

# Tracking Air Pollution in the City of Baltimore, Maryland Utilizing Light Detection and Ranging (LIDAR)

Lucas B. McCullum

**Abstract**—The Chesapeake Bay watershed is one of the most expansive bodies of water in the United States, extending from the southernmost tip of Virginia gradually reducing its reach into upstate New York. The airshed of the Chesapeake Bay, however, is influenced by nearly the entire Eastern and Midwestern United States, consequently necessitating recurrent monitoring of air contamination levels. Currently, there are limited practical methods available to keep a record of the true levels of pollution in the atmosphere. A simple approach to alleviate this limitation is the use of light detection and ranging, or lidar, by means of stationary emplacements located near major toxic hotspot zones. The efficacy of this system can be attributed to the particulates in the atmosphere that can be detected as a result of pollution deriving from different sources. Furthermore, the level of air pollution has been shown to be directly correlated to the atmospheric temperature, subsequently allowing relationships to be established between the atmospheric temperature and various forms of lidar data. Implications of a significant correlation being discovered between atmospheric temperature and any form of lidar data collected would provide insight on the effectiveness of lidar towards detecting and monitoring air pollution. In a broader sense, results from this investigation may be utilized to improve the atmospheric health and quality around the Chesapeake Bay and subsequently, the world.

**Index Terms**—Air quality, lidar, pollution, particulate matter, Baltimore.

## I. INTRODUCTION

Published in 2017, the American Lung Association (ALA) released the State of the Air (SOTA) report detailing the extent of pollution in the United States [1]. In the article, each county in the United States received a rating based on the severity of its pollution levels. According to the 2017 SOTA report, judgment of air quality is done based on two metrics, 8-hour Ozone Concentration and 24-hour PM<sub>2.5</sub> Concentration (Particulate Matter, < 2.5 microns in width). The extent of these two metrics is analyzed by the Environmental Protection Agency (EPA) and given a rating based on the severity of the concentrations. A rating of Green (Good) is the highest followed by Yellow (Moderate), Orange (Unhealthy to Sensitive Groups), Red (Unhealthy), Purple (Very Unhealthy), and Maroon (Hazardous). According to investigations conducted by the ALA between 2012 and 2014, the Baltimore metropolitan (Baltimore County and Baltimore City County) reached 34 high ozone days under the Orange rating, two under the Red, and one under Purple. Furthermore, the Baltimore metropolitan area has also reached one Orange rating for a day of high particle

pollution between the years of 2012 to 2014. The large proportion (~10%) of days spent in hazardous air quality conditions has called for an increase in the methodologies used to measure and track the prominence and growth of these pollutants [1]. One of the fastest growing technologies used in the detection of particulate matter in the atmosphere is lidar. This technology was developed in accordance with the laser in the early 1960s and provides several advantages to similar topographic data collection methods such as interferometry and radio detection and ranging (radar) [2]. The main advantages to the implementation of lidar is its effectiveness in a majority of climate conditions, ability to penetrate a majority of surfaces, and rapidity in the collection of data. Due to these factors, especially lidar's ability to rapidly collect large quantities of data, lidar is often seen as the most cost effective method of collecting and storing atmospheric data.

## II. METHODS

### A. Description of Survey Area

The Baltimore metropolitan area was traditionally founded in 1729 and is the largest city in the state of Maryland by population. It is located approximately 40 miles northeast of the capital of the United States of America, Washington D.C., on the Chesapeake Bay at the head of the Patapsco River. Baltimore is located in a major coastal plain attributing to its elevation ranging from sea level in the Inner Harbor to nearly 150 meters in the northwest outskirts. The total area of the metropolitan is 238.5 km<sup>2</sup> with 29.2 km<sup>2</sup> (12.2%) contributed to water coverage and the rest dedicated to land. Baltimore is located in a temperate climate zone with average temperatures ranging from -1°C in January to 32°C in July [3]. Precipitation is, for the most part, divided uniformly throughout the year with both snowfall and rainfall falling with dependence to the ambient atmospheric temperature.

### B. Methodologies for Acquiring Data

The primary method for acquiring data for this project is the National Aeronautics and Space Administration's (NASA) open-source federated network for Micro-Pulse Lidar (MPL) data collection titled the Micro-Pulse Lidar Network (MPLNET). MPLNET is a contributing network to the World Meteorological Organization's (WMO) Global Atmospheric Watch (GAW) and Aerosol Lidar Observation Network (GALION) aiding to continuously collect data deriving from MPL systems established around the world. The purpose of the data collection focuses on international aerosol and cloud research, climate change and air quality studies, and independent data collected to help support both

Manuscript received May 30, 2017; revised September 26, 2017.

L. B. McCullum is with the University of Maryland, Baltimore County, USA (e-mail: lumccul3@gmail.com).

present and future NASA sub-orbital missions. NASA's Goddard Space Flight Center (GSFC) was the selected source of data due to its proximity to Baltimore and quantity of actively available data [4]. The second primary data source was provided by the National Oceanic and Atmospheric Administration's (NOAA) National Centers for Environmental Information. This source provided data in the version of a comma-separated values (CSV) file on the minimum ambient atmospheric temperature, maximum ambient atmospheric temperature, and average ambient atmospheric temperature for the time period extending from January 1<sup>st</sup>, 2012 until December 31<sup>st</sup>, 2012 at eleven different locations in the state of Maryland. The site chosen for this investigation was Baltimore – Washington International Airport in Baltimore, Maryland due to its proximity to the area of investigation.

### C. Methodologies for Analyzing Data

The software that was used to analyze and interpret the data for this project was MATLAB (matrix laboratory) R2016a (for academic use). MATLAB was created by the technical computing privately-held firm MathWorks in 1984 with the intent to enable the design and development of algorithms, data analytics, visualizations, and numerical computations [5]. The primary tools of the MATLAB software that were utilized were the graphing and statistical analysis capabilities as well as the matrix operation functions. Microsoft Excel 2013 was used to import the CSV files from NOAA and remove data from the dates in which MPLNET could not provide in order to unify the analysis of the data.

### D. Description of Data

The dataset provided by MPLNET provided extensive data collected by the lidar station located at GSFC with 7 dimensions, 53 variables, and 39 attributes. The principal variables used in this investigation consist of the “time”, “altitude”, “nr”, and “nr\_err”.

The “time” variable provides times with a temporal resolution of minutes based on the Gregorian calendar measured in time since 12:00 Coordinated Universal Time (UTC), or 08:00 Eastern Daylight Time (EDT) adjusted for the district of Baltimore, Maryland. In a similar manner to the “time” variable, the “altitude” variable provides altitude values in units of kilometers (km) as measured as height above sea level with 399 values ranging from 0.0875 km to 29.9375 km at a corresponding spatial resolution of 0.075 km. As the primary data used in this investigation, the “nr” variable provides values for the normalized relative backscatter in units of  $\text{Mhz}\cdot\text{km}^2/\mu\text{J}$ . Normalized relative backscatter refers to the normalization of the reflectance energy observed when the energy pulse emitted by the Lidar station encounters a form of particulate matter and diffusely reflects in the opposite direction of its emission and is measured at the emission site upon its return. The final variable investigated in this report is “nr\_err” which corresponds to the uncertainty pertaining to the normalized relative backscatter measured in  $\text{Mhz}\cdot\text{km}^2/\mu\text{J}$ . Errors in the normalized relative backscatter derive from the noise during measurement and insufficient calibration [6]. The values for all of the stated variables are organized in matrices with

dimensions of  $399 \times 1440$  where the row and column vectors pertain to the corresponding altitude and time of day respectively.

## III. RESULTS AND DISCUSSION

The primary metric when analyzing climate is ambient atmospheric temperature and must be the principal variable that is studied when investigating the effects of climate on air particulate concentration. In this period of investigation, the ambient atmospheric temperature in Baltimore ranged from  $-10.56^\circ\text{C}$  to  $40.00^\circ\text{C}$  with a nearly uniform unimodal distribution. By conducting a Pearson Product-Moment Correlation Coefficient (PPMCC) analysis between the Normalized Relative Backscatter (NRB) at a specific altitude and time of day with the corresponding ambient atmospheric temperature on that day, one can observe the strength of the effects of ambient atmospheric temperature on NRB. Following the execution of this analysis, it is paramount to visualize the results in order to visually inspect any trends (Fig. 1).

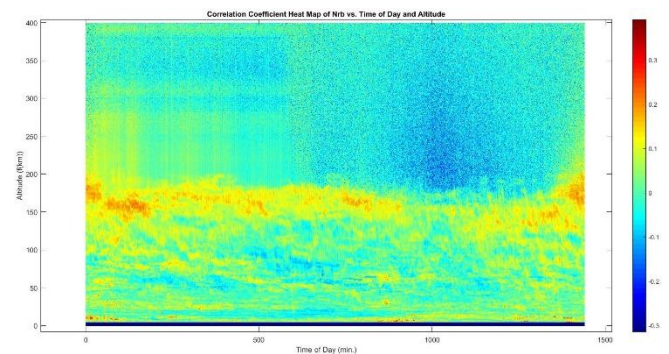


Fig. 1. Heat map representation of the correlation coefficient between NRB and average daily temperature vs. altitude.

It can be seen that the highest levels of correlation coefficients appear in horizontal bands as altitude increases until an altitude of approximately 15 km, indicating that the effects of ambient atmospheric temperature on NRB are felt most strongly in intervals of altitude until a critical altitude is reached. It can also be seen that the altitudes with the strongest correlation coefficients appear just below the critical altitude of 15 km as well as just above sea level at less than 1 km. To confirm these propositions, a graphical representation of the averaged correlation coefficient with respect to time of day is needed to determine the bands of altitudes with the greatest intensity. Following this analysis, it can be confirmed that the strongest correlations exist just below the critical altitude from 10 – 15 km as well as fairly uniform intervals of altitude increasing from sea level (Fig. 2).

In order to quantify the trends in correlation coefficient that are suspected between NRB and ambient atmospheric temperature, the most effective method of visualization is to create a plot of correlation coefficient with respect to altitude. Furthermore, to increase the level of clarity, a color map (darker color implying earlier in the day) has been applied to differentiate the values of the correlation coefficients at different times of day with respect to the altitude (Fig. 3).

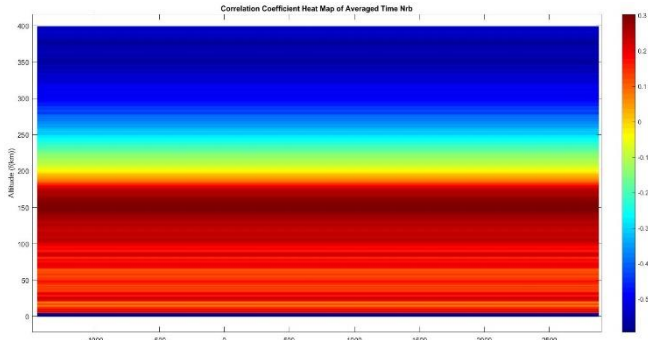


Fig. 2. Heat map representation of NRB with respect to averaged time of day and altitude.

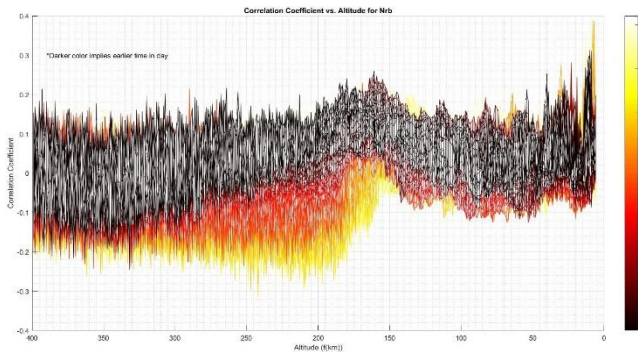


Fig. 3. Scatter plot representation of the correlation coefficient between NRB and average daily temperature vs. altitude.

It should be noted that due to software restrictions and matrix orientation, the altitude increases from the right side of the plot to the left. Continuing past this inconvenience, it can be seen that the correlation between NRB and ambient atmospheric temperature appears to decrease as the time of day increases (again, beginning at 08:00 EDT), especially in the range of 10 – 20 km in altitude. To mitigate any possible sources of error for these results, it is essential to analyze the dataset pertaining to the NRB error. This dataset will provide the associated errors computed with each of the corresponding NRB values. As with the NRB data, the first step to analyze the relationship between NRB error and ambient atmospheric temperature is to determine the corresponding PPMCC generated values. Following the execution of this analysis, it is important to visualize the results by, in this case, generating a scatter plot of each correlation coefficient with respect to its corresponding altitude and time of day (Fig. 4).

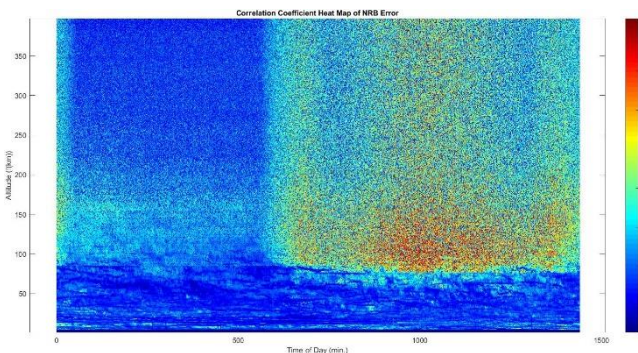


Fig. 4. Heat map representation of the correlation coefficient between NRB error and average daily temperature vs. altitude.

Following the execution of this analysis, there are several interesting features to notice. The first is that the correlation between NRB error and ambient atmospheric temperature

appears to sharply increase above the altitude of around 7 – 8 km. Secondly, the correlation between NRB error and ambient atmospheric temperature appears to sharply decrease during the hours of 08:00 EDT and 16:00 – 18:00 EDT. In similar examination as NRB, the next step in order to quantify the relationship between NRB error and ambient atmospheric temperature is to generate a scatter plot of the correlation coefficients with respect to its corresponding altitude. Furthermore, a color map (darker color implying earlier in the day) has been applied to differentiate the values of the correlation coefficients at different times of day with respect to the altitude (Fig. 5).

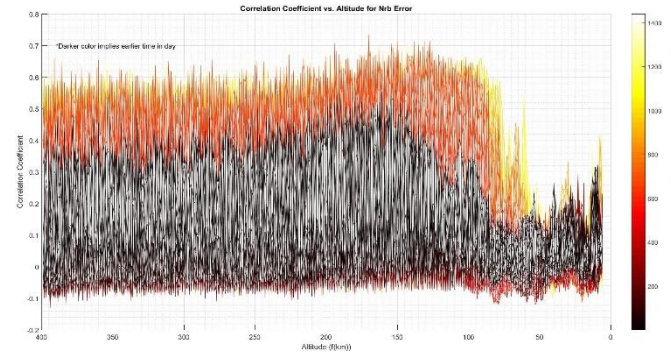


Fig. 5. Scatter plot representation of the correlation coefficient between NRB error and average daily temperature vs. altitude.

As with the plot generated for the correlation coefficients corresponding to the NRB values, due to software restrictions and matrix orientation, the values of the altitude increases from the right side to the left side. It can be seen that, unlike the results derived from analyzing NRB, the correlation between NRB error and ambient atmospheric temperature appears to increase as the time of day increases. Furthermore, this trend is seen most prominently after an altitude of around 6 km is reached with the highest correlation coefficients reaching nearly 0.75. In order to further analyze the dataset for NRB, a matrix of the variance for each of the NRB values was generated and then correlated with the ambient atmospheric temperature to generate PPMCC values for each altitude. By plotting the values of the correlations coefficients with respect to their corresponding altitudes, an analysis can be done on any trends that appear (Fig. 6).

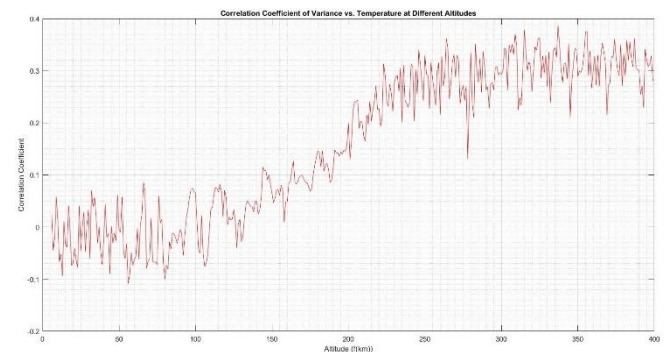


Fig. 6. Line plot of the correlation coefficient between NRB variance vs. altitude.

After brief observation, it can be noted that the variance in the NRB values at lower altitudes is much less strongly correlated to the variance in the NRB values at higher altitudes. This is in agreement with the NRB error analysis done in Fig. 4. In order to investigate the validity of the



correlation coefficient values reached after conducting the PPMCC for both NRB and NRB error versus the ambient atmospheric temperature, a single case analysis was done regarding their respective altitudes and times of day. This analysis yields the greatest and least correlation coefficient values as well as their corresponding altitudes and times of day. In the case of investigating the correlation between NRB and ambient atmospheric temperature, the greatest value of correlation coefficient is 0.3918 at 0.6125 km altitude and 23:47 EDT, and the least value of correlation coefficient is -0.3144 at 18.1625 km altitude and 00:27 EDT. These values assist in confirming that the effects of ambient atmospheric temperature on NRB are more directly correlated at lower altitudes than at higher altitudes. In the case of investigating the correlation between NRB error and ambient atmospheric temperature, the greatest value of correlation coefficient is 0.7343 at 12.7625 km altitude and 19:36 EDT, and the least value of correlation coefficient is -0.1308 at 29.5625 km altitude and 10:44 EDT. These values further assist in confirming that the effects of ambient atmospheric temperature and NRB error are seen more strongly at lower altitudes.

#### IV. CONCLUSION

In their paper titled "Influence of Temperature, Relative Humidity and Seasonal Variability on Ambient Air Quality in a Coastal Urban Area", Jayamurugan, R., Kumaravel, B., Palanivelraja, S., & Chockalingam, M. P. establish a correlation between ambient atmospheric temperature and concentrations of greenhouse gases [7]. The conclusion reached in their paper agrees with the results derived here through analysis of lidar data, such as NRB and NRB error. Furthermore, in accordance to the results reached following the investigation of NRB data, a stronger correlation between ambient atmospheric temperature and particulate matter is seen at lower altitudes. Respectively, NRB error has been shown to exhibit a higher correlation with ambient atmospheric temperature at higher altitudes and periods of

night, thus reducing the concern for undependable NRB data. Additionally, it has been shown that the variance in NRB values is more strongly correlated with ambient atmospheric temperature at higher altitudes, supporting the conclusion reached through the analysis of NRB error. Although the results derived at in this report may prove significant towards climate change efforts, further investigation must be conducted in order to generalize the findings derived at in this paper to a larger body of geographic locations and time periods.

#### ACKNOWLEDGMENT

Acknowledgments for this report can be credited to the University of Maryland, Baltimore County and its course INDS 430 – An Interdisciplinary Approach to Satellite and Geospatial Fusion lead by Dr. McAlpine and Dr. Montgomery.

#### REFERENCES

- [1] State of the Air, American Lung Association, 2017.
- [2] G. G. Goyer and R. Watson, "The laser and its application to meteorology," *Bulletin of the American Meteorological Society*, vol. 44, no. 9, pp. 564–575, September 1963.
- [3] Baltimore, Maryland. (February 24, 2017) [Online]. Available: [http://www.newworldencyclopedia.org/entry/Baltimore, Maryland](http://www.newworldencyclopedia.org/entry/Baltimore,_Maryland)
- [4] (February 4, 2017). [Online]. Available: <https://mplnet.gsfc.nasa.gov/>
- [5] Company Overview. (April 18, 2017). [Online]. Available: <https://www.mathworks.com/company/aboutus.html>
- [6] R. Francesc *et al.*, "Backscatter error bounds for the elastic lidar two-component inversion algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, 2012.
- [7] R. Jayamurugan *et al.*, "Influence of temperature, relative humidity and seasonal variability on ambient air quality in a coastal urban area," *International Journal of Atmospheric Sciences*, 2013, pp. 1-7.



**Lucas B. McCullum** is currently working towards a degree in mechanical engineering, applied mathematics, and financial economics from the University of Maryland, Baltimore County (UMBC). He plans to pursue a master's degree in either efficient energy systems or applied physics. His current research focuses on energy harvesting and biomedical engineering.