

AOPP Summer Student Project: Updating
the Weather Station and Aerosol Data
Processing
Report by Ryan Elliot

Abstract

This summer project was given as group of several short parts, these were to:

- Improve the current plots used by the EODG to present both recent and historical weather data.
- Read and log data from the sub-department UV meter and present this on the EODG website.
- Implement an instantaneous in-house version of NASA's AERONET algorithm to process data received from the Cimel sun photometer on the roof of the sub-department.

Chapter 1

EODG Weather Station Website Plots

1.1 Introduction

The EODG maintains a website with current weather statistics as well as 2 sets of plots: one set for the last few days and one set for the last few weeks or months. This can be found at <http://eodg.atm.ox.ac.uk/eodg/weather/weather.html>. The website also contains an archive of plots from previous days, weeks, months, and years, as well as an "all time" plot covering all the data available. Data is available in a data repository on the website which is updated daily.

The python scripts that create the plots are called `update_webpage.py` and `update_history.py` and can be found at the online repository:

https://gitlab.physics.ox.ac.uk/povey/aopp_weather_station

1.2 Objectives

The main objectives for this section were:

- Replace the wind direction line plot with a plot that presents the data in a more intuitive way.
- Review and improve recent weeks/months plots.
- Replace wind direction plots in historical weather archive.
- Fix the all-time plot to stop the overcrowding of data-points.

1.3 Implementation

In the course of implementing these objectives, several changes were made to the scripts:

- adjust the calendar date selector for the historical plots to include a drop-down menu for selecting the month and year to be displayed for selection.
- Replace the wind speed and wind direction graphs with one wide graph displaying both wind speed as a line and direction as both a scatter plot and an arrow with the most common wind direction.
- Add a drop-down menu to the historical day plots date-picker to make month selection quicker and easier.
- Adjust the resampling range for the all-time plot to monthly (mean/min/max/sum as appropriate).
- Edit the colour maps for the min max plots so that they are an appropriate colour for each variable i.e. not all rainbows.
- Ensure that axis labels on the current weather page do not overlap with the plotting area of adjacent plots.
- X-axis ticks for the recent weeks summary plot are set to Mondays.

1.4 Results

The old plots can be seen in 1.1 and 1.2.

The new plots (1.3 and 1.4) offer clearer data presentation, note the new, wide, twin-axis graph plotting both wind direction and speed against time as well as an arrow that qualitatively indicates wind direction and speed on the new current weather plot and similar arrows on the summary plot.

Figure 1.1: The old "current weather" plot.

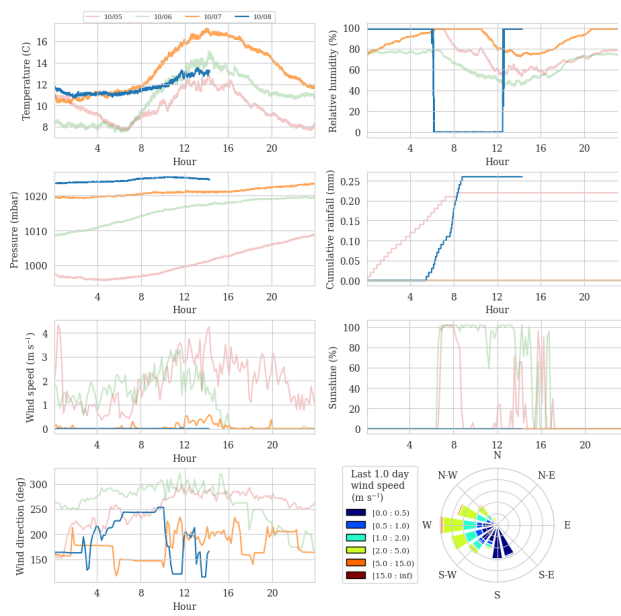


Figure 1.2: The old plot for recent weeks/months.

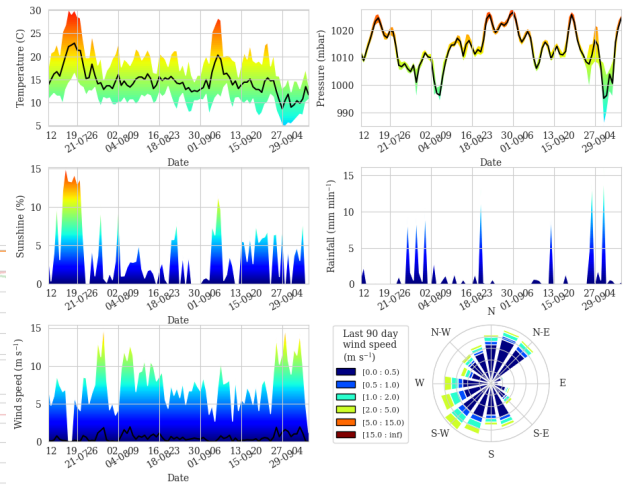


Figure 1.3: The new "current weather" plot.

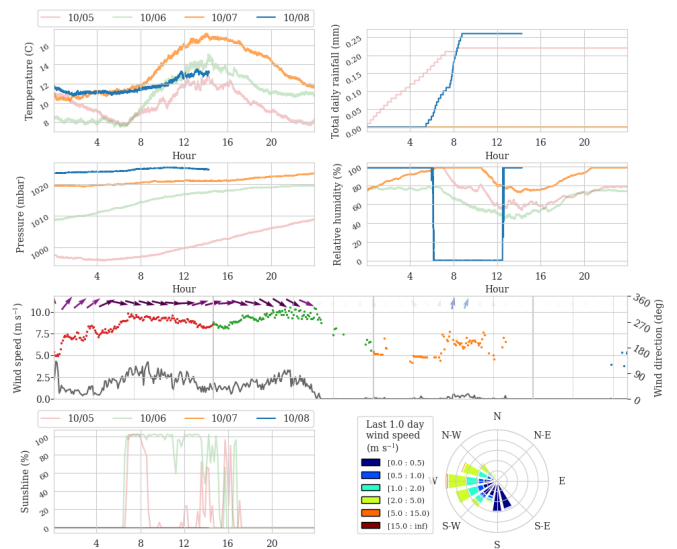
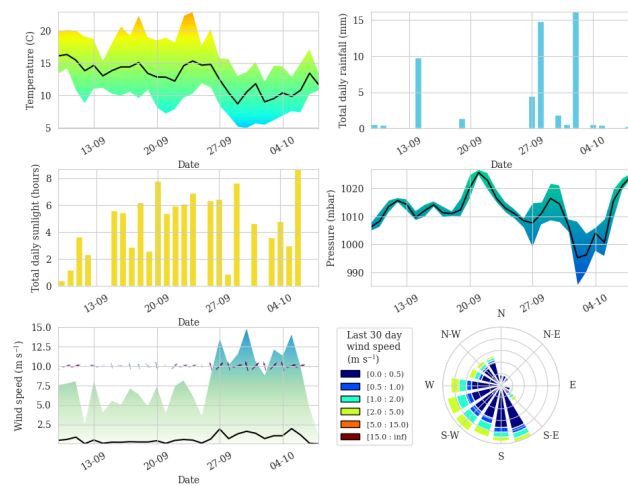


Figure 1.4: The new plot for recent weeks/months.



Chapter 2

Reading live data from UV meter

2.1 Introduction

On the roof of the AOPP sub-department there is a UV meter that is read via a serial port on the Linux machine atmlxdt055. The UV meter can be communicated with using Minicom over ssh. This gives the option to display the live output from the machine. The output from this is in a "human friendly" format designed to be read in the command shell by eye rather than as a simple line format e.g. each value separated by a comma. The output is constantly updated.

This posed difficulties with both reliably taking enough output from a port that is constantly sending data, and extracting the data from this. The data then needed to be recorded in files for each data in a similar way to how weather data is recorded for the data repository on the weather station website. In keeping with this, the code that reads and records the UV output was added to the python script `update_weather.py` and does so using classes and functions from the file `instromet.py`. The scripts are maintained in the repository: https://gitlab.physics.ox.ac.uk/povey/aopp_weather_station. How this was accomplished is laid out below.

2.2 Objectives

The main objectives for this section were:

- Create a python class for the serial port with functions that open the port, read the live output and extracts the data from the output.
- Implement a routine that takes the data and writes it to the current day's UV data file.

- Wrap said routine in an "if" loop that makes sure that no processes are currently using the UV meter.

2.3 Implementation

In the course of implementing these objectives, several changes were made to the scripts:

- In `instromet.py`, a class was created called `UVmeter` as an extension to the `Serial` class from the `serial` library. Within this class a function called `takedata` was defined with input arguments `self` (the `UVmeter` object itself), `last_line` (intended to be the last observation from the current UV data file as a pandas dataframe object with 1 row and a header), and `msg_len` (length, in characters, of live display from the UV meter, by default this is 1051). This function reads 2 messages worth of characters, takes the first instance of a full message and partitions this appropriately to take the characters containing data and outputting them as a pandas dataframe with a single row.
- Include a routine within `update_weather.py` that: 1) Checks if a process id (PID) file exists. If not, or if a process with the same PID as is in the file is not currently running, make a PID file and continue (else skip reading the meter); 2) Check if there exists a file for the current day, if so then the last line from that is retrieved, if not then a file is made and the "last line" is taken to have values of 0; 3) Read the data and calculate "Since Last" field using retrieved previous line; 4) Delete the PID file

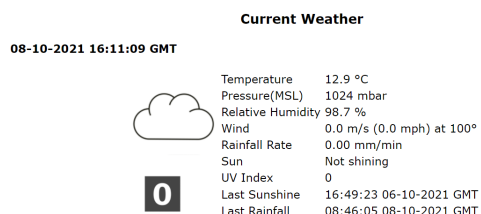
and return the data. If the function fails to properly read the serial port then then the a row of data is recorded with the date and time and data values of -999.

- Modify the script `update_webpage.py` to adjust the EODG weather website to display the current UV index, including an image of a coloured square with a with the UV index in white.

2.4 Results

The newly created files containing UV data can be found at <http://weatherstationdata.physics.ox.ac.uk/2021/>. The current UV data is displayed on the weather website <http://eodg.atm.ox.ac.uk/eodg/weather/weather.html> and can be seen in 2.1 and shows a new line added that displays the present UV index along with an image of text on a coloured square displaying the same information.

Figure 2.1: EODG weather website updated to include a line with current UV index and a coloured square displaying the same information.



Chapter 3

First Implementation of Real-Time Version of the AERONET Aerosol Optical Depth Algorithm

3.1 Introduction

When sunlight passes through the atmosphere, some photons are scattered and absorbed by whatever is in the atmosphere, meaning that the intensity of light of all wavelengths is attenuated exponentially by a frequency dependant quantity which is "Total Optical Depth" (TOD) multiplied by "relative optical air mass" (which is the ratio of the actual quantity of air light from the sun travels through to that when the sun is directly overhead i.e. the minimum). The equation relating the voltage gain on the photometer in each wavelength band ($V(\lambda)$), the corresponding calibration coefficient ($V_0(\lambda)$), the ratio of mean to actual Earth-Sun distance (d), total optical depth (τ_{tot}) and relative optical air mass (m) is [7] (More information can also be found in [9]):

$$V(\lambda) = V_0(\lambda)d^2 \exp(-\tau_{\text{tot}}(\lambda)m) \quad (3.1)$$

The intensity of light of a given wavelength at the Earth's surface can be measured using a sun photometer with narrow bandpass filter to select wavelengths of light. The voltage gain of the sun photometer is proportional to the incident radiation and transmittance is calculated by dividing the gain by a calibration coefficient equal to what the gain would be without/outside the atmosphere.

The atmosphere is made up of various constituent parts which have their own frequency dependant contributions (excluding clouds), includ-

ing from atmospheric gases, Rayleigh scattering, and, importantly for this project, aerosols [7]:

$$\tau_{\text{tot}}(\lambda) = \tau_{\text{aerosol}}(\lambda) + \tau_{\text{water}}(\lambda) + \tau_{\text{Rayleigh}}(\lambda) + \tau_{\text{O}_3}(\lambda) + \tau_{\text{NO}_2}(\lambda) + \tau_{\text{CO}_2}(\lambda) + \tau_{\text{CH}_4}(\lambda) \quad (3.2)$$

The contributions for ozone, NO_2 are calculated as [7][15][17]:

$$\tau_{\text{O}_3}(\lambda) = \sigma_{\text{O}_3}(\lambda) \cdot H_{\text{O}_3} \cdot \frac{m_{\text{O}_3}}{m} \quad (3.3)$$

$$\tau_{\text{NO}_2}(\lambda) = \sigma_{\text{NO}_2}(\lambda) \cdot H_{\text{NO}_2} \quad (3.4)$$

where $\sigma(\lambda)$ is the collisional cross-section as a function of wavelength, H is the total column height and m_{O_3} is the relative optical air mass relating to ozone.

The contribution for Rayleigh scattering may be calculated using the methods found in Bodhaine et al [3]. Table 3.1 illustrates which non-aerosol components of total optical depth amount to appreciable contributions (while the rest are considered to be negligible). A similar set-up is used for our algorithm although the Oxford sun photometer doesn't have a 1640nm filter so we need not consider the carbon dioxide or methane optical depths.

3.2 Objectives

The main objectives for this section were:

Wave-length (nm)	Water	Rayleigh	NO ₂	O ₃	CO ₂	CH ₄
340		✓	✓	✓		
380		✓	✓			
440		✓	✓			
500		✓	✓	✓		
675		✓		✓		
870		✓				
1020	✓	✓				
1640	✓	✓			✓	✓

Table 3.1: Table of which non-aerosol components count to appreciable contributions to total optical depth in the AERONET algorithm [7]. Note that 935nm is used for calculating water column and not included in this table.

- Read the files containing solar intensity measurements.
- Calculate AOD from the intensity data.
- Clean/vet the data to remove faulty/anomalous data and cloud-contaminated data.

3.3 Implementation

3.3.1 Equipment, Data and File Format

The sun photometer on the roof of the atmospheric physics sub-department is a CIMEL Multiband Sun Photometer, a User Manual can be found online [6]. Information on the specific instrument calibration settings and filter properties can be found on the CAELIS website maintained by the Universidad de Valladolid (UVa) [13]. Notably the filters/bands used are for wavelengths of 340nm, 380nm, 440nm, 500nm, 675nm, 870nm, 935nm and 1020nm. The photometer outputs files in a ".k7" file format and in comma separated format with file extension determined by the routine that collected the data. Information on these files can be found at [5]. The main file type of interest is the ".NSU" file type. This measures the solar intensity in each wavelength band three times, with each of these being referred to as a triplet. The files do not have headers so the order of measurements is not immediately clear. The order of each line is as follows: UTC date, UTC time, $V_{1,1020nm}$, $V_{1,870nm}$, $V_{1,675nm}$, $V_{1,440nm}$, $V_{1,500nm}$, $V_{1,935nm}$, $V_{1,380nm}$, $V_{1,340nm}$, $V_{2,1020nm}$, $V_{2,870nm}$, $V_{2,675nm}$, $V_{2,440nm}$, $V_{2,500nm}$, $V_{2,935nm}$, $V_{2,380nm}$, $V_{2,340nm}$, $V_{3,1020nm}$, $V_{3,870nm}$, $V_{3,675nm}$, $V_{3,440nm}$,

$V_{3,500nm}$, $V_{3,935nm}$, $V_{3,380nm}$, $V_{3,340nm}$, temperature in centigrade; where the subscripts on the photometer voltage gain V indicate the cycle and wavelength band. Many thanks to Ilya Slutsker of NASA's Goddard Center for providing this information.

3.3.2 Loading Data and Calculating AOD

The data processing and filtering is all done in a Jupyter notebook (named `read_files.ipynb`) and uses the libraries `matplotlib`, `numpy`, `pandas`, `Skyfield` [11] and `datetime`. Various constants are then defined as dicts where wavelengths form the index. Several functions are also defined. Many of these functions are used to calculate the Rayleigh optical depth according to the method outlined in Bodhaine et al [3].

Without going into any excess detail, the data is load and AOD is calculated as follows:

- The .NSU file is loaded using `pandas.read_csv` with each column being given an appropriate, systematic title and the date and time columns combined to make a "DatetimeIndex". Columns are then made containing the year, month, day, hour, minute and second of the time entries.
- Tools from the Skyfield library are then used to calculate the apparent solar zenith angle at the time of measurement, as well as the ratio between the actual Earth-Sun distance to the mean (i.e. distance divided by 1 AU).
- Total ozone and nitrogen dioxide column heights are calculated by using linear interpolation against mean data for the mean across recent years [2][4](2017-20 for nitrogen dioxide, 2016-2018 for ozone).
- Optical air mass and ozone optical air mass is calculated [12] [10].
- For each individual intensity measurement: correct for temperature variation using the method in the 2020 UVa CAELIS paper in using temperature [8] using the filter specific temperature coefficients on the CAELIS website [13], solves equation 3.1 using the values

of m and d as calculated earlier. AOD is then calculated by subtracting the estimated values of the other contributions, with the ozone and nitrogen dioxide absorption cross section being calculated from a mean over the filter range of data available on HITRAN [14] (except water vapour). A mean and standard deviation is calculated for the temperature adjusted voltage gains and a mean, standard deviation and range (max - min) is calculated for AOD at each wavelength too.

- Ångström exponents are calculated for certain pairs of wavelengths using ratios of wavelengths and corresponding mean AODs [16].
- The total column of water vapour is calculated using the method outline in Giles et al [7]. The AODs and Ångström exponents where this would make a difference to the value are then recalculated, however sadly the CAELIS website does not have the relevant coefficients for any of the other filters and so must be set to zero, removing the effect of water on calculated AOD for those channels.

3.3.3 Screening Data

Data then has to be screened for instrument failures and cloud contamination. Cloud contaminated data must be removed as cloud optical depth varies considerably over both space and time (although it is exactly this property which helps us remove them). The screening in the AERONET algorithm removes (or allows) measurements through at certain quality levels [7]; this algorithm, due to its intended use as a real-time indicative tool, adds flags to each line of data for every test it fails and outputs two files, one with all the data, including flags, and one file including only un-flagged data (i.e. data that passes every test). The tests implemented in this version are laid out below and have been selected from Giles et al [7] because they are relevant to the real-time nature of the task:

- Flag/remove data where the optical air mass is greater than 7.
- Flag/remove data where the voltage standard deviation is greater than 0.16 times the voltage mean.

- Flag/remove data for which any AOD value is NaN or infinite.
- Flag/remove data for which $V(\lambda)/V_0(\lambda) < 1/1500$
- Flag/remove triplets where AOD for 675nm, 870nm or 1020nm have a range greater than both 0.01 and 0.015 times the mean at that wavelength.*
- Keep lines of data for where the Ångström exponent for 440nm to 870nm is between -1 and 3, flag/remove the others.
- Temporal variation: For the 500nm wavelength, if the difference in AOD between two consecutive un-flagged measurements is greater than 0.01 times the time gap between the measurements in minutes then flag/remove the line of data containing the higher AOD. Repeat this until no more data is flagged.*
- If a line of data has no other data within one hour either side of it then remove/flag it.*

A high AOD retention protocol is in place to make sure that legitimately high AOD events such as smog or ash are not dismissed as clouds. For readings where the AODs for 870nm and 1020nm are greater than 0.5 and 0.0 (sic) respectively and the Ångström exponent for 675nm to 1020nm is between 1.2 and 3.0 the data is protected from the cloud screening protocols (indicated above with a *).

3.4 Results

3.4.1 Accuracy of Calculated AOD

This comparison is done using data collected on 08/09/2021 (the 251st day of the year) as this has a large number of data that both passes through this algorithm un-flagged and appears in AERONET level 1.5 data.

The data from this algorithm shows very strong linear correlation with its AERONET counterparts and the gradient of the regression line through all the points is within an acceptable distance from 1

for this use. The data for shorter wavelengths generally exhibits more correlation and is more accurate compared to the AERONET data than longer wavelengths.

Data Field	Gradient	y-Intercept	Pearson Coeff
AOD 1020nm	1.156	-0.021	0.917
AOD 870nm	1.112	-0.025	0.933
AOD 675nm	1.077	-0.010	0.972
AOD 500nm	1.040	-0.010	0.993
AOD 440nm	1.033	-0.020	0.994
AOD 380nm	1.031	-0.028	0.991
AOD 340nm	1.063	-0.044	0.987
Precipitable Water	1.033	-0.119	0.998

Table 3.2: The linear regression coefficients of the data produced by the new algorithm against the AERONET data and the Pearson correlation coefficient between the two sets.

Table 3.2 confirms this and shows that the intercept value is always negative. This shows that the new algorithm likely overestimates some of the other optical depth components compared to the AERONET algorithm.

3.4.2 Comparison of Data Screening with AERONET Quality Levels 1.0 and 1.5

AERONET Level	Flagged	Un-Flagged	Total
Level 1.5	35	209	244
Level 1.0 Only	174	8	182
No Level (0)	717	0	717
Total	926	217	1143

Table 3.3: A summary of data quality levels for data taken from 00:00:00 1 Sept 2021 to 16:48:09 21 Sept 2021. A very high percentage (over 96%) for data that passes the algorithm un-flagged is considered level 1.5 by AERONET, with no data below level 1.0 passing through.

Table 3.3 was created by analysing data passed through the AERONET algorithm (data available

at [1]) and data passed through the new "In-House" algorithm. The data was taken from 00:00:00 1 Sept 2021 to 16:48:09 21 Sept 2021.

The first thing to note is that all data rejected before level 1.0 of the AERONET algorithm was flagged by our new algorithm. Over 96% of the data that passes through un-flagged is of level 1.5 which indicates a mostly successful removal of cloud filtered data.

35 out of 244 (c. 14%) level 1.5 lines of data are flagged by the new algorithm. This shows that the new algorithm is possibly harsher than the AERONET algorithm, this could be in part due the fact that some checks in the AERONET algorithm may remove data from only one wavelength (e.g. the voltage threshold filter) while the rest of the row remains and can still then be classed as level 1.0 or level 1.5 data, meanwhile our algorithm adds a flag to the row of data meaning that the whole row may be considered removed despite failing the exact same test in the exact same way. This harshness could be alleviated by considering the number and nature of flags on each row of data.

3.5 Next Steps

As has been noted in the previous section, this "In-House" version of the AERONET algorithm is reasonably good at calculating AOD and filtering contaminated data. There are two instantly apparent issues that this algorithm suffers.

Firstly, the comparisons between calculated AOD values show that there is a non-zero y-intercept value in the regression statistics. This indicates that some of the fixed optical depths have been miscalculated. This could be either the nitrogen dioxide, ozone or Rayleigh scattering contributions. These could be improved by considering a more sophisticated interpolation technique, or considering more years in the monthly mean. One of the cross section measurements for ozone was unavailable on HITRAN and would be small so was omitted, consideration of this measurement improve the accuracy.

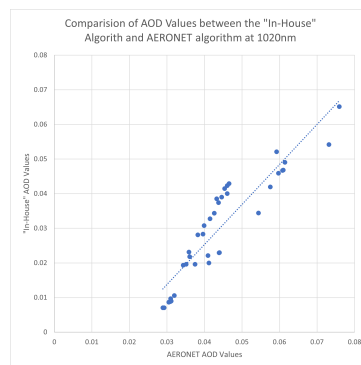
Secondly, some of the comparisons the points split into two streams. This could possibly due to a time dependant splitting caused by the calculation of the solar zenith angle. At present this is the apparent solar zenith angle i.e. the angle at which the

photometer will observe the sun to be at. It may be more appropriate to use the actual solar zenith angle as this would be the more physical angle.

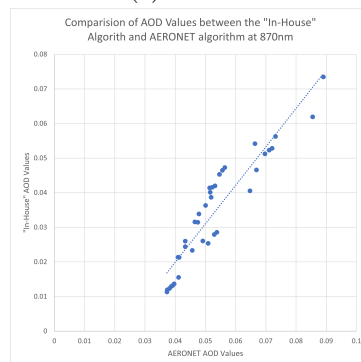
3.6 Summary

The new real-time AOD calculating and filtering algorithm based on the version used by NASA is effective at calculating a "quick number" for the aerosol optical depth over several wavelengths, with correlation coefficients between calculated AOD values and AERONET data being greater than 0.91 at all wavelengths, and is not overly lenient when it comes to filtering data for clouds as no level-0 AERONET data passed through this algorithm. There is room for improvement in calculating non-aerosol optical depth contributions and the optical air mass of optical depth components.

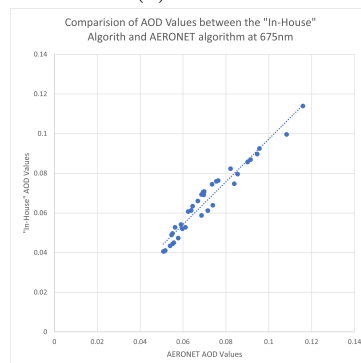
Figure 3.1: The plots in 3.1a to 3.1g compare the AOD for each wavelength as calculated by this algorithm and compares it with that which was calculated by the AERONET Algorithm while 3.1h compares the precipitable water values.



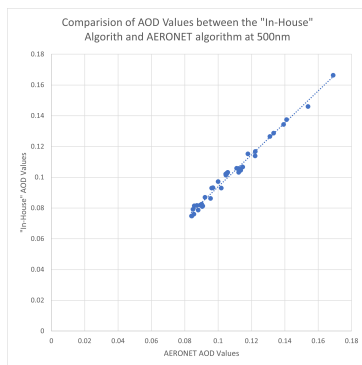
(a) 1020nm



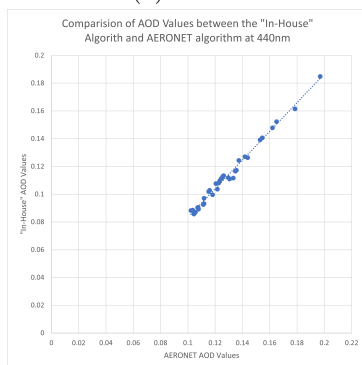
(b) 870nm



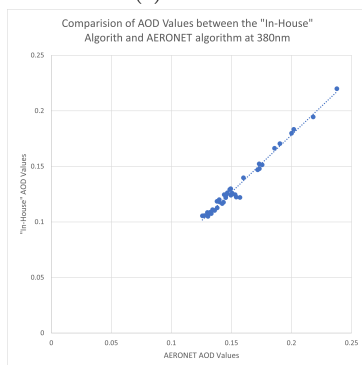
(c) 675nm



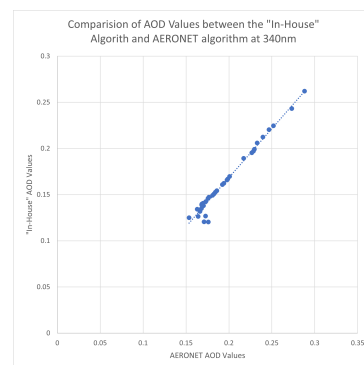
(d) 500nm



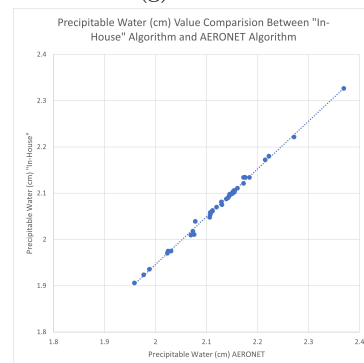
(e) 440nm



(f) 380nm



(g) 340nm



(h) Precipitable Water

Bibliography

- [1] *AERONET Aerosol Optical Depth Data Display Interface, Site: Oxford*. URL: https://aeronet.gsfc.nasa.gov/cgi-bin/data_display_aod_v3?site=Oxford&nachal=0&year=2021&month=9&aero_water=0&level=1&if_day=0&if_err=0&place_code=10&year_or_month=0.
- [2] Nickolay A. Krotkov et al and the OMI core team. *OMI/Aura NO2 Cloud-Screened Total and Tropospheric Column L3 Global Gridded 0.25 degree x 0.25 degree V3*. 2019. DOI: <https://doi.org/10.5067/Aura/OMI/DATA3007>.
- [3] Barry A. Bodhaine et al. “On Rayleigh Optical Depth Calculations”. In: *Journal of Atmospheric and Oceanic Technology* 16.11 (1999), pp. 1854–1861. DOI: 10.1175/1520-0426(1999)016<1854:ORODC>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/atot/16/11/1520-0426_1999_016_1854_orodc_2_0_co_2.xml.
- [4] European Union Copernicus Climate Change Service. *Ozone monthly gridded data from 1970 to present (Deprecated 2020-12-21)*. 2020. URL: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-ozone?tab=overview>.
- [5] Cimel Electronique. *K8 Specification Version 1.05*. URL: <https://www.cimel.fr/wp-content/uploads/2021/06/Photometer-K8-specifications-V1.05.pdf>.
- [6] Cimel Electronique. *Multiband photometer CE318, User’s Manual V.1.0*. URL: http://www-loa.univ-lille1.fr/documents/PHOTONS/Support_photons/DOCUMENTATIONS/CE318_N_Photometer_UserManual_V1-0.pdf.
- [7] D. M. Giles et al. “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”. In: *Atmospheric Measurement Techniques* 12.1 (2019), pp. 169–209. DOI: 10.5194/amt-12-169-2019. URL: <https://amt.copernicus.org/articles/12/169/2019/>.
- [8] R. González et al. “Daytime and nighttime aerosol optical depth implementation in CÆLIS”. In: *Geoscientific Instrumentation, Methods and Data Systems* 9.2 (2020), pp. 417–433. DOI: 10.5194/gi-9-417-2020. URL: <https://gi.copernicus.org/articles/9/417/2020/>.
- [9] B.N. Holben et al. “AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization”. In: *Remote Sensing of Environment* 66.1 (1998), pp. 1–16. ISSN: 0034-4257. DOI: [https://doi.org/10.1016/S0034-4257\(98\)00031-5](https://doi.org/10.1016/S0034-4257(98)00031-5). URL: <https://www.sciencedirect.com/science/article/pii/S0034425798000315>.
- [10] Fritz Kasten. “A new table and approximation formula for the relative optical air mass”. In: (1965), pp. 206–223. DOI: 10.1007/BF02248840. URL: <https://doi.org/10.1007/BF02248840>.
- [11] Brandon Rhodes. *Skyfield: Generate high precision research-grade positions for stars, planets, moons and Earth satellites*. URL: <http://rhodesmill.org/skyfield/>.
- [12] Johannes Staehelin et al. “Total ozone observation by sun photometry at Arosa, Switzerland”. In: *Optical Engineering* 34.7 (1995), pp. 1977–1986. DOI: 10.1117/12.204689.

- URL: <https://doi.org/10.1117/12.204689>.
- [13] Universidad de Valladolid. *OMI/Aura NO₂ Cloud-Screened Total and Tropospheric Column L3 Global Gridded 0.25 degree x 0.25 degree V3*. URL: <http://www.caelis.uva.es/rima/site/show/Oxford>.
 - [14] Various contributors. *HITRANonline*. URL: <https://hitran.org/>.
 - [15] Wikipedia contributors. *Absorption cross section* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 5-October-2021]. 2021. URL: https://en.wikipedia.org/w/index.php?title=Absorption_cross_section&oldid=1041224246.
 - [16] Wikipedia contributors. *Angstrom exponent* — *Wikipedia, The Free Encyclopedia*. 2020. URL: https://en.wikipedia.org/w/index.php?title=Angstrom_exponent&oldid=961396610.
 - [17] Wikipedia contributors. *Optical depth* — *Wikipedia, The Free Encyclopedia*. 2021. URL: https://en.wikipedia.org/w/index.php?title=Optical_depth&oldid=1045769522.