Aerosol Size Distribution Using Sun-Photometer AOD data of Five wavelengths and Artificial Neural Network

HAMED PARSIANI[1], ANDRES BONILLA[2]
Department of Electrical and Computer Engineering
University of Puerto Rico, Mayagüez (UPRM) Campus
PO Box 9000, Mayagüez , P.R. 00681-9000
PUERTO RICO
[1]parsiani@ece.uprm.edu, [2]andresbonilla@gmail.com
http://ece.uprm.edu/noaa-crest/

Abstract: -Aerosol size distribution (ASD) is an integral parameter in regional atmospheric models [1]. The LIDAR laboratory at UPRM will provide Puerto Rico a means of measuring ASD, and therefore improve these models. This project intends to develop a method of obtaining ASD with the use of a Sun-photometer data, local CIMEL data obtained from *AERONET* (Aerosol Robotic Network) [2], and an *artificial neural network* (ANN). The Sun-photometer used is an instrument that measures *Aerosol Optical Depth* (AOD) at five wavelengths of 380, 440, 500, 675, and 870 nms . A feed-forward, back-propagation artificial neural network was used to map the underlying pattern between AOD and ASD in southwestern Puerto Rico. After training, the network will produce ASD outputs based on new AOD inputs measured locally at UPRM with the Sun-photometer. It was determined that accurate predictions of ASD (MSE of 10^{-4}) could be made depending on the size of the AOD data pool selected.

Key-Words: Aerosol Size Distribution, Aerosol Optical Depth, Artificial Neural Network, AERONET

1 Introduction

Puerto Rico has been chosen as a test-bed for the study of hydrological functions of aerosols and cloud formations in the tropical coastal regions. Clouds transport aerosols that drastically affect global weather phenomena. The coastal aerosol boundary typical to tropical coastal areas adds a factor of complexity into the understanding of these cloud dynamics. The major complication is the presence of several types of aerosols such as sea spray, rock debris, sand dust, some anthropogenic in nature, such as smoke from incomplete combustion. In order to study the impact of each type of aerosol on the environment, it is necessary to first be able to tell which aerosols are present at a given moment in a given area.

ASD can help experts narrow down the possible types of aerosols present, discriminating by size. ASD can be calculated from LIDAR measurements; this will be one of the main products of the UPRM LIDAR lab, which is almost complete. In order to validate the UPRM LIDAR, local ASD will need to be obtained by alternate means; this is the focus of this research. The ASD over the AERONET station at nearby La Parguera might be close enough to provide readings that we can assume to be the same as Mayaguez ASD. A previous investigation showed that AOD patterns in Mayaguez and La Parguera are extremely similar, and since we are deriving ASD from AOD, it makes sense to think that the ASD between both sites is also roughly the same [3]. If Parguera and Mayaguez ASD data can be considered one data set, then the sites can account for holes in each

other's data (when readings are not taken), and AERONET readings could be used to validate the Mayaguez LIDAR readings.

The most commonly used methods for obtaining ASD directly from AOD are based on mathematical inversion techniques [4,5], which require aerosols to be spherical particles. A method is proposed here to calculate ASD which is available from AERONET [Appendix A] and train an ANN with AOD. The ASD data is taken at nearby La Parguera by AERONET. If the training data is representative of a consistent pattern between AOD and ASD, the network should produce accurate ASD predictions when presented with an AOD input vector measured at the LIDAR site with our local Sun-photometer [Appendix B]. In the past, neural networks have been used to retrieve atmospheric data such as ozone [6] and water properties [7] by using large databases of observed readings as training and testing data, but the procedure has never been employed to obtain ASD from AOD. If the ASD produced from Mayaguez AOD inputs is close enough to La Parguera ASD readings, then AERONET ASD can be used directly for Mayaguez, and data can be generated using the network on days for which AERONET does not have readings.

In order to ensure that the La Parguera AERONET site will generate AOD data similar to that of Mayaguez, an analysis of the spatial variation in AOD between both sites was made [3]. Fig. 1 shows the results for an AOD comparison over 2007.

Green lines represent Mayaguez AOD readings, and Red lines represent La Parguera AOD readings. Clearly, both sites have consistently similar AOD patterns. Fig. 2 shows the 4 test sites for an AOD variability test, the results of which are in Fig. 3. Site 1 is Mayaguez, and site 4 is La Parguera. Sites 2 and 3 are simply points along a straight line between site 1 and 2. For this test, readings were taken at each site with a portable sun-photometer and compared to the AERONET AOD reading which was taken at La Parguera at the same time [3]. As the readings approach La Parguera, the Mean Square Error decreases, but just barely. The error trend is almost a horizontal line.

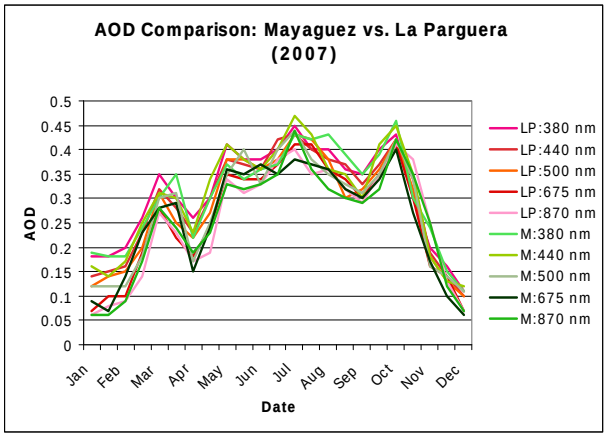

Fig. 1: AOD comparison: Mayaguez vs La Parguera

Fig. 2: AOD spatial variability test sites

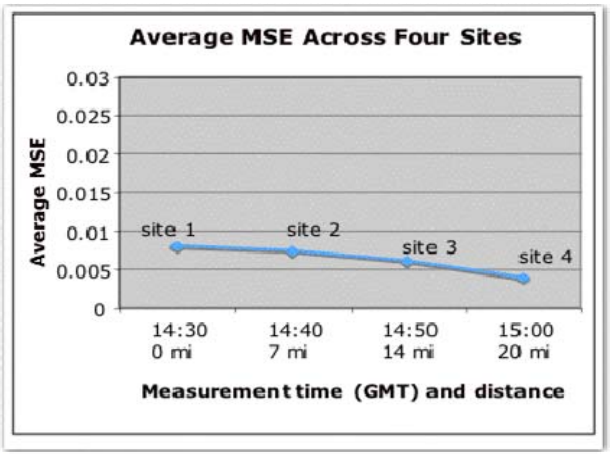

Fig. 3: AOD spatial variability test results

2 Artificial Neural Network Design

Artificial neural networks are commonly used to solve problems that involve nonlinear estimation. They are powerful tools in signal processing, due to strong numerical approximation capabilities [8]. The network's processing power rests in its distributed nature. Connections between neurons can be represented with matrices, and network execution consists of simple matrix operations [9]. In this case, we are trying to solve an ill-posed problem, based on a pattern between AOD and ASD. In order to start off with a well-implemented, efficient neural modeling environment, Matlab's Neural Network Toolbox was used to create the network. The artificial neural network used to estimate ASD is a feed-forward, backpropagation network with an input vector of size five, corresponding to the five wavelengths of an AOD reading.

Backpropagation networks have the largest number of successful applications in their field, and have almost become the standard for modeling, forecasting, and classification domains [10]. The network contains a hidden layer of 13 neurons, and an output layer of 22 neurons, which correspond to the 22 points on an AERONET ASD plot. The number of neurons in the hidden layer was experimented with in order to maximize network performance. Increasing the number of neurons in the hidden layer lowered the training error, but not significantly by any means. The increase in size also caused the network to take much longer to train, becoming less feasible as more neurons were added. Each layer in a neural network has a transfer function associated with it. In this case, the hidden layer has a log-sigmoid transfer function, and the output layer a linear transfer function.

A non-linear transfer function was selected for the hidden layer to allow the network to predict a non-linear pattern.

3 Network Implementation

In order to generate training data for the neural network, a data pool is necessary. Each element in the data pool consists of a pair, p(x, y), where x is the average AOD of one day, and y is the size distribution for that day, both provided by AERONET for the same day. In order to make the data provided by AERONET compatible with the ANN software provided by Matlab, a Java parser was written to translate the AERONET data into Matlab matrices that the ANN can interpret. The parser basically takes any AERONET AOD and ASD data that is given to it and finds dates for which valid AOD and ASD readings were taken. It then creates data points (p(x,y) pairs) in the form of an input matrix consisting of AOD readings, and a corresponding target matrix consisting of the ASD readings that complete the data point pairs.

In order to find the optimum number of neurons in the hidden layer, the network was tested with various different hidden layers, ranging from 13 neurons to 20 neurons. Fig. 4 shows a plot resulting from these tests.

As expected, the training error is almost constant, but slightly decreasing. The testing error, which represents the network's behavior against unknown data points, reaches a minimum at 17 hidden neurons, and then begins to increase. This trend is usually seen as a result of the network over-fitting the data (learning the noise as well as the actual data). Fig. 5 shows the final network architecture.

The network was trained with the popular Levenbergh-Mardquat training algorithm [11]. In order to ensure basic network functionality, it was tested with AOD readings that were part of the training set. The ANN performed as expected, consistently producing an ASD that matched the ASD provided by AERONET for the day of the reading used, with an error of epsilon (the smallest numerical value Matlab can represent). Of course, the network is expected to estimate ASD for AOD readings that are not part of the training set.

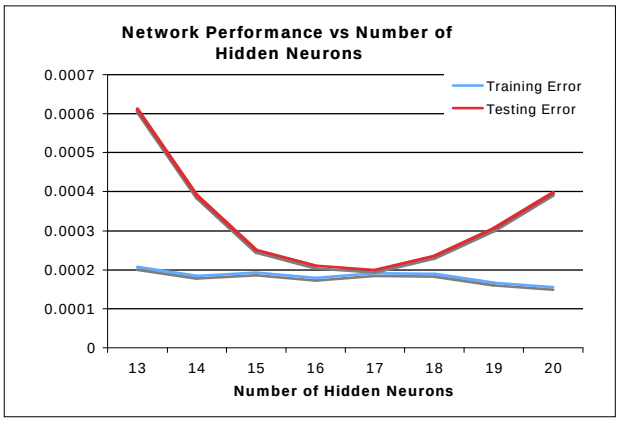

Fig. 4: Network Performance vs Number of Neurons in the hidden Layer

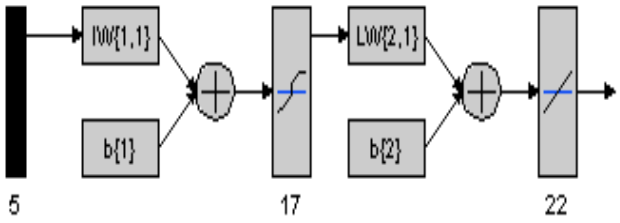

Fig. 5: Network Architecture

The first attempts at testing the network with unknown AOD inputs were done using the entire data pool as the network's training set. In other words, all the data points available were used for training. In an attempt to find a training set size that minimized network error, training sets were gradually made smaller, and more localized to the testing dates in question, but this procedure only increased network variability and error.

4 Results

The network was trained with all the available points from La Parguera, excluding the test points. The network was fed an AOD input measured by AERONET at La Parguera, and the ASD output was compared to that day's AERONET ASD reading. In other words, the network was asked to predict the ASD in La Parguera based on the AOD from La Parguera. The network performed with a mean squared error of the order 10^{-4}, producing ASD predictions such as those shown in Figs 6, 7 & 8.

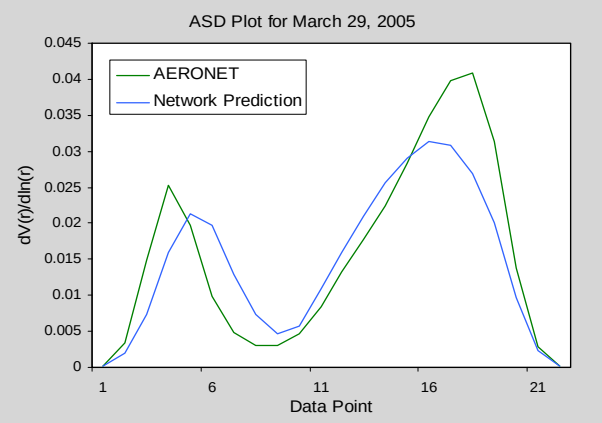

Fig. 6: ASD plot for March 29, 2005, trained with all data points

The network generally produces ASD plots that are a decent match of AERONET ASD plots.

In a separate comparison, the network was fed AOD inputs measured in Mayaguez with the sun-photometer, and the ASD output was compared to the AERONET ASD reading for that day. The network still performed with a mean squared error of the order 10^{-4}, despite using Mayaguez AOD as input, and not AERONET Parguera AOD. Figs. 9, 10 & 11 show examples of these comparisons.

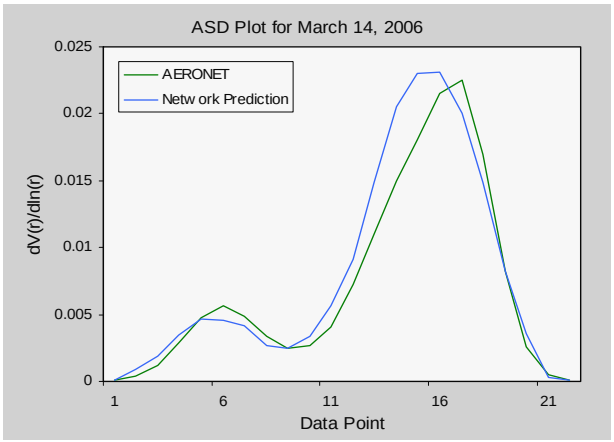

Fig. 7: ASD plot for March 14, 2006, trained with all data points

Several other training sets were experimented with. Specifically, the network was trained with data pertaining to the days before the day of the reading being predicted. As the training set became smaller, network performance became less consistent.

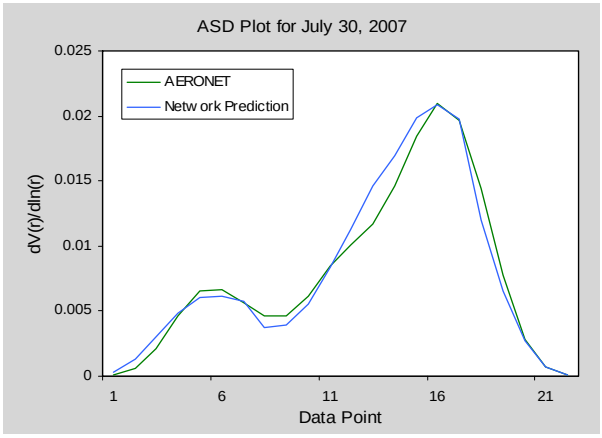

Fig. 8: ASD plot for July 30, 2007, trained with all data points

It made more accurate predictions on some dates, but also made many illogical predictions when the input data didn't resemble known input data.

The ANN's performance for any given day basically depended on whether or not the training set contains data similar to the day being predicted. If it does, then the prediction will usually be fairly accurate. However, it is impossible to know if the training set contains readings similar to the one being predicted, so limiting the data set to a monthly, weekly, or even seasonal set is out of the question. A network trained with a limited data set can achieve errors of epsilon, but this is misleading, since it is only memorizing the training data, and in a sense, becoming a sort of savant, that is, it only predicts ASD correctly when it is

similar to the ASD it was trained with; otherwise, incorrect predictions will occur.

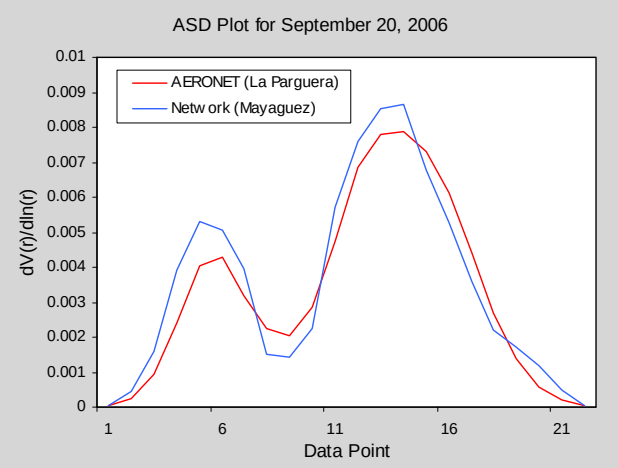

Fig. 9: ASD comparison for September 20, 2006, trained with all data points

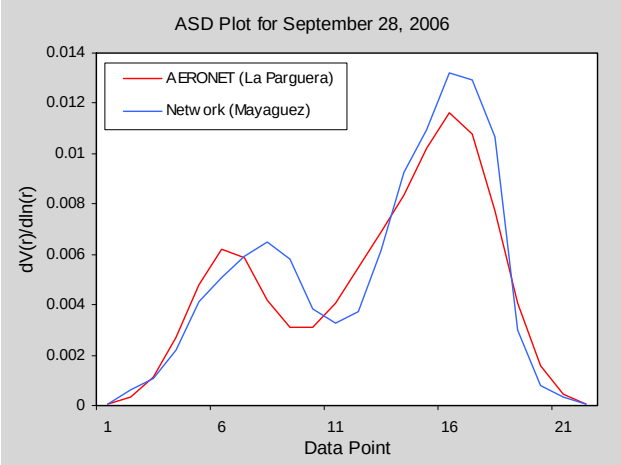

Fig. 10: ASD comparison for September 28, 2006, trained with all data points

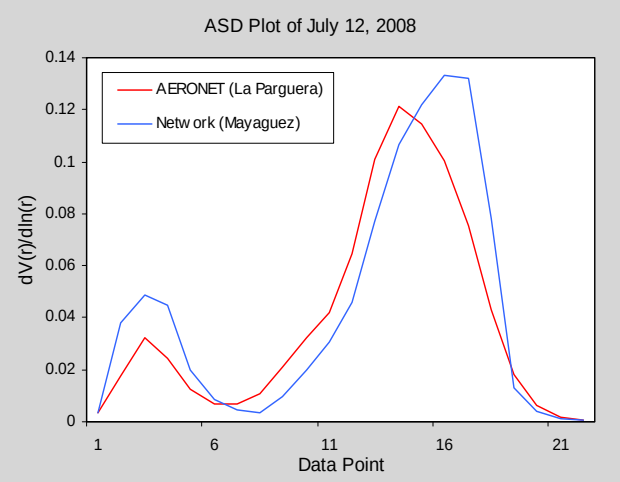

Fig. 11: ASD comparison for July 12, 2008, trained with all data points

5 Conclusion

It is clear from Figs. 6, 7 & 8 that when trained with all the data available from La Parguera and using Parguera AOD as input, the network generates accurate estimates of Parguera ASD. However, if the network had more data in its training set, its prediction error could be lowered and it could map the relationship between AOD and ASD more accurately. Figs. 9, 10 & 11 show that as with AOD, ASD between La Parguera and Mayaguez are very similar, and the similar low mean squared errors of both comparisons indicate that La Parguera ASD readings generated by the AERONET station can be used at Mayaguez, and that for days when AERONET cannot make measurements at La Parguera, the Sun-photometer at the LIDAR site can be used to gather AOD, and feed that into the ANN to generate a good estimate of the local ASD.

References:

[1] A. Blanco, F. De Tomasi, et. al, *Characterization of African dust over southern Italy*, ACP-2003.

[2] AERONET Aerosol Robotic Network, http://aeronet.gsfc.nasa.gov.

[3] Andres Bonilla, Hamed Parsiani, *Aerosol Optical Depth Spatial Variation as Sensed by a Sunphotometer and Compared to a Fixed AERONET Station*, SIP-2008 Proceedings, Aug. 18-20, 2008, Hawaii.

[4] Christine Bockman, *Hybrid regularization method for the ill-posed inversion of multiwavelength lidar data in the retrieval of aerosol size distributions*, Optical Society of America, 2001.

[5] Umberto Amato, Maria Francesca Carfora, et. al, *Objective algorithms for the aerosol problem*, Applied Optics vol. 34, 5442-5452, 1995.

[6] M. D. Muller, A. K. Kaifel, et. al, *Ozone profile retrieval from Global Ozone Monitoring Experiment (GOME) data using a neural network approach*, Journal of Geophysical Research, vol. 108, 2003.

[7] Helmut Schiller and Roland Doerffer, *Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data*, International Journal of Remote Sensing, vol. 20, 1999.

[8] Maja Sarevska, Abdel-Badeeh M. Ssalem, Nikos Mastorakis, *Null Steering Beamformer Based on RBF Neural Networks*, 12th **WSEAS** International Conference on Communications, Heraklion, Greece, July 23-25, 2008.

[9] Yelda Ozel , Irfan Guney , Emin Arca, *Neural Network Solution to the Cogeneration System by Using Coal*, 12th **WSEAS** International Conference on Circuits, Heraklion, Greece, July 22-24, 2008.

[10] Songyot Sureerattanan, Huynh Ngoc Phien, et. al, *The Optimal Multi-layer Structure of Backpropagation Networks*, Proceedings of the 7th **WSEAS** International Conference on Neural Networks, Cavtat, Croatia, June 12-14, 2006.

[11] Amir Abolfazl Suratgar, Mohammad Bagher Tavakoli, and Abbas Hoseinabadi, *Modified Levenberg-Marquardt Method for Neural Networks Training*, Proceedings of World Academy of Science, Engineering and Technology, vol. 6, 2005.

Appendix A: AERONET

The Aerosol Robotic Network is a project that provides public access to a global database of several Aerosol-related products, among them, AOD and ASD. A web application allows the user to choose from hundreds of sites around the world, and provides data display controls to choose the date or date range of interest. A simple download tool allows the user to download specific data sets.

However, AERONET AOD readings are made with an instrument called a CIMEL, not with a Sun-photometer [Appendix B]. The CIMEL is more precise than the Sun-photometer and has a scanning capability, so while the Sun-photometer analyzes a cone of air, the CIMEL rotates and scans an entire radius. However, a CIMEL instrument is much more expensive than a Sun photometer, is not portable, and must be registered with a person who keeps track of all the CIMELs.

The fact that there is a CIMEL at La Parguera makes it much less likely that Mayaguez will be authorized to install one as well, since both sites are only 20 miles apart. The high cost of the device also makes it much more difficult to acquire. This research intends to evaluate the feasibility of calculating ASD from AOD produced with a portable Sun-photometer, and not with a CIMEL. Fig. 13 depicts a CIMEL instrument like that used by AERONET at La Parguera. Readings are taken daily, but there are many gaps in the data, since, as with the Sun-photometer, clear, cloud-free skies are necessary for the instruments to take readings. This detail has proven to be one of the most limiting factors in the accuracy of the ANN predictions.

Fig. 13: CIMEL instrument

The limitation lies in the fact that the ANN requires an AOD input and an ASD output taken on the same day in order to include that day in its training set. This means a gap in either data set (AOD or ASD) is essentially a gap in both data sets as far as the ANN is concerned. This limitation will become more evident when the ANN training algorithms are discussed. Figs. 14 & 15 contain examples of AERONET AOD and ASD readings, respectively.

For the AOD readings, only the 5 wavelengths that are shared with the sun-photometer are taken into account. For the ASD, all 22 points on the plot are used.

Appendix B: Sun-photometer Theory

The instrument being used to measure AOD at UPRM is a 5-wavelength (380, 440, 500, 675, and 870 nm) MICROTOPS II portable Sun-photometer, pictured in Fig. 12. The device is equipped with 5 optical collimators with a full field view of 2.5°. Internal Baffles eliminate reflections inside the device. Each channel has a narrow-band interference filter and a photodiode suitable for the particular wavelength range. The collimators are encapsulated in a cast aluminum optical block for stability. A sun target and pointing assembly is permanently attached to the optical block and laser-aligned to ensure accurate alignment with the optical channels. When the image of the sun is centered in the bull's-eye of the sun target, all optical channels are oriented directly at the solar disk. A small amount of circumsolar radiation is also captured, but it makes little contribution to the signal. Radiation captured by the collimator and band-pass filters radiate onto the photodiodes, producing an electrical current that is proportional to the radiant power intercepted by the photodiodes. These signals are first amplified and then converted to a digital signal by a high resolution A/D converter.

The signals from the photodiodes are processed in series. However, with 20 conversions per second, the results can be treated as if the photodiodes were read simultaneously. The optical depth due to Rayleigh scattering is subtracted from the total optical depth to obtain AOD. Optical depth from other processes such as O_3 and NO_2 absorption are ignored.

An AOD reading obtained from this instrument consists of a vector of five values, which correspond to five wavelengths: 380, 440, 500, 675, and 870 nm. The device contains constants of the irradiance measured from the sun in a clean environment (taken at Mauna Loa, Hawaii). The reading is essentially a ratio of the amount of sunlight reaching the device and the constants previously mentioned. One of the drawbacks of the device is the fact that clear skies are necessary in order to take readings. Mayaguez, Puerto Rico is infamous for daily showers, but they tend to happen in the afternoon. Equation 1 shows how the device computes AOD readings from voltage measurements.

$$AOD_\lambda = \frac{\ln(V_{0\lambda}) - \ln(V_\lambda * SDCORR)}{M} - \tau_{R\lambda} * \frac{P}{P_0}$$

Eq. 1: Sun-photometer equation

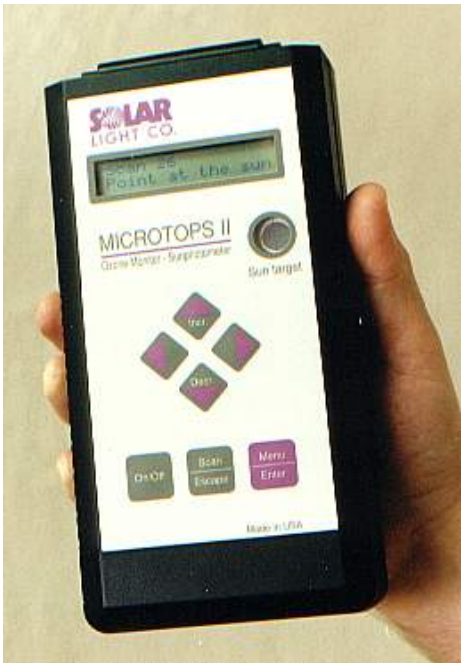

Fig. 12: MICROTOPS II portable SPM

In Eq.1, λ is the channel wavelength (380, 440, 500, 675, or 870), ln(V0λ) is the AOD calibration constant (taken at Hawaii), V_λ is the signal intensity in [mV], SDCORR is the earth-sun distance correction, M is the optical airmass, $\tau_{R\lambda}$ is the Rayleigh optical thickness, P is the station pressure, and P_0 is standard sea-level pressure.

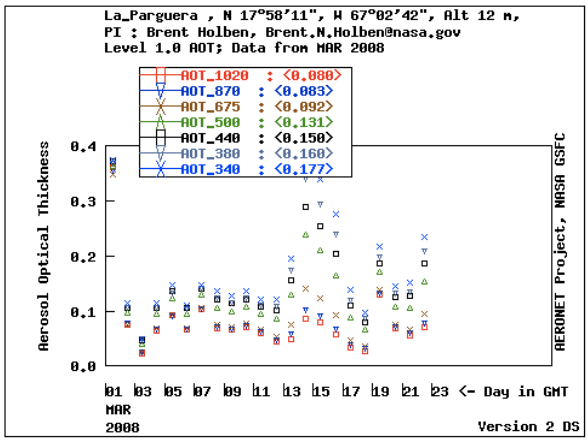

Fig. 14: AERONET AOD plot

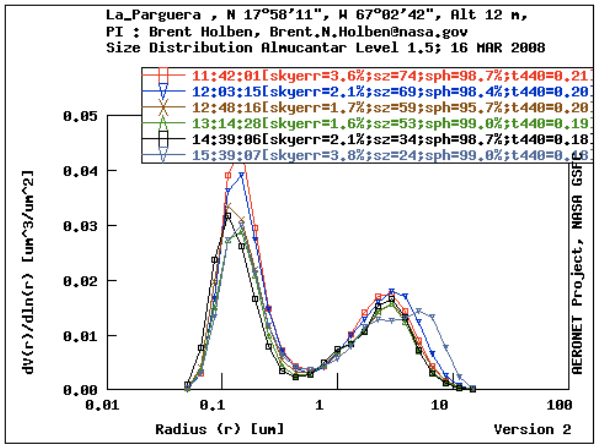

Fig. 15: AERONET ASD plot