# Description of a synthetically generated training set for remote sensing of Above Cloud Aerosols (ACA) by the NASA Research Scanning Polarimeter (RSP) for the ORACLES field campaign.

Kirk Knobelspiesse (NASA GSFC)
Michal Segal-Rozenhaimer (NASA ARC)
Daniel J. Miller (NASA GSFC)

August 8, 2019

## Introduction

This document describes a synthetically generated training set for the creation of a neural network (NN) to simultaneously retrieve aerosol and cloud optical properties from observations by the NASA Research Scanning Polarimeter (RSP, Cairns et al., 1999, 2003). The RSP is an airborne multi-angle polarimeter, that was deployed in the ObseRvations of Aerosols above CLouds and their intEractionS (ORACLES) field campaign in the SE Atlantic Ocean in the austral spring of 2016, 2017 and 2018 (Zuidema et al, 2016).

The training set was simulated by a radiative transfer software using the doubling and adding method (Hansen and Travis, 1974). This method is ideal for optically thick simulations, and can produce all potential sun and viewing geometries with little added computational expense.

This work builds upon NN's developed for the remote sensing (from RSP) of cloud optical properties alone, for which alternative algorithm validation is available. The details of these previous NN's are described in Segal-Rozenhaimer et al, 2018 and Miller, et al, in prep.

## Training set parameters

Each individual NN training set simulation is generated with different combinations of twelve scene relevant parameters. These parameters are randomly generated within a defined numerical distribution. The intent is to reduce dependence upon arbitrarily defined 'nodes' in a regularly gridded training set, and to make the training set generation unlimited. In other words, additional training set simulations presumably improve the results, but there is no minimum number of simulations needed to span the range of possible values. The distributions of the randomly generated parameter values were chosen based on the best available information about the nature of those parameters during ORACLES, including observations by other instruments that participate during the field campaign. The selection of these distributions is intended to provide the greatest amount of NN training set cases at the most likely cloud and aerosol parameter values, but not influence the distribution of results from the trained NN.

*Table 1* contains all twelve parameters that were used to generate the NN training set. These parameters, contained as a vector in the NetCDF training set files, should be used in training.

Histograms of these values are shown in *Figure 1*. Some comments about specific parameters follows.

*Table 1 NN training set simulation parameters.*

| NetCDF file tag | Parameter | Unit |
|---|---|---|
| aer_thick | Aerosol layer (physical) thickness | meters |
| aod | Aerosol optical depth (total), 555nm | - |
| cloud_top | Cloud top height | meters |
| cod | Cloud optical depth, 555nm | - |
| fmf | Aerosol fine size mode AOD fraction, 555nm | - |
| gap | Cloud top - aerosol layer bottom gap | meters |
| ImNidx_f | Aerosol fine size mode imaginary refractive index | - |
| Reff_aer | Aerosol fine size mode effective radius | microns |
| Reff_cld | Cloud droplet size distribution effective radius | microns |
| ReNidx_f | Aerosol fine size mode real refractive index | - |
| Veff_aer | Aerosol fine size mode effective variance | - |
| Veff_cld | Cloud droplet size distribution effective variance | - |

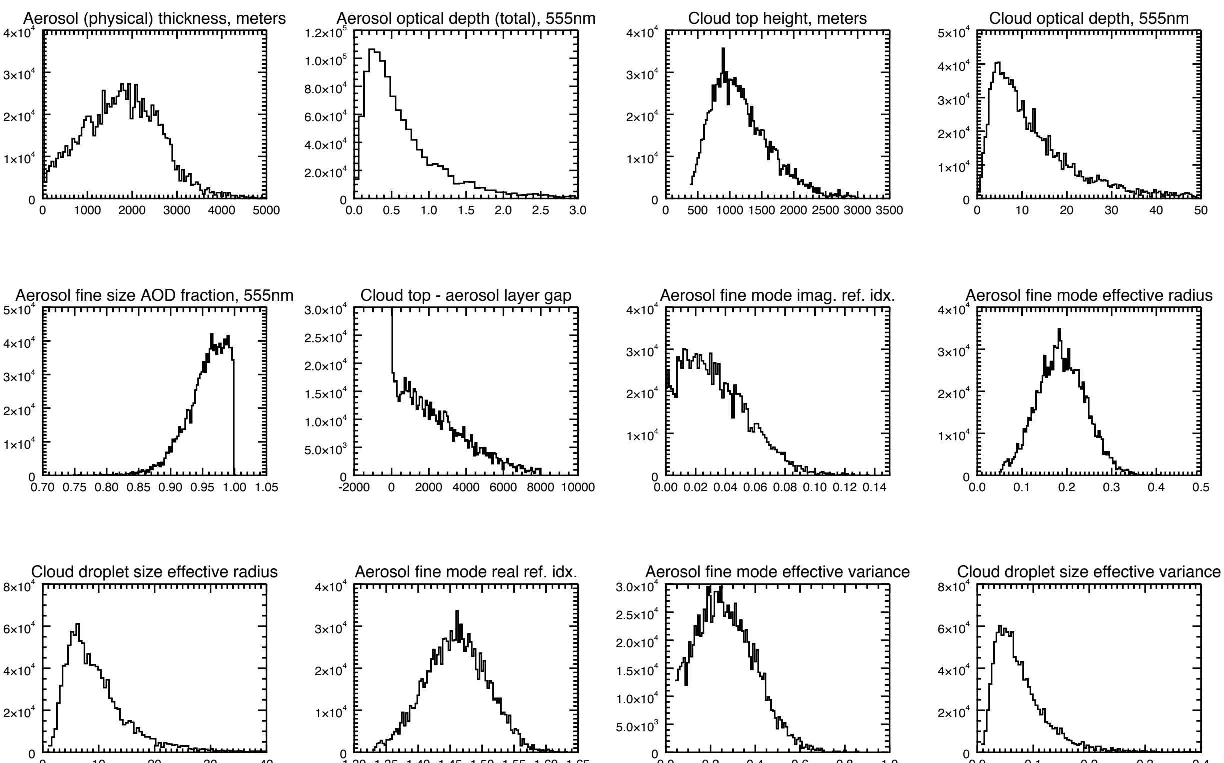


*Figure 1 NN training set histograms of randomly generated parameters.*

**aer_thick, cloud_top, gap:** Cloud top height, the physical thickness of the aerosol layer, and the gap between cloud top and aerosol layer bottom are all randomly generated parameters. The gap between the two layers was intentionally chosen to be zero in order to simulate cases with touching clouds and aerosols. 36% of our simulations had no gap. In some cases (25%) the combination of the three produced an aerosol whose top was close to the (fixed) aircraft altitude of 6100m (roughly 20,000 ft). In those cases the aer_thick parameter was reduced to constrain it by this height. This explains the somewhat non-gaussian appearance of that parameter's histogram.

**fmf**: Aerosols were defined by two size modes, fine and coarse. The aerosol fine size mode fraction is determined by taking the ratio of fine size mode aerosol aerosol optical depth to the total aerosol optical depth, at 555nm. For ORACLES, the expectation was that the primarily smoke aerosols would be dominated by the aerosol fine size mode, which is why all of the parameters defining the fine size mode (optical depth, size distribution, complex refractive index) are varied for the training set. Only minimal quantities of coarse mode aerosols are to be expected, so the ability to retrieve their optical properties is not feasible. For this reason, coarse mode aerosol optical depth is the only parameter varied in the training set, and in an indirect manner ( $AOD_{coarse} = AOD_{total}$ (1-FMF) ).

## Fixed parameters in training set

Because of the limitations of computational expense, file size, and retrieval sensitivity, not all potential scene descriptive parameters are varied in the training set. *Table 2* lists some of these fixed parameters.

*Table 2 Fixed parameters in training set*

| Parameter description | Value | Unit |
|---|---|---|
| Aircraft altitude | 6100 | Meters |
| Cloud droplet size distribution | Monomodal 2 parameter modified Gamma distribution (Hansen and Travis, 1974 eq 2.56) | |
| Aerosol size distribution | Bimodal 2 parameter lognormal distribution (Hansen and Travis, 1974 eq 2.60) | |
| Coarse size mode refractive index | 1.47-i0.01 | |
| Coarse size mode effective radius | 6.91 | microns |
| Coarse size mode effective variance | 0.867 | |
| Trace gas absorption | Neglected (to be corrected in observational data) | |
| Atmospheric pressure | 1013.25 | mbar |
| Ground temperature | 288.15˚ | K |
| Ocean surface reflectance | None | |
| Simulation geometry | Slab, plane parallel | |

**Aircraft altitude** During ORACLES, the observational aircraft (a Lockheed P-3), flew at a variety of altitudes, although the remote sensing legs were usually restricted to portions of the flight above the aerosol layer. Since the cloud layer at the bottom of the atmosphere can be, in most cases, considered optically infinite, the distance between cloud top height and the aircraft should primarily be driven by the **cloud_top** parameter. Essentially, it defines that distance = aircraft altitude – cloud top.

**Trace gases** are not accounted for in these simulations. This is in part because the RSP spectral band passes are chosen to avoid major gas absorption wavelengths. However, it is simpler, and requires less simulations, to account for the gas transmittance in the real observations, and apply a trained NN to those corrected observations.

**Ocean surface reflectance** is assumed to be black. For the ocean, this is not the worst assumption, but the primary reason is that clouds are expected to be optically thick enough above the ocean that its reflectance does not matter. For some low COD's this might not be the case, but for our purposes we do not expect to do well for COD<3 or so since 3D effects (not simulated) would also come into play.

## Other parameters

A user of this training set may wish to train against parameters not explicitly defined in *Table 1*, but those that can be derived from them. Examples might be the aircraft – cloud top distance described above, or the aerosol optical depth defined by mode rather than total aerosol optical depth and fine mode fraction. One important parameter that is not easy to derive from the parameters in *Table 1* is the Single Scattering Albedo (SSA), although this is output from intermediate files during the simulation. This parameter was also saved as the **ssa** variable in the output files, defined for the total aerosol at 555nm. SSA is the ratio of scattered to total extinction, and as such defines the aerosol absorption. In the radiative transfer model used for this analysis, aerosol absorption is defined by the imaginary refractive index. However, the amount of absorption produced for a given imaginary refractive index also depends on the real component of the refractive index and the size of the aerosol. Thus, SSA may provide a more appropriate parameter to which to train the NN, as it is presumably more orthogonal to other parameters in the space of the observations.

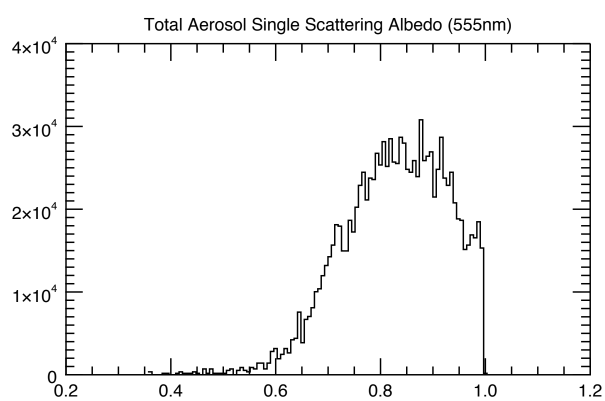


*Figure 2 Single Scattering Albedo (555nm) histogram*

## Simulated observations

Simulations are performed for seven of the RSP instrument channels (**waveln**, 0.410, 0.470, 0.555, 0.670, 0.865, 1.59, 2.26, µm) at 152 view zenith angles between +/- 60.5˚ (**vza**). Note that the user will most likely have to resample or cut the data for the latter since most observations to not span that entire range. For each of the randomly sampled parameter sets described in *Table 1*, the simulations are given for 8 solar zenith angles and 22 relative solar-view azimuth angles, specified in **sza** and **azi**, respectively. The NN is expected to be trained with geometry as an input. Unique values of these parameters are given in **uniq_sza** and **uniq_azi.**

Three radiometric variables are saved in the file: **ref_i**, **ref_q**, and **dolp**. These are the reflectance (I Stokes vector element), reflectance of the Q component of the stokes vector (defined in the solar principal plane), and the Degree of Linear Polarization. All of these have been 'standardized' in order to both ease the computational aspects of the NN creation and use, and to incorporate instrument measurement uncertainty into the NN. RSP measurement uncertainty is quite different for each of the above radiometric variables, and depends on measurement conditions. Standardization typically takes the mean of a distribution of values, and the standardization, such that

Standardized = (Value – mean) / standard deviation

In this work, we have replaced the standard deviation with the simulated measurement uncertainty in order to equally weight data from different observations (Miller et al, in prep). The RSP undertainty model is from Knobelspiesse et al, 2019.

Standardization is performed uniquely for each geometry (solar zenith and relative azimuth). Thus, to get the true simulation value for DoLP, one would use the following transform:

$DoLP_{true}$ = (**dolp** * **unc_dolp**) + **mean_dolp**

This transform would need to be applied to the output of the application of the trained NN to real data.

Figure 3 and Figure 4 are examples of the standardized I and Q reflectances for an individual simulation at a specific parameter state and geometry.

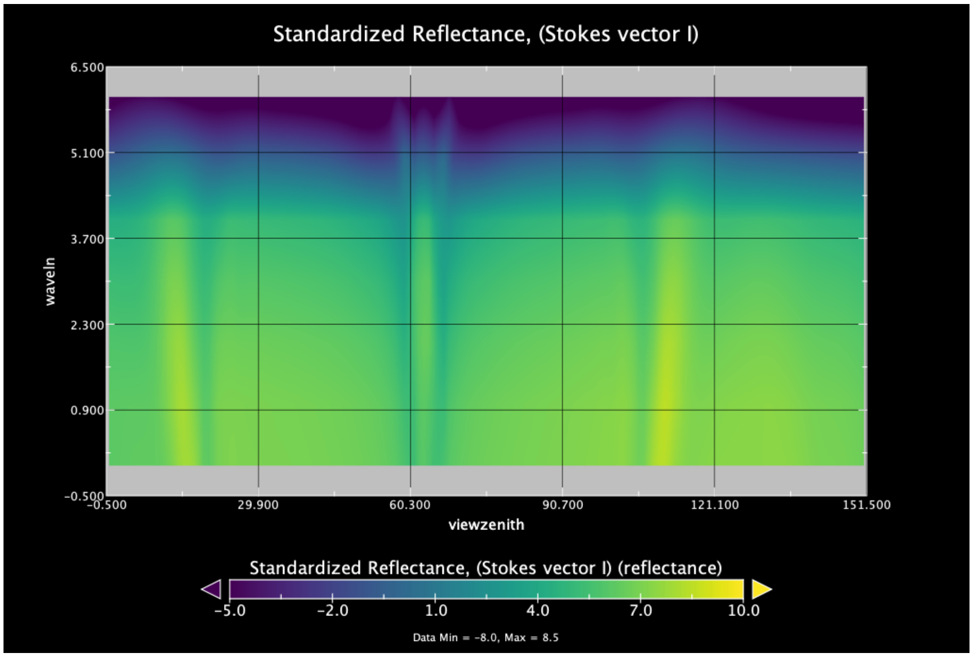

Figure 3 Sample standardized reflectance simulation

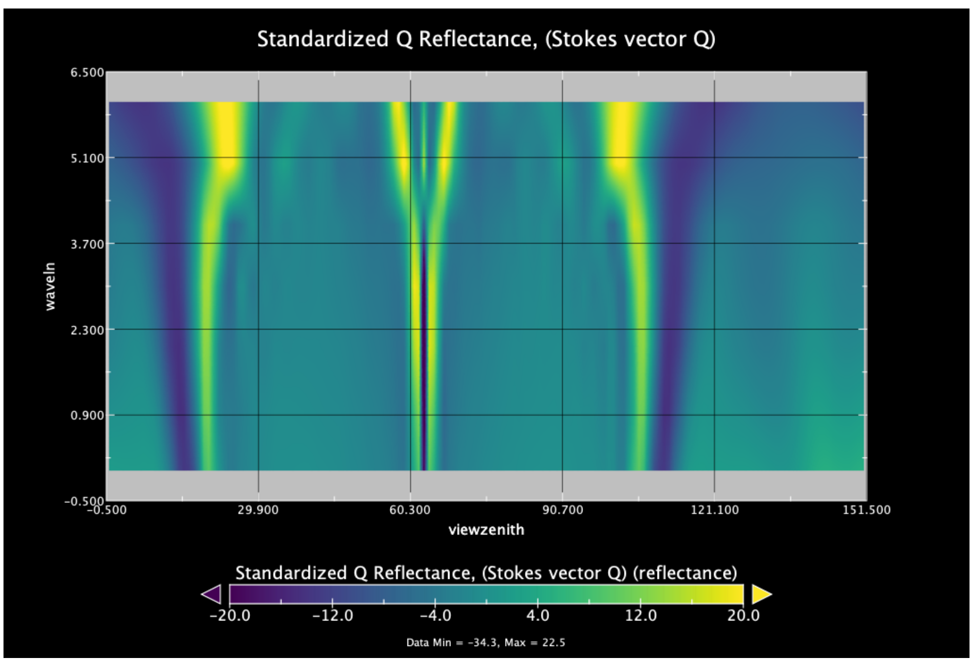

Figure 4 Sample standardized Q reflectance simulation

## Files

The following files are stored in the NASA OceanColor web repository here:
https://oceancolor.gsfc.nasa.gov/fileshare/kirk_knobelspiesse/nn_aca/

Which includes the following files:

3.39gb NN_ACA_20190805_01_std.nc
3.39gb NN_ACA_20190805_02_std.nc
3.39gb NN_ACA_20190805_03_std.nc
3.39gb NN_ACA_20190805_04_std.nc
1.07gb NN_ACA_20190805_05_std.nc

Note that these were split into smaller files to make the file size manageable. Test runs should probably be run on the last, smallest file.

**References**

Cairns B, Russell EE, Travis LD. Research scanning polarimeter: calibration and ground-based measurements. In: Goldstein DH, Chenault DB, editors. Polarization: measurement, analysis, and remote sensing, 3754; 1999. p. 186–97. *Proc. SPIE.*

Cairns B, Russell EE, LaVeigne J, Tennant P. Research scanning polarimeter and airborne usage for remote sensing of aerosols. In: Shaw JA, Tyo JS, editors. Polarization science and remote sensing, 5158; 2003. p. 33–44. *Proc. SPIE.*

Hansen, J.E., & Travis, L.D., 1974. Light scattering in planetary atmospheres. Space Sci Rev, 16(4), 527–610.

Knobelspiesse, K., Q. Tan, C. Bruegge, B. Cairns, J. Chowdhary, B. van Diedenhoven, D. Diner, R. Ferrare, G. van Harten, V. Jovanovic, M. Ottaviani, J. Redemann, F. Seidel, and K. Sinclair. 2019. "Intercomparison of airborne multi-angle polarimeter observations from the Polarimeter Definition Experiment." *Applied Optics*, 58 (3): 650.

Miller, D.J., Segal-Rozenhaimer, M., Knobelspiesse, K., et al. Low Level Liquid Cloud properties during ORACLES derived from airborne polarimetric measurements using a Neural-Network based Algorithm *in preparation*

Segal-Rozenhaimer, M., Miller, D.J., Knobelspiesse, K., Redemann, J., Cairns, B. and Alexandrov, M.D., 2018. Development of neural network retrievals of liquid cloud properties from multi-angle polarimetric observations. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *220*, pp.39-51.

Zuidema P, et al. Smoke and clouds above the southeast Atlantic: upcoming field campaigns probe absorbing aerosol's impact on climate. 2016 *Bull Am Meteor Soc*. 97:1131–5. doi:10.1175/BAMS-D-15-00082.1.