January 2021

Development Of A Deep Neural Network Based Method For Quality Control Of Modis Maiac Aerosol Data For Aerosol Modeling Applications

Joel Braxton Aldridge

DEVELOPMENT OF A DEEP NEURAL NETWORK BASED METHOD FOR QUALITY CONTROL OF MODIS MAIAC AEROSOL DATA FOR AEROSOL MODELING APPLICATIONS

by

Joel Braxton Aldridge

Bachelor of Science, University of North Dakota, 2019

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfilment of the requirements

for the degree of

Master of Science

Grand Forks, North Dakota

December
2021

This thesis, submitted by Joel Braxton Aldridge in partial fulfillment of the requirements for the Degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

_____
Dr. JiangLong Zhang

_____
Prof. Michael Poellot

_____
Dr. Matthew Gilmore

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

_____
Chris Nelson
Dean of the School of Graduate Studies

_____
Date

PERMISSION

Title	Development of a Deep Neural Network Based Method for Quality Control of MODIS MAIAC Aerosol Data for Aerosol Modelling Applications
Department	Atmospheric Sciences
Degree	Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Joel Braxton Aldridge

November 11, 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Jianglong Zhang, and my committee members, Professor Michael Poellot and Dr. Matthew Gilmore for guiding me through my research and studies. I would also like to thank my fellow students who helped work through challenges in my research, especially Jared Marquis, Nikki Marquis, and Shawn Jaker for their invaluable support and advice.

A huge thank you to my wife, Sayjen, for the encouragement, love, and motivation, and for letting me talk through challenges in my research and caring for me through the difficult times.

Thank you to my brotherhood of friends in Texas, Coleman, Cameron, Wiggles, Brasovan, Noah and Paulukaitis. You all kept me going with countless phone calls, memes, deep conversations, and visits during breaks. I also appreciate the warm welcome home and continued support.

I will always be gratful to my classmates who bonded with me over the cold weather, tough course work, and love for weather, and shared interests, especially Kyle Peterson. You have made my time as a University of North Dakota student extremely worth while and something I will be forever thankful for.

My gratitude also extends to friends across the country, and internationally who enriched my experience as a student and helped me gain a better perspective and a love for different cultures.

Finally, I would like to thank the 119th Wing North Dakota Air National Guard, especially the 176th ISRG, for bringing me to North Dakota in the first place, providing a way for me to fund my academics, and cultivating in me a hard work ethic that is defined by integrity, service, and comradery. I'll always be a Happy Hooligan at heart thanks to you all.

This thesis is dedicated to my wife Sayjen and my son August who inspire me to dream big and love deeply. Thank you for your unwavering support and motivation, you are the reason I do what I do.

ABSTRACT

Quality-assured satellite aerosol data have been shown to improve aerosol analysis and forecasts in Chemical Transport Models. However, biases present in the satellite-based aerosol data can also introduce non-negligible uncertainties in the downstream aerosol forecasts and impact model forecast accuracy. Therefore, in this study we evaluated uncertainties in Moderate Imaging Spectroradiometer (MODIS) Multi-Angle Implementation of Atmospheric Correction (MAIAC) aerosol products and developed a deep neural network (DNN) based method for quality control of Terra and Aqua MODIS MAIAC Aerosol Optical Depth (AOD) data using the version 3 level 2 AErosol RObotic NETwork (AERONET) data as the ground truth. This method is done using 14 years of Aqua MODIS (2002-2016) and 16 years of Terra MODIS (2000-2016) MAIAC data which are collocated with the AERONET observations. The resulting trained network, which is tested on one year of Aqua/Terra data, can detect and significantly reduce noisy retrieval in MAIAC AOD data resulting in an approximate 31%/27% reduction in Root-Mean-Square-Error in Aqua/Terra MODIS MAIAC AOD with an associated 14%/16% data loss. A sensitivity study performed in this effort suggests that reducing the number of output categories and hidden layers can significantly improve performance of the deep neural network in this case. This study suggests that DNN can be used as an effective method for quality control of satellite based AOD data for potential modeling applications.

CHAPTER 1

INTRODUCTION

Atmospheric aerosols, which are fine particles suspended in air, originate from natural sources such as dust, sea salt, volcanic ash, and wildfires as well as from anthropogenic sources like fossil fuel burning (Rabha and Saikia 2020). Due to their size (generally 100nm to 10μm) aerosols can remain aloft in the atmosphere for a day or several days (Williams et al. 2002). During this time, aerosols may be transported long distances, absorbing and redirecting solar/ terrestrial radiation along the way. For example, studies have shown aerosols frequently travel from the Sahara across the Atlantic Ocean and impact the United States, and winds often carry a mixture of dust and biomass burning aerosols from Asia over Japan and into the central Pacific Ocean (Chin et al. 2007).

Aerosols can absorb and redirect incoming solar and outgoing terrestrial radiation (Kaufman et al. 2002). This interaction and its effects on earth's climate are of interest to the climate science community. Besides their climate impacts, aerosol particles attenuate solar radiation and thus affect visibility (Eck et al. 1998). Near surface aerosol particles with diameters less than 2.5 μm can cause respiratory diseases (Xu et al. 2021; Maji et al. 2017; Brunekreef and Holgate 2002) and are considered a major sources of air pollution (Particulate Matter with diameters less than 2.5 μm or PM2.5). For these reasons, aerosol particles are studied by both modeling and observation-based methods.

Due to the spatial and temporal extent of aerosols, remote sensing based methods are frequently used for monitoring and quantifying the spatial coverage as well as transport of

atmospheric aerosol particles. Of the many satellite-derived aerosol properties, Aerosol Optical Depth (AOD), which measures the extinction of solar/terrestrial radiation due to scattering and absorption by aerosols, is widely used for measuring the significance of aerosol events of a region. AOD is a unitless parameter with values typically ranging from 0 to 5 (or greater, at the 550 nm spectral channel). For a pristine marine environment, AOD values are around 0.06 (at 550 nm), with AOD values of above 0.2 considered as aerosol polluted skies.

One of the direct applications of satellite derived AOD data is for assisting aerosol analyses and forecasts through aerosol data assimilation. Past studies suggest that the accuracy of aerosol analyses and forecasts from chemical transport models can be drastically improved through the assimilation of satellite derived AOD into those models. For example, Zhang et al. (2008) reported a 40% reduction in absolute error in modeled AOD at the analysis time and ~20% reduction for 48-hour forecasts. However, it was found that significant biases and uncertainties could exist in satellite derived AOD data that associate with cloud contamination, observational conditions, poor surface characterization, and the use of inaccurate aerosol model retrievals exist in the satellite aerosol retrievals (Zhang and Reid 2006). Those biases and noises in satellite AOD data can introduce non-trivial biases in downstream aerosol analyses and forecasts from chemical transport models. Still, recent studies suggest that the accuracy of aerosol analyses and forecasts from chemical transport models can be improved through the assimilation of high-quality satellite-based AOD data (Lyapustin et al. 2012; Lee 2019), or through assimilation satellite AOD that are carefully quality controlled (e.g. Zhang et al., 2008).

While AOD values are retrieved from both passive and active sensors, such as Multi-angle Imaging SpectrorRadiometer (MISR) and Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), one of the corner stone satellite AOD datasets for aerosol modeling and aerosol data assimilation are the AOD data from the Moderate Imaging Spectroradiometer (MODIS) instrument. Note that there are two MODIS instruments currently in operation, with one on board the Aqua

satellite that has an equatorial cross time of ~1:30 pm and the other on board the Terra satellite that has an equatorial cross time of ~10:30 am. Both Terra and Aqua are polar orbiting satellites that observe the entire earth every one to two days and allow for a near continuous collection of AOD on large scales which is useful for visibility, air quality and climate studies involving aerosols.

There are three standard AOD products from MODIS: the Dark Target (DT); Deep Blue (DB); and the Multi-Angle Implementation of Atmospheric Correction (MAIAC) aerosol products. AOD retrievals from the DT products are limited to non-bright surface and utilize the fact that aerosol plumes increase Top-Of-Atmosphere albedo over non-bright regions such as global oceans and vegetated areas. AOD retrievals from the DB products are derived using blue channels (e.g. ~400 nm) where bright surfaces such as desert regions at the visible channel look dark at blue channels. Note that Both the DT and DB aerosol products have been evaluated for their application in aerosol modeling and aerosol data assimilation, and data-assimilation quality DT and DB data have been developed through stringent QA and QC steps (e.g. Shi et al. 2011; Hsu et al., 2013).

The MODIS MAIAC aerosol product is a newer aerosol dataset. Different from the MODIS DT and DB aerosol data, MODIS observations are layered over time in the MODIS MAIAC aerosol retrievals to characterize the surface properties. This allows for high resolution AOD of 1km over traditional dark and vegetative surfaces as well as bright deserts (Lyapustin et al. 2011). Yet, MAIAC AOD retrievals have not been adopted for aerosol analyses and forecasts. This is because the MODIS MAIAC data have not been thoroughly evaluated for the application of AOD assimilation. Clearly, MODIS MAIAC data need to be evaluated and quality control processes need to be developed for constructing data-assimilation quality MAIAC data for use in data assimilation.

Therefore, in the first part of the study, we evaluate the uncertainties in MAIAC AOD retrievals using 16 years of Terra MODIS (2000-2016) and 14 years of Aqua MODIS (2002-2016) data. The MODIS MAIAC data are collocated with ground-based AErosol RObotic NETwork (AERONET) level 2 Version 3 data and inter-compared for examining uncertainties in MODIS

MAIAC data with respect to observed conditions for 6 selected regions including Africa, China, Europe, Mid-Asia, North America, and South America. Uncertainties in MAIAC AOD related to cloud contamination, surface characteristics, and aerosol models used in the retrieval process are explored. Knowing the uncertainties in MAIAC AOD data, an innovative neural network (NN) based approach is developed in the second part of this study for the quality control of MAIAC data. The purpose of this approach is to develop a reliable and efficient process to quality-assure MAIAC data, using a NN, for data assimilation applications.

CHAPTER 2

DATASET AND METHODOLOGY

2.1 MODIS MAIAC Data

The Moderate Resolution Imaging Spectroradiometer (MODIS) instrument is a passive-based imager onboard the Aqua and Terra satellites. The MODIS instrument provides total earth coverage every 1 to 2 days in 36 spectral bands with resolutions from 1km to 250m depending on the band. MODIS provides large swaths (~2,330km) of observable data via a rotating mirror design. The combination of total earth coverage, large scanning swaths, and multiple band options makes it ideal for the characterization and observation of atmospheric aerosols (Levy et al. 2010).

The retrieval of aerosol properties using the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm is based on the MODIS Level 1B data. MAIAC is an encoded generic algorithm which uses pixel- and image-based processing with a time series analysis of MODIS data to produce more accurate aerosol retrievals as well as other atmospheric products (Lyapustin et al. 2011; 2012; 2018). The algorithm corrects for atmospheric effects and detects aerosols by placing MODIS data in a fixed grid at 1km resolution in a running que (Wolfe et al., 1998). The MAIAC algorithm stores from 4 (at poles) to 16 (at the equator) days of past observations in operational memory, supplying a detailed knowledge of the surface conditions. This allows for the observation of an area over time using polar-orbiting satellites as if they were geostationary and benefitting from the higher resolution imagery due to the lower orbiting altitude. Using multiple days allows for the correction of atmospheric effects by providing a way to filter out

transient conditions such as weather systems and clouds. Additionally, bidirectional reflectance (BR) can be collected from the multiple days of MODIS data as many different angles of each location are collected (Lyapustin et al. 2011). MAIAC retrieves the BR factor using the time series of MODIS data in an image-based versus a pixel-based process. This allows the precise characterizing the surface reflectance during the aerosol retrieval process (Lyapustin et al. 2011). The spectral regression coefficient (SRC) is also derived by relating the BR factor for the blue (0.47μm) and shortwave infrared (2.1μm) bands. Using all of this, AOD values are thus derived, based on the multi-day que of MODIS data, for a given area after characterizing the BR and SR factors. The MAIAC aerosol retrieval algorithm can provide higher spatial resolution retrievals, on the order of 1 km, compared to traditional methods at 3-10 km resolution aerosol products from MODIS DT and DB.

For processing, the MAIAC data from the polar orbiting satellites (Aqua and Terra) are split. The Aqua dataset downloaded for this study includes data from 2002-2016 and the Terra dataset from 2000-2016. The geographic areas covered by both satellites in this study are Africa, China, Europe, Mid-Asia, North America, and South America. MAIAC products are downloaded from the Level 1 and Atmospheric Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) site. As an example, Figure 1a shows the Terra MODIS true color image over Africa on August 4, 2020. Smoke plumes over central Africa are clearly visible. Figure 1b shows the corresponding MAIAC AOD retrievals. Note that no retrievals are available over MODIS detected cloudy regions. Still, retrievals are available over the Saharan desert, where the surface is bright at the visible spectrum.

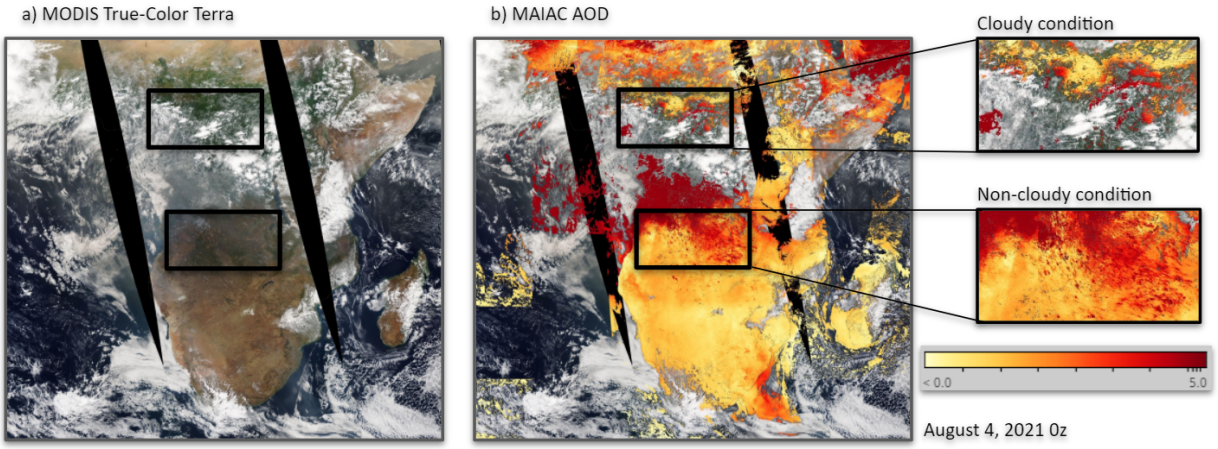

Figure 1: Example Modis True-color image (a) with corresponding MAIAC AOD overlay (b) under partly cloudy (top) and clear (bottom) conditions on August 4, 2021 (burning season) over central and southern Africa. These images are obtained from the NASA worldview webpage (https://worldview.earthdata.nasa.gov/).

2.2 AERONET Data

The use of MAIAC data in this project is evaluated against the AErosol RObotic NETwork (AERONET) data (Holben et al. 1998). AERONET is a global, ground-based sun photometer network which provides 15 minute AOD measurements at eight wavelengths (340nm, 380nm, 440nm, 500 nm, 675 nm, 870 nm, 1020 nm, and 1640 nm) for direct sun (solar irradiance) along with other optical properties (Holben et al. 1998). The eight wavelengths are measured via a rotating filter wheel and a pre-programed sequence from 7:00 am to 7:00 pm at each location (local time) across the globe. AOD is calculated using spectral extinction of the direct solar beam per wavelength referencing Beer-Bouguer's Law. Estimates are made for Rayleigh scattering, ozone absorption, and pollution gases and used to remove attenuation and isolate the AOD (System Description - Aerosol Robotic). This study uses Version 3 level 2.0 (cloud screened and quality assured) AERONET data. Version 3 implements a full automation of cloud screening and instrument anomaly quality control which allows for near real time data (within one month) and estimated uncertainty (Giles et al. 2019). The uncertainties in AERONET AOD are on the order of 0.02 (Giles et al. 2019).

2.3 Collocation of MAIAC and AERONET

The MAIAC AOD observations are validated using the AERONET AOD observations. This is done via a comparison of collocated observations. For a MAIAC and AERONET observation to be considered collocated, they must be within a predetermined spatiotemporal window of 0.3 degrees distance and $\pm$30 minutes. Additionally, collocated observations are compared on similar wavelengths even though AERONET and MAIAC AOD are not collected in the same wavelengths. MAIAC AOD is collected at 470nm and 550nm wavelengths. This study uses only the 550nm wavelength as it is one of the most commonly used wavelengths for quantifying aerosol properties. To compare using the same wavelength, AERONET data that fall within the spatiotemporal window are examined for valid data. While AERONET collects AOD at eight wavelengths, certain wavelengths for a given observation may be flagged as invalid by the AERONET V3 automated quality control process (System Description - Aerosol Robotic). Once the nearest two valid AERONET wavelengths to compare with MAIAC AOD of 550nm are found, a conversion using the Angstrom exponent (α), which is commonly used to highlight the wavelength reliance of AOD and to gain information on the aerosol size distribution, is performed (Eck et al. 1999). Here, the AOD value for a given wavelength is interpolated assuming the Angstrom exponent (α) remains unchanged. Given two AERONET AOD measurements (τ) at different wavelengths (λ), the α value can be derived using equation 2.1.

$$\alpha = -\frac{ln\left(\frac{\tau_1}{\tau_2}\right)}{ln\left(\frac{\lambda_1}{\lambda_2}\right)} \qquad (2.1)$$

The AERONET AOD at 550nm is thus derived by reversing equation 2.1 and by assuming α is a constant within the given wavelength range. This calculation is performed for all valid collocated data.

The collocated data, consisting of twenty-nine variables (Table 1), are formatted into Network Common Data Form (NetCDF) files, with one NetCDF file for each year containing all of the

respective collocated files for that year. Some of the key parameters listed in Table 1 are latitude and longitude for both AERONET and MAIAC, AOD taken at a variety of wavelengths, and the viewing geometry variables for MAIAC. Each NetCDF file contains one year of collocated observations for a single study region. The NetCDF files contain all information collected by MODIS MAIAC observations for both 470 nm and 550nm and all measurements from AERONET for each of the eight collection wavelengths. This NetCDF database serves as the input data for the Neural Network and was chosen so that large amounts of data could be stored and easily referenced in single files. The NetCDF format is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. It is also a community standard for sharing scientific data maintained by Unidata for programming interfaces such as Python (Unidata).

Table 1: Collocated file contents, including both AERONET (left) and MAIAC (right)

AERONET	**MAIAC**
Julian Day	Latitude
AOD 1640	Longitude
AOD 1020	AOD 470
AOD 870	AOD 550
AOD 675	AOD Uncertainty
AOD 667	Fine Mode Fraction
AOD 555	Column Water Vapor
AOD 551	Injection Height
AOD 532	QA rating
AOD 531	AOD Model
AOD 500	cos Solar Zenith Angle
Water (cm)	cos Viewing Zenith Angle
Latitude	Relative Azimuth Angle
Longitude	Scattering Angle
	Glint Angle

2.4 Methods for Analyzing MAIAC and AERONET data

Prior to developing a deep-learning network-based quality control method, an analysis of the MAIAC AOD data is performed to gain a comprehensive understanding of the data, Initially, a one-to-many analysis (one AERONET observation is compared to multiple MAIAC observations that fit in the collocation window) is applied to evaluate the performance of MAIAC AOD data against AERONET. This is done by taking any MAIAC data points which fall within the spatiotemporal window of an AERONET measurement and comparing them all to that one AERONET measurement for a given time. A similar analysis method was also taken by Cheng et al. (2012) to compare the AOD from multiple satellites with AERONET as well as by Kokhanovsky et al. (2007) to obtain the difference in instantaneous AOD retrievals from different algorithms. The one-to-many results can reveal occurrence and extent to which the MAIAC and AERONET AOD differ. Only one year (2015) of data is used for this analysis due to the large data volume.

Next, a one-to-one analysis is performed. In this case, an AERONET AOD observation is compared to the mean value of all MAIAC AOD retrievals within the collocation bounds. This allows for the comparison of two single values, namely AERONET and mean MAIAC AOD for a given time. The one-to-one analysis is performed based on region and satellite (Aqua and Terra) and performed for the entire time frame of downloaded data (2002-2016 Aqua, 2000-2016 Terra). In addition, the linear correlation between the mean MAIAC (AOD_{MAIAC}) and AERONET ($AOD_{AERONET}$) AOD, as well as the Root Mean Square Error (RMSE) of the MAIAC AOD are computed. The MAIAC AOD RMSE values are computed using Equation 2.2.

$$RMSE = \sqrt{\frac{1}{n}\Sigma_n (AOD_{AERONET} - AOD_{MAIAC})^2} \qquad (2.2)$$

Both RMSE and linear correlation are also used for evaluating the performance of the DNN in a later section. An increase in linear correlation and a decrease in RMSE between MAIAC and AERONET after the DNN process is indicative of the DNN properly filtering the MAIAC data.

2.5 Neural Network

2.5.1 The Basic Structure of a Deep Neural Network System

An innovative neural network (NN) based approach is developed for the quality control of MAIAC data. This approach uses a NN which is trained on a combined MAIAC and AERONET dataset to select reliable MAIAC retrievals. This NN based implementation uses the basic principles of a NN where input values are passed to a layer made of connected nodes, which give an output to the next layer, and finally render an outcome, in this case, a classification. A Deep NN (DNN, Figure 2) is any network of neurons (nodes) which has more than one layer. At each layer a number of nodes exist with each having three basic attributes: inputs, weights, and a bias. These three values are used to pass information along a unique weighted connection between that node and the next layer of nodes. As that information is being passed it goes through the activation function for the layer it is leaving. The DNN uses this process of passing inputs to nodes, calculating outputs, and then passing them through activation functions along weighted connections to filter information through the network. This filtering process is what can be tuned to achieve the desired effect the network has on the input data. The individual settings used to tune the DNN are referred to as parameters. The parameters used for tuning in this study not already discussed include the following. Hidden Layers is a parameter which denotes the number of model layers that pass data from the input to the output layer. Nodes refers to the number of neurons per layer that process the data. Classes is a parameter denoting the number of possible outputs. Epochs is a term to denote how many iterations occur during a given training instance where the learned weights and biases are back propagated during training.

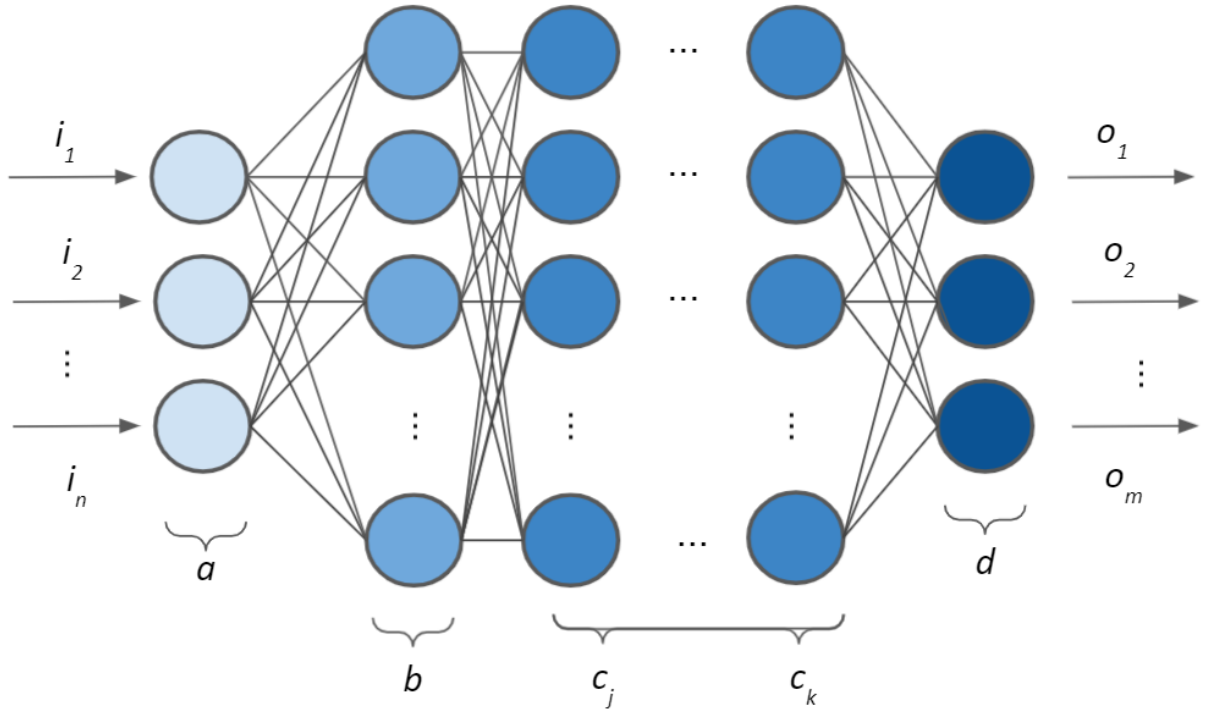

Figure 2: Structure of the Sequential Deep Neural Network where inputs i_1 through i_n (left) are given to the input layer (a) , then to the formatting layer (b) which normalizes the data for the hidden layers c_j through c_k which feed eventual outputs to output layer (d) with a number of outputs *o_1 through o_m (right).*

2.5.2 The Implementation of a DNN for Quality Control of MODIS MAIAC Aerosol Retrievals

For this study the DNN predicts the difference between a MAIAC AOD value and the collocated AERONET AOD. This is known as the target. The DNN is trained on the actual differences between a training set of MAIAC and AERONET data, then tested on new data. This training and testing is accomplished via a step by step process.

The process of this DNN is broken down into four components: initializing the data, building the network and its layers, running and evaluating the network, and making predictions. Data initialization begins with defining the input data, arbitrarily referred to as X and Y in the network code. The input values stored in X are all input data relating to MAIAC retrievals which are listed in Table 1. Y is filled with the differences between collocated AERONET and MAIAC AOD that are

linked to values in X. These differences are placed in bins according to their magnitude and sign (positive or negative) where the bins have increasing width departing from zero, as lower data density is expected for higher AOD values. Therefore, smaller differences are assigned to corresponding bins close to zero and larger differences are in bins farther from zero. This creates a distribution of differences between MAIAC and AERONET where the MAIAC values corresponding to the smallest differences (closest to zero) are considered good data. The values stored in Y are the target values the DNN is trained to predict, as the goal is for the network to correctly predict the difference between any MAIAC and AERONET AOD. This allows for any threshold to be set for the variance of MAIAC from a given AERONET site in future use. Next, the data initialization ends with X being split and normalized. X is split into a testing and training set so that a portion of the data remains unseen by the network for later testing. Normalization of the X values is done using the L2 normalization so that all values in X fall within a range while retaining distinction, making it much easier for the DNN to process.

The next component of the DNN process is the network structure. This network is built as a sequential model, meaning information is passed from start to finish though a stack of layers for every epoch. The stack of layers used in this DNN consists of an input layer, a number of hidden layers, and an output layer as shown previously in Figure 2. The input layer flattens the multidimensional matrix of input data into an array to pass through the hidden layers. The hidden layers are densely connected layers, meaning at each node the performed operation is

$$output = activation[(input \cdot weight) + bias] \tag{2.3}$$

$$activation: \quad f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{2.4}$$

where Equation 2.3 describes how the output of a node is calculated and Equation 2.4 refers to the activation function used in the calculation. The activation function used is a REctified Linear Unit (ReLU) function (Equation 2.4) which is a piecewise linear function that will output the input

directly if it is positive, otherwise, it will output zero. This function is used for its balance of having greater sensitivity than a traditional step function while avoiding the vanishing gradient issues of a sigmoid function. The ReLU activation function is used for the input layer and all hidden layers. Each layer has a specified number of nodes and is connected to the next layer. The output layer is important and is separate from the hidden layers. It specifies the number of classifications desired and frames the answer the DNN is able to provide. This requires a clear understanding of the problem being solved as well as the input data. This output layer has classifications (nodes) equal to the number of bins which are storing the difference between the MAIAC and AERONET AOD data. This is chosen so that the output layer will generate a bin location for the predicted difference.

The next component of the DNN is running and evaluating the network. This is done by choosing an optimizer and a learning rate (LR) and tracking loss and accuracy. Choices for the optimizer and learning rate are manually changed as part of the training process. For this DNN the optimizer Adam is used, which adjusts the LR throughout training with an initial LR of 0.001 (Kingma and Ba 2014). The LR determines how fast the optimal weights for the model are calculated by the optimizer. A smaller LR may lead to more accurate weights, but it may take more time to compute than is worthwhile. The DNN system can also be trapped into a regional minimum rather than the global minimum by using a small LR. Running the model is done through the compile and fit function. These initiate the sending of input data (training portions of X and Y) and the activation of the layers. At this point the DNN makes a prediction for each (attempts to fit each) MAIAC AOD value as to how close it is to the true AOD. Each prediction is checked using the collocated AERONET data. Errors in predicted values are propagated back to DNN for altering various weights in the system. The process is repeated until errors in the predicted values are smaller than a given threshold for a given set of inputs. This process is how the DNN is trained.

Once a suitable training accuracy is reached by adjusting the DNN parameters, the network may be tested. Testing is performed by evaluating the trained DNN system using the collocated MAIAC and AERONET data that are not used in the training process as described above.

Finally, using the pre-trained DNN system, predictions are achieved by passing MAIAC data into the system, and the DNN can predict the difference between the provided MAIAC and AERONET AOD. The goal is for the network to have learned how to use the other variables associated with a MAIAC AOD measurement such as MODIS latitude and longitude, and viewing geometry, to recognize valid MAIAC AOD values. A prediction estimates the difference between each MAIAC AOD value and its corresponding AERONET value. For MAIAC AOD values similar to AERONET, the DNN estimates are placed in a bin near zero (small difference) while MAIAC values dissimilar to AERONET are placed in bins farther from zero corresponding to the magnitude of the difference. Note that in the prediction process, we are not using the AERONET data; only MAIAC data are used as inputs to the system during prediction. Still, in theory, the DNN system should quantify the quality of the MAIAC AOD retrieval by predicting the difference between the given MAIAC AOD and the true AOD value (assumed based on the AERONET data).

This DNN is trained only on the collocated data from Africa. Training is performed independently for Aqua and Terra datasets and tested accordingly. While the scope of the collocated dataset includes six global regions, the scope of the DNN data is limited to Africa Aqua and Africa Terra datasets for this study.

CHAPTER 3

RESULTS

3.1 Evaluation of MAIAC AOD data against ground based AERONET data

In the first part of the study, we evaluate the uncertainties in MAIAC AOD retrievals using 16 years of Terra MODIS (2000-2016) and 14 years of Aqua MODIS (2002-2016) data and collocated Level 2 Version 3 AERONET data. First, the one-to-many analysis is performed on a single year (2015) of MAIAC AOD for all selected regions as shown in Figures 3-8. Both Terra and Aqua MODIS data are used in the analysis. Interestingly, although large regional differences can be found, for a given region, differences between Aqua and Terra MAIAC AOD data are rather marginal. Still, large spikes are found in those one-to-many analyses. This is not a surprise as for one AERONET data point, there might be multiple MAIAC AOD data with different observing conditions such as different viewing geometries and surface properties fit in the predefined collocation windows. Also, aerosol properties may vary both spatially and temporally, even with the given small collocation windows in both space and time. Still, linear or near linear patterns between MAIAC and AERONET AOD (550 nm) can be found for all regions, indicating that the MAIAC aerosol retrieval method has skill in retrieving AOD values with reasonable accuracy in all selected regions. In particular, the China region, with its complex surface features, is a problematic region for some of the other aerosol retrieval methods such as the MODIS DT method. Reasonable AOD retrievals, as suggested from Figures 3-8, indicate that the MAIAC aerosol retrieval algorithm performs well in complex surface environments such as over China.

Also, non-linear patterns between MODIS MAIAC and AEORNET AOD over Europe and South America are visible, which indicates that aerosol models used for those regions may not be optimal for aerosol retrievals over those two regions.

The Root Mean Square Error (RMSE) and Linear Correlation Coefficient (CC) are also calculated for each region and satellite for the one-to-many MODIS MAIAC and AERONET analyses as mentioned above. Previously, this evaluation has been performed on MAIAC AOD data using collocated AEROENT Version 2 (AERO V2) data (Skaer et al., 2018). Skaer et al. (2018) found that MAIAC AOD is fairly consistent with AERO V2 in the collocated dataset based on trend analysis, RMSE, and correlation coefficient values. During the completion of Skaer et al. (2018), AERONET Version 3 (AERO V3) dataset was released. The same investigation of RMSE and CC is performed in this study using the newer AERO V3 data. The resulting RMSE and CC of this investigation on AERO V3 data are shown in Table 2.

Table 2: Calculated Root Mean Square Error (RMSE) and Linear Correlation Coefficient for the raw collocated AOD data organized by region and satellite collection with 14/16 years of Aqua/Terra data.

Raw Collocated AOD Data RMSE and Correlation		
Data	RMSE	Correlation
Africa Aqua	0.124946	0.802405
Africa Terra	0.130374	0.803023
China Aqua	0.205514	0.880494
China Terra	0.195869	0.89339
Europe Aqua	0.123471	0.571263
Europe Terra	0.108113	0.664751
Mid Asia Aqua	0.178164	0.850602
Mid Asia Terra	0.170669	0.847134
North America Aqua	0.0861933	0.775441
North America Terra	0.0881225	0.799738
South America Aqua	0.0969982	0.915371
South America Terra	0.0941247	0.935363

In general, the best performance is found over South America, where the highest correlations of above 0.9 and relatively small RMSE values of less than 0.1 are found for both Terra and Aqua MAIAC AOD retrievals. The worst correlations of below 0.6 are found for both Terra and Aqua data over Europe, yet with RMSE values of around 0.1. This indicates that aerosol loading is overall low over the region. Therefore, low correlations but with marginal RMSE values in MAIAC AOD are found for this region. Highest RMSE values of above 0.16 are found over China and Mid-Asia, although reasonable correlations of above 0.8 between MAIAC and AERONET AOD are also reported. This is also not a surprise, as both regions pose challenges for aerosol retrievals from traditional passive-based aerosol retrieval methods due to high surface albedos at the visible channel for both regions. For example, high RMSE values in AOD retrievals are also expected in western and northern Asia where deserts or dry and arid or semi-arid surfaces are located. It is worth noting that correlation values of above 0.8 are found over Africa with averaged RMSE values of around 0.12-0.13. Still, we expected higher RMSE values over North Africa because of the Saharan desert.

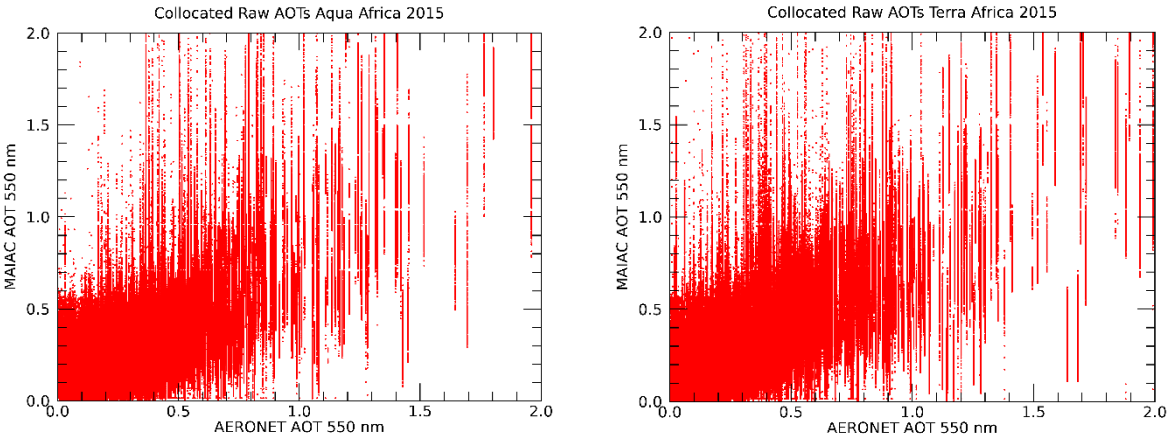

Figure 3: One-to-Many plots comparing MAIAC AOD (y-axis) vs AERONET AOD (x-axis) showing Aqua (left) and Terra (right) for Africa 2015 data.

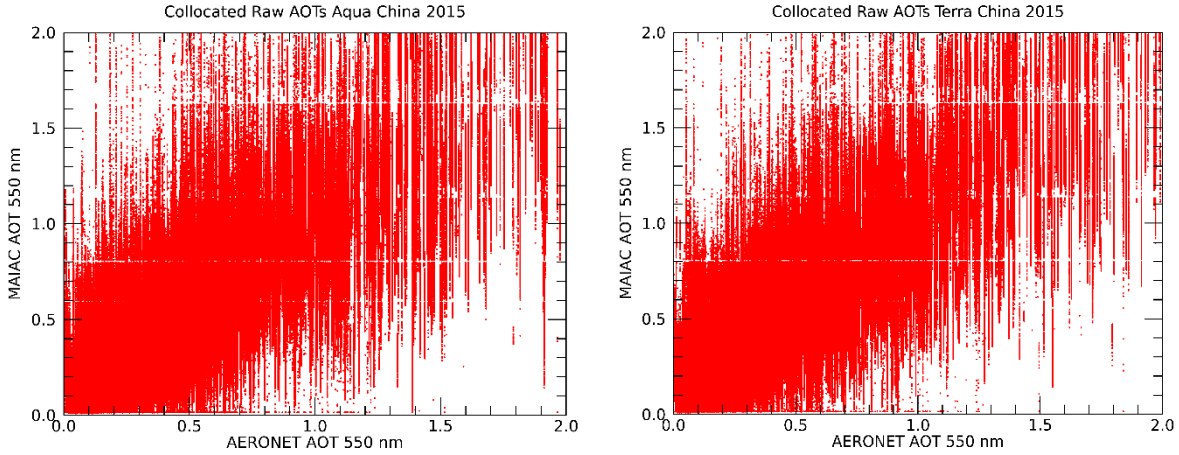

Figure 4: One-to-Many plots comparing MAIAC AOD (y-axis) vs AERONET AOD (x-axis) showing Aqua (left) and Terra (right) for China 2015 data.

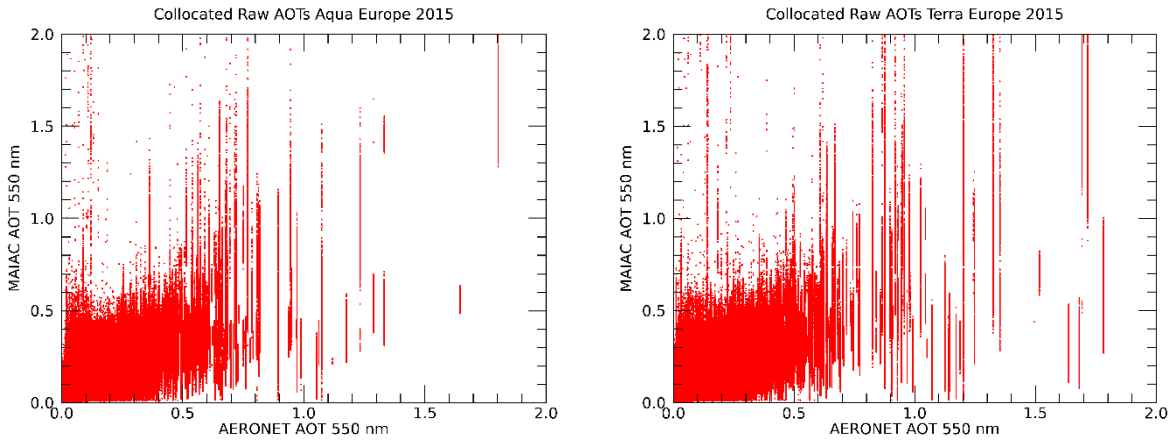

Figure 5: One-to-Many plots comparing MAIAC AOD (y-axis) vs AERONET AOD (x-axis) showing Aqua (left) and Terra (right) for Europe 2015 data.

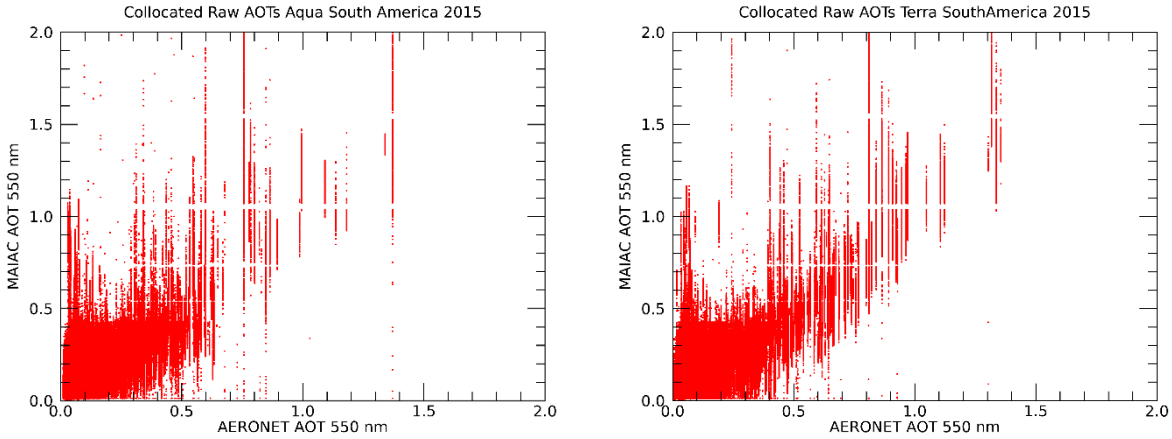

Figure 6: One-to-Many plots comparing MAIAC AOD (y-axis) vs AERONET AOD (x-axis) showing Aqua (left) and Terra (right) for South America 2015 data.

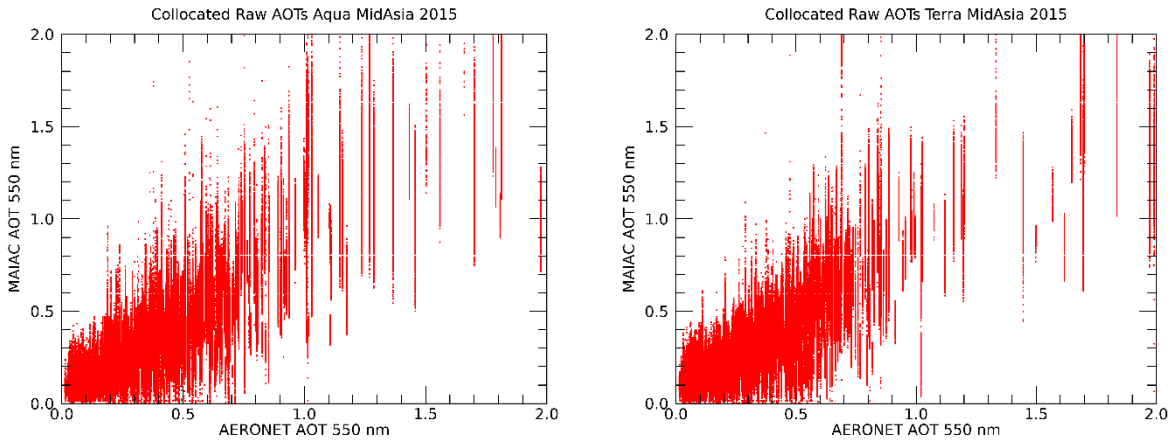

Figure 7: One-to-Many plots comparing MAIAC AOD (y-axis) vs AERONET AOD (x-axis) showing Aqua (left) and Terra (right) for Mid Asia 2015 data.

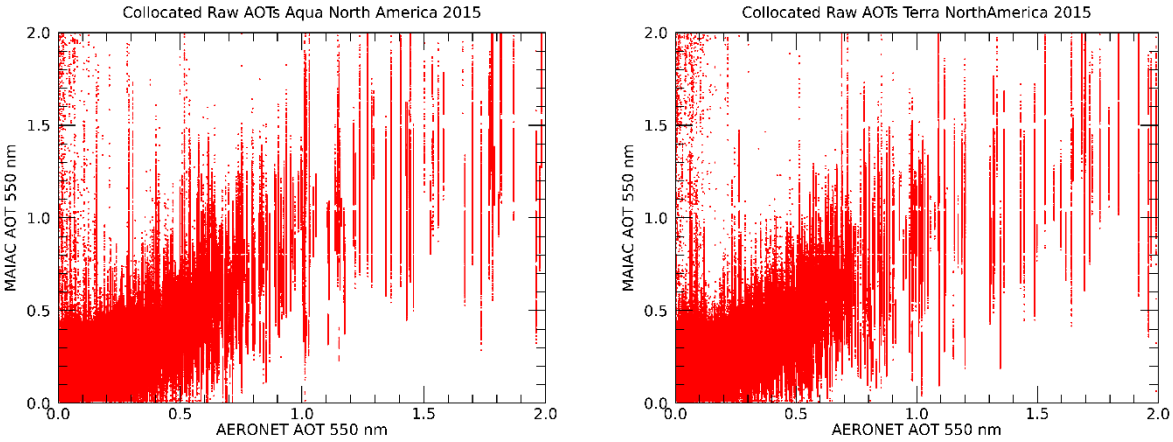

Figure 8: One-to-Many plots comparing MAIAC AOD (y-axis) vs AERONET AOD (x-axis) showing Aqua (left) and Terra (right) for North America 2015 data.

Note that the results presented above are largely consistent with those reported by Skaer et al. (2018), in which the version 2 AERONET data were used. Skaer et al. (2018) also studied the variations of MAIAC AOD retrievals as a function of observing conditions including column water vapor, fine mode fraction over ocean, cosine of solar zenith angle, cosine of solar viewing angle, relative azimuth angle, and scattering angle, using version 2 AERONET data. Although version 3 AERONET data are used in this study, no major differences are found in comparing MAIAC and AERONET AOD using version 2 and version 3 AERONET data, although better cloud screening and noise removal are expected for the Version 3 AERONET data. Thus, we expect similar relationships even with the use of the Version 3 AERONET data. Still, for the sake of completeness, we summarized findings from Skaer et al. (2018) as follows. It was found by Skaer et al. (2018) that while the number of observations is greater at smaller solar zenith angles, there is no range of solar zenith angle where MAIAC AOD has a higher or lower bias relative to AERONET data. It was also found that no viewing zenith angle promotes a high or low bias. Additionally, Skaer et al. (2018) found that no relative azimuth angles caused a high or low bias in MAIAC AOD. Finally for scattering angle, it was found that no scattering angles caused a particular higher or lower bias than the corresponding AERONET AOD.

3.2 Development of a Neural Network based method for Quality Control of MAIAC AOD

In the second part of this study, we train and test DNN to quality control the Terra and Aqua MODIS MAIAC Aerosol data. A DNN based method is developed to quantify uncertainties in MAIAC AOD as functions of observing conditions. Also, sensitivity studies are conducted to investigate the system performance with respect to parameter settings in the DNN system.

3.2.1 Performance of the Deep Neural Network for Quality Control of MODIS MAIAC AOD

The evaluation of the DNN is performed separately for the Aqua and Terra datasets. The training is initially done using one year (2014) of MODIS MAIAC Aqua data. In order to limit the time required to test the model, a small portion (30%) of the 2014 data is set aside for testing. This testing portion is untouched by the network until after the network has trained on the other 70%. Using this method, the training portion of the data is then used to adjust and tune the model before testing. The resulting trained DNN is able to begin distinguishing MAIAC AOD which are close to AERONET with moderate accuracy as seen in Figure 9. Figure 9a shows the comparison of MODIS MAIAC and AERONET AOD using the testing portion of the data which was set aside during training. Figure 9b is similar to Figure 9a but shows only the MAIAC data that are predicted to be within +/- 0.15 of AEROENT AOD by the DNN system. The DNN system settings for this initial run include 7 hidden layers with nodes for each layer equaling 512, 256, 128, 128, 128, 64, and 32 respectively, a learning rate of 0.001, 10 epochs, rectified linear activation function, and 12 output classes. Clearly, noisy data as shown in Figure 9a, are largely removed as suggested by Figure 9b, suggesting that the DNN system performs as designed.

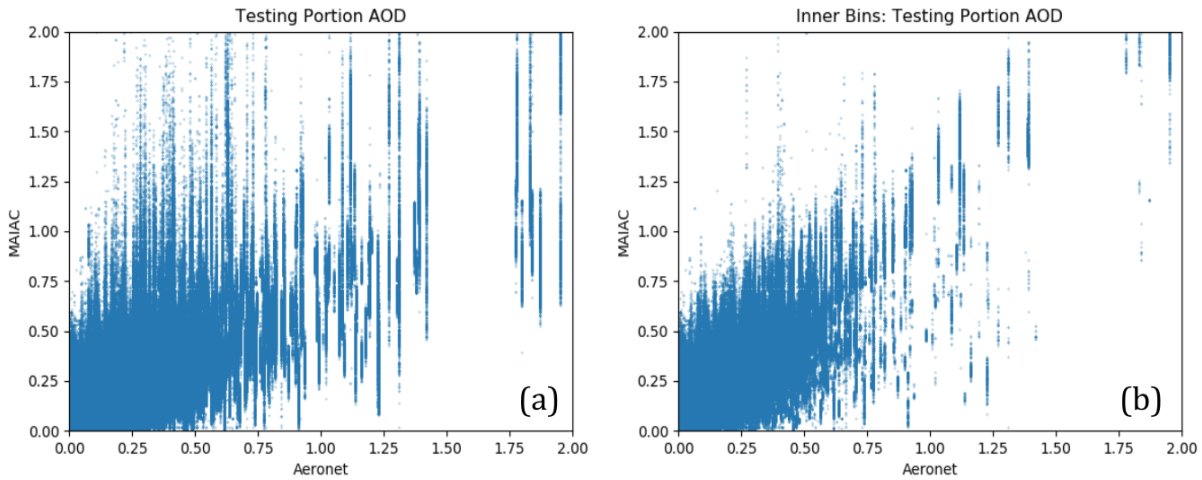

Figure 9: Initial test results of the network trained on Africa Aqua 2014 data, with the raw test data (a) compared to the network predictions (b) for which MAIAC AOD were closest to AERONET as classified by the DNN system.

While Figure 9 shows the performance of the DNN system that is trained and tested using one year of collocated MODIS MAIAC and AERONET AOD data, we also evaluate the DNN system by training with multiple years of collocated Aqua MODIS MAIAC and AERONET data and then testing the system using similar collocated data from a different year. All but one year (2016) of Africa Aqua data are used for training in this section, making an input data set of 13 years total (2002-2015) for the Aqua dataset. The DNN system is then tested using the collocated Africa Aqua and AERONET data from 2016.

The resulting trained DNN for this longer period of data can more clearly distinguish those MAIAC AOD which are close to AERONET (within +/- 0.15) as seen in Figure 10. Figure 10 shows the performance of the DNN system, on collocated MODIS and AERONET data for 2016, which is trained using the 13-year collocated Aqua MODIS and AERONET data over Africa. The red dots in Figure 10 represent the comparison of Aqua MODIS MAIAC and AERONET AOD from the

collocated 2016 data and the blue dots in Figure 10 show those data pairs that passed the filtering process based on the DNN system (predicted difference of MODIS MAIAC and AERONET AOD within +/- 0.15). In this case, the number of hidden layers is set to 3, with 512, 64 and 32 nodes in each respective layer. The other settings include a learning rate of 0.001, 20 epochs, and 9 output classes. In this case, roughly 86% of the data passed the filter with a reduction in RMSE of 31%.

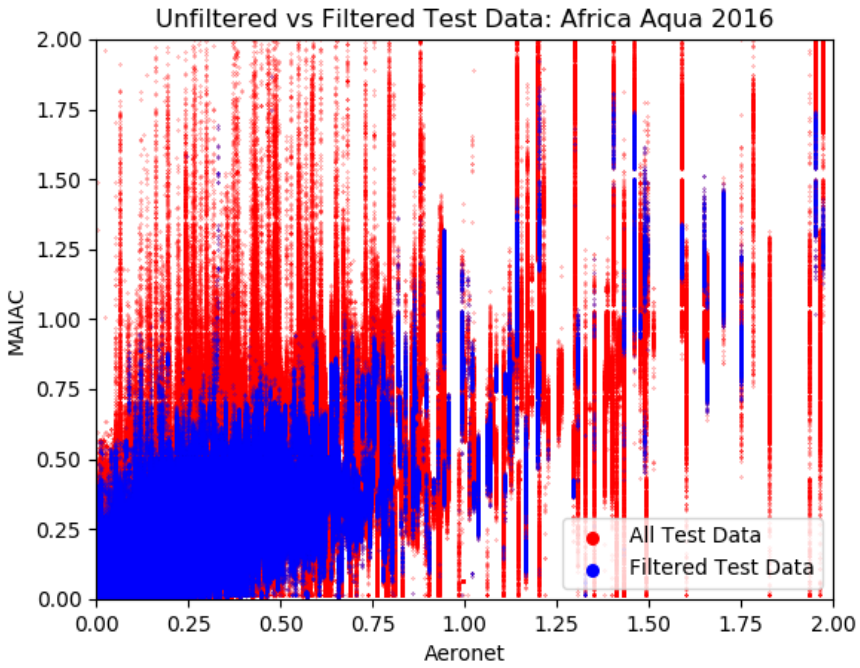

Figure 10: Filtered Aqua MAIAC AOD_{550} (blue) versus the whole 2016 test dataset (red

Performing a similar analysis on the Terra dataset we found similar results. Initially, as in Figure 9 with collocated Aqua data, the DNN analysis is done using one year (2014) of MODIS MAIAC Terra data. The resulting trained DNN can begin distinguishing Terra MAIAC AOD which are close to AERONET with moderate accuracy as seen in Figure 11. The settings for this initial DNN run are the same as those used in the initial DNN run for Aqua 2014 data. As in Figure 9, Figure 11 shows that noisy data in Figure 11a are largely removed in 11b, suggesting that the DNN system performs as designed for Terra data as well.

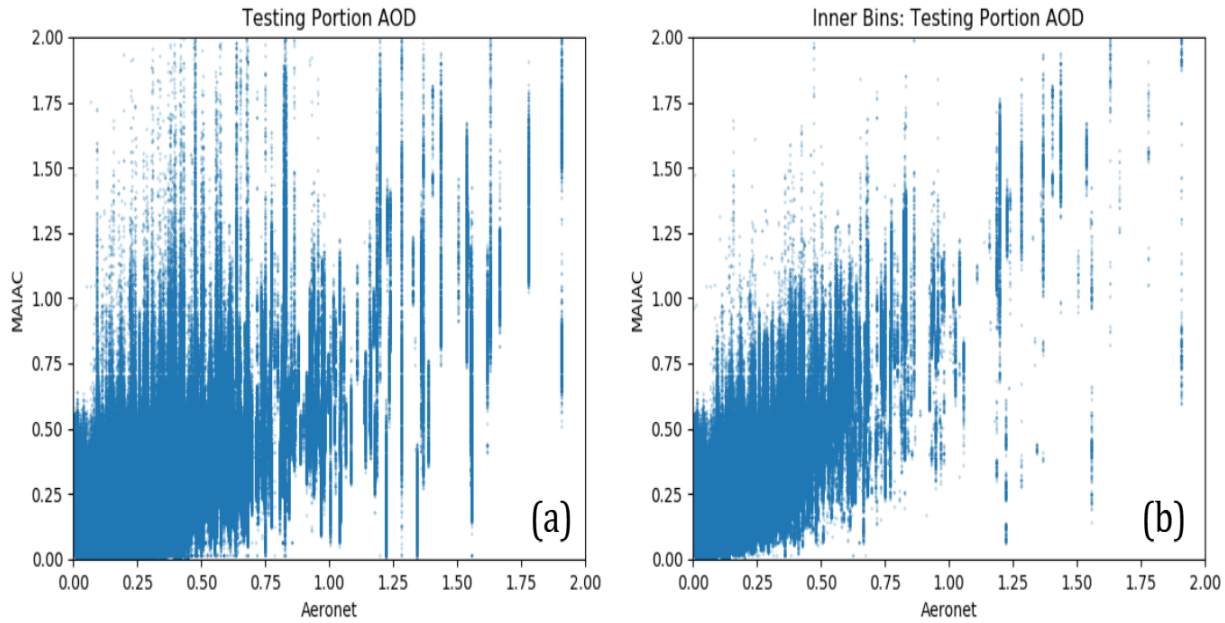

Figure 11: Initial test results of the network trained on Africa Terra 2014 data, with the raw test data (a) compared to the network predictions (b) for which MAIAC AOD were closest to AERONET as classified by the DNN system.

We also evaluate the DNN system on multiple years of collocated Terra MODIS and AERONET data using all but one year (2016) for training, making a training data set of 15 years (2000-2015). The resulting trained DNN for this longer period of Terra data is also able to clearly distinguish those MAIAC AOD which are close to AERONET (within +/- 0.15) as shown in Figure 12. In this case, the number of hidden layers is set to 3, with 512, 64 and 32 nodes in each respective layer. The other settings include a learning rate of 0.001, 20 epochs, and 9 output classes. Note that this is identical to the parameter settings for the Aqua dataset in Figure 9. In this case, roughly 84% of the data passed the filter, with a 27% reduction of RMSE.

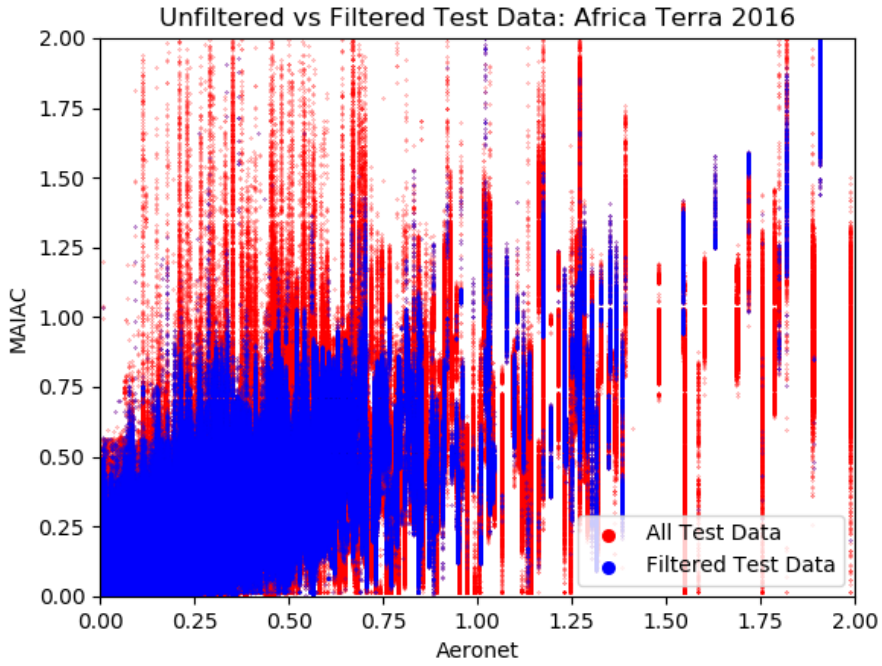

Figure 12: Filtered Terra MAIAC AOD_{550} (blue) versus the whole 2016 test dataset (red)

3.2.2 Optimization of the Performance of the Deep Neural Network

Note that the performance of the DNN system varies as functions of parameters used for the DNN system, including but not limited to different numbers of hidden layers and nodes, different learning rates, different sets of input parameters, and different quantification of outputs. To select an optimal parameter set for the DNN system, a sensitivity study is performed for both Aqua and Terra data separately. The optimization is a step-by-step process of DNN runs where one or two model parameters are adjusted for each DNN run so that the effect of each change is quantifiable. This is referred to as tuning. We use RMSE, Linear Correlation Coefficient, and the comparison of MAIAC and AERONET to quantify the tuning process.

For the Aqua dataset, the parameter settings for each step of the process are shown in Table 3 with the resulting filtered MAIAC AOD plotted in Figure 13. In Table 3, Models A-F correspond to Figure 13a-13f showing the progression of skill. Each row in Table 3 represents the parameter

values for a training instance, and the columns define the parameters. In Table 3, the column titled Model Name gives the name which connects each row in Table 3 to a plot in Figure 13. The column titled Test Year is a column displaying the year of Africa Aqua data used for testing the trained network. The column labeled as Other is for miscellaneous changes that fit in standalone categories detailed below Table 3 (Table Key). Finally, the right most two columns, RMSE and correlation display the Root Mean Square Error and correlation (respectively) discussed in chapter 3.1. The key features of Table 3 are the individual cells outlined in bold as well as the RMSE and correlation columns. The bolded cells indicate which parameter is changed for a given instance of training. The RMSE and correlation columns are the indicators of how well a certain training instance performed.

Model A shown in Table 3 gives the parameter settings for the first multi-year model run. Additionally Figure 13a shows the resulting comparison plotted in the same manner as Figure 9. Initially, the model accuracy decreased given the larger input dataset, as input parameters may be correlated and contain redundancy. Also, some input parameters may be unrelated to final outputs. The number of nodes per layer in *Model A* are in a descending gradient of 512, 256, 128, 64, and 32 nodes in each of the 5 hidden layers respectively. This resulted in an RMSE of 0.104 and a correlation of 0.815 as seen in Table 3. It is seen in Figure 13a that although some of the noisy retrievals are excluded with the use of the DNN system, there are still significant percentages of noisy retrievals that passed the DNN-based filtering. In *Model B* we set the learning rate parameter from 0.001 to 0.0001. This yields a worse performance, which is indicated by the increase in spikes in Figure 13b that are far away from the one-to-one line of the MODIS MAIAC and AERONET AOD plot. Additionally, Table 2 shows that the RMSE increased to 0.111 and correlation decreased to 0.809. In *Model C* we increased the percentage of input data used for training from 70% to 90% and therefore decreased the percentage used for testing from 30% to 10%. We found this gave slightly better performance as is found in Figure 13c indicated by fewer spikes away from the one-to-one

trend line. The RMSE and correlation in this case remained relatively the same with scores of 0.112 and 0.805 respectively.

In general, having many dimensions in the input data can introduce unnecessary complexity to the network (Verleysen et al. 2003). Therefore, in *Model D* we eliminate MODIS Latitude, MODIS Longitude, and Fine Mode Fraction (FMF) from the input data to assess the importance of having location in the input data and because FMF is a byproduct of the retrieval process. This omission is in addition to two other variables which are omitted for the entirety of training, MAIAC AOD at 470 nm and MAIAC Uncertainty. MAIAC AOD at 470 nm is omitted because the network is specifically being trained to recognize AOD at 550 nm. MAIAC Uncertainty is based only on AOD at 470 nm surface reflectance, and thus gives only a general indication of possible increase of error over brighter surfaces. Additionally, the number of Epochs is decreased from 50 to 20. The *Model D* parameters result in RMSE increasing to 0.125 and correlation decreasing to 0.785. This decrease in network skill is visible in Figure 13d with the inclusion of smaller MAIAC values at higher AEROENT values in the filtered data (blue). Based on these results, *Model E* continues the omission of Latitude, Longitude, and FMF but includes MAIAC AOD at 470 nm and MAIAC Uncertainty. This does not significantly increase network skill as Figure 13e shows only a marginal decrease in the number of low MAIAC values at high AERONET included in the filtered data. The resulting RSME of 0.122 shown in Table 2 as well as the correlation of 0.783 confirm the lack of improvement in network skill.

Therefore, in *Model F* a new approach is taken. First, MAIAC AOD at 470 nm and MAIAC Uncertainty are removed once again from the input data, while Latitude, Longitude, and FMF are re-included. In order to decrease the potential strain on the network, the complexity of the data (now 13 input dimensions) may be having on the model, the number of output classes is decreased (Verleysen et al. 2003). By decreasing in the number of output classes given to the network from 12 [(<-9999), (-9999, -1.0), (-1.0, -0.5), (-0.5, -0.05), (-0.05, -0.02), (-0.02, 0), (0, +0.02) … (+1.0,

+9999), (>+9999)] to 10 [(<-9999), (-9999, -0.15), (-0.15, -0.05), (-0.05, -0.02), (-0.02, 0), (0, +0.02) … (+0.15, +9999), (>+9999)], a significant reduction in noisy data pairs are clearly visible in Figure 13f. This is substantiated by the decrease in RMSE to 0.091 shown in Table 2 and the increase in correlation to 0.799. However, to achieve an acceptable value for correlation, a final model version is created named *Final* in Table 2 and already shown in Figure 9. To further decrease the dimensionality of the model, the number of layers is taken from 5 to 3. The finalized DNN settings include 3 hidden layers with 512, 253, and 32 nodes respectively, an output layer with 10 classifications, a learning rate of 0.001, and 20 iterations (epochs). An RMSE of 0.0857 and a correlation of 0.802 are found for the filtered data shown in Figure 9.

The RMSE and correlation achieved in this instance are in line with the desired output from the network predictions, where RMSE reduced by ~31% and correlation remained consistent compared to the initial RMSE and correlation of the data shown in the upper right of Table 2. As seen in the table, other instances of training produce higher correlation but saw a less desirable change in RMSE or in some cases an increase in RMSE.

A similar step by step process is performed for the Africa Terra dataset for collocated MODIS MAIAC and AERONET data. Table 4 and Figures 14a-14f display the progression of training for Terra data as previously detailed with the Aqua dataset. Note that while changes are made to the DNN each run only small changes are visible between each resulting plot in Figure 14. The sensitivity tests show that similar RMSE and linear correlation are achievable with the Terra dataset compared to what is seen with the Aqua data. However, Terra shows slightly worse RMSE and linear correlation in the DNN. Finally, while slight differences are seen in the RMSE and correlation, the final trained version of the DNN for Terra includes the exact same parameters as the final trained Aqua model.

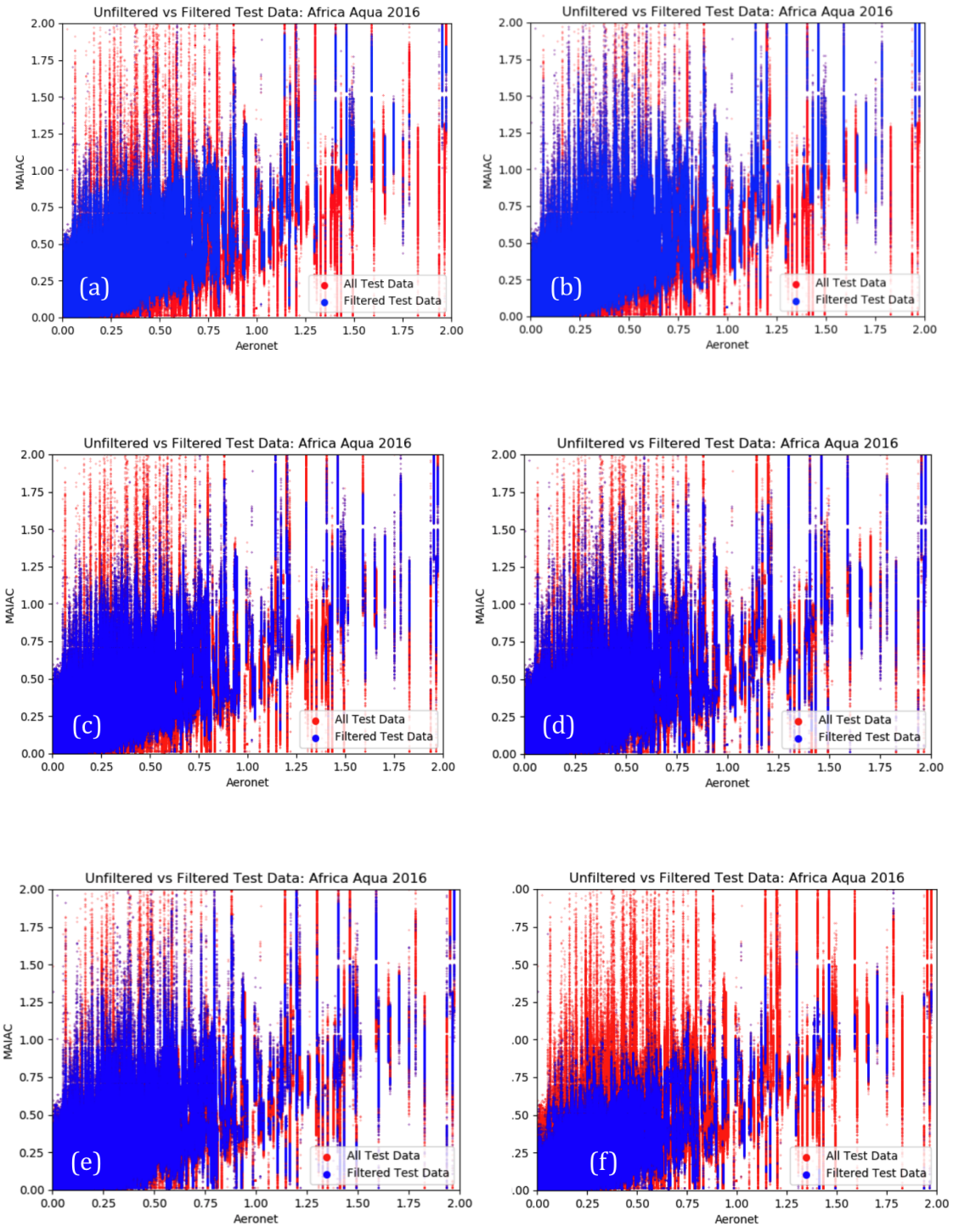

Figure 13: Africa Aqua: Filtered MAIAC AOD_{550} (blue) versus the test dataset (red) where (a) through (f) show the progression of increasing skill with each version of the trained model.

Table 3: Africa Aqua: This table shows the change in each Deep Neural Network parameter (columns) for each version of the Network Model (row) with the resulting RMSE and Linear Correlation for each Network Model in the rightmost two columns.

Sensitivity Study Progression **Row: Training Instance** **Column: Parameter Value**								**Training Data: Africa Aqua 2002-2015**			
Model Name	Hidden Layers	Nodes	Activ	Classes	Opt	LR	Epochs	Test Year	Other	RMSE	Corr
Raw	-	-	-	-	-	-	-	-	-	**0.125**	**0.802**
A	5	α*	relu	12	Adam	0.001	50	2016		0.104	0.815
B	5	α*	relu	12	Adam	0.0001	50	2016		0.111	0.809
C	5	α*	relu	12	Adam	0.001	50	2016	β*	0.112	0.805
D	5	α*	relu	12	Adam	0.001	20	2016	γ*	0.125	0.785
E	5	α*	relu	12	Adam	0.001	20	2016	δ*	0.122	0.783
F	5	α*	relu	10	Adam	0.001	20	2016	ε*	0.091	0.799
Final	3	ζ*	relu	10	Adam	0.001	20	2016		0.0857	0.802

Table Key

α* *512, 256, 128, 64, 32*

β* *Test size during training changed from 30% to 10%*

γ* *omit latitude, longitude, fine mode fraction from input data*

δ* *d* + include MAIAC AOD 470, MAIAC AOD uncertainty in input data*

ε* bins adjusted to (-9999, -0.15, -0.05, -0.02, 0, 0.02, 0.05, 0.15, 9999), e* and d* reverted

ζ* 512, 64, 32

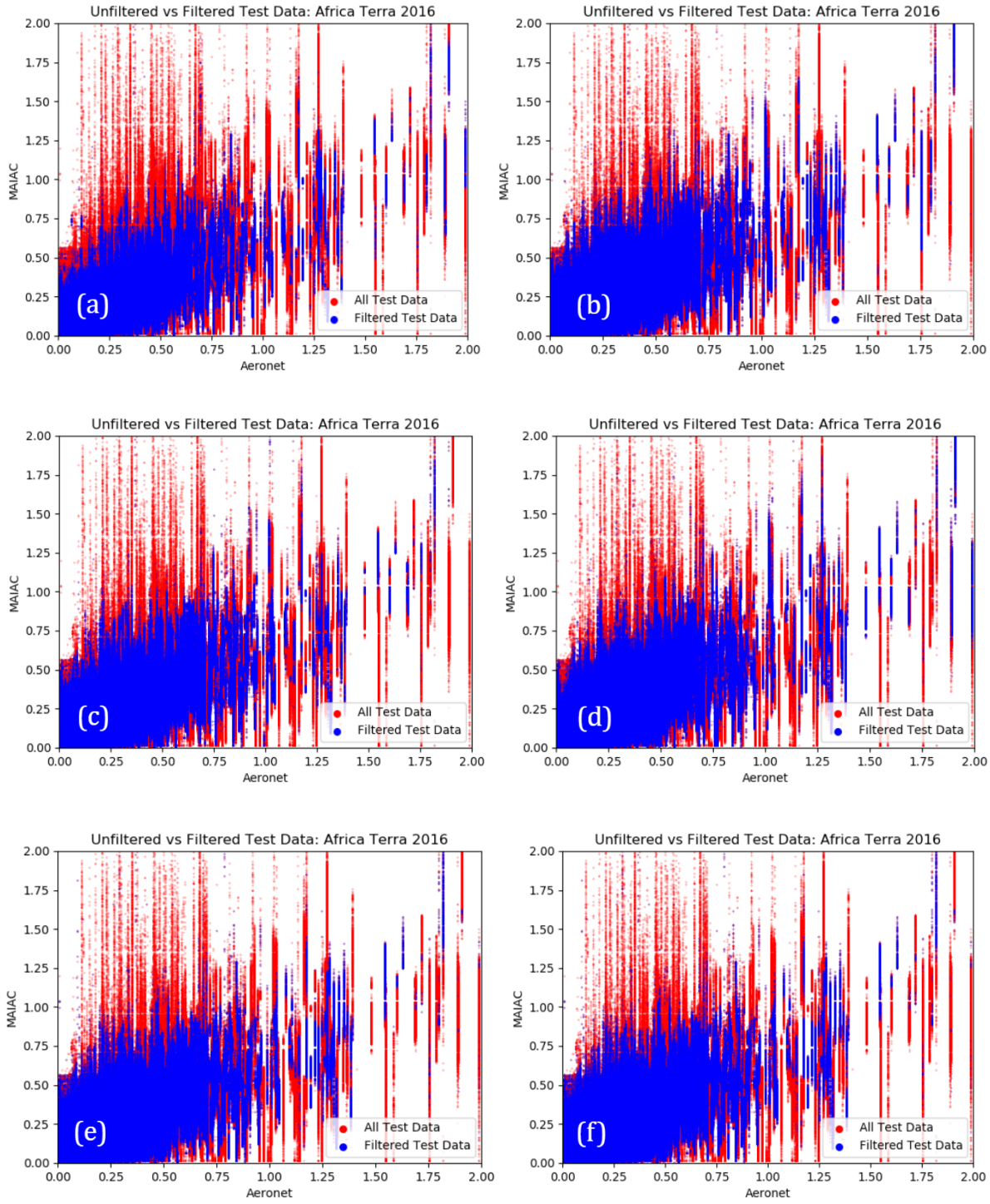

Figure 14: Africa Terra: Filtered MAIAC AOD_{550} (blue) versus the test dataset (red) where (a) through (f) show the progression of increasing skill with each version of the trained model.

Table 4: Africa Terra: This table shows the change in each Deep Neural Network parameter (columns) for each version of the Network Model (row) with the resulting RMSE and Linear Correlation for each Network Model in the rightmost two columns.

Sensitivity Study Progression **Row: Training Instance** **Column: Parameter Value**								**Training Data: Africa Terra 2000-2015**			
Model Name	Hidden Layers	Nodes	Activ	Classes	Opt	LR	Epochs	Test Year	Other	RMSE	Corr
Raw	-	-	-	-	-	-	-	-	-	**0.131**	**0.803**
A	5	**α***	relu	10	Adam	0.001	50	2016		0.103	0.797
B	3	*β**	relu	10	Adam	0.001	20	2016		0.102	0.793
C	4	γ*	relu	10	Adam	0.001	30	2016		0.096	0.785
D	3	*δ**	relu	10	Adam	0.001	30	2016		0.100	0.791
E	3	*δ**	relu	10	Adam	0.001	50	2016		0.099	0.789
Final	3	*δ**	relu	10	Adam	0.001	20	2016		0.096	0.793

Table Key
α* *512, 256, 128, 64, 32*
*β** *512, 128, 32*
γ* 512, 128, 64, 32
*δ** *512, 64, 32*

CHAPTER 4

CONCLUSIONS

Previous studies have found that aerosol analyses and forecasts can be improved through assimilation of high-quality satellite based AOD data (Lyapustin et al. 2012; Lee 2019), or through assimilation of satellite AOD that are carefully quality controlled (e.g. Zhang et al., 2008). However, it has also been found that biases in satellite aerosol data can introduce non-negligible uncertainties in the downstream aerosol analysis and forecasts (Zhang and Reid 2006). Therefore, in this study, we quantify uncertainties in the Aqua and Terra MODIS MAIAC AOD data using the version 3 level 2 AERONET data. We further develop a deep neural network-based method for quality control of MODIS MAIAC aerosol data for aerosol modeling applications. This is done using 14 years of Aqua MODIS (2002-2016) and 16 years of Terra MODIS (2000-2016) MAIAC data.

The findings of this study are as follows:

1. Validated against version 3, level 2 AERONET data, reasonable performance in MODIS MAIAC AOD retrievals is found as RMSEs in MAIAC AOD (550 nm) range from ~0.1 to 0.2 and correlations between MAIAC and AERONET AOD range between ~0.6 to 0.9. RMSE values of lower than 0.1 are found over South and North American and the region with the worst RMSE values of ~0.2 are found to be China and Mid-Asia. Given that both China and Mid-Asia are regions with bright surfaces that pose difficulties for traditional aerosol retrievals using passive sensors, the performance of MODIS MAIAC AOD over those regions is quite reasonable as well.

2. A DNN system for quality control of the MODIS MAIAC AOD data is developed. The input parameters are taken from the MODIS MAIAC dataset and the outputs are predicted differences between MAIAC and AERONET AOD, where AERONET AOD values are considered as the ground truth in this study. Using collocated MODIS MAIAC and AEORNET data from 2014, we find the DNN system has skills in detecting noisy MAIAC AOD data. We further test the DNN system using 13 years/15 years of collocated Aqua/Terra MODIS MAIAC AOD and AERONET data as training samples and use the remaining one year of collocated Aqua/Terra MODIS MAIAC AOD and AERONET data as testing samples. Our study suggests that the DNN system can detect and significantly reduce noisy retrieval in MAIAC AOD data. An approximate 31%/27% reduction in Aqua/Terra MODIS MAIAC AOD is found with the use of the DNN system, but with an approximate 14%/16% data loss.
3. To select the optimal parameter settings for the DNN system, a sensitivity study is performed. This study suggests that the reduction in number of output categories can significantly improve the performance of the DNN system. The improvement in performance of the DNN system is also observed by reducing the number of hidden layers from 7 to 3. However, marginal changes are found by altering number of notes, altering input parameters, modifying learning rates and reducing epoch times.

This study suggests that both Terra and Aqua MODIS MAIAC AOD data compare reasonably well with ground based AERONET data for all 6 study regions (Europe, Africa, China, Mid-Asia, South America, North America). Yet larger noises still exist. A DNN can be used as an effective quality control method for further reducing uncertainties and biases in MODIS MAIAC AOD data. The quality controlling method developed in this study may be used for aerosol data assimilating applications.

Myriad options exist for investigating the performance of the network given alternate starting conditions, as well as to quality control the input data. First, further sensitivity study focused on the influence of viewing geometry input parameters (solar zenith, relative azimuth) could reveal an elevated dependence on these parameters leading to the possible elimination of other input parameters and therefore a decrease dimensionality in the network which is shown to increase performance. Additionally, performing a similar process on other regions (Mid-Asia, China, North and South America) as is performed here using Africa could inform the regional dependence of such studies as this. Finally, altering the network target from finding the difference between a given MAIAC and AERONET AOD to instead predicting the corrected output MAIAC AOD based on a given input MAIAC AOD is possible with this network and could have a different result. Those remain to be studied in future efforts.

REFERENCES

Brunekreef, B., and S. T. Holgate, 2002: Air pollution and health. Lancet, 360, 1233–1242, https://doi.org/10.1016/S0140-6736(02)11274-8.

Cheng, T., H. Chen, X. Gu, T. Yu, J. Guo, and H. Guo, 2012: The inter-comparison of MODIS, MISR and GOCART aerosol products against AERONET data over China. J. Quant. Spectrosc. Radiat. Transf., 113, 2135–2145, https://doi.org/10.1016/j.jqsrt.2012.06.016.

Chin, M., T. Diehl, P. Ginoux, and W. Malm, 2007: Intercontinental transport of pollution and dust aerosols: implications for regional air quality. *Atmos. Chem. Phys.*, 7, 5501–5517, https://doi.org/10.5194/acp-7-5501-2007.

Eck, T. F., B. N. Holben, I. Slutsker, and A. Setzer, 1998: Measurements of irradiance attenuation and estimation of aerosol single scattering albedo for biomass burning aerosols in Amazonia. *J. Geophys. Res.*, 103, 31865–31878, https://doi.org/10.1029/98jd00399.

Eck, T. F., B. B. N. Holben, J. S. Reid, Dubovik, and S. Kinne, 1999: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *J. Geophys. Res. D: Atmos.,* 104349, 333–331, https://doi.org/10.1029/1999JD900923.

Falah, S., A. Mhawish, M. Sorek-Hamer, A. I. Lyapustin, I. Kloog, T. Banerjee, F. Kizel, and D. M. Broday, 2021: Impact of environmental attributes on the uncertainty in MAIAC/MODIS AOD retrievals: A comparative analysis. Atmos. Environ., 262, 118659, https://doi.org/10.1016/j.atmosenv.2021.118659.

Giles, D. M., and Coauthors, 2019: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved

cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmos. Meas. Tech*., 12, 169–209, https://doi.org/10.5194/amt-12-169-2019.Holben, B. N., and Coauthors, 1998: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. Remote Sens. Environ., 66, 1–16, https://doi.org/10.1016/S0034-4257(98)00031-5.

Hsu, N.C., Jeong, M.-J., Bettenhausen, C., Sayer, A.M., Hansell, R., Seftor, C.S., Huang, S., Tsay, C., 2013. Enhanced deep blue aerosol retrieval algorithm: the second generation. J. Geophys. Res. Atmos. 118 (16), 9296–9315. https://doi.org/10.1002/jgrd. 50712.

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

Kingma, D. P., and J. Ba, 2014: Adam: A Method for Stochastic Optimization. arXiv [cs.LG],.

Kokhanovsky, A. A., and Coauthors, 2007: Aerosol remote sensing over land: A comparison of satellite retrievals using different algorithms and instruments. *Atmos. Res*., 85, 372–394, https://doi.org/10.1016/j.atmosres.2007.02.008.Lee, H. J., 2019: Benefits of High Resolution PM2.5 Prediction using Satellite MAIAC AOD and Land Use Regression for Exposure Assessment: California Examples. *Environ. Sci. Technol*., 53, 12774–12783, https://doi.org/10.1021/acs.est.9b03799.

Levy, R. C., L. A. Remer, R. G. Kleidman, S. Mattoo, C. Ichoku, R. Kahn, and T. F. Eck, 2010: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *Atmos. Chem. Phys*., 10, 10399–10420.

Lyapustin, A., Y. Wang, I. Laszlo, R. Kahn, S. Korkin, L. Remer, R. Levy, and J. S. Reid, 2011: Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *J. Geophys. Res.*, 116, https://doi.org/10.1029/2010jd014986.

Lyapustin, A., Y. Wang, I. Laszlo, and S. Korkin, 2012: Improved cloud and snow screening in MAIAC aerosol retrievals using spectral and spatial analysis. *Atmospheric Measurement Techniques*, 5, 843–850.

Lyapustin, A., Y. Wang, S. Korkin, and D. Huang, 2018: MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, 11, 5741–5765, https://doi.org/10.5194/amt-11-5741-2018.

Maji, K. J., M. Arora, and A. K. Dikshit, 2017: Burden of disease attributed to ambient PM2.5 and PM10 exposure in 190 cities in China. *Environ. Sci. Pollut. Res. Int.*, 24, 11559–11572, https://doi.org/10.1007/s11356-017-8575-7.

Rabha, S., and B. K. Saikia, 2020: 18 - Advanced micro- and nanoscale characterization techniques for carbonaceous aerosols. Handbook of Nanomaterials in Analytical Chemistry, C. Mustansar Hussain, Ed., Elsevier, 449–472.

Shi, Y., J. Zhang, J. S. Reid, E. J. Hyer, T. F. Eck, B. N. Holben, and R. A. Kahn, 2011: A critical examination of spatial biases between MODIS and MISR aerosol products--application for potential AERONET deployment. Atmospheric Measurement Techniques, 4, 2823–2836

Skaer T., non-thesis report, An Analysis of the MAIAC AOD Product, University of North Dakota, 2018.

System Description - Aerosol Robotic Network (AERONET) Homepage. https://aeronet.gsfc.nasa.gov/new_web/system_descriptions_operation.html (Accessed February 3, 2021a).Unidata. https://www.unidata.ucar.edu/software/netcdf/ (Accessed June 30, 2021b).

Verleysen, M., D. François, G. Simon, and V. Wertz, 2003: On the effects of dimensionality on data analysis with neural networks. *Plan. Perspect.*, 105, 112.

Williams, J., M. de Reus, R. Krejci, H. Fischer, and J. Ström, 2002: Application of the variability-size relationship to atmospheric aerosol studies: estimating aerosol lifetimes and ages. *Atmos. Chem. Phys.*, 2, 133–145, https://doi.org/10.5194/acp-2-133-2002.

Xu, D., Y. Zhang, Q. Sun, X. Wang, and T. Li, 2021: Long-term PM2.5 exposure and survival among cardiovascular disease patients in Beijing, China. *Environ. Sci. Pollut. Res. Int.*, https://doi.org/10.1007/s11356-021-14043-w.

Zhang, J., and J. S. Reid, 2006: MODIS aerosol product analysis for data assimilation: Assessment of over-ocean level 2 aerosol optical thickness retrievals*. J. Geophys. Res.*, 111, https://doi.org/10.1029/2005jd006898.

Zhang, J., J. S. Reid, D. L. Westphal, N. L. Baker, and E. J. Hyer, 2008: A system for operational aerosol optical depth data assimilation over global oceans. J. Geophys. Res., 113, https://doi.org/10.1029/2007jd009065