Winter 1-13-2016

Retrival of Atmospheric Aerosol Size Distributions Using Stochastic Particle Swarm Optimization

Benjamin D. Nault-Maurer
College of Saint Benedict/Saint John's University, bnaultmaurer@gmail.com

RETRIEVAL OF ATMOSPHERIC AEROSOL SIZE DISTRIBUTIONS USING STOCHASTIC PARTICLE SWARM OPTIMIZATION

AN ALL COLLEGE THESIS

College of St. Benedict/St. John's University

In Partial Fulfillment

Of the Requirements for Distinction

In the Department of Physics

By

Benjamin Nault-Maurer

January, 2016

Project Title: Stochastic Particle Swarm Optimization in Regards to Atmospheric Aerosol Size Distributions

By Benjamin Nault-Maurer

Approved by:

_____

Dr. Matthew ArchMiller

Visiting Assistant Professor of Physics

_____

Dr. Todd Johnson

Assistant Professor of Physics

_____

Dr. Adam Whitten

Advisor, Visiting Assistant Professor of Physics

_____

Dr. Jim Crumley

Chair, Professor of Physics

_____

Dr. Emily Esch

Director, All College Thesis Program

Table of Contents

1. Abstract

A stochastic particle swarm optimization (SPSO) technique's robustness is studied in regards to atmospheric aerosol size distribution estimations for a bimodal distribution that focuses on Aitken and accumulation mode aerosols. The SPSO method is used to calculate a set of 11 aerosol optical depth (AOD) values based on a size distribution and match them to an inputted set of AOD values. This method is tested using computer generated AOD values with fixed distribution parameters, generated AOD values with varying distribution parameters, two sets of AOD measurements in clear conditions, and one set of AOD values in hazy conditions. The SPSO method is found to be robust and consistent for generated AOD values when keeping mean radii and width distribution parameters constant, but less consistent when all parameters are allowed to vary. For real data, the SPSO program was not consistent from run to run, and a process of averaging the runs is suggested. An analysis of the Mie scattering coefficient's large differences at small radii ranges and analysis of the consistency of the AOD values for the hazy conditions lead us to suggest that testing a bimodal distribution that focuses on accumulation and course mode aerosols should be done, as the major inconsistencies in the distributions occur in these small radii ranges.

2. Introduction

Atmospheric aerosols are small particulates suspended in the atmosphere. They are mostly natural materials such as small droplets of water with dissolved chemicals, dust, soot, salt, and sand, but can also be caused by humans, through industrial pollutants and vehicle emissions. They also can take part in chemical reactions and provide surfaces for absorption of gases. The aerosols' lifetime can vary from days for large particles to years for small particles. They are not uniformly distributed and move with different air masses. Therefore the size distributions of the aerosols we see over a particular location can change from minute to minute, week to week, and season to season depending on the motion of the air masses we observe. One of the most important aspects of aerosols is that they scatter and absorb light, which can play a role in the Earth's climate.

One way aerosols effect the climate is they reflect light back into space, causing less light to reach the surface of the Earth. They can also act as a base for water to condense upon, creating larger and darker clouds. These both lead to a reduction in the solar radiation that reaches the Earth's surface. In extreme cases this radiation reduction can cause major temperature drops. After the major volcanic eruption of Mount Pinatubo in 1991, so many aerosols were shot into the atmosphere that a worldwide temperature drop of about 1°F was measured in the years following the event (Newhall, Wendley and Stauffer).

Due to the effect aerosols have on the climate, it is important to study their concentrations and size distributions. However, this is difficult as their lifespan is so short. Their lifespan depends on the type of the aerosol. There are three main modes of aerosol: Aitken, accumulation, and course mode aerosols. Aitken mode aerosols are the finest of the three, and are composed mostly of sulfates, elemental carbon, and metal compounds, while accumulation mode aerosols are made up of sulfate (SO_4^-), Nitrate (NO_3^-), Ammonium (NH_4^+), Hydrogen (H^+), different metals, and many different types of

organic material. Course mode aerosols are the largest, and are composed of soil, dust, ash, salt, mold, pollen, and large chemicals such as nitrates and chlorides from HNO_3 and HCl (Wilson and Suh).

Along with varying in size, these modes have many different characteristics such as how they are emitted into the atmosphere, how long they stay there, and how they are removed from the atmosphere. Both Aitken and accumulation mode aerosols are created mostly through combustion processes, condensation, atmospheric reactions, and the evaporation of clouds where these materials were dissolved in the water vapor. Course mode aerosols tend to be created through sprays, mechanical disruption such as construction or agriculture, and also through evaporation. Aitken mode aerosols tend to stay suspended in the atmosphere anywhere from a few minutes to a few hours, while accumulation and course mode aerosols stay in the atmosphere for a few days to a few weeks. In terms of removal, Aitken mode aerosols grow in size to form accumulation mode aerosols, accumulation mode aerosols are removed through rain and dry deposition, and course mode aerosols are removed through dry deposition and fallout. The amount of these types of aerosols are determined through size distribution calculations, which can be found through measurements of solar irradiance and aerosol optical depth values (Wilson and Suh).

This study investigates atmospheric aerosols using ground based solar irradiance measurements. From these irradiance values, aerosol optical depth (AOD) can be found. Aerosol optical depth is a comparison of the solar spectrum we measured to the standard E-490 solar spectrum created by the American Society for Testing and Materials (Habte). This standard solar spectrum is used for ground based irradiance measurements as well as other projects in the atmospheric science community. AOD values are used as a standard measurement for air quality and aerosol research, as a high AOD measurement indicates more light was scattered, meaning there are more aerosols in the atmosphere.

In order to determine AOD values through irradiance measurements, we must take into account factors such as Rayleigh scattering, aerosol scattering, and absorption by gasses such as ozone and water vapor. The measured irradiance for any given wavelength can then be given by

$$L_\lambda = \frac{L_{\lambda_0}}{D^2}\left(e^{-\alpha X\mu - m\beta\frac{P}{P_0} - \delta m'}\right) \quad (2.1)$$

where L_λ is the measured irradiance, L_{λ_0} is the irradiance at the top of the Earth's atmosphere (given by the E-490 spectrum), D is the distance from the Earth to the Sun in astronomical units (AU), α is the absorption coefficient of the absorbing species in cm^{-1}, X is the total amount of absorber in a vertical column at STP, μ is the ratio of actual and vertical paths of the light through the absorbing layer, β is the Rayleigh (molecular) scattering coefficient in atm^{-1}, P is the pressure at the point of measurement, P_0 is the mean sea level pressure, δ is the aerosol (Mie) scattering coefficient, m is the optical path allowing for refraction, and m' is $sec\theta$ where θ is the solar zenith angle (London).

Table 2.1: Wavelengths Used for AOD Calculations

Number	Standard Wavelength (nm)	Wavelength Used (nm)
1	440	441
2	463	463
3	500	500
4	520	520
5	556	556
6	610	610
7	675	675
8	750	750
9	778	778
10	870	870
11	1020	1020

By choosing to look at certain wavelengths, we can avoid certain factors such as the absorption by gasses. In this study, we use 11 wavelengths that include World Meteorological Organization (WMO) standard wavelengths for AOD calculations. These wavelengths can be found in Table 2.1. The wavelengths in this table are the closest wavelengths to the standard wavelengths as we can get on our photometer. These wavelengths have been specifically chosen to avoid absorption by gasses such as

ozone and water vapor, allowing us to exclude the effect of these in our equation above. By removing this factor and solving for δ, we get

$$\delta = \frac{\ln\left(\frac{L_{\lambda_0}}{L_\lambda D^2} - \beta m \frac{P}{P_0}\right)}{m'} \tag{2.2}$$

Retrieving a size distribution with *n* unknown parameters requires *n* or more AOD values to solve the inverse problem. In choosing which wavelengths to use, we determined 11 wavelengths is the highest number of wavelengths that provide uniform intervals in the wavelength range and unique AOD values. Wavelengths that are too close together in the spectrum give similar AOD values and the size distribution program has difficulty matching a distribution to these AOD values. Also, we need to look at wavelengths that our photometer can provide accurate measurements for. If the wavelength is too far in the UV spectrum or too far in the infrared spectrum, we cannot get accurate measurements as we see a lot of noise at the edges of the photometer's range. In previous studies we have run tests using 6, 11, 12, 13, and 16 different wavelengths, and the 11 wavelengths that we use provide the most AOD values that are unique and therefore the most information that can be used to calculate size distributions.

Previously the Rayleigh scattering coefficient, β, was found through linear interpolation of calculations performed by Bucholtz (1995), but it was noticed that sometimes this yielded unrealistic negative AOD values, suggesting that interpolation of the Bucholtz values were sometimes too large. We believe the increased concentration of $C0_2$ may have been affecting this, and we have found a more accurate Rayleigh scattering coefficient calculation that takes into account air composition (Bodhaine et al.). With the new Rayleigh coefficient, we get

$$\delta = \frac{\ln\left(\frac{L_{\lambda_0}}{L_\lambda D^2}\right)}{m'} - \sigma \frac{PA}{m_a g} \tag{2.3}$$

where L_λ, L_{λ_0}, D, P, and m' are as they were before, σ is the scattering cross section per molecule, A is Avogadro's number, m_a is the mean molecular mass of air, and g is the acceleration of gravity at the

measurement site. This is the calculation we use to calculate AOD values based off our measurements of solar irradiance.

The SPSO program uses a different method to calculate AOD, as it calculates AOD based off of an aerosol size distribution, not an irradiance measurement. The SPSO program calculates AOD values based off many different types of distributions, which can be seen in Table 2.2. The AOD values are calculated by

$$\tau_c = \int N(r)\pi r^2 Q_{\text{ext}}\, dr \tag{2.4}$$

where τ_c is the calculated AOD, r is the radius of the aerosol particle, Q_{ext} is the Mie scattering coefficient given by Mie scattering theory (Bohren and Huffman), and N(r) is the number density distribution of aerosols. With integrating the size distribution and Mie extinction coefficient over all radii, you can imagine a sharp and narrow distribution could give the same AOD as a wide and shallow distribution. This will be discussed further in Section 5.

In this study the bimodal distribution is used to separate the effects of Aitken and accumulation mode aerosols, as they are both fine particles that are similar in size. There is also a trimodal distribution (which was not used in this study) discussed in Yuan et al. (2011) that takes into account the

Table 2.2: Possible Distribution Functions

Distribution	Equation
Junge	$N(r) = N_0 \beta r^{-\alpha}$
Gamma	$N(r) = N_0 r^{\beta} e^{-\alpha r}$
Log-Normal	$N(r) = \frac{N_0}{\sqrt{2\pi}\beta r} \text{Exp}\left[\frac{-\text{Log}\left[\frac{r}{\alpha}\right]^2}{2\beta^2}\right]$
Bimodal	$N(r) = \frac{N_1}{\sqrt{2\pi}\beta_1 r} \text{Exp}\left[\frac{-\text{Log}\left[\frac{r}{\alpha_1}\right]^2}{2{\beta_1}^2}\right] + \frac{N_2}{\sqrt{2\pi}\beta_2 r} \text{Exp}\left[\frac{-\text{Log}\left[\frac{r}{\alpha_2}\right]^2}{2{\beta_2}^2}\right]$
Trimodal	$N(r) = \frac{N_1}{\sqrt{2\pi}\beta_1 r} \text{Exp}\left[\frac{-\text{Log}\left[\frac{r}{\alpha_1}\right]^2}{2{\beta_1}^2}\right] + \frac{N_2}{\sqrt{2\pi}\beta_2 r} \text{Exp}\left[\frac{-\text{Log}\left[\frac{r}{\alpha_2}\right]^2}{2{\beta_2}^2}\right] + \frac{N_3}{\sqrt{2\pi}\beta_3 r} \text{Exp}\left[\frac{-\text{Log}\left[\frac{r}{\alpha_3}\right]^2}{2{\beta_3}^2}\right]$

effect of course mode aerosols. These various distributions can be used to best fit a given set of measured AOD values.

Along with AOD, it is also important to measure what types of aerosols are in the atmosphere. This is done by determining the size distribution of aerosols in the atmosphere. However, estimations of atmospheric aerosol size distributions are difficult to determine, as it is an inverse problem. A certain set of AOD values correspond to a certain size distribution, but that size distribution is not known. One method that is considered to be robust is a stochastic particle swarm optimization (SPSO) program. This program, based off the article from Yuan et al. (2010), uses a popular computer science optimization technique to optimize a size distribution function to a certain set of AOD values. This program generates many size distributions, and optimizes a fitness function that compares the AOD values of the generated distribution to the input, measured AOD values.

3. Method

3.1 Solar Irradiance Measurements

In order to calculate aerosol size distributions, we need to do many calculations to get aerosol optical depth (AOD) values to input into the SPSO size distribution program. These AOD values depend on measurements of the solar spectrum.

To take measurements of the solar spectrum, we use a Kipp and Zonen PGS-100 solar spectrophotometer (see Fig 3.1). This device uses a silicon CCD array with 2,048 pixels to measure a spectrum of the light coming from the Sun. The device is set to take a measurement of the solar spectrum every minute of every day, and each measurement contains intensity measurements of 960 wavelengths between 350-1050 nm. This allows for a resolution of about 0.5 nm per pixel. Each wavelength's

Fig. 3.1. Solar tracking mechanism and PGS-100 photometer

intensity is measured in counts, which can later be converted into units of intensity, using conversion constants specific to our device (see Fig. 3.2). In order to retrieve accurate irradiance measurements the photometer must be accurately aligned. If the photometer is misaligned, or if some object (like a cloud, bird, leaf, etc.) is in the way, we will get a spectrum that will provide inflated AOD values later on.

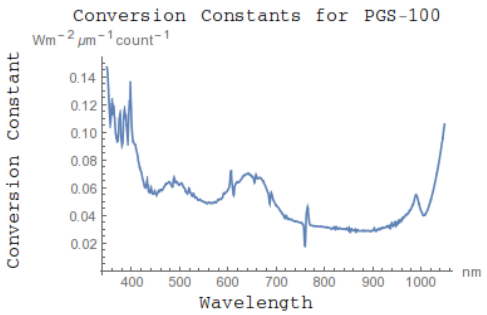

Fig. 3.2. Manufacturer supplied calibration constants

First of all, in order to get accurate measurements, we need to have a clear line of sight. Since we need light coming directly from the Sun, we need to make sure there are no clouds, haze, birds, jet contrails, etc. physically blocking the light from the photometer. Even some visible haze would diminish the spectrum intensity enough to provide inflated values later on, as the calculations depend on how much light is scattered. Haze is a boundary layer atmospheric aerosol, and our results show a combined effect of boundary layer, tropospheric, and stratospheric aerosols. So if haze is blocking some light from getting into the photometer, the spectrum we get will be diminished, and our programs will interpret this as the atmosphere containing different types of aerosols than we are expecting. This is a great example of how important visual observations of the atmosphere are while measurements are being taken, as this leads to a better interpretation of the meaning of the AOD values. In this case, we would be able to interpret the cause of the increase in AOD values was the visible haze in the atmosphere, and we would expect larger AOD values in this wavelength range.

Secondly, if the photometer is not perfectly aligned with the Sun, we will get a spectrum that is not a spectrum of the light rays coming directly from the Sun. We need to get the light rays coming directly from the Sun as those are the rays that are scattered and absorbed only due to aerosols. If we look at a spectrum that is not directly from the Sun, the light has had a chance to scatter with more particles, reflect off of different materials, and is ultimately not a measurement of the correct spectrum.

In order to take an aligned measurement, the photometer we use needs to be within 0.2° of the center of the Sun which has an angular diameter of 0.5°. This means the photometer needs to be aimed at the inner 64% of the Sun's cross sectional area.

In order to keep the photometer aligned, we have developed an automated solar tracking mechanism that aims the photometer at the Sun throughout the day. This device allows us to take consistent data as we take a measurement every minute of every day, but is another source of measurement problems as it doesn't keep the photometer perfectly aligned every time. In order to deal with this, we created a program that sorts through each data point in order to determine if the measurement is aligned or not.

For a good measurement, we look at the wavelength 782 nm, as that is the first peak on the infrared side of the oxygen absorption line. This can be seen in Fig. 3.3, a sample of an irradiance measurement on a clear day. This peak typically has the highest irradiance value for the measurement, so it makes for a good benchmark. Being the peak of the irradiance, it is sensitive to fluctuations in intensity, so we are able to set an accurate cutoff value in order to only look at the clearest measurements. If this wavelength is below 30,000 counts, we do not use the entire measurement. We decided on this cutoff value after going through many days' worth of data and looking at the count value before and after we realigned the photometer, and on clear days this wavelength's count value is above 30,000 counts.

We have built a program to automate this check. This program goes through each measurement of every day to determine whether the measurement was good or not, based off the 30,000 count cutoff at 782 nm. This program acts as

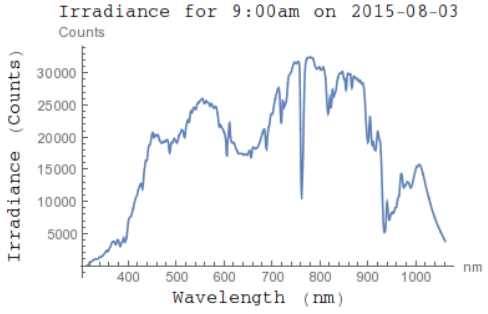

Fig. 3.3. Raw data from the PGS-100 photometer

the check for a misaligned photometer or an object blocking the sunlight's path. After running this program, we have a set of count values for the 11 wavelengths for each time throughout the day. These values will then be used to calculate aerosol optical depth (AOD).

We have built a program to calculate AOD using Equation 2.3. This program first takes the 11 wavelengths' count values and converts them to irradiance values using the wavelength and photometer specific conversion constant (see Fig. 3.4 for a completely converted spectrum). It then performs calculations based on the date and time (which are used to calculate all the required angles and the distance from the Sun to the Earth), location (which includes latitude, longitude, and height), the pressure at the time of measurement, and the current CO_2 concentration. Once these calculations are made, the program outputs an AOD value for each wavelength. Using Equation 2.3 is the most accurate method we have found to calculate AOD from measurements of solar irradiance.

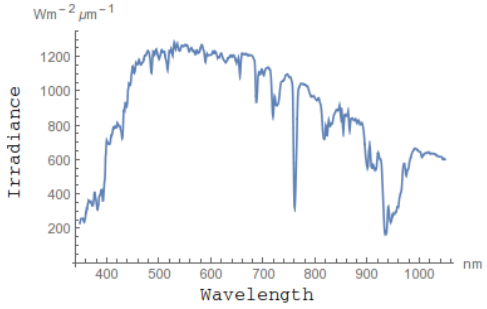

Fig. 3.4. Irradiance converted from counts to W/m2/µm using the conversation constants from Fig. 3.2

3.2 Stochastic Particle Swarm Optimization Theory

Particle swarm optimization (PSO) is an optimization technique that is capable of finding solutions without any theoretical proof. It does this by effectively searching the domain for an optimum solution, through an advanced random generation algorithm. Stochastic particle swarm optimization (SPSO), explored by Yuan et al. (2010), slightly alters this random generation method by allowing past best solutions to influence the new random solutions, as well as generating a scout particle that prevents the solution from going into a local minimum.

SPSO works by varying a swarm of particles. Each particle represents a set of parameters that define a distribution function, and there is a certain best distribution that we are searching for. To start, an initial swarm is generated and a fitness function

$$f_{\text{fitness}} = \sqrt{\sum_{i=1}^{n} \frac{[\tau_m(i) - \tau_c(i,a)]^2}{n}} \tag{3.1}$$

is calculated for each particle, where n is the number of wavelengths, τ_m is the wavelength's measured AOD value, and τ_c is the calculated AOD value for that wavelength, given a distribution from above. This fitness function is a comparison between AOD values calculated by the program for the particle's distribution and the AOD values that are measured and inputted by a user. The goal of the PSO method is to minimize this value.

The program then determines the particle with the best fitness function, and sets that particle's parameters as the global best. A particle's position contains 3 (N0, α, β), 6 (N1, α1, β1, N2, α2, β2), or 9 (N1, α1, β1, N2, α2, β2, N3, α3, β3) parameters depending on the chosen distribution. The program then generates a new swarm based off this global best value, creating a new generation of the swarm and adjusting each particle according to the position algorithm

$$X_i(t+1) = X_i(t) + c_1 r_1 [P_l(t) - X_i(t)] + c_2 r_2 \left[P_g(t) - X_i(t)\right] \tag{3.2}$$

where $X_i(t)$ is the particle's current position, c_1 and c_2 are acceleration coefficients set at c_1=c_2=1.8, r_1 and r_2 are random numbers, P_l is the local best position for that particle, and P_g is the global best position for the swarm. The subscript i represents the number of particles, and the value t is the time in terms of generations.

After adjusting the swarm, new fitness values are calculated. For each particle the new fitness value is compared to the previous fitness value. If the new fitness value is less, then the new position is

set as the local best position, otherwise the previous position remains the local best position. The overall lowest new fitness value for the entire generation is compared to the previous global fitness value. If the new value is less then it becomes the new global fitness value and that particle's position becomes the new global best position.

This algorithm adjusts the swarm to find the best fitness value, but it is possible that this swarm falls into a local minimum value, optimizing the swarm around a distribution that is not the best solution in the domain. In order to prevent this, Yuan et al. (2010) suggests adding a scout particle P_j that is randomly generated outside of the swarm each generation, allowing it to explore more of the domain than the position algorithm limits the swarm to. If this scout's fitness function is better than the swarm's global fitness value, it is set as the new global best fitness value, preventing the swarm from optimizing a local minimum. This adds the stochastic part to the PSO problem, allowing the swarm to explore the entire domain each generation in search for the best fitness value.

We have found that simply adding one scout particle per generation is not robust. To solve this problem we have added a new scout population which is generated for the first 10 generations of the swarm. For each of the ten generations a new population of 1,000 scouts is randomly generated throughout the domain. If any of these scouts have a better fitness value than the particles in the swarm, the new global best fitness value is adjusted. This is meant to replace the scout method above, and is meant to decrease the time it takes for the swarm to find an optimum solution.

A solution is reached if the parameters for the global best fitness value do not change by a certain amount for 100 generations. This delay of 100 generations allows for a more robust solution, and it also gives the swarm time to search the area around the global best solution, looking for a better solution. If there is no solution found after 3,000 generations, the best solution is output and a whole new population is created, repeating the entire process in search of a better value.

For our calculations we have set the various parameter values based off of suggestions from Yuan et al. (2010) and from our own experience. We allow only 3,000 generations so that we avoid unnecessary computation time, as we have found that it is faster to simply create a new population after 3,000 generations than it is to wait for the swarm to converge. We also have balanced swarm size versus computation time, and we use a swarm size of 50 particles as suggested by Yuan et al. (2010).

To test the reliability of this SPSO method, we used the same equations as in the SPSO program and created a program that calculates AOD values based off parameters of a size distribution. This allowed us to create a specific set of AOD values where the parameters of the size distribution were known exactly. By using these generated AOD values we were able to test settings regarding certain parameters as well as test the consistency of solutions from multiple runs of the SPSO program. We were then able to plot these distributions and analyze how different the distributions and parameters were, as well as look at how much the calculated AOD values differed from the measured AOD values.

4. Data

4.1 SPSO Algorithm on Generated Data with Set Parameters

The SPSO algorithm was successful in being consistent and accurate with generated AOD values where the distribution type and parameter values were defined. It was able to reproduce these parameters to within 0.0003% (see Fig. 4.2) of the input values when some of the parameters were defined. This shows that for multiple runs we can get reliable, consistent data.

For this test, we generated AOD values with parameters fixed at: N1 = 1.0E10 $atm^{-1}\mu m^{-2}$, N2 = 1.0E8 $atm^{-1}\mu m^{-2}$, $\alpha 1$ = 0.1 μm, $\alpha 2$ = 1.0 μm, $\beta 1 = \beta 2$ = 0.307 μm. The $\alpha 1$, $\alpha 2$, $\beta 1$, and $\beta 2$ values were

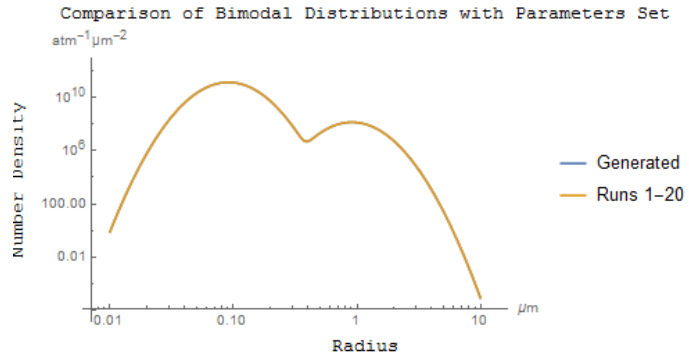

Fig. 4.1. Plots of generated and sample size distributions show consistency of solutions

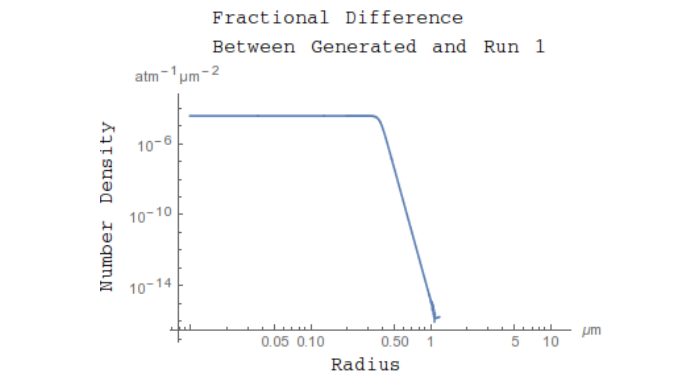

Fig 4.2. Fractional difference shows accuracy of around 0.0003%

defined as suggested in Yuan et al. (2011). In the SPSO program we fixed these parameter values as defined above, only allowing the N1 and N2 values to vary.

The results can be seen in Fig. 4.1. Through 20 runs the program was able to reproduce the N1 and N2 parameters well, giving N1 = 9.99967E9 $atm^{-1}\mu m^{-2}$ and N2 = 1.0E8 $atm^{-1}\mu m^{-2}$ each run. This shows that for this situation the SPSO program is robust, consistent, and accurate. In Fig. 4.1, the two distributions are on top of each other, showing that the parameters do not vary enough to see an altered distribution. There are only two distributions plotted as the other distributions have the same parameters and fitness value as runs 1-20.

4.2 SPSO Algorithm on Generated Data with Varying Parameters

When we tested the SPSO program on the same test data but let all the parameters vary, we got slightly less reliable data, as can be seen in Fig. 4.3 and Table 4.1.

In Fig. 4.3 there are four distributions with varying parameters, plotted with the distribution generated from the actual parameters. These four distributions were output by the SPSO program using the same generated AOD values as above. The various parameter values in Table 4.1 are similar, but

differences can be seen in Fig. 4.3. Without having the α and β parameters fixed, we do not see the same kind of consistency as we did with them fixed.

This inconsistency is troubling. The parameter values are similar and the fitness values are small, but the distributions are not the same. Although they are similar, we would hope for them to be exact like before. It can be seen that the distributions are close to the same from 0.5 µm to 3 µm, but vary at all other radii.

Table 4.1: Parameter Values for Generated AODs

	Fitness	N1 ($atm^{-1}\mu m^{-2}$)	N2 ($atm^{-1}\mu m^{-2}$)	$\alpha 1$ (µm)	$\alpha 2$ (µm)	$\beta 1$ (µm)	$\beta 2$ (µm)
Actual	0.0	1.0E10	1.0E8	0.1	1.0	0.307	0.307
Run 1	5.45768E-5	1.05711E11	8.04836E7	0.0579644	0.90252	0.16942	0.370639
Run 2	2.98691E-5	6.30367E9	7.94877E7	0.0981083	1.07063	0.24538	0.323436
Run 3	6.01875E-5	6.17708E11	9.82442E7	0.0309464	0.977901	0.230634	0.326713
Run 4	2.85774E-5	2.32441E11	4.5903E7	0.0349286	0.939994	0.360621	0.443974

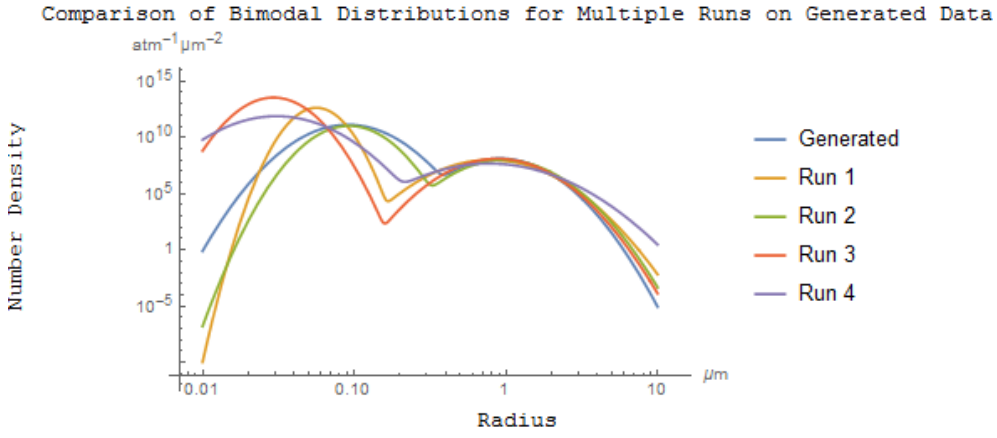

Fig. 4.3. Variety of distributions show the inconsistencies between runs of the SPSO program

4.3 Real measurements

All measurements of solar irradiance for this study were taken at the observatory on the Saint John's University campus in Collegeville, MN (45.5749135°N, 94.3959302°W). The solar tracking mechanism and photometer are installed on this observatory. Weather data was taken from the weather center on the Peter Engel Science Center at the Saint John's University Campus.

In this study we will look at data from August 3rd, 2015 for an example of a clear conditions, and August 31st, 2015 as an example of hazy conditions. This was one day where we realigned the photometer, and was one of the mornings we used to determine the cutoff value as discussed before. We know that for this entire day there were clear skies, and therefore this is a good time period to use as a test for the SPSO program on real data. A pressure of 29.94 in (1014 mb) was measured at 9:00am that morning, and at 3:00pm a pressure of 29.92 in (1013 mb) was measured.

We also looked into a hazy day in order to look at an example of AOD values and size distributions in hazy conditions. For this example, we looked at August 31st, 2015. There were scattered clouds throughout the day, but there were times where there were no clouds and there was haze. We looked through the irradiance data for this day, and decided on a count value of 12,000 counts. This cutoff allowed us to pick the times in the day that only had the haze present. Through this cutoff value we were able to get a two hour time period from 11:00am-1:00pm, so we chose a time of 12:15pm to include in this study. The pressure at this time was measured at 29.89 in (1012 mb).

The differences in the measured irradiance can be seen in Fig. 4.4 (August 3rd, 3:00pm) and Fig. 4.5 (August 31st, 12:15pm). Looking at Fig. 4.5 as compared to Fig. 4.4 and Fig. 3.4, the differences in the shape of the irradiance can be seen. The hazy conditions have a much lower irradiance in the smaller

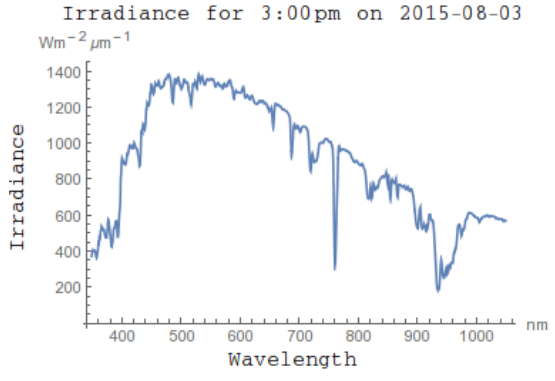

Fig. 4.4. Irradiance in clear conditions has larger irradiance values in smaller wavelengths (400-700nm)

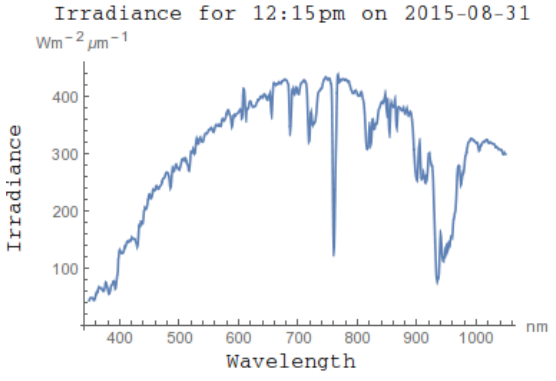

Fig 4.5. Irradiance in hazy conditions has larger irradiance in larger wavelengths (800-1000nm)

wavelengths (400-700 nm), with a higher irradiance in the larger wavelengths (800-1000 nm), and the overall irradiance is much lower than the clear conditions. This will result in much higher AOD values which we will discuss later on.

4.4 SPSO Algorithm on real data

We employed the SPSO algorithm in order to be able to retrieve aerosol size distributions for sets of AOD values that we obtained from our own measurements. In testing the real data that we have, we were unable to retrieve size distributions that had a good fitness value. We ran the SPSO program for many different types of data for this study, including a clear morning, clear afternoon, and hazy mid-day measurement. Each of these times provides a unique example of the strengths and weaknesses of this program.

Each time we ran the SPSO algorithm on the same set of data, we got similar fitness values, but the parameters for each run were different. This gave us many different distributions, and made it

Table 4.2: 8-3-2015 9:00am

	Fitness	N1 ($atm^{-1}\mu m^{-2}$)	N2 ($atm^{-1}\mu m^{-2}$)	$\alpha 1$ (μm)	$\alpha 2$ (μm)	$\beta 1$ (μm)	$\beta 2$ (μm)
Run 1	0.0170743	5.78888E14	4.73639E8	0.0468793	0.925941	0.168536	0.312546
Run 2	0.0172278	4.84439E13	8.93577E8	0.0908005	1.18766	0.114607	0.166731
Run 3	0.0170626	3.9693E13	2.31793E9	0.098212	0.97171	0.100001	0.112532
Run 4	0.0170423	2.07847E14	1.3874E8	0.0563699	0.939118	0.156688	0.484478

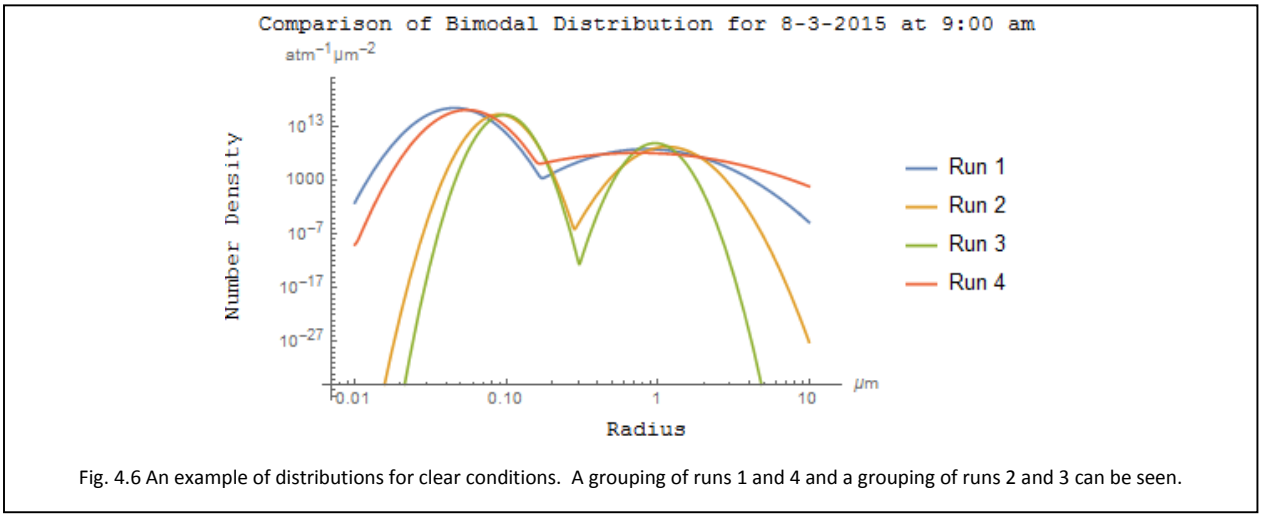

Fig. 4.6 An example of distributions for clear conditions. A grouping of runs 1 and 4 and a grouping of runs 2 and 3 can be seen.

difficult to decide which distribution was the best distribution to choose. This inconsistency shows that the program could not find the same global minimum each run, as it found a local minimum instead. Due to this, we get many different solutions for the size distribution of the given AOD values. Table 4.2 and Fig. 4.6 show the parameter values and size distributions for four runs of the SPSO program on the same set of measured AOD values from August 3rd, 2015 at 9:00am. This was an example of a clear morning's measurement. All four runs have similar fitness values, but we can see two groupings of distributions. Runs 1 and 4 are similar, and runs 2 and 3 are similar. It can be seen that runs 1 and 4 are more gradual, while runs 2 and 3 have sharper peaks around about 0.1 µm and again at 1 µm. This is an example of how different these runs can be. This is a result of the SPSO program attempting to match the different ranges of radii and the various parameters. The causes of these differences will be discussed in Section 5.

Table 4.3: 8-3-2015 3:00pm

Run	Fitness	N1 ($atm^{-1}\mu m^{-2}$)	N2 ($atm^{-1}\mu m^{-2}$)	$\alpha 1$ (µm)	$\alpha 2$ (µm)	$\beta 1$ (µm)	$\beta 2$ (µm)
1	0.0132362	2.01889E13	6.60184E8	0.0985884	0.975388	0.1	0.344916
2	0.131071	5.53359E13	1.59003E9	0.0811047	0.80604	0.104691	0.309504
3	0.0141046	7.0028E12	2.52939E8	0.121346	1.00679	0.100001	0.487518
4	0.0131847	1.62403E14	6.37111E8	0.0504367	1.07828	0.176906	0.319432

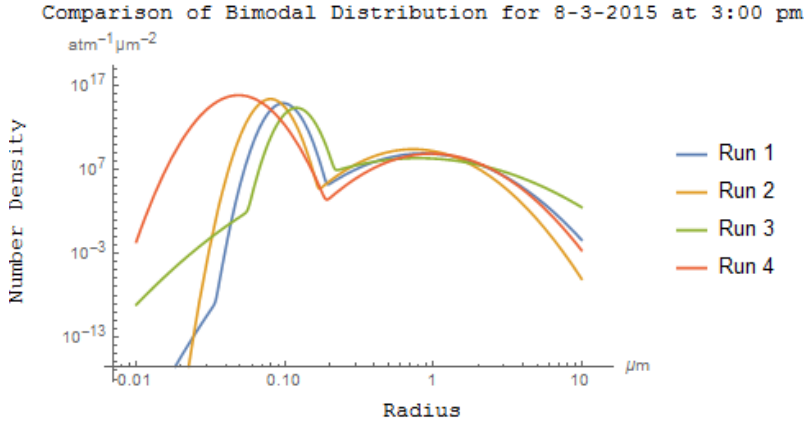

Fig. 4.7 An example of clear conditions. Similar size distributions can be seen in runs 1-3.

There are differences that can be seen in the same runs, but as discussed before, there can be noticeable differences in the AOD values and size distributions from morning to afternoon, which can be seen in Table 4.3 and Fig. 4.7. This is an example of a clear afternoon, and it can be seen these size distributions look different than the morning's distributions. Like before, there are clear groupings, as runs 1, 2, and 3 have similar peaks, with run 4 being a bit different. It is also interesting to note that run 3 has a slightly higher fitness value. Run three of this plot is interesting in that the combination of a smaller N2 factor (lower amplitude) with a larger $\beta 2$ (wider distribution) creates a bump on the distribution for the first peak.

Now we will look at an example of a hazy day's conditions, with the data represented in Table 4.4 and Fig. 4.8. It can be seen that this graph shows very similar distributions for runs 2, 3, and 4 being close together and having similar fitness values while run 1 is different and has a larger fitness value. Looking at the parameter values it can be seen that these values are much more similar than other

Table 4.4: 8-31-2015 12:15pm

	Fitness	N1 ($atm^{-1}\mu m^{-2}$)	N2 ($atm^{-1}\mu m^{-2}$)	$\alpha 1$ (μm)	$\alpha 2$ (μm)	$\beta 1$ (μm)	$\beta 2$ (μm)
Run 1	0.0658979	7.82039E17	3.5314E9	0.0155039	0.945089	0.194838	0.34925
Run 2	0.0344163	7.14553E14	2.77379E9	0.0493468	0.955903	0.195556	0.354593
Run 3	0.0333209	3.68436E14	3.46707E9	0.052197	0.997514	0.213778	0.295462
Run 4	0.0333569	3.0758E14	2.77532E9	0.0553747	1.05831	0.209123	0.304011

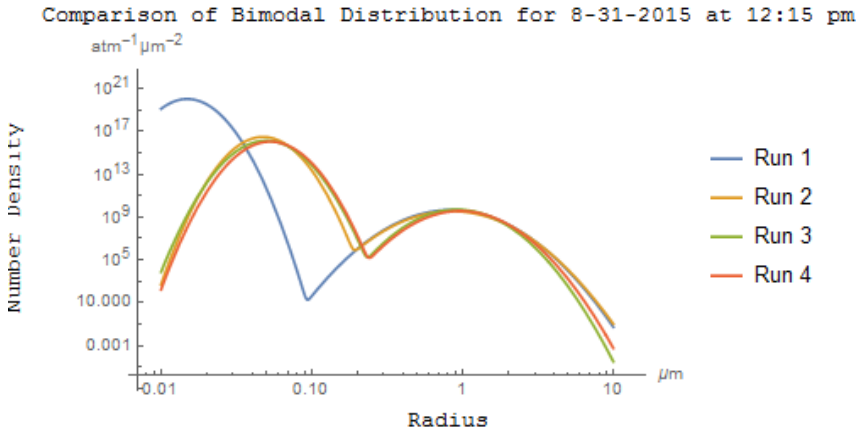

Fig. 4.8 For hazy conditions consistency can be seen in runs 2-4

parameter values that produce similar results, as in Table 4.2 or Table 4.3. The similarity of this example could be due to the increased value of the course mode aerosols due to the hazy conditions. This increase of course mode aerosols could provide information to the SPSO program that could lead to clearer results.

5. Results

From the Section 4 it can be seen that one run of the SPSO program does not provide consistent and reliable data. The retrieved distributions were different and the parameters varied enough to provide inconsistent results. Due to this, we can conclude that a single run of the SPSO program is not robust enough to provide an accurate size distribution estimation. Therefore we must find a way to determine the best method to use to get accurate size distributions with the SPSO program.

We suggest using an average of runs to determine the best size distribution. In order to retrieve accurate size distribution parameters for each set of AOD values, we suggest first running the same data through the SPSO program in order to get multiple size distributions. Next, the user must determine a best fitness value cutoff to provide the best size distributions, choosing to use only the lowest fitness values. There is usually a fitness value that can be easily identified by looking at the distributions as we did in Section 4. Next, the user must average these best fitness values' parameter values. These average fitness and parameter values will give the best estimation of aerosol size distribution. In order to assess how well this method will work, we will look at both the average size distribution parameters in addition to how well the average of the calculated AOD values differ from the measured AOD values.

First we will look at the average of the size distributions. Table 5.1 and Fig. 5.1 provide the average values of the parameters provided in the Section 4. The uncertainties are calculated by taking the standard deviation of the mean of the four runs of parameters in the Section 4. From this table it can be seen that the N1 and N2 parameters are the most uncertain, while the $\alpha 1$, $\alpha 2$, $\beta 1$, $\beta 2$ parameters do not vary as much. Also, the fitness value of the hazy conditions (8-31-2015) with all runs included is the least accurate, but with only the best three runs chosen, the fitness value uncertainty decreases. In looking at Table 5.1 and Fig. 5.1, it can be seen that between clear conditions (8-3-2015)

Table 5.1: Average Parameter Values

	Fitness	N1 ($atm^{-1}\mu m^{-2}$)	N2 ($atm^{-1}\mu m^{-2}$)	$\alpha 1$ (µm)	$\alpha 2$ (µm)	$\beta 1$ (µm)	$\beta 2$ (µm)
8-3-2015 9:00am							
Average	0.01710	2.2E+14	9.6E+08	0.07	1.01	0.13	0.27
Uncertainty	0.00004	1.3E+14	4.8E+08	0.01	0.06	0.02	0.08
8-3-2015 3:00pm							
Average	0.0134	6.1E+13	7.9E+08	0.09	0.97	0.12	0.37
Uncertainty	0.0002	3.5E+13	2.8E+08	0.01	0.06	0.02	0.04
8-31-2015 12:15pm All Runs							
Average	0.042	2.0E+17	3.14E+09	0.043	0.99	0.203	0.33
Uncertainty	0.008	2.0E+17	2.1E+08	0.009	0.03	0.005	0.02
8-31-2015 12:15pm Best Three Runs							
Average	0.0337	4.6E+14	3.01E+09	0.052	1.00	0.206	0.32
Uncertainty	0.0004	1.3E+14	2.3E+08	0.002	0.03	0.005	0.02

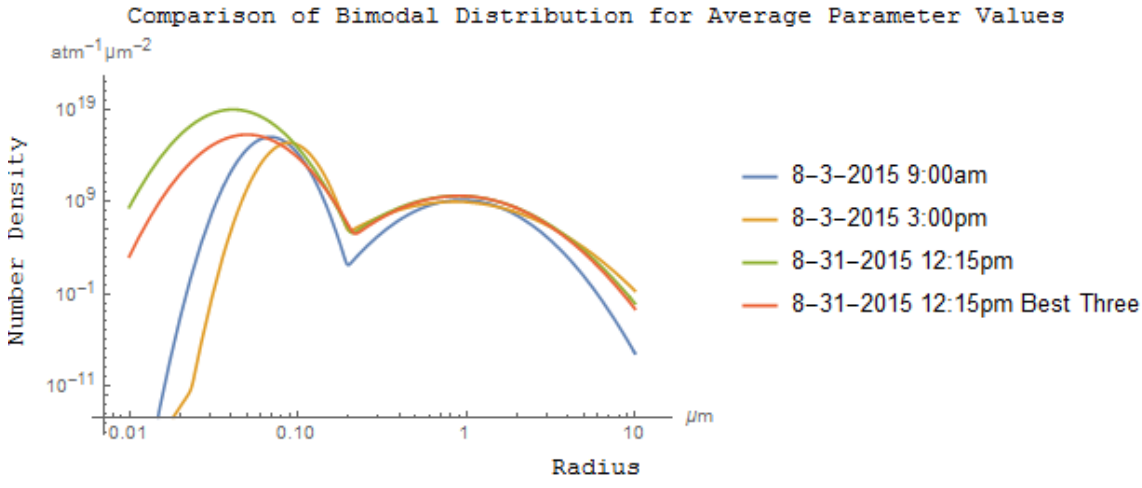

Fig. 5.1 Size Distributions for all examples are shown. Similarities between 8-3-2015 at 9:00am and 3:00pm can be seen. By only using the best three runs on 8-31-2015 a difference can be seen from using all runs

and hazy conditions (8-31-2015), the total number of Aitken mode aerosols are similar, but the Aitken mode aerosols tend to have a smaller mean radius and a larger distribution width. It can also be seen that for hazy conditions there are a higher number of accumulation mode aerosols, but the mean radius and width of the distribution is similar to that of clear conditions.

Along with looking at the various average parameters, we can look at the average AOD values. The SPSO program determines the best parameter values based off their fitness value, and this fitness value is based off the difference between the measured AOD values and the SPSO calculated AOD values. These AOD values are calculated directly from the size distribution parameters, making the difference between the measured AOD values and the calculated AOD values a good way to determine how accurate this method is.

Table 5.2: Generated Test Data

Wavelength (nm)	Generated	Average	SDOM
441	0.0147178	0.014769	9E-06
463	0.0147547	0.014782	3E-06
500	0.0148094	0.014806	4E-06
520	0.0148356	0.014820	6E-06
556	0.0148356	0.014820	6E-06
610	0.014936	0.01489	1E-05
675	0.0149906	0.014941	8E-06
750	0.01504	0.015005	4E-06
778	0.0150563	0.015031	2E-06
870	0.0150988	0.015111	7E-06
1020	0.0151553	0.01525	2E-05

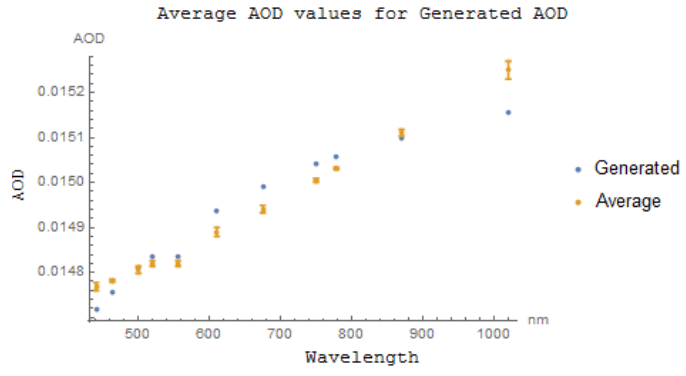

Fig. 5.2 Differences between generated and averaged AOD values and error bars can be seen. AOD values that increase with wavelength can be seen.

Table 5.2 shows the results of the average AOD values compared to the input generated or measured AOD values, depending on the situation. The uncertainty provided in the tables are the standard deviation of the mean as before. Fig. 5.2 shows the differences in the generated AOD values we discussed in Section 3.2 along with the average AOD values of the four runs discussed. Only a couple of the points fall within the error bars, which shows there is not a good fit even though Table 5.2 shows that the AOD values were replicated the best of all

the runs. The AOD values also increase with wavelength, unlike any of the measured AOD values. This is simply a result of the parameters of the generated AOD values creating a distribution that differs from our measured data. Table 5.3 and Fig. 5.3 show the AOD values for August 3rd, 2015 at 9:00am.

These generated AOD values match the measured values the least out of the three sets of measured data. Table 5.4 and Fig. 5.4 show the data for August 3rd, 2015 at 3:00pm, and Table 5.5 and Fig. 5.5 show the data for August 31st, 2015. It can be seen that the hazy conditions of August 31st, 2015 show the most accurate replication of the AOD values. The data set also contained the graphs of size distributions that looked the most consistent. Possibly with more runs we would get less consistent data, but it could also be due to the increased number of course mode aerosols in the distribution, which would play an important factor in how the AOD values are calculated in the SPSO program. The SPSO program calculates AOD values with Equation 2.4, multiplying the Mie extinction coefficient and number distribution and integrating over all radii ranges. Fig. 5.6 is a plot of what the Mie extinction coefficients look like for three wavelengths. It can be seen that for radii from 0-1.0 µm, especially between 0-0.2 µm, a small change in radii equates to a large difference in the Mie extinction factor.

Table 5.3: 8-3-2015 9:00am

Wavelength (nm)	Measured	Average	SDOM
441	0.463546	0.4535	0.0015
463	0.399695	0.40684	0.00087
500	0.330382	0.34241	0.00056
520	0.333001	0.31407	0.00081
556	0.29045	0.31407	0.00081
610	0.249844	0.21924	0.00093
675	0.175181	0.17663	0.00064
750	0.125844	0.14281	0.00028
778	0.114345	0.13308	0.00026
870	0.130308	0.10970	0.00055
1020	0.087901	0.08820	0.00077

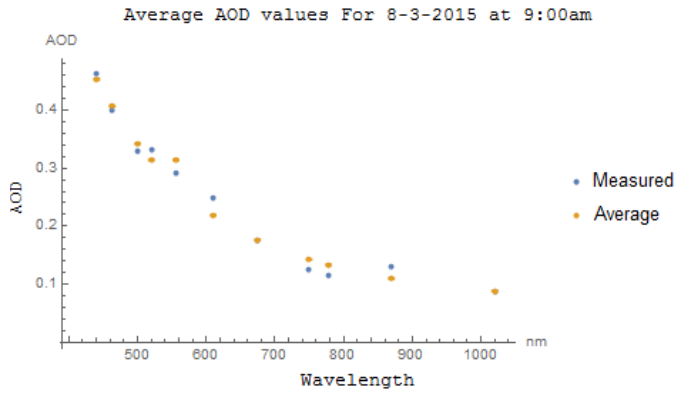

Fig 5.3 AOD values for clear conditions with AOD values that decrease with wavelength can be seen.

From 1.0-3.0 µm this difference becomes less as the Mie extinction factors

become more similar. This would provide better results for the hazy conditions as the increase in course mode aerosols would increase the contribution to the calculated AOD values from the larger radii ranges where the Mie extinction is more stable. This would lead to more stable size distributions, and more consistent results.

It can also be seen in Figures 4.3, 4.6, 4.7, 4.8, 5.1 that the distributions are more consistent in the radii ranges larger than 1.0 µm, due to this affect. This large variation at smaller radii leads the SPSO program to have difficulties in determining the best distribution as it balances the difference between a large number of small particles versus a small number of large particles, which is difficult if there are widely ranging solutions for differences in radii. When the actual atmospheric aerosol distribution consists of more particles at larger radii (accumulation and course mode), the Mie scattering coefficients are more uniform and the program can more easily estimate the correct number of particles for the distributions.

Table 5.4: 8-3-2015 3:00pm AODs

Wavelength (nm)	Measured	Average	SDOM
441	0.33556	0.3233	0.0022
463	0.298581	0.30006	0.00074
500	0.255171	0.26824	0.00068
520	0.265686	0.25399	0.0011
556	0.232832	0.25399	0.0011
610	0.222946	0.20554	0.0013
675	0.176307	0.18346	0.00083
750	0.162133	0.16591	0.00017
778	0.15542	0.16088	0.00007
870	0.171317	0.14883	0.00077
1020	0.126643	0.13781	0.0017

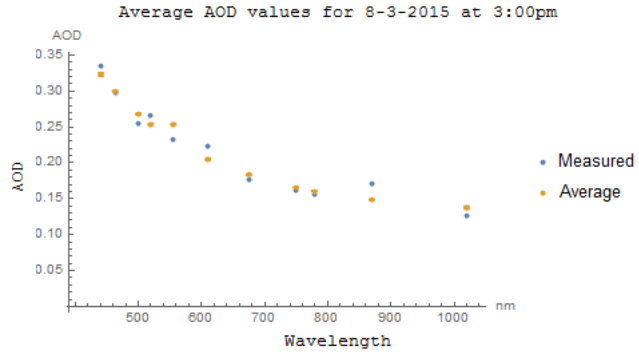

Fig 5.4 Aerosols for a clear afternoon with AOD decreasing with wavelength

This sensitivity to the number of aerosols at smaller radii also means that the SPSO program can have many different size distribution solutions for the same AOD values as the parameters can easily and quickly change. The program can hone in on a distribution for larger wavelengths,

Table 5.5: 8-31-2015 12:15pm AODs			
Wavelength (nm)	Measured	Average	SDOM
441	1.82518	1.850	0.026
463	1.71149	1.709	0.007
500	1.55132	1.518	0.011
520	1.49663	1.432	0.016
556	1.36164	1.432	0.016
610	1.146	1.135	0.019
675	1.02256	0.994	0.012
750	0.873267	0.877	0.002
778	0.826031	0.843	0.002
870	0.76442	0.756	0.013
1020	0.636971	0.670	0.027

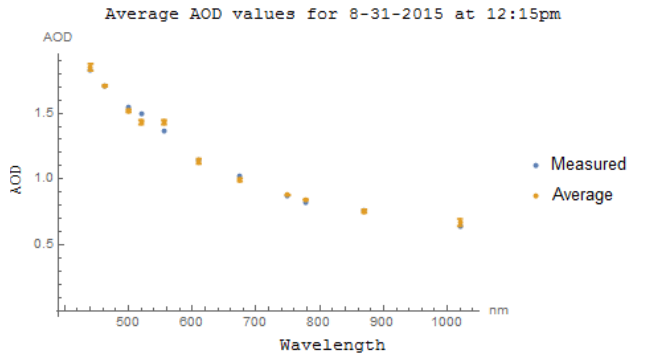

Fig 5.5 An AOD set that decreases with wavelength can be seen for hazy conditions.

corresponding to the N2, $\alpha 2$, and $\beta 2$ parameters, and the N1, $\alpha 1$, and $\beta 1$ parameters can vary greatly due to the large variations in the Mie coefficient for small changes in the radii. This can best be seen in Fig.4.3, as the distributions around 1.0 µm are consistent, and the distributions at smaller radii vary. Table 4.1 also shows more consistency in the N2, $\alpha 2$, and $\beta 2$ parameters than the N1, $\alpha 1$, and $\beta 1$ parameters. This shows that the Aitken mode aerosols are more difficult to determine, making a bimodal distribution of Aitken and accumulation mode aerosols difficult to determine.

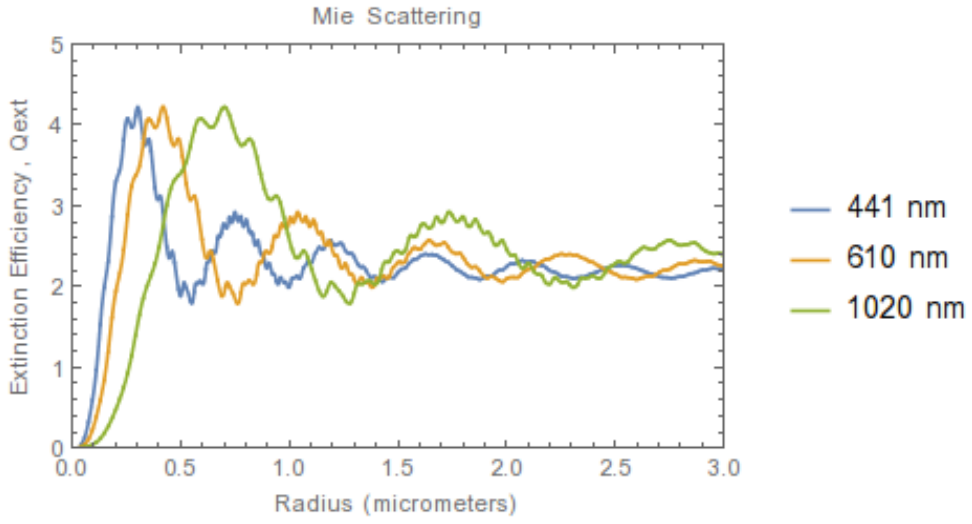

Fig 5.6 Mie extinction coefficients can be seen for three wavelengths. The large differences in smaller radii and consistency in larger radii are shown.

6. Conclusion

Using the SPSO program, a bimodal distribution focusing on Aitken and accumulation mode aerosols is difficult to determine due to the Mie scattering coefficient at small radii. Better results may be achieved using a bimodal distribution that focuses on accumulation and course mode aerosols as the program would focus on the more consistent Mie coefficients found at larger radii. This could be achieved by adjusting the program to look at only the radii range of 1.0-10.0 μm instead of the range 0.0-9.961 μm used in this study.

It should be noted that different aerosol optical depths can be fitted to different types of distributions, such as the Junge, Gamma, Log-Normal, and trimodal distributions presented in Table 2.2. This adds another complexity to size distribution calculations as each set of AOD values have a certain distribution it can be best fit to. King et al. (1978) proposes that with AOD values that decrease with increasing wavelengths are best fit to a Junge distribution, while AOD values that increase with increasing wavelength are best fit to a Log-Normal distribution. Yuan et al (2011) proposes that all distributions can be fitted with a trimodal distribution. However, due to the SPSO having a difficult time with the small radii, this distribution may lead to inconsistent results similar to what we have found with the bimodal distribution. All these distributions can be solved for using the SPSO program by simply adjusting the distribution the program calculates, leading to different sets of AOD values that are calculated.

Upon closer inspection the AOD values in Tables 5.2, 5.3, 5.4, and 5.5, it can be seen the calculated AOD values corresponding to 520 nm and 556 nm are the same value each time. This is a result we have seen in this study and believe this to be due to how similar the wavelengths are. In Section 2 we discussed how we chose these wavelengths, as we have run into this problem using more wavelengths. Due to how similar the wavelengths are, they could have been provided extremely similar

values of the irradiance, resulting in the same AOD value. To correct this, we suggest throwing out the 520 nm wavelength, resulting in a total of 10 wavelengths used. An area of further research could be to look into how well this program works with 10 wavelengths instead of the 11 used here. This still provides enough wavelengths to run a trimodal distribution while varying all parameters, as discussed by Yuan et al. (2011).

Another area of further research could be to consider using the oxygen absorption wavelength, around 773 nm, as the cutoff value used to determine the best measurements to use. At this wavelength there is a deep local minimum in the irradiance plot, as can be seen in Figures 3.3, 3.4, 4.4, and 4.5. The amount of oxygen in the atmosphere is relatively constant, so this wavelength's irradiance values shouldn't vary too much, and wouldn't be as sensitive to the small variations in the total solar irradiance. For example, if a jet contrail crossed the Sun at any given time, using a wavelength around 773 nm may count this as a good measurement while the 782 nm wavelength discussed in this study may throw it out. This may not be as appropriate if the user is looking for perfectly clear measurements, but it could be interesting to use as a cutoff to analyze aerosol optical depth fluctuations over the course of a day while still sorting out bad measurements.

In this study we have deduced that the SPSO program cannot provide consistent results for a bimodal distribution that focuses on Aitken and accumulation mode aerosols. Rather, a method of averaging many runs of the SPSO program is a more accurate estimation of the size distributions we are searching for. This is not to say the SPSO program is not a useful tool in determining aerosol size distributions, but simply that it is not robust enough to use one set of its results. We were successful in determining different distributions for clear and hazy conditions, as can be seen in Fig. 5.1. The SPSO program could produce results that were approriate for the input conditions. For the hazy condiditons we would expect a different distribution in regards to the Aitken mode aerosols, which is what we saw. We also saw consistency in results in clear conditions, seeing similar but different distributions from

morning to the afternoon. This is good as the SPSO program was able to show a noticable difference in these similar sets of AOD values, showing that it is capable of picking an appropriate distribution. Further research into a bimodal distribution looking at accumulation and course mode aerosols or a trimodal distribution looking at Aitken, accumulation, and course mode aerosols need to be done in order to determine the robustness of the SPSO program in these cases.

7. Acknowledgements

I would like to thank Dr Adam Whitten for everything he has taught me and all the guidance and assistance he has given me with this project. I would also like to thank the College of Saint Benedict and Saint John's University Summer Undergraduate Research Program and the Rooney Grant Program for funding this research through their summer research positions.

8. Bibliography

Allen, Bob. "Atmospheric Aerosols: What Are They, and Why Are They So Important?" NASA 2015. Web2015.

Bodhaine, B. A., et al. "On Rayleigh Optical Depth Calculations." *Journal of Atmospheric and Oceanic Technology*16.11 (1999): 1854-61. Print.

Bohren, Craig F., Huffman, Donald R. “Appendix A: Homogeneous Spheres.” *Absorption and Scattering of Light by Small Particles*. Wiley-VCH Verlag GmbH, 2007. 477-482. Web.

Bucholtz, A. "Rayleigh-Scattering Calculations for the Terrestrial Atmosphere." *Applied Optics* 34.15 (1995): 2765-73. Print.

Habte, Aron. "Solar Spectra: Standard Air Mass Zero." National Renewable Energy Laboratory (NREL). Web. 12/19/2015 2015.

Jorge, H. G., and J. A. Ogren. "Sensitivity of Retrieved Aerosol Properties to Assumptions in the Inversion of Spectral Optical Depths." *Journal of the Atmospheric Sciences* 53.24 (1996): 3669-83. Print.

King, Michael D., et al. "Aerosol Size Distributions Obtained by Inversion of Spectral Optical Depth Measurements." *Journal of the Atmospheric Sciences* (1978): 2153-67. Print.

London, Julius. "The Observed Distribution of Atmospheric Ozone and Its Variations." Eds. Whitten, Robert C. and Sheo S. Prasad. Ozone in the Free Atmosphere: Van Nostrand Reinhold, 1985. 11-80. Print.

Newhall, Chris, James H. Wendley, and Peter H. Stauffer. "The Cataclysmic 1991 Eruption of Mount Pinatubo, Philippines." U.S. Geological Survey 2005. Web2015.

Slusser, J., et al. "Langley Method of Calibrating UV Filter Radiometers." *Journal of Geophysical Research-Atmospheres* 105.D4 (2000): 4841-49. Print.

Whitten, A. T., C. E. Sawyer, and M. Baatar. "Comparison of Ground Based Spectrophotometer Measurements of Total Column Ozone with Satellite Data." 2009. Print.

Wilson, W. E., and H. H. Suh. "Fine Particles and Coarse Particles: Concentration Relationships Relevant to Epidemiologic Studies." *Journal of the Air & Waste Management Association* 47.12 (1997): 1238-49. Print.

Yuan, Y. A., et al. "Inverse Problem for Particle Size Distributions of Atmospheric Aerosols Using Stochastic Particle Swarm Optimization." *Journal of Quantitative Spectroscopy & Radiative Transfer* 111.14 (2010): 2106-14. Print.

Yuan, Y., et al. "Inverse Problem for Aerosol Particle Size Distribution Using SPSO Associated with Multi-Lognormal Distribution Model." *Atmospheric Environment* 45.28 (2011): 4892-97. Print.

Yuan, Y., et al. "Using a New Aerosol Relative Optical Thickness Concept to Identify Aerosol Particle Species." *Atmospheric Research* 150 (2014): 1-11. Print.