
Water Vapor-Low	0.05	19.68	-9670.85	0.02	

Variables	Adj R-squared	Intercept	Regression coefficient	p-value	Data pairs type
Water Vapor-Direct	0.04	18.82	-2308.09	0.04	

Note: * Predictor variables with either p-value>0.1

5.2.3. Multicollinearity analysis

Table 5-4 shows the predictors variables selected with variance inflation factor value (VIF<30) for each data pairs of without time commensurate and with time commensurate respectively. The predictor variable such as water vapor, water vapor-low, water vapor-direct which remained in bivariate regression (Table 5-3) were excluded due to high collinearity with other predictor variables for without time commensurate data pairs.

Variance Inflation factor of 10 or 30 characterises that the predictor variable with other remaining predictor variables (Table 3-6 and Table 3-7) has strong relationship and can together explain 90 to 97% of variance (Equation 4-1) in that predictor variable. Thus, collinearity between the variables might increases the estimate of the variance in PM2.5.

Table 5-4 Predictor variables with Variance Inflation Factor (VIF<30)

Predictor Variables	VIF	Data pairs type
Wind Direction	7.28	With time commensurate
Wind Speed	18.94	
Temperature	13.66	
Relative Humidity	11.22	
Boundary Layer Height	6.00	
Aerosol Optical Depth	4.40	
Wind Direction	2.13	Without time commensurate
Wind Speed	3.07	
Temperature	18.01	
Relative Humidity	3.19	
Boundary Layer Height	1.98	
Aerosol Optical Depth	5.64	
Surface Skin Temperature	6.16	
Retrieved Moisture-Level 18	4.77	
Retrieved Moisture-Level 19	2.02	

5.2.4. Stepwise Regression

The selected predictor variables obtained after supervised forward selection stepwise regression procedure is presented in Table 5-5. These final predictor variables were used for multiple linear regression modelling to estimate PM2.5. AOD is selected as a first variable for data pairs that are time commensurate. Temperature explained highest variance (15%) in PM2.5 for data pairs without time commensurate. Thus, temperature is considered as a first variable. The predictor variables such as wind direction, wind speed and relative humidity were excluded as they were not found to be significant (p-value>0.1) for data pairs that are made time commensurate.

Table 5-5 Selected predicted variables for multiple linear regression modelling

Predictor Variables	Data pairs type
Aerosol Optical Depth	With time commensurate
Temperature	

Predictor Variables	Data pairs type
Boundary Layer Height	
Aerosol Optical Depth	Without time commensurate
Temperature	
Boundary Layer Height	
Relative Humidity	
Wind Direction	
Surface Skin Temperature	

Table 5-5 shows that out of altogether 24 potential predictor variables (Table 3-6 and Table 3-7), mainly temperature, AOD and Boundary layer height tend to explain more variance in PM2.5 (data pairs with time commensurate). It shows that the mostly all satellite retrieved atmospheric observations from MODIS atmospheric profile product were not remained at the end of variables selection. It is because of severe multicollinearity (VIF results not shown) although they were significant at 10% significance level in bivariate regression. Also, there is only one variable namely surface skin temperature, which is the only satellite retrieved atmospheric observations from MODIS atmospheric profile product that was retained in data pairs without time commensurate.

5.2.5. Multiple Linear Regression Models

These exploratory plots were prepared to visualize the linear relationship between PM2.5 and predictor variables.

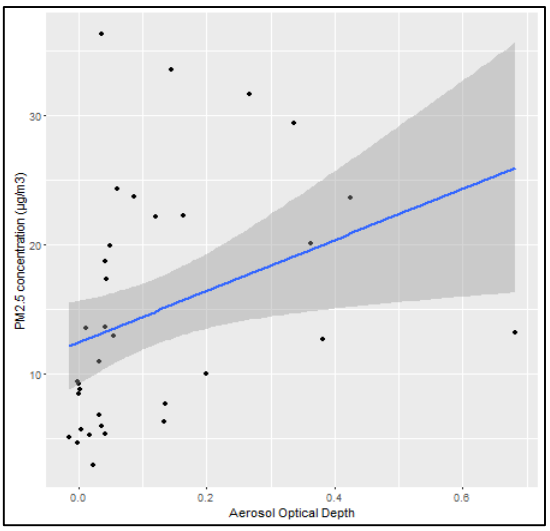


Figure 5-1 Scatterplot of AOD and PM2.5 concentrations with 90% CI across slope of a Linear Regression line

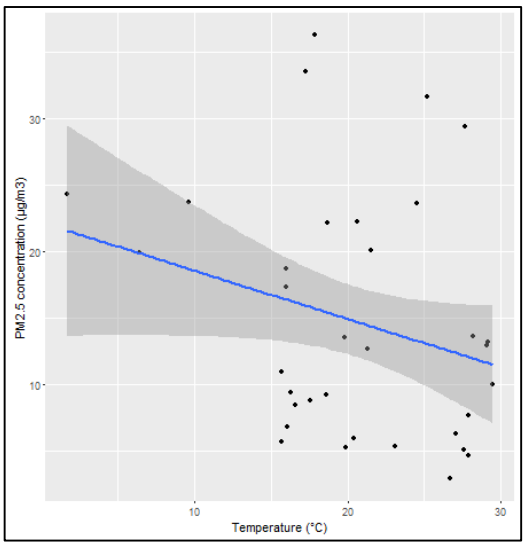


Figure 5-2 Scatterplot of Temperature and PM2.5 concentrations with 90% CI across slope of a Linear Regression line

Although, only temperature, aerosol optical depth and boundarly layer height were selected as a predictor variable for time commensurate data pairs (Table 5-5), the linear relationship of PM2.5 and relative humidity (Figure 5-4) is also plotted (which was not not found to be significant). In Figure 5-1 and Figure 5-4, AOD and relative humidity shows positive relationship with PM2.5 respectively. It shows that there are very less data points that falls within 90% confidence interval around the slope of the regerssion line, which shows that the relationship between PM2.5 and predictor variables are more complex than linear. In case of temperature and boundary layer height (Figure 5-2 and Figure 5-3), it shows negative relationship with PM2.5 respectively. It also shows that there are very less data points that falls within 90% confidence interval

around the slope of the regerssion line, which shows that the relatinoship between PM2.5 and predictor variables are more complex than linear.

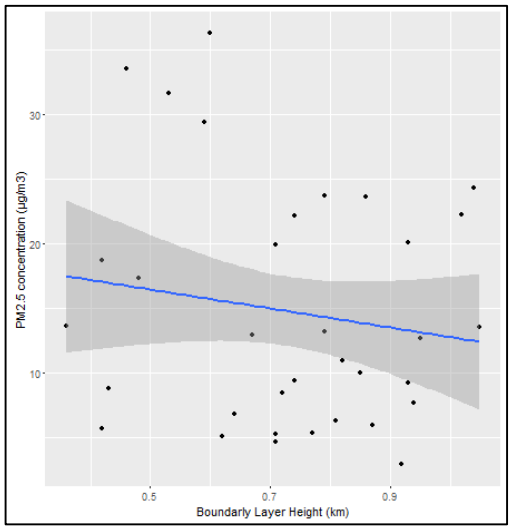


Figure 5-3 Scatterplot of Boundary Layer Height and PM2.5 concentrations with 90% CI across slope of a Linear Regression line

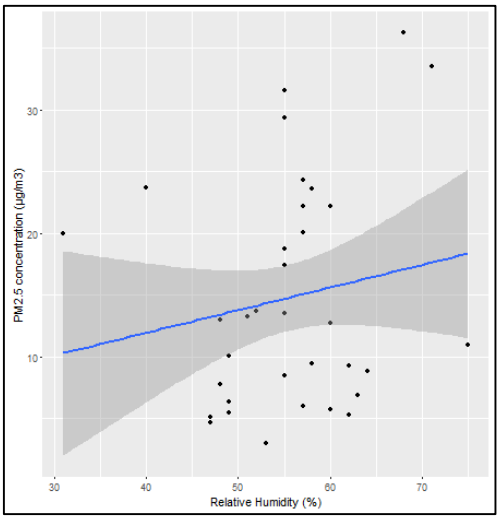


Figure 5-4 Scatterplot of Relative Humidity and PM2.5 concentrations with 90% CI across slope of a Linear Regression line

The results of a multiple linear regression analysis is presented as a summary in Table 5-6 and Table 5-7 with their statistical significane (p-value<0.1). AOD shows strong relationship with levels of PM2.5 ($p<0.001$) in regression model for AOD with time commensurate data pairs. However, estimated PM2.5 increases, when boundary layer height decreases (negative coefficient) as the dilution of PM2.5 is less in lower boundary layer height so, PM2.5 is concentrated (See Figure 5-3).

Temperature has negative regression coefficient in data pairs for both with time commensurate and without time commensurate. Generally, increase in temperature increases the PM2.5 concentration. In our study, we found increase in temperature decreases PM2.5 concentration. An increase in temperature accelerates the generation of secondary partcles near the surface. High amount of sulphate particles tend to have larger extinction coefficients in the atmosphere (Chin et al., 2002). Thus, if there is more sulfate particles in the air, the AOD would correspond to less PM2.5 concentration. It might be true that sufate particles may present in higher amount in Cabauw. as it is found that ammonium nitrate and sulfate particles comprise on average 8 µg/m3 ± 10% of the PM2.5 concentratin in the Netherlands (Matthijsen & Brink, 2007).

Table 5-6 Regression model summary with time commensurate data pairs

Predictor variable	Intercept	Regression coefficient	T-statistics	p-value	R2 (Adjusted)
Aerosol Optical Depth	211.350	32.960	3.520	0.001	29.6%
Boundary Layer Height		-13.530	-1.887	0.069	
Temperature		-0.650	-3.003	0.005	

Table 5-7 Regression model summary for without time commensurate data pairs

Predictor variable	Intercept	Regression coefficient	T-statistics	p-value	R2 (Adjusted)
Temperature	497.53	-0.81	-4.07	0.0001	29.5%

Predictor variable	Intercept	Regression coefficient	T-statistics	p-value	R2 (Adjusted)
Surface Skin Temperature		-1.67	-1.84	0.070	
Relative Humidity		0.19	2.37	0.020	
Aerosol Optical Depth		12.23	2.47	0.016	
Wind Direction		0.02	2.25	0.027	
Boundary Layer Height		-7.51	-1.98	0.052	

Strikingly, it is found that the final regression model for data pairs with time commensurate (Equation 5-1) and data pairs without time commensurate (Equation 5-2) explained similar variability in PM2.5 as 29.6% and 29.52% respectively at Cabauw. This explained variability is low compared to Schaap et al. (2009) which explained 52% of variability in PM2.5 at Cabauw for the study period August 2006 to May 2007. It should be also noted that, there were 50 data pairs that were created within the satellite overpass time of MODIS Terra and Aqua (collection 5) in Schaap et al., (2009). In comparison to current study, the data pair retrieval is low (34) with time commensurate AOD. And it is also seen that data pairs are higher without time commensurate AOD as in Schaap et al. (2009) for August 2006 to May 2007 compared to 86 data pairs for 2013-2014 in current study. It shows that there might be more cloud cover during our study period.

It is also found that the correlation of AOD and PM2.5 improved from 0.34 to 0.45 (Table 5-9) when accounting for boundary layer height with time commensurate AOD which is contrasting to Schaap et al. (2009), where accounting for BLH didn't improve correlation between AOD and PM2.5. It shows that AOD is strongly correlated with PM2.5 within the boundary layer. It also shows that our assumption of vertical distribution of particles above the boundary layer is relatively smooth is valid.

However, for data pairs without time commensurate, the correlation of AOD and PM2.5 did not improve well (Table 5-11) compared to data pairs with time commensurate. This shows that the AOD retrieved without time commensurate are less representative of the ground level PM2.5 concentrations. Also, It can be seen from scatterplot of PM2.5 and AOD (see Annex Figure 6-1), PM2.5 and BLH (see Annex Figure 6-3) that there are very less data points that falls within 90% confidence interval around the slope of the regerssion lines, which shows that the relationship between PM2.5 and predictor variables are more complex than linear.

With time commensurate data pairs,

$$PM_{2.5} = 32.96 \times AOD - 13.53 \times BLH - 0.65 \times TEMP + 211.35 \quad \text{Equation 5-1}$$

Without time commensurate data pairs,

$$PM_{2.5} = 12.23 \times AOD - 0.81 \times TEMP - 1.67 \times STEMP + 0.19 \times RH + 0.02 \times WDD - 7.51 \times BLH + 497.53 \quad \text{Equation 5-2}$$

Where, AOD is aerosol optical depth; TEMP is surface weather temperature; STEMP is surface skin temperature over land which is characterised as a function of surface air temperature, solar zenith and azimuth angles; RH is relative humidity; WDD is wind direction and BLH is boundary layer height.

The F-test result shows that F= 5.62 for 30 degrees of freedom for semi-empirical model with data pairs without time commensurate. The p-value is 0.0035 (p-value<0.1), which means that the semi-empirical model is statistically siginificant at 90% confidence interval. Similarly, the F-test result shows that F= 6.94 for 79 degrees of freedom for semi-empirical model with data pairs that are time commensurate. The p-

value is 0.0001 (p-value<0.1), which means that the semi-empirical model is statistically siginificant at 90% confidence interval.

5.2.6. Relative Importance of predictor variable

Table 5-8Table 5-8 shows that the AOD explains proportionately higher variance in regression Equation 5-1 compared to Equation 5-2 which is 18.6%. AOD is relatively explaining less variance in data pairs that are made without time commensurate. The reason could be the low correlation (0.11) between AOD and PM2.5. The true nature of relationship between AOD and PM2.5 might not be explained by linear relationship (see Figure 6-1). It may hold more complex relationship in the atmosphere, which is not well understood due to limitation of knowledge and the processes in the atmosphere about aerosol. Multiple R-squared is used to find the proportionate percentage of variance explained by predictor variables. The output of partitioning of a variance explained by a regression model is obtained with **"Relaimpo"** package in R.

Table 5-8 Results of Percentage of response variance for predictor variable

Predictor Variable	% of response variance	Multiple R-squared	Data pairs type
Aerosol Optical Depth	18.61	35.97	With time commensurate
Temperature	12.74		
Boundary Layer Height	4.63		
Temperature	13.77	34.50	Without time commensurate
Relative Humidity	7.14		
Surface Skin Temperature	4.61		
Aerosol Optical Depth	3.43		
Boundary Layer Height	3.10		
Wind Direction	2.45		

5.3. Correction of AOD for Relative Humidity using $f(RH)$

The AOD was corrected for relative humidity using simple method found in (Li et al., 2005; Tsai et al., 2011) for both data pairs (with time commensurate and without time commensurate). This method was applied when there was availability of relative humidity using Equation 4-4. Furthermore, AOD was corrected for relative humidity using advanced method when there was availability of hygroscopic growth factor, $f(RH)$ and different air mass type at Cabauw. This correction method was also used for both data pairs. During study, two sources of information to correct AOD for relative humidity using advanced method was available. An experimental fitting curve (humidogram) of $f(RH = 72\%, \lambda = 475\ \text{nm})$ only (Veefkind et al., 1996) for Petten (Equation 4-6), percentile values of $f(RH = 85\%, \lambda = 550\ \text{nm})$, average humidograms for different air mass type (Table 4-1) at Cabauw was available.

The offset values were obtained (Table 4-2) for Petten to modify humidogram to obtain $f(RH = 85\%, \lambda = 550\ \text{nm})$ after Cabauw. It is assumed (see Research assumptions) that it can be adapted with modification of offset values for different air mass type based on Zieger et al., (2013). This was carried out because instantaneous AOD was observed at 550 nm. The fitting curve at Petten was measured at 475 nm and it is approximately 90 km north-north west of Cabauw. It was necessary to modify according to measured f(RH) for different air mass types at Cabauw as the characteristics of aerosol depends upon air mass type (Wang & Christopher, 2003). The measured $f(RH)$ values were higher for air mass arriving at Cabauw from the Northern Sea or the Atlantic Ocean (Zieger et al., 2013) as it contained hygroscopic sea salt. The air mass mainly originated from industrialized area of the Ruhr area, Northern France, Southern Britain, the Netherlands and Belgium consisted of low $f(RH)$ at Cabauw. Thus, to simplify, the grouped air mass type

"maritime" and "others" (see Research assumptions) is obtained for Petten data adapted after Cabauw. The fitting curve to obtain $f(RH)$ for "maritime" and "others" is given by Equation 4-7 and Equation 4-8 respectively for Petten. Figure 4-3 shows the humidogram for grouped air mass type and other adapted air mass type at Petten after Cabauw. The two measured instantaneous values of $f(RH = 85\%, \lambda = 550\ \text{nm})$ of Cabauw is also plotted (see Figure 4-3).

The average humidogram obtained from re-plotting of extracted data (relative humidity and $f(RH)$) is shown in Figure 4-5. Again, to simplify, the grouped air mass type "maritime" and "others" (see Research assumptions) is obtained for Cabauw. The fitting curve to obtain f(RH) for "maritime" and "others" is given by Equation 4-14 and Equation 4-15 respectively. The correction of AOD for relative humidity using $f(RH)$ required selection of fitting curve of $f(RH)$ between "maritime" and "others" of either Petten or Cabauw.

5.4. Selection of fitting curve of f(RH) between "Maritime" and "Others" air mass type

5.4.1. With Conditional Statement

The 72 hr air mass back trajectories measured at Cabauw (Zieger et al., 2013) showed that for maritime air mass $f(RH)$ was higher than 2 and for others, it was less than 2. Thus, in the polynomial curve (Figure 4-6), the lower dotted line distinguishes $f(RH)$ above or below 2 as maritime and others respectively. But, there are still "other" air mass which lies above lower dotted line which can be both maritime and others respectively. Thus, upper dotted line is drawn to separate maximum value available for "others" air mass (2.57) with maritime.

It is found that 63.24% data pairs with "other" airmass $f(RH)$ values obtained with simple correction method for time commensurate data pairs, satisfies the conditional statements (4.4.1). And as per Research assumptions, wind direction at surface weather observation station at Cabauw is used as criteria to choose the air mass type for f(RH) in the current study. It would have been more certain if most of conditions were fulfilled by either of the air mass. The $f(RH)$ of **"Others"** grouped air mass type is chosen which is found similar to Schaap et al. (2009) study, where polluted continental conditions were chosen. In support, sensitivity analysis further provides information on variability of $f(RH)$ within the percentiles (25^{th}, median, 75^{th}) of grouped "maritime" and "others" air mass type in estimated PM2.5. The sensitivity analysis of $f(RH)$ for different air mass type at Cabauw and similarly, at Petten might also give more characteristics on effect of variability of $f(RH)$ on different air mass type in estimating PM2.5. However, as assumed (see Research assumptions), the air mass type were grouped to simplify the reality.

5.4.2. Sensitivity analysis of grouped "maritime" and "others" air mass type in estimated PM2.5

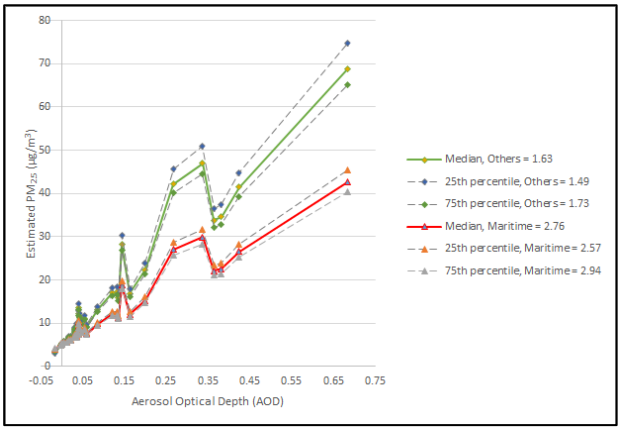


Figure 5-5 Sensitivity analysis of grouped "maritime" and "others" air mass type at Cabauw after Petten

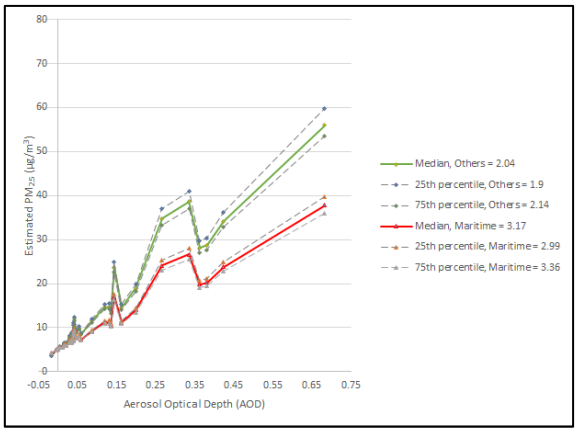


Figure 5-6 Sensitivity analysis of grouped "maritime" and "others" air mass type at Cabauw

Figure 5-5 and Figure 5-6, shows the characteristics of the effect of grouped "maritime" and "others" air mass type in estimated PM2.5 at Petten after Cabauw and at Cabauw respectively. The lower the value of $f(RH)$ for 25th percentile others air mass type (1.49), the uncertainty in the estimated PM2.5 is propagated (75 µg/m3) as shown in Figure 5-5, when AOD is high (0.69). Conversely, there is less uncertainty in estimated PM2.5 (40 µg/m3) for 75th percentile value of $f(RH)$ value (2.94) for maritime air mass type, when AOD is high as well. In Figure 5-5 at Petten after Cabauw, the variability of $f(RH)$ values between 25th and 75th percentile for maritime leads to a relatively small amount of variability in the estimated PM2.5 with respect to characteristics shown by others air mass. The estimated PM2.5 is comparatively more sensitive to variability of $f(RH)$ values between 25th and 75th percentile of other air mass type than maritime. Thus, more uncertainty in the $f(RH)$ values propagates to the estimated PM2.5, the greater the sensitivity of an estimated PM2.5 to $f(RH)$ values of others.

The $f(RH)$ values of grouped air mass types are less sensitive for lower values of AOD (<0.05) as the estimated PM2.5 is almost similar for both air mass type at Petten (Figure 5-5) and Cabauw (Figure 5-6). When AOD is higher, the variability in $f(RH)$ leads to a relatively large variability (about 10 µg/m3) in estimated PM2.5 for others air mass type at Petten compared to about 5 µg/m3 for maritime air mass type at Petten and Cabauw.

From Figure 5-5 and Figure 5-6, it can be seen that $f(RH)$ others air mass type of Cabauw is less sensitive than $f(RH)$ others air mass type at Petten in estimating PM2.5 under high observed instantaneous AOD. Thus, the increased uncertainty in $f(RH)$ for others air mass type at Cabauw leads to less uncertainty in estimated PM2.5.

The $f(RH)$ of Others air mass type found at Cabauw is chosen for correcting AOD. The result shows that the explained variability in PM2.5 is improved, when others air mass type is chosen for data pairs that are time commensurate (see Table 5-12). The consequence of the choice of air mass type is also tested by using $f(RH)$ of maritime air mass type as well found to correct AOD for relative humidity and boundary layer height. The result showed that the explained variability in PM2.5 decreased (result not shown). However, the choice of the air mass type did not improve (result not shown) the explained variability in PM2.5 for data pairs that are not time commensurate (Table 5-14). Thus, it is certain that the choice of "other" air mass should be considered for correcting AOD for relative humidity and boundary layer height.

5.5. Multiple Linear Regression Modelling Result with meteo scaled AOD

5.5.1. Correlation and Bivariate regression assessment

The correlation of the PM2.5 and meteo-scaled AOD improved from 0.34 to 0.39 using simple correction method for data pairs with time commensurate (see Table 5-9) The result is contrasting to Tsai et al., (2011) study at Taiwan which followed the method of Li et al., 2005. However, it should be noted that RH values were matched within the MODIS Terra (N=77) and Aqua (N=80) overpass time with RH only in the range between 50-65% in Tsai et al., (2011). In current study, in data pairs (N=34) consists of RH<94%. The correlation of AOD and PM2.5 did not improve for data pairs (N=86) without time commensurate AOD (Table 5-10) using simple method.

The correlation of the PM2.5 and meteo-scaled AOD improved from 0.34 to 0.41 using advanced correction method. The result is similar to Koelemeijer et al., (2006) study for Europe, where the correlation of meteo-scaled AOD and PM2.5 was improved from 0.38 to 0.59. The correlation of meteo-scaled AOD and PM2.5 did not improve for data pairs without time commensurate AOD (Table 5-10) using advanced method. The correlation between AOD and AOD normalized for BLH is improved from 0.34 to 0.45 (Table 5-9) for data pairs that are time commensurate, which is similar to Tsai et al., (2011) study in Taiwan, where AOD it was improved from 0.44 to 0.53 for MODIS Aqua.Similarly, the correlation between AOD and AOD normalized for BLH also improved from 0.11 to 0.17 (Table 5-10) for data pairs that are not time commensurate.

The meteo-scaled AOD is regressed with PM2.5 to assesess the linear relationship between meteo-scaled AOD and PM2.5. The explained variability in PM2.5 increased by 5% with meteo-scaled AOD for data pairs that are time commensurate. However, the meteo-scaled AOD did not improve explained variability in PM2.5 (result not shown) for data pairs that are not time commensurate.

Table 5-9 Result of correlation of AOD and PM2.5 for time commensurate data pairs

Dependent Variable	Predictor Variable	Correlation	Adj. R-squared	Remarks
PM2.5	AOD	0.34	9%	No correction for BLH and RH
	AOD/BLH	0.45	-	Correction for BLH only
	Meteo-scaled AOD	0.39	12.65	$f(RH)$ used in meteo-scaled AOD is obtained from Simple Method
	Meteo-scaled AOD	0.41	14.07	$f(RH)$ used is meteo-scaled AOD is obtained from advanced method

Table 5-10 Result of correlation of AOD and PM2.5 for without time commensurate data pairs

Dependent Variable	Predictor Variable	Correlation	Remarks
PM2.5	AOD	0.11	No correction for BLH and RH
	AOD/BLH	0.17	Correction for BLH only
	Meteo-scaled AOD	0.06	$f(RH)$ used in meteo-scaled AOD is obtained from Simple Method
	Meteo- scaled AOD	0.1	$f(RH)$ used is meteo-scaled AOD is obtained from advanced method

5.5.2. Using simple and advanced method for time commensurate data pairs

The summary of final multiple linear regression model for estimating PM2.5 for data pairs with time commensurate is shown in Table 5-11 (simple method) and Table 5-12 (advanced method) respectively. Boundary layer height was not found to be significant (p-value>0.1) for simple and advanced method. It could be because the AOD was already normalized for boundary layer height along with relative humidity (Equation 4-3). Thus, it was excluded as a candidate predictor variable in regression modelling with meteo-scaled AOD (Table 5-12) for advanced method. The meteo-scaled AOD shows strong relationship (positive coefficient) with estimated PM2.5 with p-value (0.001), which indicates the increase in AOD with correction for RH and BLH accounts for increase in estimated PM2.5.

The regression model (Equation 5-3) with meteo-scaled AOD and temperature using simple method explains at least 1.5 % more variability compared to 29.6 % (Equation 5-1) without correction of AOD for relative humidity and boundary layer height. This shows that the correction of AOD for relative humidity and BLH improves the variability in PM2.5. Altogether, the semi-empirical model (Equation 5-4) with meteo-scaled AOD and other predictor variables explains 4% more of variability in PM2.5 compared to AOD without correction for RH and BLH. However, the explained variability in PM2.5 is increased by 5% (Table 5-9) with only meteo-scaled AOD in bivariate regression analysis for data pairs that are time commensurate which is less than Koelemeijer et al., (2006) study for Europe, where there was 20% improvement in explained variability in bivariate regression of meteo-scaled AOD and PM2.5.

AOD is an instantaneous measurement measured in 1.2 sec with the orbital velocity of MODIS observation at 7.1 km/sec in a 10 × 10- km pixel. Since, an instantaneous AOD is matched with PM2.5 measured for 42 minutes at Cabauw air quality monitoring station during data pairing, it could be the reason that there is only 4 % increase in variability in PM2.5. Literature shows that in similar studies, area aggregated spatial matching of 5 × 5 pixels AOD is matched with hourly PM2.5 measurement time, assuming that air mass travels at 50 km/hr in troposphere before temporal matching to create data pairs of AOD and PM2.5. It should be further investigated in future that considering spatial matching and temporal matching with time commensurate AOD might improve more of the variability in PM2.5.

The total explained variability of a semi-empirical model is increased to 33.6%, 34.8% when both meteo-scaled AOD and temperature is considered, which is close 34.8% variability explained by meteo-scaled AOD alone in Koelemeijer et al., (2006). It should be noted however that the PM2.5 closer to overpass time of MODIS Terra and Aqua collection 04 was used in Koelemeijer et al., (2006). It should be also noted that the PM2.5 measurement techniques used also differs the correlation of PM2.5 and satellite retrieved AOD. So, the results may not be comparable to other studies which used different measurement techniques for PM2.5 measurements as the uncertainties in the measurement techniques may differ.

Table 5-11 Regression model summary using meteo-sclaed AOD in simple method for time commensurate data pairs

Predictor variable	Intercept	Regression coefficient	T-statistics	p-value	R2 (Adjusted)
Meteo scaled AOD	206.06	56.42	3.69	0.001	31.1%
Temperature		-0.67	-3.1	0.004	

Regression model for Meteo-scaled AOD with simple method,

$$PM_{2.5} = 56.42 \times AOD^{*} - 0.67 \times TEMP + 206.06 \qquad \text{Equation 5-3}$$

Where, AOD^{*} is the meteo scaled AOD (corrected AOD for $f(RH)$ and BLH) and TEMP is surface weather temperature.

Table 5-12 Regression model summary using meteo-scaled AOD in advanced method for time commensurate data pairs

Predictor variable	Intercept	Regression coefficient	T-statistics	p-value	R2 (Adjusted)
Meteo scaled AOD	210.78	28.65	3.90	0.0001	33.6%
Temperature		-0.68	-3.22	0.003	

Regression Model for meteo-scaled AOD with advance method ($f(RH)$),

$$PM_{2.5} = 28.65 \times AOD^{*} - 0.68 \times TEMP + 210.78 \quad \text{Equation 5-4}$$

Where, AOD^{*} is the meteo scaled AOD (corrected AOD for $f(RH)$ and BLH) and $TEMP$ is surface weather temperature.

The F-test result shows that F= 9.34 for 31 degrees of freedom for semi-empirical model with data pairs without time commensurate. The p-value is 0.0006 (p-value<0.1), which means that the semi-empirical model is statistically siginificant at 90% confidence interval. Similarly, the F-test result shows that F= 7.79 for 80 degrees of freedom for semi-empirical model with data pairs that are time commensurate. The p-value is 0.0001 (p-value<0.1), which means that the semi-empirical model is statistically siginificant at 90% confidence interval.

5.5.3. Using simple and advanced method for without time commensurate data pairs

On the other hand, the summary of multiple linear regression model for estimating PM2.5 using data pairs without time commensurate is shown in Table 5-13 and Table 5-14 respectively. The final regression model with meteo-scaled AOD with simple (Equation 5-5) and advanced method (Equation 5-6) did not improve the initial explained variability (29.5%) in PM2.5 (Table 5-7) for without time commensurate data pairs which is contrasting to the current study results for with time commensurate data pairs. This could be that the AOD suffered from hygroscopic growth of particles. Due to hygroscopic growth of particles, the size of the aerosol may increase from 2-10 times (Malm et al., 2000) and accounts for increase in extinction efficiencies of particles. Thus, PM2.5 observed would be lower with the same amount of AOD in the atmosphere (Liu et al., 2005). However, the correction of AOD for RH also didn't improve the explained variability in PM2.5. It may be also that the relationship of PM2.5 with selected predictor variables couldn't be explained by linear relationship and may hold complex relationship.

Table 5-13 Regression model summary using meteo-sclaed AOD in simple method for without time commensurate data pairs

Predictor variable	Intercept	Regression coefficient	T-statistics	p-value	R2 (Adjusted)
Temperature	476.91	-0.81	-3.99	0.001	28%
Surface Skin Temperature		-1.61	-1.81	0.07	
Relative Humidity		0.28	3.46	0.001	
Meteo scaled AOD		18.69	2.47	0.02	
Wind Direction		0.02	2.19	0.03	

Without time commensurate data pairs, for meteo scaled AOD with simple method,

$$PM_{2.5} = 18.69 \times AOD^{*} - 0.81 \times TEMP + 0.28 \times RH + 0.02 \times WDD - 1.61 \times STEMP + 476.91 \quad \text{Equation 5-5}$$

Where, AOD^{*} is the meteo scaled AOD (corrected AOD for $f(RH)$ and BLH), TEMP is surface weather temperature, RH is relative humidity, WDD is wind direction, and STEMP is surface skin temperature over land which is characterised as a function of surface air temperature, solar zenith and azimuth angles.

Table 5-14 Regression model summary using meteo-scaled AOD in advanced method for without time commensurate data pairs

Predictor variable	Intercept	Regression coefficient	T-statistics	p-value	R2 (Adjusted)
Temperature	465.28	-0.80	-3.99	0.00	29%
Surface Skin Temperature		-1.54	-1.74	0.09	
Relative Humidity		0.26	3.22	0.00	
Meteo scaled AOD		9.91	2.55	0.01	
Wind Direction		0.02	2.21	0.03	

Without time commensurate data pairs, for meteo scaled AOD with advanced method,

$$PM_{2.5} = 9.91 \times AOD^{*} - 0.80 \times TEMP + 0.26 \times RH + 0.02 \times WDD - 1.54 \times STEMP + 465.28 \quad \text{Equation 5-6}$$

Where, $AOD^{*}$is the meteo scaled AOD (corrected AOD for $f(RH)$ and BLH), TEMP is surface weather temperature, RH is relative humidity, WDD is wind direction, and STEMP is surface skin temperature over land which is characterised as a function of surface air temperature, solar zenith and azimuth angles.

5.6. Performance result of regression models for PM2.5 estimation

Table 5-15 shows the summary of the statistics of PM2.5 estimation performance of regression models (Equation 5-1 to Equation 5-6) for estimating PM2.5. The slope and correlation is higher for data pairs that are time commensurate and corrected with advanced method among all data pairs as presented below. Thus, it shows that the regression model for predictor variables meteo-scaled AOD and temperature with time commensurate data pairs performs better compared to data pairs without time commensurate in estimating PM2.5. However, the accuracy of the satellite estimate of PM2.5 at a location depends upon robustness of regression relationship between AOD and PM2.5, which depends upon retrieval accuracy of AOD, PM2.5 measurement and surface weather observations. Since, the retrieval accuracy is about ±20% of AERONET AOD, which need to be validated for current studies, which depends upon factors such as density, aerosol composition, hygroscopicity etc. at a location. The uncertainty of PM2.5 measurements at Cabauw air quality monitoring station is about 11% for daily averages for TEOM (reference) method. However, in current study, hourly average PM2.5 is used, thus the uncertainty increases for hourly measurement, which could be the reason for RMSE of about 7 μg/m3.

Slope is less than 1, which means that AOD and temperature is not a perfect predictor. The reason could be that since the interaction between predictor variable is not considered, which could influence the variability of PM2.5 at a location. Also, the accounting of time discrepancies does not show significant improvement in variability in PM2.5, which shows that a lot more improvement is required in the current method. There could be presence of multiple aerosol layer in the atmosphere in a vertical column of

atmosphere at Cabauw, which could be the reason for correlation to be about 0.61, as total AOD would remain same in the vertical column but PM2.5 and AOD might correlate less at the surface.

Table 5-15 Statistics of PM2.5 estimation performance of regression models

Data pair type	Equation	Mean a	Mean b	Pearson's r$\pm\sigma$	Mean RMSE$\pm\sigma$
Without AOD correction for $f(RH)$	Equation 5-1	10.03	0.38	0.53±0.43	7.94±3.46
AOD corrected using $f(RH)$ (simple method)	Equation 5-3	9.56	0.44	0.60±0.40	7.82±4.01
AOD corrected using $f(RH)$ (advance method)	Equation 5-4	9.18	0.46	0.61±0.40	7.6±3.85
Without AOD correction for $f(RH)$	Equation 5-2	9.85	0.29	0.48±0.31	6.76±1.8
AOD corrected using $f(RH)$ (simple method)	Equation 5-5	10.03	0.28	0.48±0.31	6.79±1.8
AOD corrected using $f(RH)$ (advance method)	Equation 5-6	10.00	0.28	0.48±0.31	6.78±1.81

Note: Mean a and b are the intercept and slope from the linear regression equation $Estimated\ PM2.5 = a + b \times observed\ PM2.5$, where estimated PM2.5 is obtained from the corresponding linear models respectively in each different data pairs cases. Mean r and σ is the average values of all the correlation coefficients and their standard deviation. Mean RMSE is the mean of the root mean square error of the estimated and observed PM2.5 where σ denotes the standard deviations of all the RMSE values.

6. CONCLUSION AND RECOMMENDATION

Hypothesis 1 conclusion:

1. The present study created data pairs that are time commensurate when the satellite overpasses the air quality monitoring station at Cabauw. It is found that there is significant difference between semi-empirical model with time commensurate data pairs and without time commensurate data pairs at 90% confidence interval. Thus, the null hypothesis is rejected.

2. However, the correlation between AOD and PM2.5 is increased, when data pairs are time commensurate whereas there is weak correlation in data pairs that are not time commensurate. Thus, the study concludes that the variability in PM2.5 is improved, when AOD is corrected for time discrepancies with other predictor variables.

However, the true nature of relationship between AOD and PM2.5 couldn't be well understood in depth. Thus with current limitations, the following recommendation can be made:

1. The validation study of AOD with AERONET need to be carried out to assure about retrieval accuracy of AOD at a location. The global validation study of AOD for MODIS collection 005 shows the retrieval accuracy of $0.05 \pm 15\% \times$AOD over land (Levy et al., 2007) compared to AERONET AOD. However, this might not be the case for the study period which need to be validated (see Research assumptions).
2. It is necessary to have more in depth understanding of the relationship between PM2.5 and AOD, which should be studied to enhance the modeling, which could improve the explain variability in PM2.5.
3. More in depth understanding of relationship between PM2.5 and other predictors variable should be studies which would improve the modeling effort to improve the explain variability in PM2.5.
4. The interactions between predictors variables need to be understood and considered in modelling, which is unclear in this research due to lack of background knowledge on the air quality system dynamics and processes of the atmosphere. Thus, the understanding of those interactions and consideration into modeling might improve the explained variability in PM2.5.
5. The change in atmosphere may change AOD values within satellite overpass time. Thus, more advanced interpolation techniques may assist to more accurately interpolate to consider the true nature of relationship between satellite retrieved instantaneous AOD and PM2.5.
6. As found in similar studies (Liu et al., 2005; Guo et al., 2014) AOD varies with seasons. So, the seasonal variability of AOD could explain more variability in some seasons in PM2.5 better than the other. Thus, the longer study period may be considered.

Hypothesis 2 conclusion:

1. The present study found that, there is significant difference between the semi-empirical model with data pairs that are time commensurate with meteo-scaled AOD and data pairs without time commensurate meteo-scaled AOD at 90% confidence interval. Thus, the null hypothesis is rejected.

Thus, for future work following recommendation can be made:

1. It is assumed that boundary layer is a well-mixed layer (see Research assumptions) and AOD values are normalized using boundary layer height. There could be conditions when there might be the

residual layer in the atmosphere during satellite overpass time, which reduces the correlation of boundary layer AOD with PM2.5. Thus, consideration of the vertical distribution of particles above the boundary layer may account for residual layer as found in study (Tsai et al., 2011) which might improve the correlation between AOD and PM2.5.

LIST OF REFERENCES

Aryal, R. K., Lee, B. K., Karki, R., Gurung, A., Baral, B., & Byeon, S. H. (2009). Dynamics of PM2.5 concentrations in Kathmandu Valley, Nepal. *Journal of Hazardous Materials*, *168*(2-3), 732–738. doi:org/10.1016/j.jhazmat.2009.02.086

Brunekreef, B., & Holgate, S. T. (2002). Air Pollution and Health, *360*, 1233–1242. doi:org/10.1016/S0140-6736(02)11274-8

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., … Nakajima, T. (2002). Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements. *Journal of the Atmospheric Sciences*, *59*(3), 461–483. doi:org/10.1175/1520-0469(2002)059

Chow, J. C., Engelbrecht, J. P., Freeman, N. C. G., Hisham Hashim, J., Jantunen, M., Michaud, J. P., … Zhu, T. (2002). Chapter one: Exposure measurements. *Chemosphere*, *49*(9), 873–901. doi:org/10.1016/S0045-6535(02)00233-3

Chu, D. a., Kaufman, Y. J., Zibordi, G., Chern, J. ., Mao, J., Li, C., & Holben, B. N. (2003). Global monitoring of air pollution over land from the Earth Observing System-Terra Moderate Resolution Imaging Spectroradiometer (MODIS). *Journal of Geophysical Research*, *108*(D21), 1–18. doi:org/10.1029/2002JD003179

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 027–046. doi:org/10.1111/j.1600-0587.2012.07348

Engel-Cox, J. a., Hoff, R. M., Rogers, R., Dimmick, F., Rush, A. C., Szykman, J. J., … Zell, E. R. (2006). Integrating lidar and satellite optical depth with ambient monitoring for 3-dimensional particulate characterization. *Atmospheric Environment*, *40*(40), 8056–8067. doi:org/10.1016/j.atmosenv.2006.02.039

Engel-Cox, J. a., Holloman, C. H., Coutant, B. W., & Hoff, R. M. (2004). Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment*, *38*(16), 2495–2509. doi:org/10.1016/j.atmosenv.2004.01.039

European Parliament, & Council of the European Union. (2008). Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. *Official Journal of the European Communities*, 1–43.

Feng, X., Wu, B., & Yan, N. (2015). A Method for Deriving the Boundary Layer Mixing Height from MODIS Atmospheric Profile Data. *Atmosphere*, *6*(9), 1346–1361. doi:org/10.3390/atmos6091346

Gatebe, C. K., King, M. D., Si-Chee, T., Ji, Q., Arnold, G. T., & Li, J. Y. (2001). Sensitivity of off-nadir zenith angles to correlation between visible and near-infrared reflectance for use in remote sensing of aerosol over land. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(4), 805–819. doi:org/10.1109/36.917901

Giri, D. (2006). Ambient air quality of Kathmandu valley as reflected by atmospheric particulate matter concentrations (PM 10), *3*(4), 403–410.

Grömping, U. (2006). Relative importance for linear regression in R: the package relaimpo. *Journal Of Statistical Software*, *17*(1), 139–147. doi:org/10.1016/j.foreco.2006.08.245

Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, *61*(2), 139–147. doi:org/10.1198/000313007X188252

Grömping, U. (2010). Relaimpo. R package version 2.2-2. Retrieved from https://cran.r-project.org/web/packages/relaimpo/relaimpo

Guo, Y., Feng, N., Christopher, S. a., Kang, P., Zhan, F. B., & Hong, S. (2014). Satellite remote sensing of fine particulate matter (PM2.5) air quality over Beijing using MODIS. *International Journal of Remote Sensing*, *35*(17), 6522–6544. doi:org/10.1080/01431161.2014.958245

Gupta, P. (2008). *Particulate Matter air quality assessment over south east United States using Satellite retrieved and ground measurements*. The University of Alabama in Huntsville, The Department of Atmospheric Science. PhD Dissertation.

Gupta, P., & Christopher, S. a. (2008). Seven year particulate matter air quality assessment from surface and satellite measurements. *Atmospheric Chemistry and Physics Discussions*, *8*(1), 327–365. doi:org/10.5194/acpd-8-327-2008

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), 1–13. doi:org/10.1029/2008JD011496

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. doi:org/10.1016/j.atmosenv.2006.03.016

Gurung, A., & Bell, M. L. (2013). The state of scientific evidence on air pollution and human health in Nepal. *Environmental Research*, *124*, 54–64. doi:org/10.1016/j.envres.2013.03.007

Haij, M. de, Wauben, W., & Baltink, H. K. (2007). *Continuous mixing layer height determination using the LD-40 ceilometer: a feasibility study*. De Bilt.

Hoff, R. M., & Christopher, S. a. (2009). The A&WMA 2009 Critical Review -- Remote Sensing of Particulate Pollution from Space: Have We Reached the Promised Land? *Journal of the Air & Waste Management Association*, *59*(6), 645–675. doi:org/10.3155/1047-3289.59.6.645

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, a., … Smirnov, a. (1998). AERONET - A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*(1), 1–16. doi:org/10.1016/S0034-4257(98)00031-5

Hu, X., Waller, L. a, Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G., Estes, S. M., … Liu, Y. (2013). Estimating ground-level PM(2.5) concentrations in the southeastern U.S. using geographically weighted regression. *Environmental Research*, *121*, 1–10. doi:org/10.1016/j.envres.2012.11.003

Hu, X., Waller, L. a., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., … Liu, Y. (2014). Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sensing of Environment*, *140*, 220–232. doi:org/10.1016/j.rse.2013.08.032

Hutchison, K. D., Faruqui, S. J., & Smith, S. (2008). Improving correlations between MODIS aerosol optical thickness and ground-based PM2.5 observations through 3D spatial analyses. *Atmospheric Environment*, *42*(3), 530–543. doi:org/10.1016/j.atmosenv.2007.09.050

Ichoku, C., Chu, D. A., Mattoo, S., Kaufman, Y. J., Remer, L. A., Tanre, D., … Holben, B. N. (2002). A spatio-temporal approach for global validation and analysis of MODIS aerosol products. *Geophysical Research Letters*, *29*(12), 1–4. doi:org/10.1029/2001GL013206

Justice, E., Huston, L., Krauth, D., Mack, J., Oza, S., Strawa, A. W., … Schmidt, C. (2009). Investigating Correlations Between Satellite-Derived Aerosol Optical Depth and Ground Pm 2.5 Measurements in California ' S San. Paper presented at ASPRS 2009 Annual Conference. Baltimore, Maryland

Kaufman, Y. J., Tanré, D., & Boucher, O. (2002). A satellite view of aerosols in the climate system. *Nature*, *419*(6903), 215–223. doi:org/10.1038/nature01091

Kloog, I., Nordio, F., Coull, B. a., & Schwartz, J. (2012). Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the mid-atlantic states. *Environmental Science and Technology*, *46*(21), 11913–11921. doi:org/10.1021/es302673e

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. doi:org/10.1016/j.atmosenv.2006.04.044

Kumar, N. (2010). A Hybrid Approach for Predicting PM2.5 Exposure. *Environmental Health Perspectives*, *118*, A425. doi:org/10.1289/ehp.1002706

Kumar, N., Chu, A. D., Foster, A. D., Peters, T., & Willis, R. (2011). Satellite Remote Sensing for Developing Time and Space Resolved Estimates of Ambient Particulate in Cleveland, OH. *Aerosol Science and Technology*, *45*(9), 1090–1108. doi:org/10.1080/02786826.2011.581256

Lang, J., Cheng, S., Li, J., Chen, D., Zhou, Y., Wei, X., … Wang, H. (2013). A monitoring and modeling study to investigate regional transport and characteristics of PM2.5 pollution. *Aerosol and Air Quality Research*, *13*(3), 943–956. doi:org/10.4209/aaqr.2012.09.0242

Lee, H. J., Liu, Y., Coull, B. a., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict PM2.5 concentrations. *Atmospheric Chemistry and Physics*, *11*(15), 7991–8002. doi:org/10.5194/acp-11-7991-2011

Lee, S.-J., Serre, M., van Donkelaar, A., Martin, R. V., Burnett, R. T., & Jerrett, M. (2012). Comparison of Geostatistical Interpolation and Remote Sensing Techniques for Estimating Long-Term Exposure to Ambient PM2.5 Concentrations across the Continental United States. *Environmental Health Perspectives*, (September 2015). doi:org/10.1289/ehp.1205006

Levy, R. C. (2009). The dark-land MODIS collection 5 aerosol retrieval: algorithm development and product evaluation. In *Satellite Aerosol Remote Sensing over Land* (pp. 19–68). Springer Berlin Heidelberg.

Levy, R. C., Remer, L. A., & Dubovik, O. (2007). Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land. *Journal of Geophysical Research Atmospheres*, *112*(13), n/a–n/a. doi:org/10.1029/2006JD007815

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. doi:org/10.5194/acp-10-10399-2010

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*(13), n/a–n/a. doi:org/10.1029/2006JD007811

Levy, R., Remer, L., & Tanré, D. (2009). Algorithm for Remote Sensing of Tropospheric Aerosol over Dark Targets from MODIS: Collections 005 and 051: Revision 2; Feb 2009. *Download from Http://…*, 1–96. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.386.980&rep=rep1&type=pdf

Li, C., Jieta, M., Lau, A. K. H., Zibing, Y., Meihua, W., & Xiaoyang, L. (2005). Application of MODIS satellite products to the air pollution research in Beijing. *Science in China(Series D:Earth Sciences)*, *48*(March 2016), 209–219. doi:org/10.1360/05yd0395

Liu, Y., Franklin, M., Kahn, R., & Koutrakis, P. (2007). Using aerosol optical thickness to predict ground-level PM2.5 concentrations in the St. Louis area: A comparison between MISR and MODIS. *Remote Sensing of Environment*, *107*(1-2), 33–44. doi:org/10.1016/j.rse.2006.05.022

Liu, Y., Paciorek, C. J., & Koutrakis, P. (2009). Estimating regional spatial and temporal variability of PM2.5 concentrations using satellite data, meteorology, and land use information. *Environmental Health Perspectives*, *117*(6), 886–892. doi:org/10.1289/ehp.0800123

Liu, Y., Sarnat, J. A., Kilaru, A., Jacob, D. J., & Koutrakis, P. (2005). Estimating ground-level PM(2.5) in the eastern united states using satellite remote sensing. *ENVIRONMENTAL SCIENCE & TECHNOLOGY*, *39*(9), 3269–3278. doi:org/Doi 10.1021/Es049352rn

Luo, Y., Trishchenko, A. P., Latifovic, R., & Li, Z. (2005). Surface bidirectional reflectance and albedo properties derived using a land cover-based approach with Moderate Resolution Imaging Spectroradiometer observations. *Journal of Geophysical Research D: Atmospheres*, *110*(1), 1–17. doi:org/10.1029/2004JD004741

Lyapustin, a. I. (2001). Three-dimensional effects in the remote sensing of surface albedo. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(2), 254–263. doi:org/10.1109/36.905233

Malm, W. C., Day, D. E., & Kreidenweis, S. M. (2000). Light scattering characteristics of aerosols at ambient and as a function of relative humidity: Part II - A comparison of measured scattering and aerosol concentrations using statistical models. *Journal of the Air & Waste Management Association*, *50*(5), 701–709. doi:org/10.1080/10473289.2000.10464114

Matthijsen, J., & ten Brink, H. M. (2007). *PM 2.5 in the Netherlands: Consequences of the new European air quality standards*.

McIDAS-V User ' s Guide. (2015). Retrieved from http://www.ssec.wisc.edu/mcidas/software/v/index.html

Monna, W., & Bosveld, F. (2013). *In Higher Spheres: 40 years of observations at the Cabauw Site*. De Bilt.

NASA. (2016a). Earthdata-Frequently Asked Questions: MODIS Near Real-Time Data. Retrieved April 5, 2016, from https://earthdata.nasa.gov/faq

NASA. (2016b). LAADS WEB Home: Search for Data Products. Retrieved December 5, 2016, from https://ladsweb.nascom.nasa.gov/data/search.html

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. doi:org/10.1007/s11135-006-9018-6

Panday, A. K., Prinn, R. G., & Schar, C. (2009). Diurnal cycle of air pollution in the Kathmandu Valley, Nepal: 2. Modeling results. *Journal of Geophysical Research Atmospheres*, *114*(21). doi:org/10.1029/2008JD009808

Pelletier, B., Santer, R., & Vidot, J. (2007). Retrieving of particulate matter from optical measurements: A semiparametric approach. *Journal of Geophysical Research: Atmospheres*, *112*(2007), 1–10.

doi:org/10.1029/2005JD006737

Pope, C. A. I., & Dockery, D. W. (2006). Health Effects of Fine Particulate Air Pollution: Lines that Connect, 709–742.

Putero, D., Cristofanelli, P., Marinoni, A., Adhikary, B., Duchi, R., Shrestha, S. D., … Bonasoni, P. (2015). Seasonal variation of ozone and black carbon observed at Paknajol, an urban site in the Kathmandu Valley, Nepal. *Atmospheric Chemistry and Physics*, *15*(24), 13957–13971. doi:org/10.5194/acp-15-13957-2015

Quinn, G. P., & Keough, M. J. (2002). *Experimental Design and Data Analysis for Biologists*. *Experimental design and data analysis for biologists* (Vol. 277). doi:org/10.1016/S0022-0981(02)00278-2

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., … Holben, B. N. (2005). The MODIS Aerosol Algorithm, Products, and Validation. *Journal of the Atmospheric Sciences*, *62*(4), 947–973. doi:org/10.1175/JAS3385.1

Remer, L. a., Wald, a. E., & Kaufman, Y. J. (2001). Angular and seasonal variation of spectral surface reflectance ratios: implications for the remote sensing of aerosol over land. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(2), 275–283. doi:org/10.1109/36.905235

Remer, L., Kaufman, Y., Tanré, D., Mattoo, S., Li, R., Martins, J. V., … Koren, I. (2002). *Collection 005 Change Summary for MODIS Aerosol (04 _ L2) Algorithms*.

RIVM. (2016). Home- National Institute for Health and the Environment. doi:org/10.1017/CBO9781107415324.004

Sahu, S. K., Beig, G., & Parkhi, N. S. (2011). Emissions inventory of anthropogenic PM2.5 and PM10 in Delhi during Commonwealth Games 2010. *Atmospheric Environment*, *45*(34), 6180–6190. doi:org/10.1016/j.atmosenv.2011.08.014

Schaap, M., Apituley, A., Timmermans, R. M. A., Koelemeijer, R. B. A., & Leeuw, G. de. (2009). Exploring the relation between aerosol optical depth and PM 2.5 at Cabauw, the Netherlands. … *and Physics*, 909–925. doi:org/10.5194/acp-9-909-2009

Seemann, S. W., Borbas, E. E., Li, J., Menzel, W. P., & Gumley, L. E. (2006). Modis Atmospheric Profile Retrieval. *Modis Atbd*, *Version 6*(October), 40. Retrieved from \Biblioteca_Digital_SPR\Seemann2006_ATBD.pdf

Seinfeld, J. H., & Pandis, S. N. (2006). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Hoboken, N. J.: John Wiley & Sons.

Shrestha, H. K. (2015). Govt to install 11 air quality monitoring stations | Nepal Mountain News. Retrieved from http://www.nepalmountainnews.com/cms/2015/07/22/govt-to-install-11-air-quality-monitoring-stations/

Space Science and Engineering Center. (2000). *The VisAD Java Component Library Developers Guide*.

Stull, R. B. (1988). *An Introduction to Boundary Layer Meteorology* (Vol. 13). Dordrecht, The Netherlands: Kluwer Academic Publishers. doi:org/10.1007/978-94-009-3027-8

Tian, J., & Chen, D. (2010). A semi-empirical model for predicting hourly ground-level fine particulate matter (PM2.5) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sensing of Environment*, *114*(2), 221–229. doi:org/10.1016/j.rse.2009.09.011

Tsai, T.-C., Jeng, Y.-J., Chu, D. A., Chen, J.-P., & Chang, S.-C. (2011). Analysis of the relationship between MODIS aerosol optical depth and particulate matter from 2006 to 2008. *Atmospheric Environment*, *45*(27), 4777–4788. doi:org/10.1016/j.atmosenv.2009.10.006

United States Environmental Protection Agency. (2015). EPA Home- Fine particle (PM2.5) designations: Frequent Questions. Retrieved August 18, 2015, from http://www3.epa.gov/pmdesignations/faq.htm#1

US EPA. (2015). EPA Home-Particulate Matter: Basic Information. Retrieved January 15, 2016, from http://www3.epa.gov/pm/basic.html

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. doi:org/10.1289/ehp.0901623

van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research: Atmospheres*, *111*(21), 1–10. doi:org/10.1029/2005JD006996

Veefkind, J. P., Hage, J. C. H. Van Der, & Brink, H. M. (1996). Nephelometer derived and directly measured aerosol optical depth of the atmospheric boundary layer. *Atmospheric Research*, *41*, 217–228. doi:org/10.1016/0169-8095(96)00011-7

Vienneau, D., de Hoogh, K., Beelen, R., Fischer, P., Hoek, G., & Briggs, D. (2010). Comparison of land-use regression models between Great Britain and the Netherlands. *Atmospheric Environment*, *44*(5), 688–696. doi:org/10.1016/j.atmosenv.2009.11.016

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. doi:org/10.1029/2003GL018174

Wang, J., & Martin, S. T. (2007). Satellite characterization of urban aerosols: Importance of including hygroscopicity and mixing state in the retrieval algorithms. *Journal of Geophysical Research*, *112*(D17), D17203. doi:org/10.1029/2006JD008078

WHO | Ambient (outdoor) air quality and health. (n.d.). Retrieved June 27, 2015, from http://www.who.int/mediacentre/factsheets/fs313/en/

World Health Organization. (2006). WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment. *Geneva: World Health Organization*, 1–22. Retrieved from http://whqlibdoc.who.int/hq/2006/WHO_SDE_PHE_OEH_06.02_eng.pdf?ua=1

Zhang, A., Qi, Q., Jiang, L., Zhou, F., & Wang, J. (2013). Population Exposure to PM2.5 in the Urban Area of Beijing. *PLoS ONE*, *8*(5). doi:org/10.1371/journal.pone.0063486

Zieger, P., Fierz-Schmidhauser, R., Weingartner, E., & Baltensperger, U. (2013). Effects of relative humidity on aerosol light scattering: Results from different European sites. *Atmospheric Chemistry and Physics*, *13*(21), 10609–10631. doi:org/10.5194/acp-13-10609-2013

Zieger, P., Weingartner, E., Henzing, J., Moerman, M., De Leeuw, G., Mikkil, J., … Baltensperger, U. (2011). Comparison of ambient aerosol extinction coefficients obtained from in-situ, MAX-DOAS and LIDAR measurements at Cabauw. *Atmospheric Chemistry and Physics*, *11*(6), 2603–2624. doi:org/10.5194/acp-11-2603-2011

APPENDIX

A. MODIS products used, their scientific data set and description

MODIS Products, Parameters and their Description:

ID	Processing Level	Product type	Scientific Data Set	Key Parameters	Units	Data Type	Raw data (range)	fill value	Scale factor	Offset value	Physical Value (range)	QAC flag	Comment
MOD04/MYD04	2	Aerosol	Optical_Depth_Land_And_Ocean	AOT at 0.55 µm for both ocean (best) and land (corrected)	none	16-bit integer	-100, 5000	9999	0.001	0	-0.05, 5	QAC≥ 0 (Ocean), QAC≥ 1 (Land)	For C005, QAC= 0 (bad quality), QAC= 3 (good quality)
-	-	-	-	-									
MOD07/MYD07	2	Atmospheric Profiles	Brightness_Temperature	Brightness Temperature	K	16-bit integer	0, 20000	-32768	0.00999	-15000	150, 350		
MOD07/MYD07	2	Atmospheric Profiles	Skin_Temperature	Surface Temperature	K	16-bit integer	0, 20000	-32768	0.00999	-15000	150, 350		
MOD07/MYD07	2	Atmospheric Profiles	Surface_Pressure	Surface Pressure	hPa	16-bit integer	8000, 11000	-32768	0.1	0	800, 1100		
MOD07/MYD07	2	Atmospheric Profiles	Surface_Elevation	Surface Elevation	m	16-bit integer	-400, 8840	-32768	1	0	-400, 8840		
-	-	-	-	-									
MOD07/MYD07	2	Atmospheric Profiles	Tropopause_Height	Tropopause Height	hPa	16-bit integer	10, 11000	-32768	0.1	0	1, 1100		
MOD07/MYD07	2	Atmospheric Profiles	Lifted_Index	Lifted Index	K	16-bit integer	-2000, 4000	-32768	0.00999	0	-20, 40		
MOD07/MYD07	2	Atmospheric Profiles	K_Index	K_Index	K	16-bit integer	11500, 20000	-32768	0.00999	-15000	268, 340		
MOD07/MYD07	2	Atmospheric Profiles	Water_Vapor	Total Column Precipitable Water Vapor - IR Retrieval	cm	16-bit integer	0, 20000	-9999	0.001	0	0, 20		
MOD07/MYD07	2	Atmospheric Profiles	Water_Vapor_Direct	Total Column Precipitable Water Vapor - Direct IR Retrieval	cm	16-bit integer	0, 20000	-9999	0.001	0	0, 20		
MOD07/MYD07	2	Atmospheric Profiles	Water_Vapor_Low	Precipitable Water Vapor Low - IR Retrieval	cm	16-bit integer	0, 20000	-9999	0.001	0	0, 20		
MOD07/MYD07	2	Atmospheric Profiles	Water_Vapor_High	Precipitable Water Vapor High - IR Retrieval	cm	16-bit integer	0, 20000	-9999	0.001	0	0, 20		
-	-	-	-	-									
MOD07/MYD07	2	Atmospheric Profiles	Retrieved_Temperature_Profile	Retrieved Temperature Profile	K	16-bit integer	0, 20000	-32768	0.00999	-15000	150, 350		
MOD07/MYD07	2	Atmospheric Profiles	Retrieved_Moisture_Profile	Retrieved Dew Point Temperature Profile	K	16-bit integer	0, 20000	-32768	0.00999	-15000	150, 350		
MOD07/MYD07	2	Atmospheric Profiles	Retrieved_Height_Profile	Retrieved Geopotential Height Profile	m	16-bit integer	-32500, 32500	-32768	1	-32500	0, 65000		

B. Scatterplot of PM2.5 with predictor variables

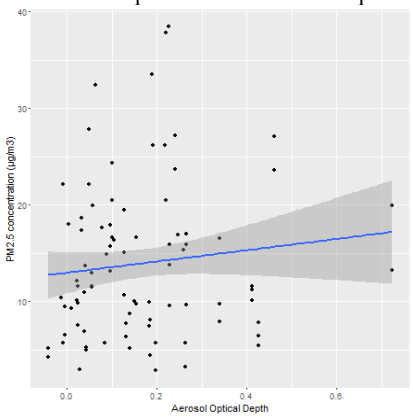


Figure 6-1 Scatterplot of AOD and PM2.5 concentrations with 90% CI across slope of a Linear Regression line

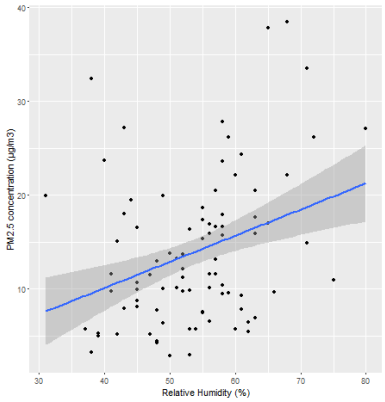


Figure 6-2 Scatterplot of RH and PM2.5 concentrations with 90% CI across slope of a Linear Regression line

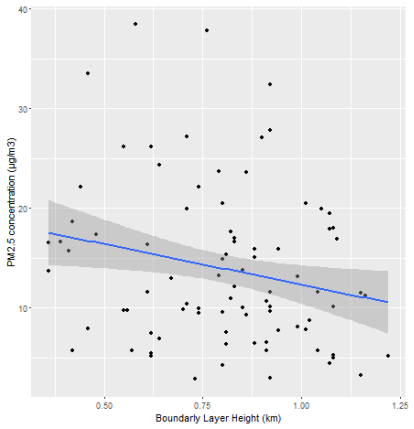


Figure 6-3 Scatterplot of BLH and PM2.5 concentrations with 90% CI across slope of a Linear Regression line

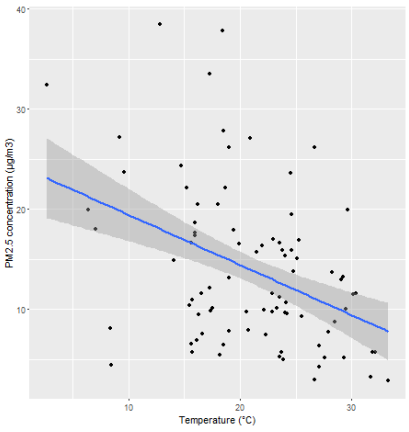


Figure 6-4 Scatterplot of Temperature and PM2.5 concentrations with 90% CI across slope of a Linear Regression line

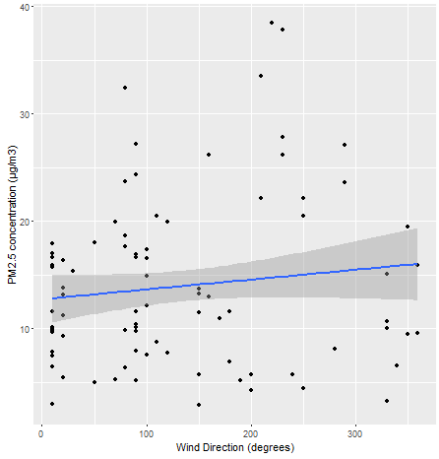


Figure 6-5 Scatterplot of wind direction and PM2.5 concentrations with 90% CI across slope of a Linear Regression line

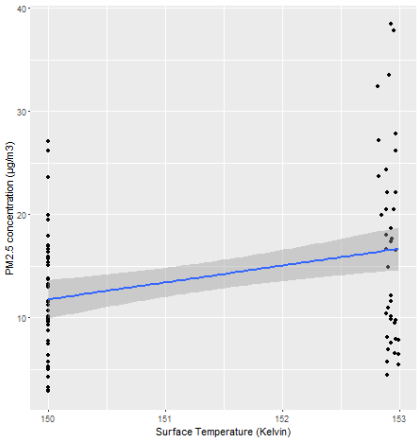


Figure 6-6 Scatterplot of Surface skin temperature and PM2.5 concentrations with 90% CI across slope of a Linear Regression line