2-2021

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

Mohammed Alqawba

Norou Diawara

Kwasi G. Afrifa

Mohamed I. Elbakary

Mecit Cetin

See next page for additional authors

Authors

Mohammed Alqawba, Norou Diawara, Kwasi G. Afrifa, Mohamed I. Elbakary, Mecit Cetin, and Khan Iftekharuddin

Transactions on Engineering and Computer Science

Research| Vol 2 Iss 1

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

Mohammed Alqawba[1,2], Norou Diawara[2]*, Kwasi G. Afrifa[3], Mohamed I. Elbakary[4], Mecit Cetin[3], and Khan M. Iftekharuddin[3]

[1]Department of Mathematics, College of Science and Arts, Qassim University, Ar Rass, Saudi Arabia

[2]Department of Mathematics and Statistics, Old Dominion University, Hampton Boulevard, Norfolk, VA, USA

[3]Vision Laboratory, Electrical and Computer Engineering Department, Old Dominion University, Hampton Boulevard, Norfolk, VA, USA

[4]Department of Civil and Environmental Engineering, Old Dominion University, Hampton Boulevard, Norfolk, VA, USA

*Corresponding author: Diawara N, Department of Mathematics and Statistics, Old Dominion University, Hampton Boulevard, Norfolk, VA, USA; E-mail: NDiawara[At]odu[DOT]edu

Received: January 15, 2021; Accepted: February 20, 2021; Published: February 28, 2021

Abstract

In this article, we present 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 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 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 detailed statistical comparisons and analyses shows 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 efficient under appropriate functional distributions. The articleoffers a method for natural ordering of the aerosol types.

Keywords: ***Remote sensing and sensors; Lidar; Aerosol detection.***

Citation: Alqawba M, Diawara N, Afrifa KG, et al. Statistical analysis and comparison of optical classification of atmospheric aerosol lidar data. Trans Eng Comput Sci. 2021;2(1):119.

1. Introduction

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-related health effects to Earth's energy balance. 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 [1].

This article 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). 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. The 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 the literature [2] and [3]. 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, generalized linear model (GLM) analysis and regression tree are then used to infer whether there are significant differences between the approaches adopted by [2] and [3] to identify the aerosols and then recommend which approach is suitable for specific aerosols from the measured LIDAR data.

The first classification method as described in [2] used intensive parameters of aerosol (LIDAR ratio and backscatter color ratio) which vary with aerosol type. 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. Another parameter that is used is backscatter color ratio, which is defined by the ratio of aerosol backscatter coefficient of the 532 nm to that of the 1064 nm channel. Backscatter color ratios are inversely related to aerosol particle sizes [4] and [5]. Another intensive parameter, the depolarization ratio is used to distinguish between fine aerosol particles such as dust [6]. Low values of depolarization ratio usually indicate the presence of spherical particles [7] and [8]. High values of depolarization ratio are also indicative of the presence of pure dust. The fourth aerosol optical parameter is the 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 [9]. The combination of these aerosol optical parameters is used to indicate the likely presence of specific aerosol particles in the environment. For this first classification method, the optical parameters used are the LIDAR ratio and backscatter color ratio.

The second classification method to identify aerosols in the atmosphere as described in [3] is based on their spectral properties. A paradigm based on scattering Angstrom exponent (SAE), absorption Angstrom exponent (AAE), extinction Angstrom exponent (EAE), and single scattering albedo (SSA) are found to be suitable to identify the presence of aerosols such as soot, biomass burning, dust and organic particles. Absorption coefficient decreases monotonically with wavelength and it is approximated by a power law expression which is described by an absorption Angstrom exponent

(AAE) [1]. SAE and AAE are used to distinguish between dust, urban aerosol and biomass burning [10]. Also [11] proposed the use of spectral SSA variability to differentiate between dust and black carbon absorption. The separation of days with minimal pollution from polluted days is achieved using SAE and AAE with SSA as performed in [12].

Some combinations of the intensive parameters AAE, EAE, SSA, and SAE are used to optically distinguish between aerosols. Some of those classifications of aerosols are based on EAE vs. AAE plots. It is advantageous in using EAE as it takes into account SAE and AAE, however it also fails in separating particle size effects attributed to SAE from particle composition attribute to AAE. To distinguish particle size from particle composition SAE is used instead of EAE [13]. Through the combined analysis of spectral optical properties, the method used in [3] provides understanding of the aerosol composition and particle size. This suggested combination of parameters aid in distinguishing between dust and polluted dust, polluted and clean days or showing the contribution of absorption in classifying aerosol particles. Aerosol particles have spectral optical properties which contain information on their size and composition. The knowledge of these information help in the classification of aerosols. The approach based on SAE, AAE, EAE, and SSA is used to classify the aerosols found in the atmosphere.

The types of aerosols that we intend to identify are NO_x, SO_x, CO and soot which are likely to be observed in an urban environment. The two factors that are employed in the analysis of the data collected are the altitude and duration. The altitude is the vertical distance the LIDAR pulses reached before they are scattered by the aerosols in the atmosphere. The duration is also the amount of time used by the LIDAR to collect data from the atmosphere.
The statistical analysis of the data asks if the layers for these aerosol types change or are independent under some random mechanism.

The rest of this article is organized as follows. In Section 2, we introduce the methods of classifications. In Section 3, we review analysis of variance (ANOVA) and GLM procedure. Section 4 presents the regression tree. In Section 5, we apply GLM and regression tree to the collected data and the results are presented in Section 6. Finally, Section 7 concludes this article.

2. Proposed Analyses

The analysis for the classification of aerosols was based on the measured optical properties from the LIDAR in Vision Lab at Old Dominion University [14]. We used two classification schemes on the collected data. The data were collected at several locations to cover the study area. The data were collected on one day and then we repeated the collection on another day. The two classification schemes are then analyzed with the aid of Analysis of Variance (ANOVA) to classify the particles in the atmosphere into their component aerosols. While applying the first classification method [2], we used only the LIDAR ratio and backscatter color ratio. This is due to the fact that such LIDAR does not generate depolarization ratio and spectral depolarization ratio. 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:

$$S_a = \frac{\sigma_a^{532}}{\beta_a^{532}} \qquad (1)$$

where S_a is the lidar ratio, σ_a^{532} is the aerosol extinction coefficient at 532nm and β_a^{532} is the backscatter coefficient at 532nm. The backscatter color ratio is defined as the ratio of backscattering coefficient at 532nm to 1064 nm.

$$BCR = \frac{\beta_a^{532}}{\beta_a^{1064}} \qquad (2)$$

where BCR is the backscatter color ratio, β_a^{532} is the aerosol backscatter coefficient at 532 nm and β_a^{1064} is the aerosol backscatter coefficient at 1064 nm.

For the second method of aerosols classification [3], we calculate the parameters SAE, AAE, EAE and SSA which are then combined to identify the presence of aerosols in the data. The four parameters are obtained through the following equations. To obtain the extinction Angstrom exponent, the equation used is

$$EAE = -\frac{ln\left(\frac{\sigma_{532}}{\sigma_{1064}}\right)}{ln\left(\frac{532}{1064}\right)} \qquad (3)$$

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

$$SAE = -\frac{ln\left(\frac{\beta_{532}}{\beta_{1064}}\right)}{ln\left(\frac{532}{1064}\right)} \qquad (4)$$

where β_{532} is the backscatter coefficient at 532nm and β_{1064} is the backscatter coefficient at 1064 nm. The sum of absorption and scattering is the extinction. From equation below we obtain the absorption Angstrom exponent,

$$AAE = -\frac{ln\left(\frac{\gamma_{532}}{\gamma_{1064}}\right)}{ln\left(\frac{532}{1064}\right)} \qquad (5)$$

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

$$SSA = \frac{\beta_{532}}{\beta_{532}+\gamma_{532}} = \frac{\beta_{532}}{\sigma_{532}}. \qquad (6)$$

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. In other words, we use GLM of the ANOVA technique and regression tree as statistical analysis tools for data analysis. The analysis is performed by assigning classification labels or the number 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. Due to the dominance of aerosol value of 5, we decided to remove it from the model and build predictions on the remaining data.

3. Analysis of Variance (Anova)

Analysis of Variance (ANOVA) is a common statistical method for testing the differences among group of means [15]. 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 a classifier and test significant relationship between the variables of altitude and duration.

The Generalized Linear Models (GLM) procedure was adopted for analysis [16-18]. We modeled the identified aerosols and classification profiles using such procedure. Our goal is to find if there exists significant difference between the methods used to record the aerosols based on the altitudes and duration that have been conducted. This model is based on the fact that the classification is a function of many factors and nuisance or errors. A mixed effect model was first used and this model may be described as follows:

$$Y_{ij} = \mu + \alpha_i + \delta_j + \epsilon_{ij} \qquad (7)$$

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 $j = 1,2, \dots ,128$, and 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. For example, $n_1 = 63$ means the number of aerosols profile measurements in a given duration is 63 at location 1. Altitude and duration at location 1 span to 3840 meters and 14 minutes, respectively. This is because every altitude level is 30 meters apart from each other and the number of aerosols profile measurements is 63 within the 14 minutes duration.

The model can be expressed as,

$$Y = X\varphi + \varepsilon \qquad (8)$$

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:

$$\hat{\varphi} = (X'X)^{-1}X'Y \qquad (9)$$

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. Because of that, we also adopt another prediction technique called regression tree.

4. Regression Tree

Although ANOVA is typically the first choice for prediction, an alternative method to perform prediction is in ensemble learning methods, which include classification and regression tree. The latter were first introduced by [19]. Several techniques are proposed as in [20-22]. However, the uncertainty analysis is present in one way or another. Solutions

are provided, "under trade-off between accuracy and meaningful solution" [22]. Such concerns prompted us to analyze the data as it is. 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 $i = 1, 2, \ldots, N$, where $\mathbf{x}_i = (x_{i1}, x_{2i})'$, x_{i1} is the ith observation from the first predictor, x_{i2} is the ith observation from the second predictor, and y_i is the ith 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 $R_1, \ldots, R_M$ and the response is modeled as a constant c_j for $j = 1, \ldots, M$:

$$f(x) = \sum_{j=1}^{M} c_j \, I\big(x \in R_j\big) \qquad (10)$$

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

$$\hat{c}_j = \text{ave}\big(y_i \big| x_i \in R_j\big) \qquad (11)$$

Therefore, the sum of square errors for a tree is

$$S = \sum_{j=1}^{M} \sum_{i=1}^{n_j} \big(y_i - \hat{c}_j\big)^2 \qquad (12)$$

For $M = 2$, an algorithm suggested by [20] 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

$$R_1(k, s) = \{x | x_k \leq s\}, \;\; R_2(k, s) = \{x | x_k > s\} \qquad (13)$$

Then choose the variable k and split point s that minimizes

$$\sum_{x_i \in R_1(k,s)} (y_i - \hat{c}_1)^2 + \sum_{x_i \in R_2(k,s)} (y_i - \hat{c}_2)^2 \qquad (14)$$

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 validated 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 the testing sample takes the rest, i.e. 25% of the data.

5. Application of Statistical Methods

To validate the difference in the two classification methods in aerosol, GLM and regression tree models are considered. For our analysis, we use the identified aerosols: NO_x, SO_x, CO, Soot and No Aerosol by assigning values of 1, 2, 3, 4 and 5 to them respectively. For the aerosol described as "No Aerosol", it means there is the absence of the four originally described aerosols. The first classification method to collect the data is based on the calculated LIDAR ratio and backscatter color ratio. It is completed by assigning a particular aerosol to a location and duration when the aerosol optical parameter condition for that aerosol is met as seen from Table1.

Table 1: Method 1.

Aerosol	Optical Parameter	
	S_a	BCR
NOx	70 - 80	3.3±10%
Sox	70 - 100	3.3±15%
CO	43 - 52	0.7±10%
Soot	60 - 65	1.4±10%

The classified aerosols for LIDAR remote sensing data using Method 1 is shown in Fig.1.

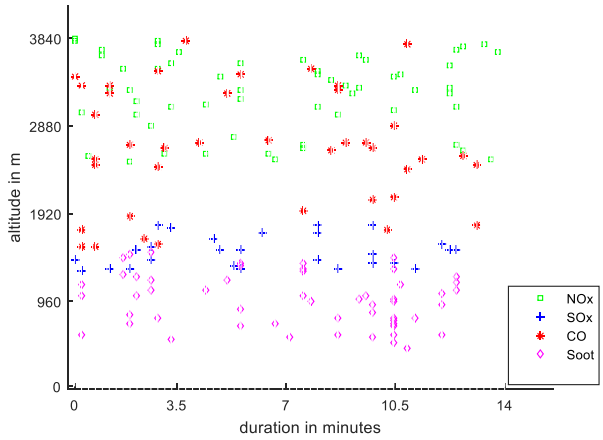

Fig. 1. Aerosols detected using method 1.

Fig. 1 shows the aerosols – NO_x, SO_x, CO, and soot as classified from our data using the first classification approach. The second classification method is done by using SSA, SAE, AAE, and EAE as our parameters. From Table 2, the aerosols are assigned when the parameter conditions for that particular aerosol are met.

Table 2: Method 2.

Aerosol	Spectral Parameter			
	SSA	AAE	EAE	SAE
NOx	> 0.85	> 2.5	1.8- 2	1.5-3.5
SOx	> 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 the LIDAR data using Method 2 are shown in Fig.2.

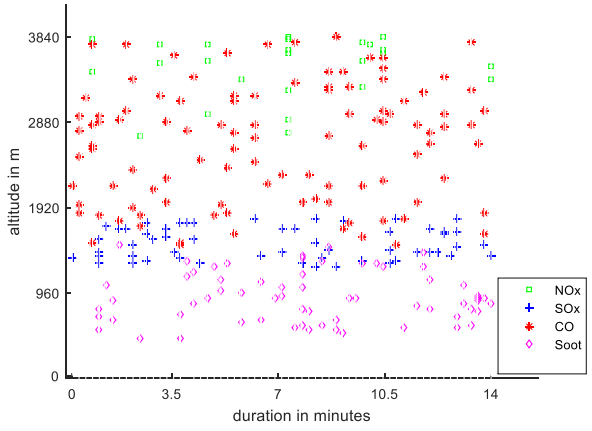

Fig. 2. Aerosols Detected using method 2.

Fig. 2 shows the aerosols – NO_x, SO_x, CO, and soot as classified from our data using the second classification approach. 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 the 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 [23].

6. Results

The results of GLM and regression tree models are presented in this section. The figures displayed in this section show the effects of altitud and duration on the model. The tables of the GLM and regression tree results show that the variable altitude has more of an effect on the differences in the aerosols identified than the duration. Summaries of effects of altitude and duration are presented.

6.1 Altitude coefficient model

The GLM model in Equation 8 shows that the altitude has a significant effect on the types of observed aerosols as seen in Figs. 3 and 4. There appears to be a clear separation of the aerosols based on altitude and a mix of the aerosols when compared with duration. These findings match the results that are published by [2] 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 observed that the dominant aerosol types in terms of aerosol optical depth vary significantly with altitude.

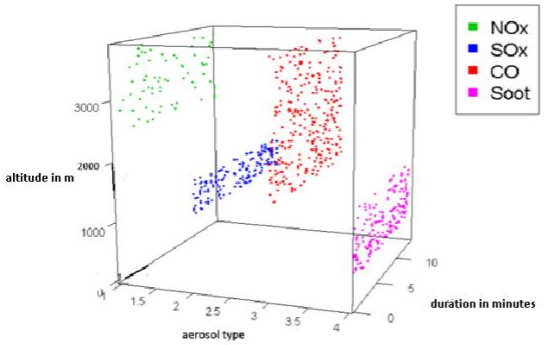

Fig. 3. 3-D view of detected aerosols using method 1.

6.2 Duration varying coefficient model

From Figs. 3 and 4, we observe that duration does not have a significant effect on the type of aerosols classification found in the environment. The plot is quite scattered and does not lead to any pattern.

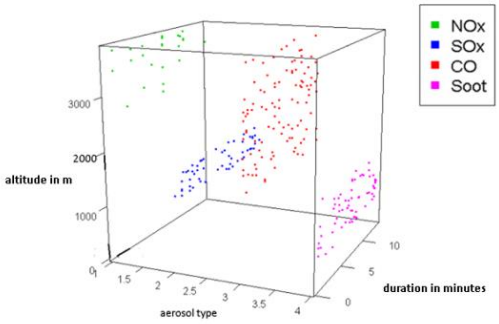

Fig. 4. 3-D view of detected aerosols using method 2.

The following subsections show detail statistical comparison of the two methods (1 and 2) on multiple datasets, which are collected in different days at ODU campus.

6.3 First dataset

The first dataset is collected on July 21, 2017 at ODU campus. The ANOVA Tables for each of the method types are displayed in Tables 3 and 4.

Table 3. ANOVA for Method 1.

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 4. ANOVA for Method 2.

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 analogous analysis shows that in both methods, duration is not found to be a significant indicator of aerosol due to its larger p-value, whereas altitude is due to its smaller p-value as observed in all the ANOVA tables regardless of date and location. Altitude has associated significant p-value reported that is less than 0.05, compared to p-values that are higher than 0.05 for duration. The results show 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 uncertainty 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. This suggests that some of the aerosol at value 4 may be classified as value 5 and vice versa and then the method has less misclassification at aerosol value 5 than method 2.

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. 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 variable importance is as follows: the most important variable is altitude with a value equals 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 equals to 94 which is much greater than 6 the value corresponding to duration.

Tables of classification/count of misclassification were also computed and they indicate 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 and Method 2. 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 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. In aerosol value 4 both methods have some errors, but Method 1 performs better. As in the ANOVA analysis, note that prediction is made for aerosol value 5 even though it was not part of the selection. This again suggests that some of the aerosol at value 4 should have been classified as aerosol value 5.

6.4 Second dataset

Similar analysis was performed on the second dataset collected on August 16, 2016. The ANOVA tables for each of the method types showed significance only for the Altitude variable.

Additionally, the use of the regression tree technique showed similar results as the ANOVA analysis and 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.

6.5 Third dataset

Additionally, the third dataset is collected on August 17, 2016 was similarly analyzed. The ANOVA tables for each of the method types were studied and results were comparable to the first and second datasets. Similarly, the classification/count of misclassification based on the ANOVA analysis using the first method and second method were computed.

The use of the regression tree technique showed similar results as the ANOVA analysis as seen in earlier and 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.

6. Conclusion

The statistical analysis of the two distinct classification methods showed that results and predicted aerosol values were overall equivalent. However, although predictions always lead to misclassifications, method 1 performs better at the classifications of SO_x particle and CO particle. Both methods 1 and 2 perform equally well for NO_x and Soot. Based on this observation, the use of optical parameters of LIDAR (i.e. ratio and backscatter color ratio) to classify these aerosols in the environment and also the use of other parameters (i.e. SSA, SAE, EAE and AAE) to also perform the classification need adjustments. Thus, the optical parameters obtained from the LIDAR data are appropriate for the classification of some aerosols in the environment. The analysis provided consistent 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. Three datasets of LIDAR data were considered for this analysis and altitude was of significant interest within the datasets. In the use of classification and regression tree to find predictive accuracy, classification under Method 1 appears to have a higher accuracy than the classification under Method 2, except for the third dataset. Although we have presented a practical way of describing the data, other algorithms such as the Bayesian methods or the Zero inflated Poisson may be considered as other alternatives for accurate classification of the aerosols.

7. Conflict of Interest

All financial, commercial or other relationships that might be perceived by the academic community as representing a potential conflict of interest must be disclosed. If no such relationship exists, authors will be asked to confirm the following statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

8. Funding and Acknowledgments

This work was supported in part by a Grant # GG11746-146796-01, from the Department of Transportation and the Mid-Atlantic Transportation Sustainability University Transportation Center to the Vision Lab at ODU.

REFERENCES

1. Seinfeld, JH and Pandis SP. Atmospheric Chemistry and Physics, 2nd edn., John Wiley, New York, USA, 2006.
2. Burton SP, Ferrare RA, Hostetler CA, et al. Aerosol classification using airborne high spectral resolution lidar measurements – methodology and examples. Atmos Meas Tech. 2012;5:73-98.
3. Costabile F, Barnaba F, Angelini F, et al. Identification of key aerosol populations through their size and composition resolved spectral scattering and absorption. Atmos Chem Phys. 2013;13:2455-2470.
4. Sugimoto N, Matsui I, Shimizu A, et al. Observation of dust and anthropogenic aerosol plumes in the Northwest Pacific with a two-wavelength polarization lidar on board the research vessel Mirai. Geophys Res Lett. 2020;29:1901.
5. Sasano Y, and Browell EV. Light-Scattering characteristics of various aerosol types derived from multiple wavelength lidar observations. Appl Optics. 1989;28:1670-1679.
6. Omar AH, Winker DM, Kittaka C, et al. The CALIPSO automated aerosol classification and lidar ratio selection algorithm. J Atmos Ocean Tech. 2009;26:1994-2014.
7. Sugimoto N and Lee CH. Characteristics of dust aerosols inferred from lidar depolarization measurements at two wavelengths. Appl Optics. 2006;45:7468-7474.
8. Murayama T, Masonis SJ, Redemann J, et al. An intercomparison of lidar-derived aerosol optical properties with airborne measurements near Tokyo during ACE-Asia. J Geophys Res Atmos. 2003;108:8651.
9. Somekawa T, Yamanaka C, Fujita M, et al. A new concept to characterize nonspherical particles from multi-wavelength depolarization ratios based on T-matrix computation. Part Part Syst Charact. 2008;25:49-53.
10. Clarke A, McNaughton C, Kapustin V, et al. Biomass burning and pollution aerosol over North America: organic components and their influence on spectral optical properties and humidification response. J Geophys Res. 2007;112(D12).
11. Bergstrom RW, Pilewskie P, Russell PB, et al. Spectral absorption properties of atmospheric aerosols. Atmos Chem Phys. 2007;7:5937-5943.
12. Gyawali M, Arnott WP, Zaveri RA, et al. Photoacoustic optical properties at UV, VIS, and near IR wavelengths for laboratory generated and winter time ambient urban aerosols. Atmos Chem Phys. 2012;12:2587-2601.
13. Rusell PB, Bergstrom RW, Shinozuka Y, et al. Absorption angstrom exponent in aeronet and related data as an indicator of aerosol composition. Atmos Chem Phys. 2010;10:1155-1169.

14. Yousef AH, Iftekharuddin K, and Karim M. Towards aerosols lidar scattering plots clustering and analysis. SPIE Defense, Security, and Sensing. International Society for Optics and Photonics. 2013;8718: 87180F-87180F.
15. Heiberger RM and Holland B. Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R, and SAS, Springer Science Business Media LLC, 2004.
16. Gelman A and Hill J. Data Analysis using Regression and Multilevel/Hierarchical Models, Cambridge University Press, Cambridge, U.K., 2007.
17. Faraway JJ. Linear Models with R, CRC Press, Boca Raton, FL: USA, 2005.
18. Faraway JJ. Extending the linear model with R: Generalized linear, mixed effects, and nonparametric regression models, CRC Press, Boca Raton, FL, USA, 2005.
19. Breiman L. Classification and regression trees, the Wadsworth statistics/probability series. Belmont, Calif.: Wadsworth International Group, 1984.
20. Hastie T, Tibshirani R, and Friedman JH. The elements of statistical learning data mining, inference, and prediction: With 200 full-color illustrations, Springer series in statistics, New York: Springer, 2001.
21. Dang R, Yang Y, Hu XM, et al. A review of techniques for diagnosing the atmospheric boundary layer height (ABLH) using aerosol lidar data. Remote Sens. 2009;11(13):1590. [Online]. Available: https://doi.org/10.3390/rs11131590.
22. Zhou T and Popescu SC. Bayesian-decomposition of full waveform LiDAR data with uncertainty analysis. Remote Sens Environ. 2017;200:43-62.
23. SAS Institute Inc. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc, 2011.

Norou Diawara research interests include applied statistics, and it can be classified into two main areas: (1) Spatio-temporal estimation methods and (2) Discrete choice modelling. It includes estimation of the spatio-temporal behaviors, estimation of time to event, and visualization of time series data. With the high volume of data, we bring analytic prospects in medical and engineering applications.

Citation: Alqawba M, Diawara N, Afrifa KG, et al. Statistical analysis and comparison of optical classification of atmospheric aerosol lidar data. Trans Eng Comput Sci. 2021;2(1):119.