


5-2018

Statistical Analysis and Comparison of Optical Classification of Atmospheric Aerosol Lidar Data

Kwasi Gyening Afrifa

Old Dominion University, kafri002@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/ece_deng_projects

 Part of the [Atmospheric Sciences Commons](#), and the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Afrifa, Kwasi G.. "Statistical Analysis and Comparison of Optical Classification of Atmospheric Aerosol Lidar Data" (2018). Doctor of Engineering (D Eng), doctoral_project, Electrical/Computer Engineering, Old Dominion University, DOI: 10.25777/wscd-7r39 https://digitalcommons.odu.edu/ece_deng_projects/1

This Doctoral Project is brought to you for free and open access by the Electrical & Computer Engineering at ODU Digital Commons. It has been accepted for inclusion in Electrical & Computer Engineering Projects for D. Eng. Degree by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

**STATISTICAL ANALYSIS AND COMPARISON OF OPTICAL
CLASSIFICATION OF ATMOSPHERIC AEROSOL LIDAR DATA**

by

Kwasi Gyening Afrifa

B.S. May 2005, Kwame Nkrumah University of Science & Technology, Ghana

M.S. October 2010, University of Windsor, Canada

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF ENGINEERING

ELECTRICAL AND COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY

May 2018

Approved by:

Khan M. Iftekharuddin (Director)

Xin Chunsheng (Member)

Norou Diawara (Member)

ABSTRACT

STATISTICAL ANALYSIS AND COMPARISON OF OPTICAL CLASSIFICATION OF ATMOSPHERIC AEROSOL LIDAR DATA

Kwasi Gyening Afrifa

Old Dominion University, 2018

Director: Dr. Khan M. Iftekharuddin

This dissertation presents a new study for the analysis and classification of atmospheric aerosols in remote sensing LIDAR data. Information on particle size and associated properties are extracted from these remote sensing atmospheric data which are collected by a ground-based LIDAR system. This study first considers optical LIDAR parameter-based classification methods for clustering and classification of different types of harmful aerosol particles in the atmosphere. Since accurate methods for aerosol prediction behaviors are based upon observed data, computational approaches must overcome design limitations, and also consider appropriate calibration and estimation accuracy. Consequently, two statistical methods based on generalized linear models (GLM) and regression tree techniques are used to further analyze the performance of the LIDAR parameter-based aerosol classification methods. The goal of this part of GLM and regression tree analyses is to compare and contrast distinct classification data schemes, and compare the results with the measured aerosol reflection data in the atmosphere. The detail statistical comparison and analysis show that the optical methods adopted in this study for classification and prediction of various harmful aerosol types such as soot, carbon monoxide (CO), sulfates (SO_x) and nitrates (NO_x) are effective.

Copyright, 2018, by Kwasi G. Afrifa, All Rights Reserved.

*To my family who has sustained me with love and affections and have shepherded me through all
strifes in my life, this dissertation is humbly dedicated to.*

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr. Khan M. Iftekharuddin for his guidance throughout my graduate experience and for his constructive comments and suggestions. I am truly thankful to Dr. Jiang Li and Dr. Diawara for serving on my committee. My profound appreciation goes to Dr. Mohammed Elbakary for his time, support, and advice toward the completion of my doctoral project research. I would also like to thank all my colleagues at the Old Dominion Vision Laboratory for all the nuggets of knowledge they provided to me and all the special conversations we had with each other. Lastly and most importantly, I would like to express my gratitude to my family and friends for their love, encouragement, and emotional support.

NOMENCLATURE

AAE	Absorption Angstrom Exponent
ANOVA	Analysis Of Variance
AOD	Aerosol Optical Depth
APD	Avalanche PhotoDiode
BCR	Backscatter Color Ratio
CAL	Compact Aerosol LIDAR
CALIOP	Cloud-Aerosol LIDAR with Orthogonal Polarization
CALIPSO	Cloud-Aerosol LIDAR and Infrared Pathfinder Satellite Observation
CPM	Channel PhotoMultiplier
DIAL	Differential Absorption LIDAR
EAE	Extinction Absorption Exponent
EPA	Environmental Protection Agency
ESB	Engineering Systems Building
FOV	Field Of View
FWHM	Full-Width-at-Half-Maximum
GLM	Generalized Linear Model

HSRL	High Spectral Resolution Lidar
LIDAR	LIght Detection and Ranging
MODIS	MODerate resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NEB	New Education Building
PM	Particulate Matter
SAE	Scattering Angstrom Exponent
SAS	Statistical Analysis System
SSA	Single Scattering Albedo

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	x
LIST OF FIGURES.....	xiii
 Chapter	
1. INTRODUCTION.....	1
1.1 Aerosol Measurement Techniques.....	3
1.2 Problem Statement.....	5
1.3 Contributions.....	6
1.4 Organization of the Research Project.....	7
2. BACKGROUND REVIEW.....	8
2.1 Rayleigh Scattering.....	9
2.2 Mie Scattering.....	11
2.3 Atmospheric Absorption.....	12
2.4 Lidar Equation.....	13
2.5 Analytic Solutions to the Lidar Equation.....	14
3. LIDAR EXPERIMENTAL SETUP.....	17
3.1 Chapter Overview.....	17
3.2 Description of Lidar.....	17
3.3 Methodology for Lidar Experiments.....	21
3.4 Analysis of Variance (ANOVA).....	24
3.5 Regression Tree.....	25
3.6 Application of Statistical Methods.....	27

	Page
4. LIDAR EXPERIMENTAL RESULTS.....	32
4.1 Chapter Overview.....	32
4.2 Results of ANOVA Analysis.....	32
4.3 Results of Regression Tree Analysis.....	40
4.4 Discussion of Results.....	47
5. CONCLUSIONS.....	55
5.1 Future Work.....	56
BIBLIOGRAPHY.....	58
VITA.....	64

LIST OF TABLES

Table	Page
1. Summary of LIDAR system.....	18
2. Summary of LIDAR data acquired during 2016 – 2017 on the Old Dominion University campus.....	21
3. System efficiencies and calibration constants for CAL.....	22
4. Method 1.....	28
5. Method 2.....	29
6. ANOVA for Method 1 – August 16, 2016 at ESB	32
7. ANOVA for Method 2 – August 16, 2016 at ESB	33
8. Count of Misclassification for Method 1 and proportions are given in the parentheses – August 16, 2016 at ESB.....	33
9. Count of Misclassification for Method 2 and proportions are given in the parentheses – August 16, 2016 at ESB.....	34
10. ANOVA for Method 1 – August 17, 2016 at ESB	34
11. ANOVA for Method 2 – August 17, 2016 at ESB	34
12. Count of Misclassification for Method 1 and proportions are given in the parentheses – August 17, 2016 at ESB.....	35
13. Count of Misclassification for Method 2 and proportions are given in the	

Page

parentheses – August 17, 2016 at ESB.....	35
14. ANOVA for Method 1 – July 21, 2017 at Constant Hall	36
15. ANOVA for Method 2 – July 21, 2017 at Constant Hall	36
16. Count of Misclassification for Method 1 and proportions are given in the parentheses – July 21, 2017 at Constant Hall.....	36
17. Count of Misclassification for Method 2 and proportions are given in the parentheses – July 21, 2017 at Constant Hall.....	37
18. ANOVA for Method 1 – July 21, 2017 at Rogers Hall	37
19. ANOVA for Method 2 – July 21, 2017 at Rogers Hall	37
20. Count of Misclassification for Method 1 and proportions are given in the parentheses – July 21, 2017 at Rogers Hall.....	38
21. Count of Misclassification for Method 2 and proportions are given in the parentheses – July 21, 2017 at Rogers Hall.....	38
22. ANOVA for Method 1 – July 21, 2017 at NEB	39
23. ANOVA for Method 2 – July 21, 2017 at NEB	39
24. Count of Misclassification for Method 1 and proportions are given in the parentheses – July 21, 2017 at NEB.....	39
25. Count of Misclassification for Method 2 and proportions are given in the parentheses – July 21, 2017 at NEB.....	40

26. Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – August 16, 2016 at ESB	41
27. Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses– August 16, 2016 at ESB	41
28. Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – August 17, 2016 at ESB	42
29. Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses – August 17, 2016 at ESB	43
30. Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – July 21, 2017 at Constant Hall	44
31. Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses – July 21, 2017 at Constant Hall	44
32. Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – July 21, 2017 at Rogers Hall	45
33. Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses – July 21, 2017 at Rogers Hall	46
34. Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – July 21, 2017 at NEB	47
35. Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses – July 21, 2017 at NEB.....	47

LIST OF FIGURES

Figure	Page
1. Diagram of CAL LIDAR system.....	19
2. Diagram of CAL receiver box.....	20
3. Aerosols Detected Using Method 1(July 21, 2017 Rogers Hall).....	28
4. Aerosols Detected Using Method 2(July 21, 2017 Rogers Hall).....	30
5. 3-D View of Detected Aerosols Using Method 1.....	48
6. 3-D View of Detected Aerosols Using Method 2.....	49

CHAPTER 1

INTRODUCTION

Aerosols are substances consisting of liquid and solid particles suspended in air. Aerosol particles play a central role in the atmosphere. Changes of their physical and chemical properties induces feedback mechanisms, with combined impacts ranging from air pollution and related health effects. They represent very key components in the determination of air quality. Aerosols are categorized based on their source: anthropogenic aerosols, which are produced by man-made processes; and natural aerosols, which obtained from wind uplifting of particles, erosion, or other natural processes. Aerosols are sometimes categorized as primary or secondary particles. Primary particles are particles that are emitted directly into the air, such as volcanic ash and dust. Secondary particles are formed in the atmosphere from chemical reactions of gases and subsequent particle-forming processes like condensation, coagulation and aggregation. Some examples of secondary particles are nitrates, sulfates and organic carbon. Aerosols are classified based on the aerosol diameter by the Environmental Protection Agency (EPA). Fine particles or Particulate Matter (PM) with aerosol diameter less than 2.5 micrometers are classified as PM_{2.5}, while coarse particles with aerosol diameters less than 10 micrometers are classified as PM₁₀. With the use of the best estimate, anthropogenic contributions represent 12.5% of the total aerosol budget and sulfates represent more than 42% of all anthropogenic aerosols.

It is essential to identify the types of aerosols due to the effects they have on climate, health, and the environment. Governmental agencies understand the need to deal with air pollution not only on a local basis, but on a regional basis because aerosols from an area are easily transported

to another region. Lots of people in the United States live in counties where the PM_{2.5} concentration regularly exceeded the EPA safe limits. Territorial aerosol distributions can be affected by sources such as industry and transportation, as well as the physical geography, terrain and transport of aerosols from outside the territory.

Efforts have been made to model and monitor PM_{2.5} mass concentration in urban areas by the EPA. In-situ measurements of PM_{2.5} mass concentration at point locations are monitored throughout the United States by ground stations. The Moderate resolution imaging spectroradiometer (MODIS) instrument, aboard the Aqua and Terra satellite, provides daily measurements with large spatial coverage of the aerosol optical depth (AOD), a quantity related to the aerosol column amount and the aerosol ability within the column to scatter and absorb sunlight [1]. Light Detection and Ranging (LIDAR) is a powerful tool for atmospheric aerosol profiling because it resolves the vertical distribution of an atmospheric column. Any correlation between MODIS AOD and ground station PM_{2.5} monitors will be difficult without the use of LIDAR or a similar instrument capable of measuring aerosols as a function of altitude.

The buildup of pollution in Donora near Pittsburgh, PA in 1948 and another in London, England in 1952 led to the deaths of people that could be linked to the inhalation of particulate matter and other health related issues. These pollution events led to the passage of state and federal laws like the Air Pollution Control Act of 1955 and the Clean Air Act of 1963. Small-sized aerosols, can become embedded in the lungs after inhalation. Exposure to PM_{2.5} is related to premature death, lung disease, decreased lung function, asthma attacks, irregular heartbeat, respiratory and cardiovascular disease; mainly in children, older adults, and people with pre-existing lung and heart disease [2]. Some studies also provide evidence that PM₁₀ is also associated with mortality and morbidity [3].

Aerosols are also known to impact climate and the environment by means of absorption and/or scattering of visible and infrared radiation leading to cooling or warming of the atmosphere. The size, shape and composition of the aerosols are considered to determine their effect on the earth's radiation budget, which is the balance between the incoming radiation from the sun and the outgoing longwave and shortwave radiation from the earth. Sulfates produced from the burning of fossil fuels are the biggest anthropogenic aerosol contribution to the earth's radiation budget, which cools the planet. Black carbon also produced by the burning of fossil fuels produces a warming effect. Tropospheric aerosols contain sulfates, nitrates, carbon monoxide and soot with their sizes spanning over more than four orders of magnitude, from a few nanometers to several micrometers [4].

1.1 Aerosol Measurement Techniques

Passive remote sensing provides the advantage of large spatial coverage even though it is generally less accurate and has variable vertical resolution depending on the altitude and technique. When light interacts with the aerosol particles in the atmosphere, it would be attenuated due to extinction, a combination of scattering and absorption. The optical depth, τ , through a path length, L , is defined by the relation:

$$\tau(\lambda, L) = \int_0^L \kappa(\lambda, z) dz \quad (1)$$

where λ is the wavelength, κ is the extinction coefficient, and z is the range. The extinction coefficient is a measure of light removed over the path length, either by scattering or absorption and is equivalent to

$$\kappa = \sum_j (N_{abs,j} \sigma_{abs,j} + N_{scat,j} \sigma_{scat,j}) \quad (2)$$

where $N_{abs,j}$ and $N_{scat,j}$ are the number concentrations of the various absorbing and scattering species, and $\sigma_{abs,j}$ and $\sigma_{scat,j}$ are the absorption and scattering cross sections. The passive sensors use an external source of radiation to calculate the optical depth or other atmospheric quantities.

Unlike passive sensors, active sensors provide their own source of radiation (e.g. laser) to determine atmospheric properties. They have the advantage of improved vertical resolution through the atmosphere. A LIDAR emits short, laser pulses into the atmosphere. The light emitted from the laser is scattered and absorbed by molecules and aerosols in the atmosphere. Scattered light returning to the LIDAR is collected by a telescope and transmitted to a detector. The return signal is measured as a function of time, therefore, distance from the LIDAR, and used to determine range-resolved atmospheric absorption and scattering properties [5].

Doppler LIDAR is used to measure wind speeds along the laser beam path, by determining the frequency shift of backscattered light. Differential Absorption LIDAR (DIAL) is used to determine concentrations of atmospheric gases by emitting an on-line wavelength, which is absorbed by the gas, and an off-line wavelength which is not absorbed. Raman LIDAR can be used to detect molecular species and temperature in addition to the backscatter signal by making use of Raman wavelength shifts attributable to the specific molecule being measured. An elastic backscatter LIDAR operates under the assumption that the light scattered from molecules and aerosols has the same wavelength as the emitted light. Under this assumption, a priori knowledge of the atmosphere must be assumed or obtained from an external source in order to obtain the extinction. In contrast, the High Spectral Resolution LIDAR (HSRL) resolves the molecular and

aerosol signals independently using Doppler broadening, which removes the need for a priori knowledge.

LIDAR measurements are made from various platforms. Ground based LIDAR systems provide point source data with high temporal coverage. Aircraft LIDAR can be used to determine aerosol transport with high spatial resolution. Sparse global coverage of aerosols can be obtained from spaceborne LIDAR systems such as the Cloud-Aerosol LIDAR with Orthogonal Polarization (CALIOP) aboard the Cloud-Aerosol LIDAR and Infrared Pathfinder Satellite Observation (CALIPSO) satellite LIDAR system. The LIDAR measurements in this research project come primarily from a ground based LIDAR system. The fundamental benefit of using LIDAR in atmospheric measurements is the ability to determine the vertical profile of aerosol layers. Another benefit of LIDAR aerosol measurements is the possibility for higher spatial resolution than MODIS and similar instruments. Due to the LIDAR being an active sensor, it also provides the opportunity for nighttime measurements, whereas passive sensors require a natural light source.

1.2 Problem Statement

The goal of this research project is to classify the types of aerosols found in a heavy traffic area in Hampton Roads near the campus of Old Dominion University (ODU) with the usage of the Compact Aerosol LIDAR (CAL) which is a sophisticated ground based LIDAR system. This area is chosen for the study because of its proximity to Virginia International Sea Port, where many diesel engine trucks travel on the streets near ODU campus. Several experiments are done on campus on different days and locations to obtain LIDAR data which are to be used to identify

the types of aerosols in the atmosphere. The experiments are proposed to achieve the following goals related to problem statements.

The first goal of this project is to obtain a model to classify aerosols using aerosol optical parameters. Two different methods for the classification events need to be addressed without compromising the performance of the identification procedure. Experiments performed are used to demonstrate the potency of these identification processes.

The second goal is to apply classification techniques on these procedures to help distinguish between different aerosols. Analyze classification approaches using statistical analysis for predictive behavior. This is achieved by applying Generalized Linear Model (GLM) and regression tree technique to analyze distinct classification approaches. This enables the investigation of temporal variation of recognized aerosols in the atmosphere.

In general, this would enable the combined usage of aerosol parameters to identify aerosols in the atmosphere when all aerosol intensive parameters cannot be deduced from data from our LIDAR. This new scheme would determine aerosols in the atmosphere and the most prevalent which causes pollution in the environment.

1.3 Contributions

This study first employs a sophisticated ground-based LIDAR system to acquire atmospheric aerosol reflection data over a heavy traffic area in Hampton Roads near the campus of Old Dominion University (ODU). The sophisticated ground-based LIDAR system is deployed to collect aerosol data in the atmosphere for analysis and classification of harmful aerosol particles. Two-step aerosol classification is performed using the remote sensing data. The first step involves aerosol particle type classification from the measured remote sensing data with the help

of two well-known methods in literature [6, 7]. It was observed that these two optical LIDAR parameter-based approaches did not produce the same aerosol particle classification results. Consequently, in the second step, statistical and regression analysis techniques were proposed to ascertain which of the two classification approaches should be preferred in identifying each specific aerosol. In order to accomplish this objective, GLM analysis and regression tree are then used to infer whether there are significant differences between the approaches adopted by [6] and [7] to identify the aerosols and then recommend which approach is suitable for specific aerosols from the measured LIDAR data.

1.4 Organization of the Research Project

The rest of the research project is organized as follows. Chapter 2 provides a background review which discusses the interaction of light and matter, the LIDAR equation, and LIDAR measurements. Chapter 3 describes the experimental setup, the methodology of obtaining the aerosol parameters, the review of statistical analysis tools such as regression tree and GLM which is an ANOVA technique and their application on the data. Chapter 4 presents the LIDAR measurement results and the interpretations for the results obtained after using the statistical analysis tools. Chapter 5 of this research project concludes with a summary of the results and future work.

CHAPTER 2

BACKGROUND REVIEW

The concept of LIDAR demonstrations were initially performed by the use of searchlight beams to measure air density profiles [8, 9, 10]. Following the invention of the laser, the first LIDAR was made and used to measure aerosol layers in the upper atmosphere [11]. Subsequently, LIDAR has been used to determine concentrations of aerosols and various atmospheric constituents including wind speed, temperature, water vapor, ozone, and to visualize atmospheric dynamics, and physical and chemical processes.

A basic LIDAR consists of:(i) a laser transmitter, (ii) a receiver, and (iii) a data acquisition system. The laser emits pulses into the atmosphere from the transmitter, and the light is scattered and/or absorbed. The combined effect of removing light from the beam by scattering and absorption is known as extinction. The Beer-Lambert law gives the intensity of light at a particular wavelength, I_λ , passing through an atmospheric layer with an extinction coefficient κ_λ as:

$$I_\lambda = I_{o,\lambda} \exp\left(-\int_0^L \kappa_\lambda(z) dz\right) \quad (3)$$

where $I_{o,\lambda}$ is the initial light intensity and L is the thickness of the layer.

A fraction of light from the initial beam is scattered in all directions as the pulse propagates through the atmosphere, some of which is scattered directly back to the receiver where it is

detected by a detector. The output of the detector is an electrical signal that is digitized and stored by the data acquisition system.

The strength of the interaction of light with atmospheric aerosols depends on their size, shape, type, and composition. The size parameter, α , is helpful in determining this interaction and is given by

$$\alpha = \frac{2\pi r}{\lambda} \quad (4)$$

where r is the particle radius and λ is the lidar wavelength. Particles with a size parameter much less than unity exhibit Rayleigh scattering, while a scattering parameter $0.1 \leq \alpha \lesssim 50$ indicates Mie scattering [12]. Another important quantity which determines how light scatters from aerosols is the index of refraction, m , which can be a complex number. The real part of the index of refraction, m_r , is equivalent to the ratio of the speed of light in a vacuum to the phase velocity of electromagnetic radiation in the medium, V_p ,

$$m_r = \frac{c}{V_p} \quad (5)$$

The imaginary part of the index of refraction, m_i , is related to the ability of the media to absorb electromagnetic radiation. For non-absorbing materials, m_i is equal to zero. The imaginary component also affects how the aerosol scatters radiation.

2.1 Rayleigh Scattering

The theory of Rayleigh scattering is used primarily to describe atmospheric molecular scattering; however, the principle also applies to very small aerosols. A uniform electromagnetic field exists around the aerosols due to the particles being very small in comparison to the

wavelength. The electromagnetic field produces an induced dipole in the aerosols which oscillates at the same frequency as the incident light. The dipole then reradiates this scattered light in all directions. If the incident light is unpolarized, then the intensity of the scattered light is given by:

$$I(\theta) = \frac{\pi^2(m^2 - 1)^2 I_0}{2N^2 R^2 \lambda^4} (\cos^2 \theta + 1) \quad (6)$$

where N is the number density of the scatterers and R is the distance from the particle [13]. The term $\cos^2 \theta + 1$ represents the two polarization components of the scattered light. The perpendicular component is independent of the angle, and the component parallel to the scattering plane is symmetric; thus there is equal scattering in the forward and backward directions. The Rayleigh backscattering cross section, which is a measure of the amount of scattering by one particle at $\theta = \pi$, is widely used in the LIDAR field. The backscattering cross section,

$$\sigma_m = \frac{\pi^2(m^2 - 1)^2}{N^2 \lambda^4} \quad (7)$$

has units of length squared. An important parameter used in LIDAR measurements is the molecular backscattering coefficient,

$$\beta_m = N \sigma_m^R \quad (8)$$

which determines the intensity of light scattered at $\theta = \pi$ and has units of inverse length. In addition to the symmetric nature of Rayleigh scattering, another important feature is the variation of the intensity of scattered radiation with λ^{-4} . The intensity of light scattered at shorter wavelengths is much larger than that of longer wavelengths.

2.2 Mie Scattering

As the size of the aerosols approaches the wavelength of the incident radiation, scattering will change greatly. The theory of Mie scattering is used to describe scattering by aerosols. In this process, waves from the scattered radiation will interfere constructively in some cases, and destructively in other cases. The resulting can be expressed as an infinite series of Bessel functions and Legendre polynomials and solved with the help of computers [14, 15]. For an initially unpolarized light source, the intensity of scattered light due to Mie scattering from a spherical aerosol is commonly expressed as

$$I(\theta) = \frac{\lambda^2(i_1 + i_2)}{8\pi^2 R^2} I_0 \quad (9)$$

where i_1 and i_2 are the Mie intensity parameters for the perpendicular and parallel polarization components, respectively [16]. The Mie intensity parameters are functions of the index of refraction, size parameter, and scattering angle. For small values of α , the Mie intensity parameters resemble the parallel and perpendicular polarization components of Rayleigh scattering. Unlike the case of light scattering by molecules, Mie scattering is non-symmetric. As the size parameter increases, there is an increase in forward scattering and the intensity of forward scattering becomes much greater than the intensity of back scattering. Mie scattering is less dependent on the wavelength, when compared to Rayleigh scattering. The relationship between the total scattering coefficient for an aerosol particle and the wavelength of incident light is

$$\beta_a = \frac{\text{Constant}}{\lambda^\eta} \quad (10)$$

where η is the Angstrom exponent, which is another measure of the particle size [5]. The Angstrom exponent can range from a value of 0 for large particles to a value of 4 for purely molecular scattering.

In tropospheric LIDAR applications, Mie scattering takes place mostly within the boundary layer, close to the surface or in dense aerosol or cloud layers aloft, while Rayleigh scattering dominates in the upper troposphere, at altitudes over approximately 6km, depending on the latitude.

2.3 Atmospheric Absorption

Light at certain wavelengths may be absorbed by gases and aerosols in the atmosphere. The imaginary component of the index of refraction relates to the ability of a substance to absorb radiation. Molecules can exist at only certain energy levels due to a combination of their electron configuration, rotational state, and vibrational state. Photons can be absorbed or emitted resulting in a discrete change of energy level, ΔE , as long as the photon frequency meets the condition

$$\Delta E = h\nu \quad (11)$$

where h is Planck's constant and ν is the photon frequency [13]. However, most of the common elastic scattering LIDAR wavelengths are not in the range of molecular absorption lines hence absorption is normally not an issue. The LIDAR measurements which will be presented as part of this project were conducted at wavelengths of 532 nm and 1064 nm, where absorption is negligible and where aerosol scattering dominates.

2.4 Lidar Equation

The elastic LIDAR equation defines the power returned to the LIDAR receiver, $P(z)$, and is given as a function of the distance from the transmitter, that is range, z :

$$P(z) = \frac{C}{z^2} [\beta_m(z) + \beta_a(z)] T^2 \quad (12)$$

where C is the calibration constant, β_m and β_a are the backscattering coefficients for molecules and aerosols, respectively, and T^2 is the two-way transmittance. The calibration constant can be expressed as

$$C = P_0 \frac{c\tau}{2} A\eta \quad (13)$$

P_0 is the power of a single laser pulse, τ is the temporal pulse width, A is the area of the lidar telescope, and η is the overall system efficiency which includes the efficiency of all transmitter and receiver optics, as well as the detection efficiency. The two-way transmittance is given by

$$T^2 = \exp\left(-2 \int_0^z [\kappa_m(z') + \kappa_a(z')] dz'\right) \quad (14)$$

where κ_m and κ_a are the molecular and aerosol extinction coefficients, respectively. Equation 12 assumes single scattering. Multiple scattering can occur in optically dense atmospheres (e.g. fog or clouds) when forward scattered light, through subsequent scattering events, returns to the LIDAR receiver. This effect is reduced by narrowing the beam divergence and receiver Field Of View (FOV). Also, scattered light from the solar and lunar background can contribute to additional power detected by the LIDAR; however, most of this signal is removed by

background subtraction during data processing, as well as by using narrow-band interference filters in front of the detectors and properly baffled receiving optics.

2.5 Analytic Solutions to the Lidar Equation

The LIDAR equation cannot be solved for an elastic backscatter because there are too many unknowns. However, by using a priori knowledge from independent sources, a solution can be obtained. The molecular properties, β_m and κ_m can be obtained from atmospheric soundings or look-up tables. Then a number of analytic methods [17, 18, 19] can be used to determine the solutions for β_a and κ_a .

Two useful quantities for obtaining a solution to the LIDAR equation are the molecular extinction-to-backscatter ratio,

$$S_m = \frac{\kappa_m}{\beta_m} = \frac{8\pi}{3} sr \quad (15)$$

And the aerosol extinction-to-backscatter ratio,

$$S_a = \frac{\kappa_a}{\beta_a} \quad (16)$$

which depend on the size, shape, and composition of the aerosol particles, as well as the wavelength of the incident light. In actuality, due to the heterogeneity of aerosols, S_a has been shown to vary with altitude [20]; however, due to inherent limitations with elastic backscatter LIDAR, it is assumed to have a constant value within an atmospheric column. The aerosol extinction-to-backscatter ratio is commonly referred as the LIDAR ratio. In practice, a priori knowledge is used to determine the value of S_a or independent measurements of the aerosol optical depth (AOD) are used to constrain its value.

Substituting the values of S_m and S_a into Equation 12,

$$P(z)z^2 = C[\beta_m(z) + \beta_a(z)] \exp\left(-2 \int_0^z [S_m \beta_m(z') + S_a \beta_a(z')] dz'\right) \quad (17)$$

and then by introducing the parameters,

$$X(z) = P(z)z^2 \quad (18)$$

$$Y(z) = S_a[\beta_m(z) + \beta_a(z)] \quad (19)$$

the LIDAR equation can be rewritten as

$$X(z) = \frac{C}{S_a} Y(z) \exp\left(-2 \int_0^z [S_m \beta_m(z') + Y(z') - S_a \beta_a(z')] dz'\right) \quad (20)$$

Rearranging the terms gives

$$S_a X(z) \exp[-2(S_a - S_m) \int_0^z \beta_m(z') dz'] = C Y(z) \exp[-2 \int_0^z Y(z') dz'] \quad (21)$$

Logarithms are taken on both sides of the equation and differentiated with respect to z ,

$$\frac{d \ln(S_a X(z) \exp[-2(S_a - S_m) \int_0^z \beta_m(z') dz'])}{dz} = \frac{1}{Y(z)} \frac{dY(z)}{dz} - 2Y(z) \quad (22)$$

The left hand side is a function of z containing values that are all either known or assumed, so

Equation 21 is in the form of the Bernoulli equation. Using the boundary condition

$$Y(z^*) = S_a[\beta_m(z^*) + \beta_a(z^*)] \quad (23)$$

where z^* is some reference altitude where the value of β_a can be assumed, the LIDAR equation has the solution

$$\beta_a(z) = \frac{X(z)\exp[-2(S_a - S_m) \int_{z^*}^z \beta_m(z')dz']}{\frac{X(z^*)}{\beta_m(z^*) + \beta_a(z^*)} - 2S_a \int_{z^*}^z X(z')\exp[-2(S_a - S_m) \int_{z^*}^{z'} \beta_m(z'')dz'']dz' - \beta_m(z)} \quad (24)$$

We chose an altitude of z^* where aerosol scattering is negligible compared to molecular scattering, thus $\beta_a(z^*) = 0$ and then used iterative procedure to determine the value of $\beta_a(z)$ at all other altitudes. Similarly, the measured LIDAR backscatter profile can be normalized with respect to the computed molecular backscatter profile to determine the scattering ratio. A scattering ratio of unit indicates pure molecular scattering, and values above unity indicate the presence of clouds or aerosols.

After the determination of $\beta_a(z)$, the aerosol extinction, κ_a , and LIDAR derived AOD, τ , can be determined using the equations:

$$\kappa_a(z) = S_a\beta_a(z) \quad (25)$$

and

$$\tau = \int_0^{z^*} \kappa_a(z)dz \quad (26)$$

CHAPTER 3

LIDAR EXPERIMENTAL SETUP

3.1 Chapter Overview

This chapter discusses the design and protocol for the experiments conducted in this report examining the analysis of classification of aerosols. The system was setup on the Old Dominion University campus in order for the LIDAR to have an unimpeded access to the skies. The ground based LIDAR is able to provide us the temporal variation of LIDAR measurements. In this chapter, the LIDAR system used in this project report will be described in Section 3.2. The methodology for conducting LIDAR experiments and calculating parameters are described in Section 3.3. The ANOVA method to analyze the classified results will be explained in Section 3.4 and the regression tree technique to analyze the classified results will also be explained in Section 3.5.

3.2 Description of LIDAR

The LIDAR used for the experiments is the Compact Aerosol LIDAR (CAL) system which was built in the Science Directorate at NASA Langley Research Center and is described in [21]. A summary of the key system characteristics is shown in Table 1. The LIDAR system is mounted on an aluminum frame as shown in Figure 1. The LIDAR system has a mass of 115 kg and dimensions of 108 cm x 64 cm x 53 cm. The laser (Big Sky/Quantel CFR 200) is a frequency doubled Nd:AG, 20 Hz, with 1.5 mrad divergence output. The 532 nm (80 mJ,

nominal) and residual 1064 nm (60 mJ, nominal) pulses are transmitted into the atmosphere using a steerable 45° turning mirror.

The aerosol LIDAR receiver uses a 30.5 cm diameter (f/2) telescope using a parabolic mirror with a 1.6 mrad FOV. A 1 mm diameter optical fiber is mounted at the focal point of the

TABLE 1: Summary of LIDAR system

Laser			
	Manufacturer	Big Sky (Quantel)	
	Repetition rate	20 Hz	
	Beam divergence	1.5 mrad	
	Pulse width	11 ns, FWHM	
	1064 nm energy	60 mJ	
	532 nm energy	80 mJ	
Telescope			
	Diameter	0.30 m	
	FOV	1.6 mrad	
Filters			
	1064 nm	1.0 nm, FWHM (66.5% T)	
	532 nm	0.5 nm, FWHM (78% T)	
Detectors			
	1064 nm channel	APD (analog)	Perkin Elmer (C30955E)
	532 nm channels	CPM (analog)	Perkin Elmer (MH943)
		CPM (photon counting)	Perkin Elmer (MP943)

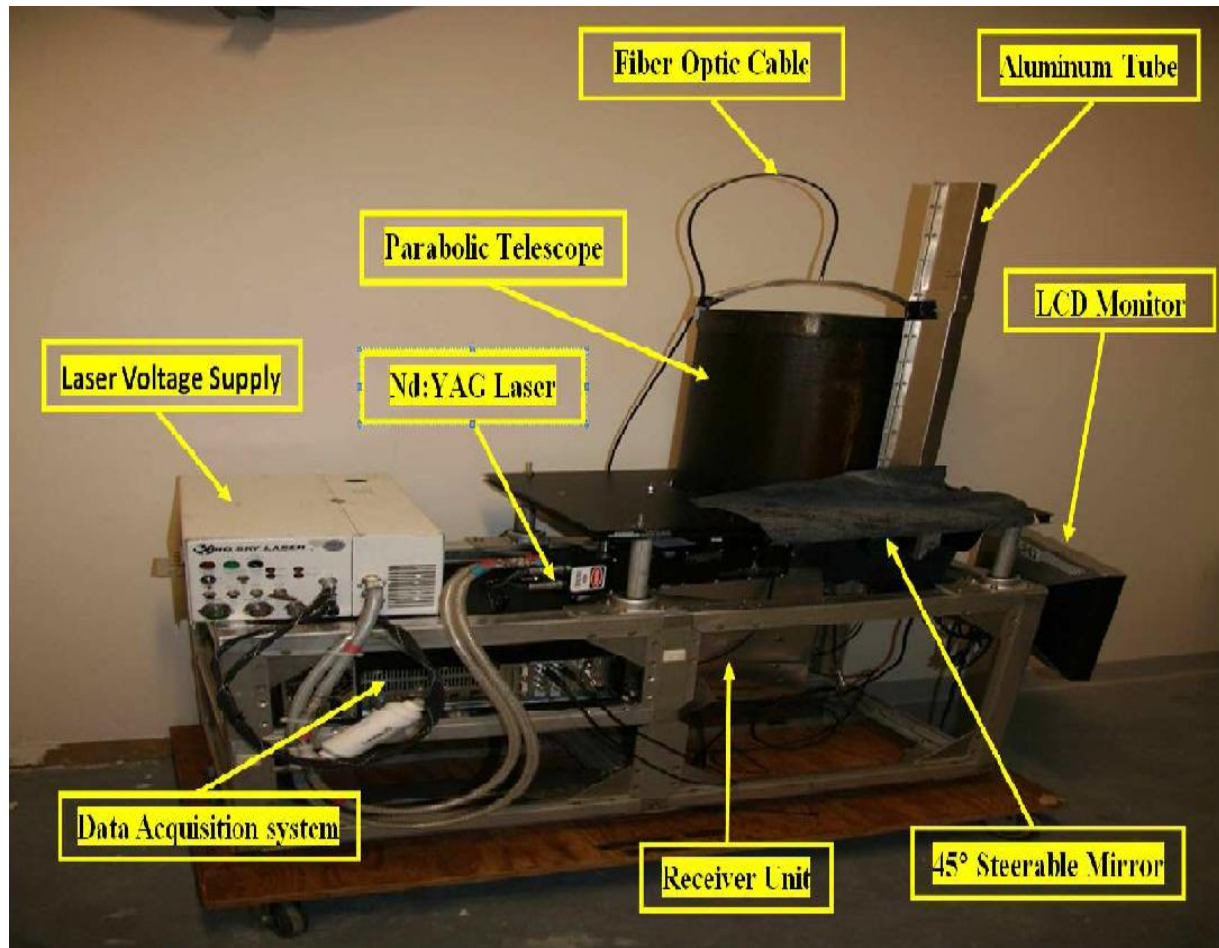


FIG. 1: Diagram of CAL LIDAR system

telescope mirror which passes the light into the receiver box as shown in Figure 2 [22]. The received light is collimated and split into 532 nm and 1064 nm channels, with a further split of the 532 nm channel into photon counting (10%) and analog (90%) signal channels. The 1064 nm signal passes through a 66.5%T, 1 nm Full-Width-at-Half-Maximum (FWHM) interference filter, focuses onto an APD detector (Perkin Elmer C30955E), and is then amplified (Femto DHPVA-100). The 532 nm signal passes through a 78%T, 0.5 nm FWHM interference filter

before being split into separate analog (CPM Perkin Elmer MH943) and photon counting channels (CPM Perkin Elmer MP943).

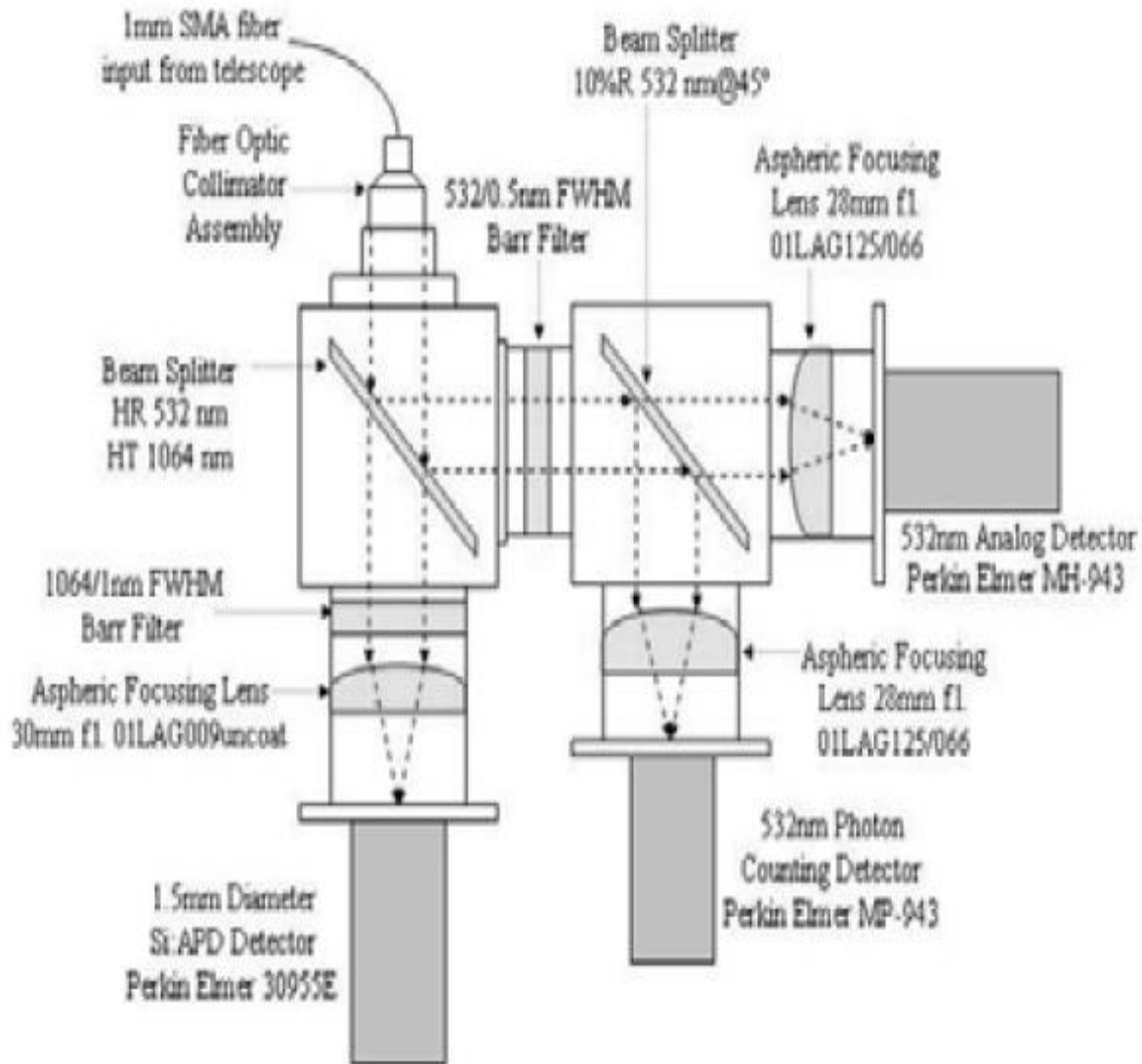


FIG. 2: Diagram of CAL receiver box

The 532 nm and 1064 nm analog channels are filtered electro-optically by a 1.5 MHz filter and sampled at a rate of 5MHz using a 14-bit digitizer (Gage Applied Technologies CS1450).

The data files are then averaged over 2 seconds intervals and stored on the computer using a

custom LabView program. The 532 nm photon counting channel is sent to a multi-channel scaler (Ortec MCS-pci) where typical integration times of 10s (200 ns dwell) are used before storing the waveform.

3.3 Methodology for Lidar Experiments

Table 2 shows the summary of the collected data during this study. The LIDAR data collection was conducted during the daytime. The data collections were performed at the same location on August 16, 2016 and August 17, 2016; and on July 21, 2017 at various locations on the Old Dominion University campus.

TABLE 2: Summary of LIDAR data acquired during 2016-2017 on the Old Dominion University campus

Experiment Date	Location	Time	Average Outside Temperature	Weather Conditions
August 16, 2016	ESB	11:04 - 12:14	88°F	Mostly Sunny skies
August 17, 2016	ESB	12:46 – 13:59	90°F	Clear Sunny skies
July 21, 2017	Constant Hall	13:42 – 13:57	93°F	Sunny skies
July 21, 2017	Rogers Hall	15:34 – 15:49	91°F	Sunny skies
July 21, 2017	NEB	16:45 – 17:00	91°F	Sunny skies

At altitudes above 5km, noise contamination exceeds the backscattering signal for the two 532nm channels. The theoretical LIDAR power was obtained using Equation 17 with $\beta_a = 0$. The calibration constant Equation 13 was calculated using values from Table 1 and the overall system efficiency which was taken as the product of the receiver efficiency given in Section 3.2 and the quantum efficiency of the detector. The values for the overall system efficiencies and resulting calibration constants for each LIDAR channel is given in Table 3. The molecular

backscattering coefficient was calculated using a model for the molecular number density found in [5].

TABLE 3: System efficiencies and calibration constants for CAL

Channel	System efficiency	Calibration constant ($\times 10^3, Wm^3$)
1064nm	17.6%	112
532nm analog	6.1%	52.1
532nm photon counting	0.76%	6.44

3.3.1 Calculation of Aerosol Optical Parameters for Method 1

The first classification method as described in [6] used intensive parameters of aerosol (LIDAR ratio and backscatter color ratio) which vary with aerosol type. The LIDAR ratio, S_a is calculated on the 532nm channel as the ratio of extinction coefficient to the backscatter coefficient. The expression for S_a is as follows:

$$S_a = \frac{\kappa_a^{532}}{\beta_a^{532}} \quad (27)$$

where S_a is the LIDAR ratio, κ_a^{532} is the aerosol extinction coefficient at 532nm and β_a^{532} is the backscatter coefficient at 532nm. The aerosols have optical parameters, such as LIDAR ratio, which varies with aerosol size, shape and composition. Aerosols found in the atmosphere have low values of LIDAR ratios for coarse mode particles and higher LIDAR ratios for small and highly absorbing mode particles [6]. Another parameter that is used was backscatter color ratio. The backscatter color ratio is defined as the ratio of backscattering coefficient at 532nm to 1064nm. This is expressed as:

$$BCR = \frac{\beta_a^{532}}{\beta_a^{1064}} \quad (28)$$

where BCR is the backscatter color ratio, β_a^{532} is the aerosol backscatter coefficient at 532nm and β_a^{1064} is the aerosol backscatter coefficient at 1064nm. Backscatter color ratios are inversely related to aerosol particle sizes [23, 24].

3.3.2 Calculation of Spectral Parameters for Method 2

The second classification method to identify aerosols in the atmosphere is done based on their spectral optical properties which are absorption and scattering. The second scheme is mostly described by [7]. We calculate the parameters SAE, AAE, EAE and SSA to be combined to identify the presence of aerosols from the data obtained from our LIDAR. The four parameters were obtained through the following equations. To obtain the extinction Angstrom exponent, the equation used is

$$EAE = -\frac{\ln\left(\frac{\kappa_{532}}{\kappa_{1064}}\right)}{\ln\left(\frac{532}{1064}\right)} \quad (29)$$

where κ_{532} is the extinction coefficient at 532nm and κ_{1064} is the extinction coefficient at 1064nm. For the scattering Angstrom exponent, the next equation is used:

$$SAE = -\frac{\ln\left(\frac{\beta_{532}}{\beta_{1064}}\right)}{\ln\left(\frac{532}{1064}\right)} \quad (30)$$

where β_{532} is the backscatter coefficient at 532nm and β_{1064} is the backscatter coefficient at 1064nm. The sum of absorption and scattering is the extinction. Absorption coefficient decreases monotonically with wavelength and it is approximated by a power law expression which is

described by an absorption Angstrom exponent (AAE) [25]. From equation below we obtain the absorption Angstrom exponent

$$AAE = - \frac{\ln\left(\frac{\gamma_{532}}{\gamma_{1064}}\right)}{\ln\left(\frac{532}{1064}\right)} \quad (31)$$

where γ_{532} is the absorption coefficient at 532nm and γ_{1064} is the absorption coefficient at 1064nm. The single scattering albedo is found as the ratio of the backscatter coefficient to the extinction coefficient

$$SSA = \frac{\beta_{532}}{\beta_{532} + \gamma_{532}} = \frac{\beta_{532}}{\kappa_{532}} \quad (32)$$

3.4 Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a statistical method used to test differences between multiple means of data [26]. The inferences about the means are made by analyzing variance. This statistical method for making simultaneous comparisons among multiple means yields values that can be tested to build classification and test significant relationship between variables. The Generalized Linear Models (GLM) procedure is the proposed analysis method adopted [27, 28, 29]. We modeled the identified aerosols and activity profiles using such procedure. This model is based on the fact that the intensity is a function of many sources and nuisance or errors. A mixed effect model was first used. Such a model can be described as follows:

$$Y_{ij} = \mu + \alpha_i + \delta_j + \epsilon_{ij} \quad (33)$$

where Y_{ij} corresponds to the recorded aerosol at the i – th altitude level and at the j – th duration, μ is the overall intensity, α_i represents the effect due to altitude, and δ_j represents the effect due to duration for each method, ϵ_{ij} represents the error terms, j defined as the number of altitude levels with $j = 1, 2, \dots, 128$. i defines the number of aerosols profile measurements in a given duration $i = 1, 2, \dots, n_k$ where $k = 1, 2, 3, 4$, and 5 denotes the locations of data collection and n_k is number of aerosols profile measurements in location k . Thus $n_1 = 63$ at location 1 (Rogers Hall July 21, 2017), $n_2 = 61$ at location 2 (Constant Hall July 21, 2017), $n_3 = 60$ at location 3 (NEB July 21, 2017), $n_4 = 286$ at location 4 (ESB August 16, 2016) and $n_5 = 251$ at location 5 (ESB August 17, 2016). The model can be expressed as

$$Y = X\varphi + \varepsilon \quad (34)$$

where Y is the vector of measured intensity by altitude and duration, X is the design matrix, φ is the vector of regression coefficients and ε is the vector of error terms. Assumptions are made that the errors are normally distributed. Estimation of the parameter set φ is such that

$$\hat{\varphi} = (X'X)^{-1}X'Y \quad (35)$$

assuming that the design matrix X is invertible. If not, using the generalized equation will allow us to obtain solution, even though they will not be unique. Normality is a strong assumption made and transformation techniques may be considered to achieve such assumptions and reduce biases.

3.5 Regression Tree Analysis

Although ANOVA is typically the first choice for prediction, an alternative method to perform prediction is regression tree, which was first introduced by [30]. The data here consists of the two

predictors altitude and duration and the response aerosol values. Suppose the data has N observations: that is, (\mathbf{x}_i, y_i) for $i = 1, 2, \dots, N$, where $\mathbf{x}_i = (x_{i1}, x_{i2})'$, x_{i1} is the i th observation from the first predictor, x_{i2} is the i th observation from the second predictor, and y_i is the i th observation from the response. To grow the tree, the algorithm needs to decide on the important variables (predictors) and splits points. If the predictors are equally important, the choice would be arbitrary. Now, suppose the data is partitioned into M regions R_1, \dots, R_M and the response is modeled as a constant c_j for $j = 1, \dots, M$:

$$f(x) = \sum_{j=1}^M c_j I(x \in R_j) \quad (36)$$

If the adopted criterion is minimizing the sum of squares $\sum (y_i - f(x_i))^2$, it can be shown that the best choice of c_j is the average of y_i in R_j . That is,

$$\hat{c}_j = \text{ave}(y_i | x_i \in R_j) \quad (37)$$

Therefore, the sum of square errors for a tree is

$$S = \sum_{j=1}^M \sum_{i=1}^{n_j} (y_i - \hat{c}_j)^2 \quad (38)$$

For $M = 2$, an algorithm suggested by [31] stated that starting with all of the data, consider a splitting variable k and a split point s , and define the pair of the half-planes

$$R_1(k, s) = \{x | x_k \leq s\}, \quad R_2(k, s) = \{x | x_k > s\} \quad (39)$$

Then choose the variable k and split point s that minimizes

$$\sum_{x_i \in R_1(k,s)} (y_i - \hat{c}_1)^2 + \sum_{x_i \in R_2(k,s)} (y_i - \hat{c}_2)^2 \quad (40)$$

After finding k and s , the data is partitioned into the two resulting regions. The same splitting process is repeated again on each of the two regions, then on all the resulting regions until the region has values that are the same.

The above model will be trained on a sub-sample of the data known as the training sample. Then, it will be tested on a testing sample, which is what is left of the data after selecting the training sample. Typically, the training sample takes 75% of the data, and of course the testing sample takes the rest, i.e. 25% of the data.

3.6 Application of Statistical Methods

After using the two methods to obtain classified aerosols in the atmosphere, the results from the two methods are compared to identify the differences or similarities of the aerosols. Since sampling methodology of data collection is complex and classification can be misleading, initiatives are employed in statistical analysis to extract the most reliable information from data through the model and its parameters. We use GLM of the ANOVA technique and regression tree as the statistical analysis tools. To validate the difference in the two classification schemes in aerosol, GLM and regression tree models are considered. The analysis is performed by assigning classification labels or counts of 1, 2, 3, 4 and 5 to the aerosols - NO_x , SO_x , CO, soot and No Aerosol found in our remote sensing data, respectively. For the aerosol described as No Aerosol, it means there is the absence of the four originally described aerosols. Due to the dominance of aerosol value of 5, we decided to remove it from the model and build predictions on the remaining data. The first classification using the calculated LIDAR ratio and backscatter color

ratio is done by assigning a particular aerosol to a location and duration when the aerosol optical parameter condition for that aerosol is met as seen from Table 4.

TABLE 4: Method 1

Aerosol	Optical Parameter	
	S_a	BCR
NO _x	70 - 80	$3.3 \pm 10\%$
SO _x	70 - 100	$3.3 \pm 15\%$
CO	43 - 52	$0.7 \pm 10\%$
Soot	60 - 65	$1.4 \pm 10\%$

The classified aerosols from our LIDAR using Method 1 is shown in Fig.3 below.

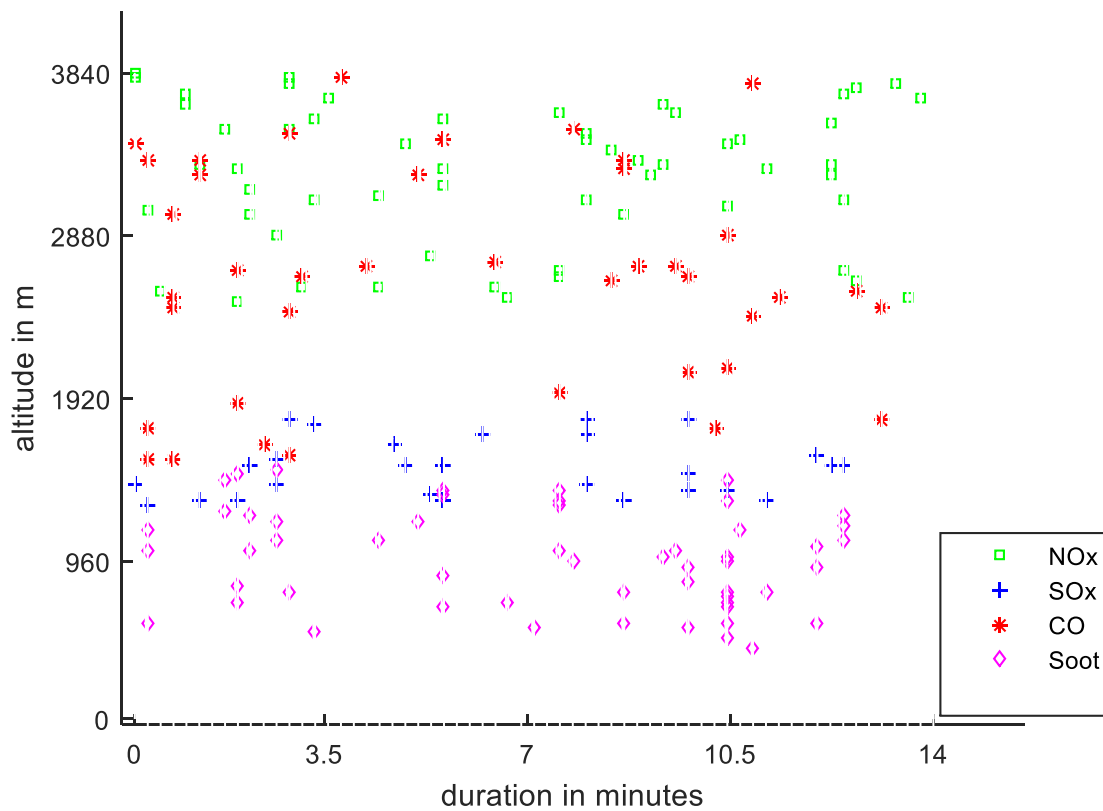


FIG.3: Aerosols Detected Using *Method 1* (July 21, 2017 Rogers Hall)

The second classification method was done by using SSA, SAE, AAE and EAE as our parameters. From Table 5, the aerosols were assigned when the parameter conditions for that particular aerosol were met.

TABLE 5: Method 2

Aerosol	Spectral Parameter			
	<i>SSA</i>	<i>AAE</i>	<i>EAE</i>	<i>SAE</i>
NO _x	> 0.85	> 2.5	1.8- 2	1.5-3.5
SO _x	> 0.95	≈ 2	1.5– 1.9	0.5-3
CO	< 0.85	< 2	1.75- 2.1	1-3
Soot	< 0.8	< 1.5	< 2	≈ 4

The classified aerosols from our LIDAR using method 2 is shown in Fig.4. below.

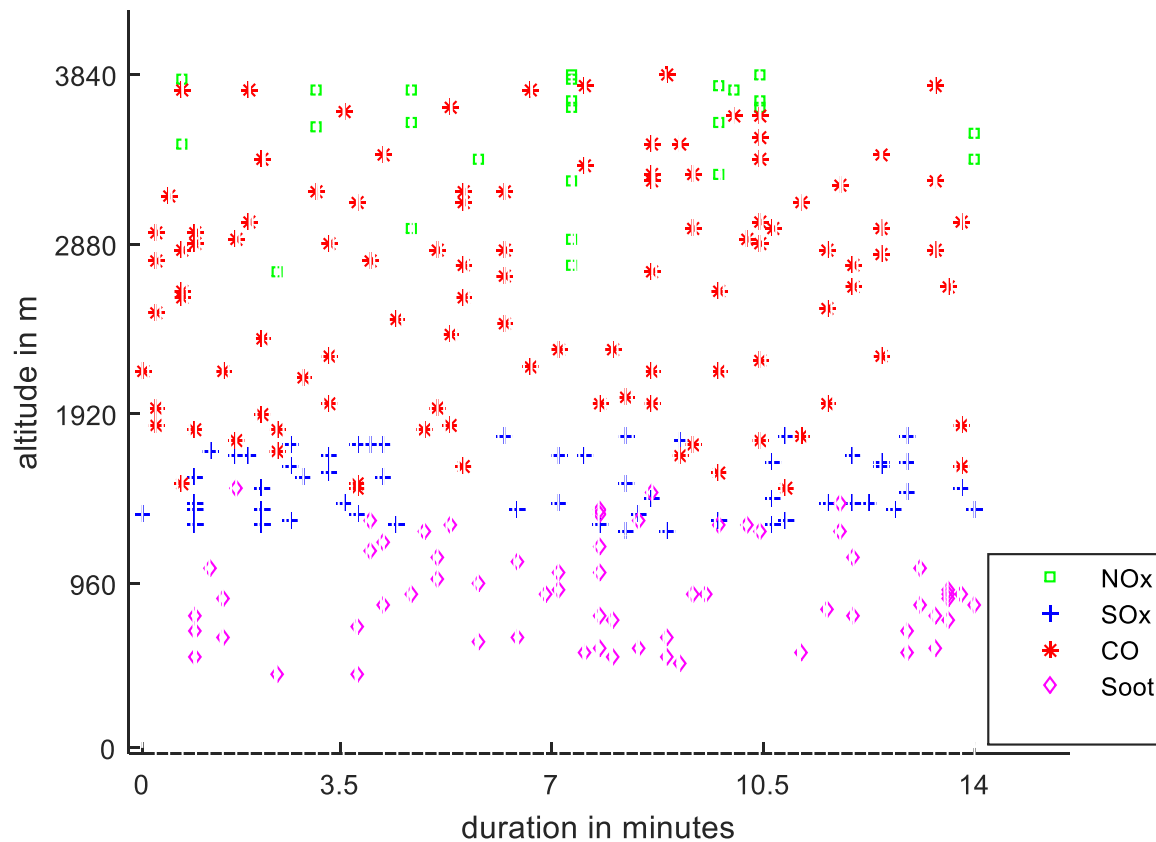


FIG. 4: Aerosols Detected Using *Method 2* (July 21, 2017 Rogers Hall)

The plots of the aerosol distribution show a wide and scattered representation. Identification and estimation of aerosols is highly unpredictable because there is no clear separation based on duration and altitude.

Although standard algorithms are used, the analysis encounters major challenges in scaling up to massive datasets with most of them being at aerosol value 5. Because of that, thinning is applied to the data, and values of 5 have to be removed. The predictions are made on the remaining observations. Even though some of the observations can be 5, the predictions will be made on the first four values of aerosols. The GLM is extended to both methods of classification. GLM

analysis is performed to find the significance of altitude and duration. The analysis is performed using the SAS 9.3®software [32].

CHAPTER 4

LIDAR EXPERIMENTAL RESULTS

4.1 Chapter Overview

In this chapter, the LIDAR data analysis for this study will be presented. The measurements taken by this ground-based Compact Aerosol LIDAR (CAL) pointed to the atmosphere is used to compute optical parameters and then used to identify aerosols. Two classification methods described are then analyzed by the ANOVA approach and Regression Tree. In Section 4.2, the results of the ANOVA analysis will be shown. Also in Section 4.3, the results of Regression Tree analysis will be shown. There will then be a discussion of the results in Section 4.4.

4.2 Results of ANOVA Analysis

The ANOVA analyses for the classification of aerosols using different methods for the various dates and locations are shown in this section.

4.2.1 August 16, 2016 at ESB

The ANOVA tables for each of the method types are displayed in Tables 6 and 7.

TABLE 6: ANOVA for Method 1 – August 16, 2016 at ESB

Source	DF	SS	MS	F	P value
Altitude	1	494.97	494.97	821.33	<0.0001
Duration	1	0.79	0.79	1.31	0.2518
Error	2381	1434.9	0.60	X	X
Total	2383	1930.66	X	X	X

TABLE 7: ANOVA for Method 2 – August 16, 2016 at ESB

Source	DF	SS	MS	F	P value
Altitude	1	154.59	154.59	298.71	<0.0001
Duration	1	0.05	0.05	0.09	0.764
Error	964	498.89	0.518	X	X
Total	966	653.53	X	X	X

The classification/count of misclassification based on the ANOVA analysis using the first method is shown in Table 8.

TABLE 8: Count of Misclassification for Method 1 and proportions are given in the parentheses – August 16, 2016 at ESB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	162 (100)	0 (0)	0 (0)	0 (0)	0 (0)	162
	2	0 (0)	609 (100)	0 (0)	0 (0)	0 (0)	609
	3	0 (0)	55 (6)	752 (82.1)	109 (11.9)	0 (0)	916
	4	0 (0)	0 (0)	0 (0)	697 (100)	0 (0)	697
Total	162	664	752	806	0	2384	

The classification/count of misclassification based on the ANOVA analysis for the second method is presented in Table 9.

TABLE9: Count of Misclassification for Method 2 and proportions are given in the parentheses
– August 16, 2016 at ESB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	51 (70.8)	21 (29.2)	0 (0)	0 (0)	0 (0)	72
	2	0 (0)	157 (100)	0 (0)	0 (0)	0 (0)	157
	3	0 (0)	0 (0)	374 (71.6)	148 (28.4)	0 (0)	522
	4	0 (0)	0 (0)	0 (0)	216 (100)	0 (0)	216
Total		51	178	374	364	0	967

4.2.2 August 17, 2016 at ESB

The ANOVA tables for each of the method types are displayed in Tables 10 and 11.

TABLE 10: ANOVA for Method 1 – August 17, 2016 at ESB

Source	DF	SS	MS	F	P value
Altitude	1	401.02	401.02	828.81	<0.0001
Duration	1	0.87	0.87	1.81	0.1792
Error	1814	877.71	0.48	X	X
Total	1816	1279.60	X	X	X

TABLE 11: ANOVA for Method 2 – August 17, 2016 at ESB

Source	DF	SS	MS	F	P value
Altitude	1	164.98	164.98	451.64	<0.0001
Duration	1	0.95	0.95	2.60	0.1071
Error	1007	427.90	0.65	X	X
Total	1008	593.83	X	X	X

The classification/count of misclassification based on the ANOVA analysis using the first method is shown in Table 12.

TABLE 12: Count of Misclassification for Method 1 and proportions are given in the parentheses – August 17, 2016 at ESB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	12 (92.3)	0 (0)	0 (0)	0 (0)	1 (7.7)	13
	2	0 (0)	481 (100)	0 (0)	0 (0)	0 (0)	481
	3	1 (0.2)	50 (9.4)	481 (90.4)	0 (0)	0 (0)	532
	4	0 (0)	0 (0)	1 (0.1)	790 (99.9)	0 (0)	791
Total	13	531	482	790	1	1817	

The classification/count of misclassification based on the ANOVA analysis for the second method is presented in Table 13.

TABLE 13: Count of Misclassification for Method 2 and proportions are given in the parentheses – August 17, 2016 at ESB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	9 (90)	1 (10)	0 (0)	0 (0)	0 (0)	10
	2	0 (0)	157 (100)	0 (0)	0 (0)	0 (0)	157
	3	0 (0)	16 (3.4)	441 (94.4)	10 (2.1)	0 (0)	467
	4	0 (0)	0 (0)	1 (0.3)	375 (99.7)	0 (0)	376
Total	9	174	442	385	0	1010	

4.2.3 July 21, 2017 at Constant Hall

The ANOVA tables for each of the method types are displayed in Tables 14 and 15.

TABLE 14: ANOVA for Method 1 – July 21, 2017 at Constant Hall

Source	DF	SS	MS	F	P value
Altitude	1	59.27	59.27	91.69	<0.0001
Duration	1	1.26	1.26	1.96	0.1624
Error	600	387.88	0.646	X	X
Total	602	448.41	X	X	X

TABLE 15: ANOVA for Method 2 – July 21, 2017 at Constant Hall

Source	DF	SS	MS	F	P value
Altitude	1	45.254	45.254	64.36	<0.0001
Duration	1	0.716	0.716	1.0181	0.3136
Error	379	155.73	0.63	X	X
Total	381	201.70	X	X	X

The classification/count of misclassification based on the ANOVA analysis using the first method is shown in Table 16.

TABLE 16: Count of Misclassification for Method 1 and proportions are given in the parentheses – July 21, 2017 at Constant Hall

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	36 (63.2)	21 (36.8)	0 (0)	0 (0)	0 (0)	57
	2	0 (0)	138 (100)	0 (0)	0 (0)	0 (0)	138
	3	0 (0)	0 (0)	293 (99.7)	1 (0.3)	0 (0)	294
	4	0 (0)	0 (0)	0 (0)	114 (100)	0 (0)	114
Total	36	159	293	115	0	603	

The classification/count of misclassification based on the ANOVA analysis for the second method is presented in Table 17.

TABLE 17: Count of Misclassification for Method 2 and proportions are given in the parentheses – July 21, 2017 at Constant Hall

True Value	Predicted Value					Total
	1	2	3	4	5	
1	46 (100)	0 (0)	0 (0)	0 (0)	0 (0)	46
2	0 (0)	102 (100)	0 (0)	0 (0)	0 (0)	102
3	0 (0)	7 (4.2)	159 (95.8)	0 (0)	0 (0)	166
4	0 (0)	0 (0)	0 (0)	68 (100)	0 (0)	68
Total	46	109	159	68	0	382

4.2.4 July 21, 2017 at Rogers Hall

The ANOVA tables for each of the method types are displayed in Tables 18 and 19.

TABLE 18: ANOVA for Method 1 – July 21, 2017 at Rogers Hall

Source	DF	SS	MS	F	P value
Altitude	1	134.17	134.17	206.00	<0.0001
Duration	1	0.26	0.26	0.39	0.53
Error	657	427.90	0.65	X	X
Total	659	562.33	X	X	X

TABLE 19: ANOVA for Method 2 – July 21, 2017 at Rogers Hall

Source	DF	SS	MS	F	P value
Altitude	1	56.05	56.05	88.54	<0.0001
Duration	1	0.46	0.46	0.73	0.39
Error	246	155.73	0.63	X	X
Total	248	212.24	X	X	X

The classification/count of misclassification based on the ANOVA analysis for the first method is shown in Table 20.

TABLE 20: Count of Misclassification for Method 1 and proportions are given in the parentheses – July 21, 2017 at Rogers Hall

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	58 (87.9)	8 (12.1)	0 (0)	0 (0)	0 (0)	66
	2	0 (0)	148 (100)	0 (0)	0 (0)	0 (0)	148
	3	0 (0)	17 (6.1)	242 (87.4)	18 (6.5)	0 (0)	277
	4	0 (0)	0 (0)	0 (0)	144 (85.2)	25 (14.8)	169
Total		58	173	242	162	25	660

Also, the classification/count of misclassification based on the ANOVA analysis for the second method is presented in Table 21.

TABLE 21: Count of Misclassification for Method 2 and proportions are given in the parentheses – July 21, 2017 at Rogers Hall

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	25 (100)	0 (0)	0 (0)	0 (0)	0 (0)	25
	2	0 (0)	55 (100)	0 (0)	0 (0)	0 (0)	55
	3	0 (0)	4 (3.8)	93 (88.6)	8 (7.6)	0 (0)	105
	4	0 (0)	0 (0)	0 (0)	49 (76.6)	15 (23.4)	64
Total		25	59	93	57	15	249

4.2.5 July 21, 2017 at NEB

The ANOVA tables for each of the method types are displayed in Tables 22 and 23.

TABLE22: ANOVA for Method 1 – July 21, 2017 at NEB

Source	DF	SS	MS	F	P value
Altitude	1	139.85	139.85	248.33	<0.0001
Duration	1	0.13	0.13	0.2312	0.6308
Error	545	306.93	0.563	X	X
Total	547	446.91	X	X	X

TABLE23: ANOVA for Method 2 – July 21, 2017 at NEB

Source	DF	SS	MS	F	P value
Altitude	1	101.58	101.58	218.58	<0.0001
Duration	1	0.00	0.00	0.0001	0.9924
Error	337	156.61	0.465	X	X
Total	339	258.19	X	X	X

The classification/count of misclassification based on the ANOVA analysis using the first method is shown in Table 24.

TABLE 24: Count of Misclassification for Method 1 and proportions are given in the parentheses – July 21, 2017 at NEB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	31 (100)	0 (0)	0 (0)	0 (0)	0 (0)	31
	2	0 (0)	134 (100)	0 (0)	0 (0)	0 (0)	134
	3	0 (0)	24 (12.4)	170 (87.6)	0 (0)	0 (0)	194
	4	0 (0)	0 (0)	0 (0)	189 (100)	0 (0)	189
Total	31	158	170	189	0	548	

The classification/count of misclassification based on the ANOVA analysis for the second method is presented in Table 25.

TABLE 25: Count of Misclassification for Method 2 and proportions are given in the parentheses – July 21, 2017 at NEB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	21 (100)	0 (0)	0 (0)	0 (0)	0 (0)	21
	2	0 (0)	50 (100)	0 (0)	0 (0)	0 (0)	50
	3	0 (0)	15 (10.6)	117 (83)	9 (6.4)	0 (0)	141
	4	0 (0)	0 (0)	0 (0)	128 (100)	0 (0)	128
Total		21	65	117	137	0	340

4.3 Results of Regression Tree Analysis

The Regression Tree analyses for the classification of aerosols using different methods for the various dates and locations are shown in this section.

4.3.1 August 16, 2016 at ESB

We begin by training regression tree models on 75% of the full data from both methods, i.e. 1788 and 725 data points from Method 1 and Method 2, respectively. Then, we test the model prediction accuracy on the testing samples which are 596 and 242 data points from Method 1 and Method 2, respectively.

The use of the regression tree technique shows similar results as the ANOVA analysis as seen in the Tables. Table 26 shows a classification/count of misclassification that is a cross-tabulation

that indicates the model incorrectly classified records in the testing data, which is a simple random sample of 25% (596) of the full data from Method 1.

TABLE 26: Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – August 16, 2016 at ESB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	29 (100)	0 (0)	0 (0)	0 (0)	0 (0)	29
	2	0 (0)	158 (100)	0 (0)	0 (0)	0 (0)	158
	3	0 (0)	0 (0)	242 (100)	0 (0)	0 (0)	242
	4	0 (0)	0 (0)	0 (0)	147 (88)	20 (12)	167
Total		29	158	242	147	20	596

Also, Table 27 shows the classification/count of misclassification corresponding to Method 2.

TABLE 27: Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses – August 16, 2016 at ESB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	22 (91.7)	2 (8.3)	0 (0)	0 (0)	0 (0)	24
	2	0 (0)	35 (100)	0 (0)	0 (0)	0 (0)	35
	3	0 (0)	0 (0)	123 (100)	0 (0)	0 (0)	123
	4	0 (0)	0 (0)	0 (0)	53 (88.3)	7 (11.7)	60
Total		22	37	123	53	7	242

4.3.2 August 17, 2016 at ESB

We begin by training regression tree models on 75% of the full data from both methods, i.e. 1362 and 757 data points from Method 1 and Method 2, respectively. Then, we test the model prediction accuracy on the testing samples which are 455 and 253 data points from Method 1 and Method 2, respectively.

The use of the regression tree technique shows similar results as the ANOVA analysis as seen in the Tables. Table 28 shows a classification/count of misclassification that is a cross-tabulation that indicates the model incorrectly classified records in the testing data, which is a simple random sample of 25% (455) of the full data from Method 1.

TABLE 28: Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – August 17, 2016 at ESB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	3 (100)	0 (0)	0 (0)	0 (0)	0 (0)	3
	2	0 (0)	127 (100)	0 (0)	0 (0)	0 (0)	127
	3	0 (0)	0 (0)	137 (100)	0 (0)	0 (0)	137
	4	0 (0)	0 (0)	0 (0)	171 (91)	17 (9)	188
Total	3	127	137	171	17	455	

Also, Table 29 shows the classification/count of misclassification corresponding to Method 2.

TABLE 29: Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses – August 17, 2016 at ESB

	Predicted Value					Total	
		1	2	3	4		5
True Value	1	2 (100)	0 (0)	0 (0)	0 (0)	0 (0)	2
	2	17 (35.4)	31 (64.6)	0 (0)	0 (0)	0 (0)	48
	3	0 (0)	0 (0)	122 (100)	0 (0)	0 (0)	122
	4	0 (0)	0 (0)	0 (0)	71 (87.7)	10 (12.3)	81
Total		19	31	122	71	10	253

4.3.3 July 21, 2017 at Constant Hall

We begin by training regression tree models on 75% of the full data from both methods, i.e. 452 and 286 data points from Method 1 and Method 2, respectively. Then, we test the model prediction accuracy on the testing samples which are 151 and 96 data points from Method 1 and Method 2, respectively.

The use of the regression tree technique shows similar results as the ANOVA analysis as seen in the Tables. Table 30 shows a classification/count of misclassification that is a cross-tabulation that indicates the model incorrectly classified records in the testing data, which is a simple random sample of 25% (151) of the full data from Method 1.

TABLE 30: Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – July 21, 2017 at Constant Hall

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	8 (100)	0 (0)	0 (0)	0 (0)	0 (0)	8
	2	0 (0)	33 (100)	0 (0)	0 (0)	0 (0)	33
	3	0 (0)	6 (6.9)	81 (93.1)	0 (0)	0 (0)	87
	4	0 (0)	0 (0)	0 (0)	21 (91.3)	2 (8.7)	23
Total		8	39	81	21	2	151

Also, Table 31 shows the classification/count of misclassification corresponding to Method 2.

TABLE31: Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses – July 21, 2017 at Constant Hall

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	5 (62.5)	3 (37.5)	0 (0)	0 (0)	0 (0)	8
	2	0 (0)	30 (100)	0 (0)	0 (0)	0 (0)	30
	3	0 (0)	13 (30.2)	30 (69.8)	0 (0)	0 (0)	43
	4	0 (0)	0 (0)	0 (0)	12 (80)	3 (20)	15
Total		5	46	30	12	3	96

4.3.4 July 21, 2017 at Rogers Hall

We begin by training regression tree models on 75% of the full data from both methods, i.e. 495 and 187 data points from Method 1 and Method 2, respectively. Then, we test the model

prediction accuracy on the testing samples which are 165 and 62 data points from Method 1 and Method 2, respectively.

The use of the regression tree technique shows similar results as the ANOVA analysis as seen in the Tables. Table 32 shows a classification/count of misclassification that is a cross-tabulation that indicates the model incorrectly classified records in the testing data, which is a simple random sample of 25% (165) of the full data from Method 1.

TABLE 32: Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – July 21, 2017 at Rogers Hall

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	16 (100)	0 (0)	0 (0)	0 (0)	0 (0)	16
	2	0 (0)	48 (100)	0 (0)	0 (0)	0 (0)	48
	3	0 (0)	0 (0)	68 (100)	0 (0)	0 (0)	68
	4	0 (0)	0 (0)	0 (0)	28 (84.8)	5 (15.2)	33
Total	16	48	68	28	5	165	

Similarly, Table 33 shows the classification/count of misclassification corresponding to Method 2.

TABLE 33: Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses – July 21, 2017 at Rogers Hall

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	6 (85.7)	1 (14.3)	0 (0)	0 (0)	0 (0)	7
	2	0 (0)	8 (88.9)	1 (11.1)	0 (0)	0 (0)	9
	3	0 (0)	4 (14.8)	23 (87.4)	0 (0)	0 (0)	27
	4	0 (0)	0 (0)	0 (0)	14 (73.7)	5 (26.3)	19
Total		7	9	27	14	5	62

4.3.5 July 21, 2017 at NEB

We begin by training regression tree models on 75% of the full data from both methods, i.e. 411 and 255 data points from Method 1 and Method 2, respectively. Then, we test the model prediction accuracy on the testing samples which are 137 and 85 data points from Method 1 and Method 2, respectively.

The use of the regression tree technique shows similar results as the ANOVA analysis as seen in the Tables. Table 34 shows a classification/count of misclassification that is a cross-tabulation that indicates the model incorrectly classified records in the testing data, which is a simple random sample of 25% (137) of the full data from Method 1.

TABLE 34: Count of Misclassification of test sample using the Regression Tree Technique for Method 1 and proportions are given in the parentheses – July 21, 2017 at NEB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	4 (80)	1 (20)	0 (0)	0 (0)	0 (0)	5
	2	0 (0)	36 (100)	0 (0)	0 (0)	0 (0)	36
	3	0 (0)	0 (0)	55 (100)	0 (0)	0 (0)	55
	4	0 (0)	0 (0)	0 (0)	36 (87.8)	5 (12.2)	41
Total		4	37	55	36	5	137

Also, Table 35 shows the classification/count of misclassification corresponding to Method 2.

TABLE 35: Count of Misclassification of test sample using the Regression Tree Technique for Method 2 and proportions are given in the parentheses – July 21, 2017 at NEB

	Predicted Value					Total	
	1	2	3	4	5		
True Value	1	4 (80)	1 (20)	0 (0)	0 (0)	0 (0)	5
	2	0 (0)	10 (83.3)	2 (16.7)	0 (0)	0 (0)	12
	3	0 (0)	4 (10.8)	33 (89.2)	0 (0)	0 (0)	37
	4	0 (0)	0 (0)	0 (0)	31 (100)	0 (0)	31
Total		4	15	35	31	0	85

4.4 Discussion of Results

In this chapter, the results of the models are presented. The figures displayed in this section are to show the effects of altitude coefficients and duration varying coefficients on the model.

The Tables of the GLM and Regression Tree results show that altitude has more effect on the

differences in the aerosols identified. Summary of each of the effects of altitude and duration are presented.

4.4.1 Altitude Coefficients Model

The GLM model in Equation 34 shows that the altitude has a significant effect on the types of observed aerosols as seen in Figs. 5 and 6. There appears to be a clear separation of the aerosols based on altitude (Figs. 5 and 6), and a mix of the aerosols when compared with duration. These findings match the results that are published by [6] which shows that the classification results are used together with the measurements of aerosol optical depth to apportion the aerosol optical depth among the various aerosol types. It is then observed that the dominant aerosol types in terms of aerosol optical depth vary significantly with altitude.

4.4.2 Duration Varying Coefficients Model

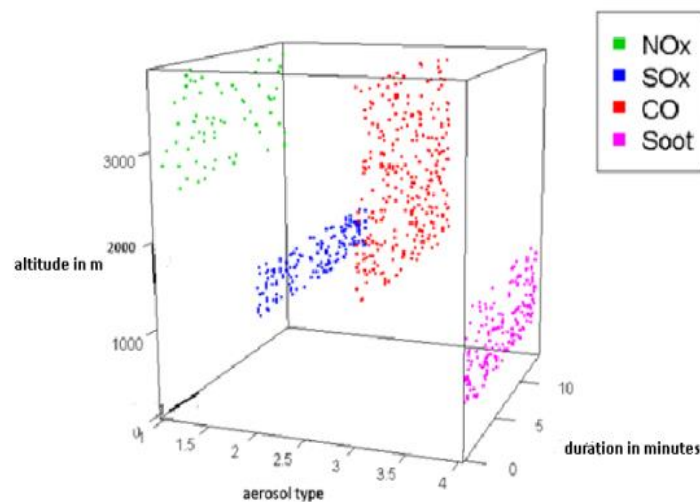


FIG. 5: 3-D View of Detected Aerosols Using Method 1

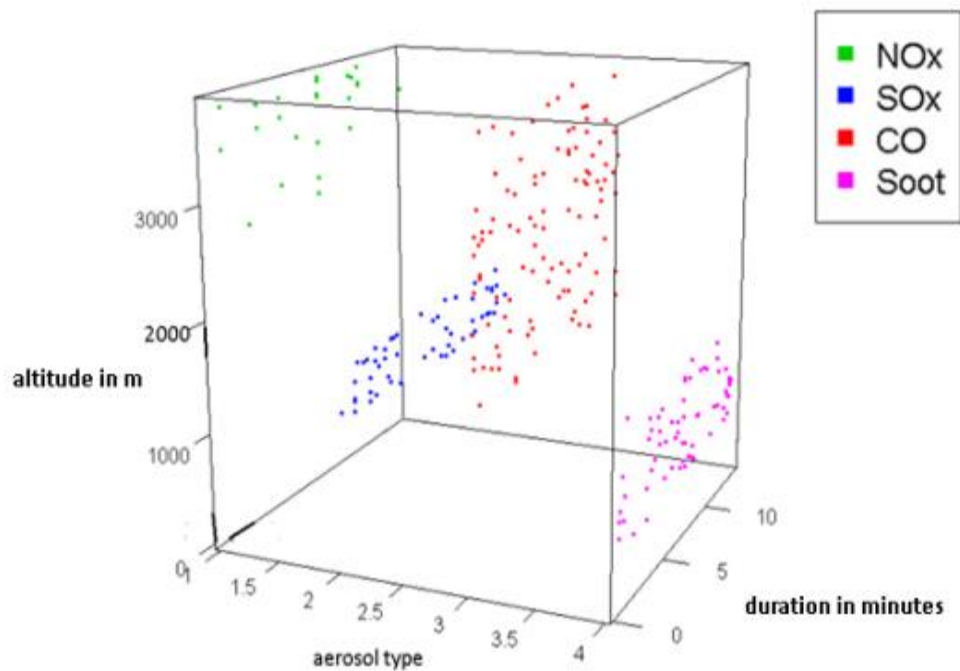


FIG. 6: 3-D View of Detected Aerosols Using Method 2

From Fig.5. and Fig. 6., it was observed that duration does not have a significant effect on the type of aerosols classification found in the environment. The plots are quite scattered and do not lead to any pattern. The figures also show that the aerosols are categorized by the altitudes they are found. Thus it proves that altitude is a significant factor in finding aerosols in the atmosphere. The analogous analysis shows that in both methods, duration is not found to be a significant indicator of aerosol, whereas altitude is as observed in all the ANOVA Tables irrespective of the date and/or location the LIDAR data was collected. This significance is observed in the p-values from the ANOVA tables for both altitude and duration. The smaller the p-value the more significant the factor is. The regression tree technique shows similar results as the ANOVA

analysis, that is, in both methods altitude significantly affects the types of observed aerosols whereas the duration is not significant. This is known by the value of the variable importance and the larger the value is the more significant that factor is.

4.4.3 Comparison of Analysis

For the analysis from August 16, 2016 at ESB, the misclassification of aerosol values for Methods 1 and 2 are displayed in Tables 8 and 9 respectively. It appears that the proportions of correct classification are high for both methods. Overall, the use of Methods 1 and 2 did not show any significant difference in predicting aerosol value 2 (SO_x) and aerosol value 4 (soot). However, Method 1 showed more accuracy for aerosol value 1 (NO_x) than Method 2. Classifications at values 2 (SO_x) and 4 (soot) are perfect for both methods. At aerosol value 3 (CO), both methods have errors but they seem to predict equally well. It is to be noted that predictions are made for aerosol values 2 (SO_x) and 4 (soot) when Method 2 is used when aerosol values 1 (NO_x) and 3 (CO) respectively are expected. That forces us to think that some of the aerosol at values 2 (SO_x) and 4 (soot) can or should be classified as values 1 (NO_x) and 3 (CO) and vice versa, hence Method 1 has less misclassifications than Method 2.

The variable importance is as follows: the most important variable is altitude with a value equals to 100 that is much greater than the one corresponding to the duration variable, which is 0. Similarly, Method 2 shows that the most important variable is altitude with a value equals to 100 which is much greater than 0 the value corresponding to duration. Both tables 26 and 27 show high proportions of correct classification. For Method 1, the error rate, or the proportion of incorrectly classified aerosols levels is 0.03, while the proportion of incorrectly classified aerosols levels for Method 2 is 0.04. Although the misclassification corresponding to the first

method is much smaller than the one from the second method, both are acceptable, and the difference might be due the smaller training sample of Method 2 compared to the one from Method 1. The first three aerosol values of Method 1 shows perfect accuracy while Method 2 has some errors especially at aerosol values 1 (NO_x) and 4 (soot). In aerosol value4 (soot) both methods have some errors and performs equally as well as each other. It can be seen that prediction is made for aerosol value 5 (No Aerosol) even though it was not part of the selection. This again suggests that some of the aerosol at value 4 (soot) should have been classified as aerosol value 5 (No Aerosol).

In respect to the analysis from August 17, 2016 at ESB, the misclassification of aerosol values for Methods 1 and 2 are displayed in Tables 12 and 13 respectively. It appears that the proportions of correct classification are high for both methods. Overall, the use of Methods 1 and 2 did not show any significant difference in predicting aerosol value 2 (SO_x). At aerosol value 1 (NO_x), both methods have errors but they seem to predict equally well. Classification at value 2 (SO_x) is perfect for both methods. At aerosol value 3 (CO), both methods have errors but Method 2 seem to predict better than Method 1. At value 4 (soot), both methods almost perfectly predict.

The variable importance is as follows: the most important variable is altitude with a value equal to 100 that is much greater than the one corresponding to the duration variable, which is 0. Similarly, Method 2 shows that the most important variable is altitude with a value equal to 98 which is much greater than 2 the value corresponding to duration. Both Tables 28 and 29 show high proportions of correct classification. For Method 1, the error rate, or the proportion of incorrectly classified aerosols levels is 0.04, while the proportion of incorrectly classified aerosols levels for Method 2 is 0.11. Although the misclassification corresponding to the first

method is much smaller than the one from the second method, both are acceptable, and the difference might be due the smaller training sample of Method 2 compared to the one from Method 1.

With the analysis from July 21, 2017 at Constant Hall, the misclassification of aerosol values for Methods 1 and 2 are displayed in Tables 16 and 17 respectively. It appears that the proportions of correct classification are high for both methods. Overall, the use of Methods 1 and 2 did not show any significant difference in predicting aerosol value 2 (SO_x) and aerosol value 4 (soot). Also, Method 1 showed less accuracy for aerosol value 1 (NO_x) than Method 2. Classification at value 2 (SO_x) is perfect for both methods. At aerosol value 3 (CO), both methods have errors but they seem to predict equally well.

The variable importance is as follows: the most important variable is altitude with a value equal to 95 that is much greater than the one corresponding to the duration variable, which is 5. Similarly, Method 2 shows that the most important variable is altitude with a value equal to 95 which is much greater than 5 the value corresponding to duration. Both Tables 30 and 31 show high proportions of correct classification. For Method 1, the error rate, or the proportion of incorrectly classified aerosols levels is 0.05, while the proportion of incorrectly classified aerosols levels for Method 2 is 0.20. Although the misclassification corresponding to the first method is much smaller than the one from the second method, both are acceptable, and the difference might be due to the smaller training sample of Method 2 compared to the one from Method 1. The first two aerosol levels of Method 1 show perfect accuracy while Method 2 also has perfect accuracy at aerosol level 2 (SO_x). In aerosol levels 3 (CO) and 4 (soot) both methods have some errors, but Method 1 performs significantly better.

Also for the analysis from July 21, 2017 at Rogers Hall, the misclassification of aerosol values for Methods 1 and 2 are displayed in Tables 20 and 21 respectively. It appears that the proportions of correct classification are high for both methods. Overall, the use of Methods 1 and 2 did not show any significant difference in predicting aerosol value 2 (SO_x). However, Method 1 showed less accuracy for aerosol value 1 (NO_x) than Method 2. Classification at value 2 (SO_x) is perfect for both methods. At aerosol value 3 (CO), both methods have errors but they seem to predict equally well. At value 4 (soot), Method 1 works better. It is to be noted that prediction is made for aerosol value 5 (No Aerosol) even though it was not part of the selection. That forces us to think that some of the aerosol at value 4 (soot) can or should be classified as value 5 (No Aerosol) and vice versa.

The variable importance is as follows: the most important variable is altitude with a value equal to 98 that is much greater than the one corresponding to the duration variable, which is 2. Similarly, Method 2 shows that the most important variable is altitude with a value equal to 94 which is much greater than 6 the value corresponding to duration. Both Tables 32 and 33 show high proportions of correct classification. For Method 1, the error rate, or the proportion of incorrectly classified aerosols levels is 0.03, while the proportion of incorrectly classified aerosols levels for Method 2 is 0.18. Although the misclassification corresponding to the first method is much smaller than the one from the second method, both are acceptable, and the difference might be due to the smaller training sample of Method 2 compared to the one from Method 1. The first three aerosol values of Method 1 show perfect accuracy while Method 2 has some errors especially at aerosol value 3 (CO). In aerosol value 4 (soot) both methods have some errors, but Method 1 performs better. As in the ANOVA analysis, it can be seen that prediction is made for aerosol value 5 (No Aerosol) even though it was not part of the selection. This again

suggests that some of the aerosol at value 4 (soot) should have been classified as aerosol value 5 (No Aerosol).

Lastly with the analysis from July 21, 2017 at NEB, the misclassification of aerosol values for Methods 1 and 2 are displayed in Tables 24 and 25 respectively. The proportions of correct classification are high for both methods. The use of Methods 1 and 2 did not show any significant difference in predicting aerosol value 1 (NO_x), aerosol value 2 (SO_x) and aerosol value 4 (soot). However, Method 2 showed less accuracy for aerosol value 3 (CO) than Method 1. Classification is perfect for both methods at values 1 (NO_x), 2 (SO_x) and 4 (soot).

The variable importance is as follows: the most important variable is altitude with a value equal to 99 that is much greater than the one corresponding to the duration variable, which is 1. Similarly, Method 2 shows that the most important variable is altitude with a value equal to 92 which is much greater than 8 the value corresponding to duration. Both Tables 34 and 35 show high proportions of correct classification. For Method 1, the error rate, or the proportion of incorrectly classified aerosols levels is 0.04, while the proportion of incorrectly classified aerosols levels for Method 2 is 0.08. The misclassification corresponding to the first method is much smaller than the one from the second method, both are acceptable, and the difference might be due the smaller training sample of Method 2 compared to the one from Method 1. The aerosol values 2 (SO_x) and 3 (CO) of Method 1 show perfect accuracy while Method 2 has some errors except at aerosol value 4 (soot) where it also has perfect accuracy. In aerosol value 1 (NO_x) both methods have some errors, but they seem to predict equally well.

CHAPTER 5

CONCLUSIONS

This dissertation proposes experiments which are performed on different days and locations to obtain LIDAR data to be used to identify the types of aerosols in the atmosphere. The work demonstrates computational models to help obtain optical and spectral parameters to help classify aerosols. Another contribution of this dissertation project is that it applies classification techniques to distinguish between different aerosols.

It also analyzes classification approaches using statistical analysis for predictive behavior. Extensive statistical analyses and comparison of data from both methods reveal significant differences in the prediction of aerosol values using the two classification methods described with the aid of statistical analysis tools such as GLM and the Regression Tree. The overall contributions of this dissertation are summarized and further discussed below.

First, the extensive aerosol parameters such as backscatter and extinction coefficients are retrieved from the LIDAR data collected as seen in Chapter 2 with the help of retrieval methods by Fernald or Klett. Second, the optical and spectral aerosol parameters were computed using the extensive parameters and then combined to obtain two different classification methods as described in Chapter 3 in order to help classify the various identified aerosols found in the atmosphere around the Old Dominion University campus. Third, the statistical analysis tools such as GLM and Regression Tree techniques are introduced and implemented by modeling the identified aerosols and activity profiles. These tools also enable the prediction of aerosols when the two classification methods are used.

Finally, from the use of p-values from the ANOVA tables and the variable importance of the Regression Tree technique it is observed that altitude is more significant than duration in the identification of aerosols in the atmosphere. This means that to predict the presence of aerosols, it is more dependent on the altitude one is observing but not on the duration of observance.

Interestingly, it appears that the proportions of classification are high for both methods which imply both methods are likely to correctly predict the type of aerosols identified in the atmosphere. Using the count of misclassification tables, it is observed that to assume the presence of an identified aerosol with value 1 (NO_x), both methods were equally adept in predicting it. To assume the presences of aerosol with value 2 (SO_x) and aerosol with value 3 (CO) in the atmosphere, Method 1 was likely to predict them more often as compared to Method 2. Lastly, to assume the presence of aerosol of value 4 (soot), the two methods were equally likely to predict its presence.

Method 1 is thus more appropriate for the classification of the aerosols than Method 2 due to its lower misclassification errors irrespective of the statistical analysis tool which is used and/or the date and location which was analyzed. The analysis provided concurrent estimations of the underlying aerosol distributions between the methods and locations. As illustrated, the functional connectivity between the regions is shown to provide evidence that parameters are significantly different.

5.1 Future Work

In this dissertation, aerosols in the atmosphere are classified by the use of two classification methods, which are found to be appropriate for prediction of the types of aerosol being sought. There are however some limitations to the classification of aerosols using the methods suggested

in this dissertation and the statistical description of the data. Future studies are necessary to address these limitations.

5.1.1 Methods Used

For the classification methods used, to be more specific for Method 1, a LIDAR with depolarizers on both the 532nm and 1064nm channels would be very useful in computing optical parameters such as depolarization ratio and depolarization spectral ratios which when combined with LIDAR ratio and backscatter color ratio would improve the classification process. This is because more parameters help in describing the different characteristics between the aerosols. Depolarization ratio is used to distinguish between fine aerosol particles [33]. Low values of depolarization ratio usually indicate the presence of spherical particles [34, 35] while high values will indicate non-spherical particles. Depolarization spectral ratio is found to be dependent on particle size in the case of ice particles and on mixing ratio, and spherical and non-spherical particle sizes in mixtures of dust and non-spherical particles [36]. A procedure can be pursued to integrate Methods 1 and 2 together to perform the classification of aerosols.

5.1.2 Types of Statistical Analysis Techniques

Although a practical way of describing the data has been presented, other algorithms such as the Bayesian methods or the Zero inflated Poisson would be alternatives for the research design for the innovative and efficient classification of the aerosols.

BIBLIOGRAPHY

- [1] D.A., Chu, Y.J., Kaufman, G., Zibordi, J.D., Chern, J., Mao, C., Li, and B.N., Holben, “Global Monitoring of Air Pollution Over Land From the Earth Observing System-Terra Moderate Resolution Imaging Spectroradiometer (MODIS).” *Journal of Geophysical Research* **108**(D21), 2003.
- [2] P., Velasco, P., Gaffney, M., Khedia, P., Lee, D., Mazzera, E., Miller, L., Negrete, and A., Panson, “2002 Annual Report on the Air Resources Board’s Fine Particulate Matter Program.” *California Air Resources Board*, 2003.
- [3] B., Brunekreef, and S.T., Holgate, “Air Pollution and Health.” *The Lancet* **360**(9341), pp. 1233-1242, 2002.
- [4] Seinfeld, J.H., and Pandis, S.P., *Atmospheric Chemistry and Physics: Air Pollution to Climate Change*, 2nd edition, John Wiley, New York, USA, 1232, 2006.
- [5] Kovalev, V.A., and Eichinger, W.E., *Elastic Lidar: Theory, Practice, and Analysis Methods*, John Wiley and Sons, 2004.
- [6] S.P, Burton, R.A., Ferrare, C.A., Hostetler, J.W., Hair, R.R., Rogers, M.D., Obland, C.F., Butler, A.L., Cook, D.B. Harper, and K.D., Froyd, “Aerosol classification using airborne High Spectral Resolution Lidar measurements – methodology and examples”, *Atmospheric Measurement Techniques*, **5**, pp. 73-98 (2012).

- [7] F., Costabile, F., Barnaba, F. Angelini, and G.P., Gobbi, “Identification of key aerosol populations through their size and composition resolved spectral scattering and absorption”, *Atmospheric Chemistry and Physics*, **13**, pp. 2455-2470 (2013).
- [8] E.H., Synge, “A method of investigating the higher atmosphere.” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **9**(60), pp. 1014-1020, 1930.
- [9] E.O., Hulbert, “Observations of a searchlight beam to an altitude of 28 kilometers.” *Journal of the Optical Society of America* **27**(11), pp. 377-382, 1937.
- [10] E.A., Johnson, R.C., Meyer, R.E., Hopkins, and W.H., Mock, “The measurement of light scattered by the upper atmosphere from a search-light beam.” *Journal of the Optical Society of America* **29**(12), pp. 512-517, 1939.
- [11] G., Fiocco, and L.D., Smullin, “Detection of scattering layers in the upper atmosphere (60-140 km) by optical radar.” *Nature***199**(4900), pp. 1275-1276, 1963.
- [12] Wallace, J.M., and Hobbs, P.V., *Atmospheric Science: An Introductory Survey*, 2nd edition. Academic Press, 2006.
- [13] Measures, R.M., *Laser Remote Sensing: Fundamentals and Applications*, Krieger Publishing Company, 1984.
- [14] McCormick, M.P., (April 1967) *Laser backscatter measurements of the lower atmosphere*.

(PhD thesis), The College of William and Mary, Williamsburg, VA.

- [15] W.J., Wiscombe, "Improved Mie scattering algorithms." *Applied Optics* **19**(9), pp. 1505-1509, 1980.
- [16] Hinds, W.C., *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles*, 2nd edition. John Wiley and Sons, 1999.
- [17] J.D., Klett, "Stable analytical inversion solution for processing lidar returns." *Applied Optics* **20**(2), pp. 211-220, 1981.
- [18] F.G., Fernald, "Analysis of atmospheric lidar observations: some comments." *Applied Optics* **23**(5), pp. 652-653, 1984.
- [19] F.G., Fernald, B.M., Herman, and J.A., Reagan, "Determination of aerosol height distribution by lidar." *Journal of Applied Meteorology* **11**, pp. 482-489, 1972.
- [20] J.W., Hair, C.A., Hostetler, A.L., Cook, D.B., Harper, R.A., Ferrare, T.L., Mack, W., Welch, L.R., Isquierdo, and F.E., Hovis, "Airborne high spectral resolution lidar for profiling aerosol optical properties." *Applied Optics* **47**(36), pp. 6734-6752, 2008.
- [21] R.J., DeYoung, W.B., Grant, and K., Severance, "Aerosol Transport in the California Central Valley Observed by Airborne Lidar." *Environmental Science & Technology* **39**(21), pp. 8351-8357, 2005.

- [22] C., Gili, and R.J., DeYoung, “A compact efficient lidar receiver for measuring atmospheric aerosols.” *Technical Report NASA/TP-2006-213950*, NASA, 2006.
- [23] N., Sugimoto, I., Matsui, A., Shimizu, I., Uno, K., Asai, T., Endoh, and T., Nakajima, “Observation of dust and anthropogenic aerosol plumes in the Northwest Pacific with a two-wavelength polarization lidar on board the research vessel Mirai”, *Geophysical Research Letters*, **29**, pp. 1901, 2002.
- [24] Y. Sasano, and E.V., Browell, “Light-Scattering Characteristics of Various Aerosol Types Derived from Multiple Wavelength Lidar Observations”, *Applied Optics*, **28**, pp. 1670-1679, 1989.
- [25] R.W., Bergstrom, P., Pilewskie, P.B., Russell, J., Redemann, T.C., Bond, P.K., Quinn, and B., Sierau, “Spectral absorption properties of atmospheric aerosols”, *Atmospheric Chemistry and Physics*, **7**, pp. 5937-5943, 2007.
- [26] R.M., Heiberger, and B., Holland, *Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R, and SAS*, (Springer Science Business Media LLC, 2004).
- [27] A. Gelman, and J., Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, Cambridge, UK, 2007).

- [28] J.J., Faraway, Linear Models with R (CRC Press, Boca Raton, FL., 2005).
- [29] J.J., Faraway, Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models (CRC Press, Boca Raton, FL, 2005).
- [30] L., Breiman, Classification and regression trees (The Wadsworth statistics/probability series). Belmont, California: Wadsworth International Group, (1984).
- [31] Hastie, Tibshirani, Friedman, Tibshirani, Robert, & Friedman, J. H., The elements of statistical learning data mining, inference, and prediction: With 200 full-color illustrations (Springer series in statistics). New York: Springer, 2001.
- [32] SAS Institute Inc. 2011. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc.
- [33] A.H., Omar, D.M, Winker, C., Kittaka, M.A., Vaughn, Z.Y., Liu, Y.X., Hu, C.R.,Trepte, R.R., Rogers, R.A., Ferrare, K.P., Lee, R.E., Kuehn, and C.A., Hostetler, “The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm”, Journal of Atmospheric and Oceanic Technology, **26**, pp. 1994-2014, 2009.
- [34] N. Sugimoto, and C.H., Lee, “Characteristics of dust aerosols inferred from lidar depolarization measurements at two wavelengths”, Applied Optics, **45**, pp. 7468-7474, 2006.
- [35] T., Murayama, S.J., Masonis, J., Redemann, T.L., Anderson, B., Schmid, J.M., Livingston,

P.B, Russell, B., Huebert, S.G., Howell, C.S., McNaughton, A., Clarke, M., Abo, A., Shimizu, N., Sugimoto, M., Yabuki, H., Kuze, S., Fukagawa, K., Maxwell-Meier, R.J., Weber, D.A, Orsini, B., Blomquist, A., Bandy, and D., Thornton, “An intercomparison of lidar-derived aerosol optical properties with airborne measurements near Tokyo during ACE-Asia”, *Journal of Geophysical Research.-Atmos.*, **108**(D23), Art. No. 800, doi:10.1029/2002JD003259, 2003.

- [36] T., Somekawa, C., Yamanaka, M., Fujita, and M.C., Galvez, “A new concept to characterize nonspherical particles from multi-wavelength depolarization ratios based on T-matrix computation”, *Particle & Particle System Characterization*, **25**, pp. 49-53, 2008.

VITA

Kwasi G. Afrifa

Department of Electrical and Computer Engineering

Old Dominion University

Norfolk, VA 23529

Kwasi G. Afrifa received his M.S. degree in Electrical and Computer Engineering from the University of Windsor, Canada in 2010 and the B.S. degree in Electrical and Electronic Engineering from the Kwame Nkrumah University of Science and Technology, Ghana in 2005. In Fall 2012, he started his D.Eng. program in the ECE Vision Lab at Old Dominion University. The motivation of his current research is to provide a way of finding aerosols in the atmosphere with the use of a ground-based LIDAR device. His research interests include Signal Processing, Machine Learning and Computer Vision.