Retrievals of aerosol optical depth over the western North Atlantic Ocean during ACTIVATE

Leong Wai Siu[1], Joseph S. Schlosser[2,3], David Painemal[2,4], Brian Cairns[5], Marta A. Fenn[2,4], Richard A. Ferrare[2], Johnathan W. Hair[2], Chris A. Hostetler[2], Longlei Li[6], Mary M. Kleb[2], Amy Jo Scarino[2,4], Taylor J. Shingler[2], Armin Sorooshian[1,7], Snorre A. Stamnes[2], and Xubin Zeng[1]

[1]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, Arizona
[2]NASA Langley Research Center, Hampton, Virginia
[3]NASA Postdoctoral Program, NASA Langley Research Center, Hampton, Virginia
[4]Analytical Mechanics Associates, Inc., Hampton, Virginia
[5]NASA Goddard Institute for Space Studies, New York, New York
[6]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York
[7]Department of Chemical and Environmental Engineering, University of Arizona, Tucson, Arizona

Correspondence: Leong Wai Siu (leongwaisiu@arizona.edu)

Abstract. Aerosol optical depth was retrieved from two airborne remote sensing instruments, the Research Scanning Polarimeter (RSP) and Second Generation High Spectral Resolution Lidar (HSRL-2), during the NASA Aerosol Cloud meTeorology Interactions oVer the western ATlantic Experiment (ACTIVATE). The field campaign offers a unique opportunity to evaluate an extensive 3-year dataset under a wide range of meteorological conditions from two instruments on the same platform.

5　However, a longstanding issue in atmospheric field studies is that there is a lack of reference datasets for properly validating field measurements and estimating their uncertainties. Here we address this issue by using the triple collocation method, in which a third collocated satellite dataset from the Moderate Resolution Imaging Spectroradiometer (MODIS) is introduced for comparison. HSRL-2 is found to provide a more accurate retrieval than RSP over the study region. The error standard deviation of HSRL-2 with respect to the ground truth is 0.027. Moreover, this approach enables us to develop a simple, yet

10　efficient, quality control criterion for RSP data. The physical reasons for the differences in two retrievals are determined to be cloud contamination, aerosols near surface, multiple aerosol layers, absorbing aerosols, non-spherical aerosols, and simplified retrieval assumptions. These results demonstrate the pathway for optimal aerosol retrievals by combining information from both lidar and polarimeter for future airborne and satellite missions.

1　Introduction

15　Aerosol particles constitute an important component of the Earth's atmosphere by altering its radiative energy balance. Aerosols impact our climate system directly by scattering and absorbing solar and terrestrial radiation and indirectly by serving as cloud condensation nuclei (CCN) and ice nuclei (IN) (Lohmann and Feichter, 2005; Pöschl, 2005). The change in CCN and IN concentrations can influence the distribution of liquid water, ice, and mixed-phase clouds and the frequency of precipitation. The complex interaction of aerosols with clouds, radiation, and meteorology makes it difficult to probe the feedback and

20 response of our climate system under a perturbation of anthropogenic or natural aerosols (Bellouin et al., 2020; Chen et al., 2014). In the most recent Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC), aerosols still represent the largest uncertainty in the global radiative forcing (IPCC, 2023).

To disentangle aerosol–cloud interactions, it is imperative to collect adequate observations for robust statistics across a wide range of cloud regimes. The recent Aerosol Cloud meTeorology Interactions oVer the western ATlantic Experiment (ACTI-

25 VATE) epitomizes a coordinated effort to respond to this need (Sorooshian et al., 2023). Being one of the National Aeronautics and Space Administration (NASA) Earth Venture Suborbital-3 (EVS-3) missions, ACTIVATE has three objectives: First, quantify the underlying relationship between the aerosol number concentration N_a, CCN concentration, and cloud drop number concentration N_d. Second, develop a better process-level representation of cloud properties and aerosol–cloud interactions in a hierarchy of numerical models. Finally, evaluate state-of-the-art remote sensing instruments that are built for retrieving the

30 aerosol and cloud properties (Sorooshian et al., 2019).

Aerosol optical depth (AOD) is a vertically integrated quantity of aerosol extinction coefficient, representing the column aerosol loading. It is one of the most widely used parameters to monitor the long term evolution of aerosols at both global and regional scales (Andreae, 2009; Seinfeld et al., 2016). It is also important to evaluate AOD retrievals and their associated uncertainties. An uncertainty of AOD retrievals of 0.01 could lead to an uncertainty of aerosol radiative forcing of ∼0.5

35 W m^{-2} (Chylek et al., 2003; Hansen et al., 1995; Mishchenko et al., 2004). Two approaches are regularly used to quantify uncertainties of AOD retrievals: Propagated (prognostic) uncertainty and truth-based (diagnostic) uncertainty (Gao et al., 2022; Sayer et al., 2020). The former is a theoretical estimate which is obtained by considering measurement uncertainties, forward model uncertainties, and a priori assumptions via minimizing a suitable cost function. The latter is to compare the retrieval with a reference dataset. The two approaches are complementary to each other as the reference dataset can assess the theoretical

40 estimate using appropriate statistical methods (Gao et al., 2022).

This work is focused on the second approach as two remote sensing instruments, the Research Scanning Polarimeter (RSP) and Second Generation High Spectral Resolution Lidar (HSRL-2), retrieved AOD over the western North Atlantic Ocean during ACTIVATE. Before proceeding, however, there is a catch. That is, how do we objectively assess these two datasets? Ideally, the error of a dataset should be quantified by comparing it with a 'ground truth' (Caires and Sterl, 2003). Practically,

45 the 'ground truth' for AOD measurements is replaced by measurements from some ground-based instruments such as sun photometers in the AErosol RObotic NETwork (AERONET) (Holben et al., 1998). But ground-based observations are unavailable over the study region, primarily an open ocean. Furthermore, all instruments and retrievals are subject to errors consisting of systematic bias and random noises (Stoffelen, 1998).

In this study, we propose to use a statistical method called Triple Collocation (TC) which partially circumvents the afore-

50 mentioned issues. A third dataset is introduced and the error characteristics of each dataset are determined with respect to the unknown ground truth. The most accurate dataset is then used to help design a simple, yet efficient, data flag to improve the quality of the other datasets. The physical processes that lead to the difference between RSP and HSRL-2 retrievals are discussed.
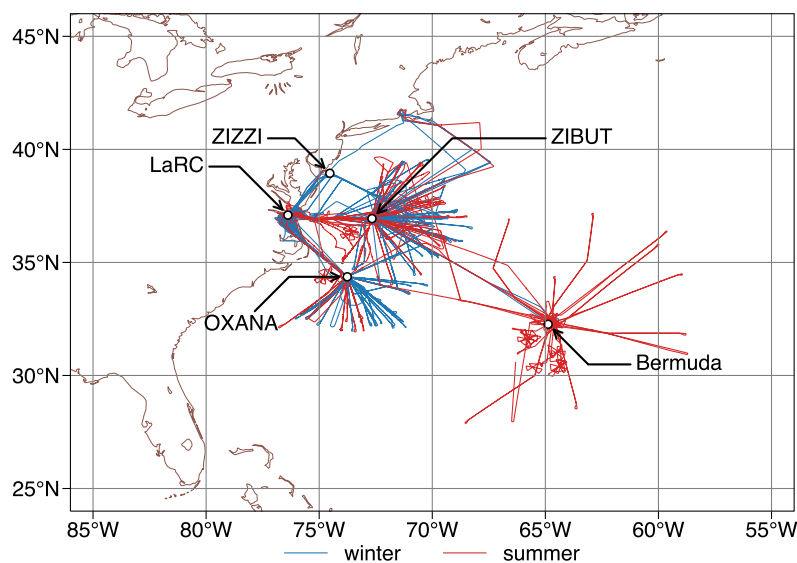
Figure 1. The ACTIVATE study region and NASA Langley King Air flight tracks between 2020 and 2022. Summer and winter deployments are indicated in red and blue, respectively. The two campaign bases, NASA Langley Research Center (LaRC) and Bermuda, and three commonly used waypoints OXANA, ZIBUT, and ZIZZI are annotated with arrows.

2 Data

2.1 Overview of the ACTIVATE field campaign

The ACTIVATE field campaign started in February 2020 and concluded in June 2022 with six deployments. There were winter and summer deployments each year, roughly covering November to March and May to September, respectively (Sorooshian et al., 2023). While it is recommended to explicitly label the year and season of each deployment, for simplicity we refer to the three winter and three summer deployments collectively as the 'winter' and 'summer' deployments.

Two aircraft, the King Air (high-flying) and Falcon (low-flying) from the NASA Langley Research Center (LaRC), were used to carry out spatially coordinated flights over the western North Atlantic Ocean. The total number of flights for the King Air and Falcon are 174 and 168, respectively, with 162 being joint flights. The remote sensing instruments in this study were flown on the King Air which logged a total of 592 flight hours (Sorooshian et al., 2023).

Most of the 3–4 hour flights were based at NASA LaRC, with a notable exception being 17 flights based in Bermuda during June 2022 (Figure 1). Due to air traffic considerations associated with military restricted areas, the aircraft usually had to transit through one of three waypoints OXANA, ZIBUT, and ZIZZI.

2.2 Research Scanning Polarimeter

The Research Scanning Polarimeter (RSP) is a remote sensing instrument that uses three pairs of optical assemblies to simultaneously measure linear polarization of the intensity at four polarization azimuths and nine spectral channels centered at 410, 470, 555, 670, 865, 960, 1590, 1880, and 2264 nm for each scene (Cairns et al., 1999, 2003). The first three components of the Stokes vector (I, Q, and U) are then derived from the measurement. RSP operates by continuously scanning the field of view (14 mrad) over $\pm 55°$ from viewing zenith (~ 140 views) along the flight track, which samples at $0.8°$ interval (Alexandrov et al., 2012; Cairns et al., 2003). Each RSP scan consists of measurements from all different viewing angles. That is, each target at the cloud top or ground surface is swept from different scans at different viewing angles. These actual scans are assembled into virtual scans for each target at nadir (Alexandrov et al., 2012).

The aerosol properties are retrieved using the RSP Microphysical Aerosol Properties from Polarimetery (MAPP) algorithm version 1.48 (Stamnes et al., 2018; Schlosser et al., 2022). The algorithm uses an optimal-estimation method to minimize a cost function χ^2 which includes the data term and a priori term. The current version assumes that the aerosol size distribution is bimodal: one fine mode and one coarse mode (Stamnes et al., 2018). Aerosol properties such as AOD, refractive index, and effective radius are retrieved separately for the two modes. The coarse mode is assumed to be composed of non-absorbing sea salt particles with a refractive index equivalent to that of water. The coarse-mode aerosol top height is assumed to be 1 km while the fine-mode aerosol top height is retrieved (Schlosser et al., 2022). The retrievals are considered successful when the final normalized cost function of the data term χ' is below 0.15. The theoretical $\pm 1\sigma$ accuracy of the MAPP algorithm for AOD is ± 0.02 (Stamnes et al., 2018).

In this study, we use the RSP level 2 aerosol product which includes column-averaged aerosol optical and microphysical parameters and ocean color parameters. The RSP data are averaged horizontally over 10 scans from level 1C product, with a temporal resolution of ~ 9.3 s and an along-track spatial resolution of ~ 1 km.

2.3 Second Generation High Spectral Resolution Lidar

The Second Generation High Spectral Resolution Lidar (HSRL-2) is an airborne multi-wavelength instrument which uses the HSRL technique to separate the aerosol and molecular backscatter signals at 355 and 532 nm and the standard backscatter lidar technique at 1064 nm. HSRL-2 can vertically resolve the backscatter coefficient and depolarization at 355, 532, and 1064 nm and extinction coefficient at 355 and 532 nm (Burton et al., 2018; Hair et al., 2008; Ferrare et al., 2023). Because of the incorporation of three backscatter and two extinction measurements, the instrument is also referred to as a "$3\beta + 2\alpha$" lidar (Burton et al., 2016). The field of view of HSRL-2 is 1 mrad (Schlosser et al., 2022). AOD is derived from the aerosol backscattering coefficient using the difference in the molecular return signals (Hair et al., 2008).

The performance of the HSRL-2 measurements have been analyzed with various instruments. For example, Shinozuka et al. (2013) showed that the root mean square difference of 532 nm AOD between the Spectrometer for Sky-Scanning, Sun-Tracking Atmospheric Research (4STAR) and HSRL-2 is 0.01 during the 2012 Two-Column Aerosol Project (TCAP) campaign. Sawamura et al. (2017) showed outstanding agreement between the AERONET and HSRL-2 from over 300 profiles collected during

100   the 2013 Deriving Information on Surface Conditions from COlumn and VERtically Resolved Observations Relevant to Air Quality (DISCOVER-AQ) campaign. Stamnes et al. (2018) found that the root mean square deviations between RSP and HSRL-2 AOD at 532 nm are ∼0.07 during TCAP and ∼0.04 during the 2014 Ship-Aircraft Bio-Optical Research (SABOR) campaign.

   In this study, we use the 10-s aerosol profile product which has a spatial resolution of ∼1 km. We also use the following
105   derived products: Mixing layer height (MLH) and aerosol identifier (ID). Mixing layer height is estimated from the sharp gradients of HSRL-2 aerosol backscatter profiles at 532 nm (Scarino et al., 2014). A Haar wavelet transform is applied to the vertical profiles to identify the transition zone above the top of the aerosol mixed layer. MLH is defined as the altitude of the maximum value after the wavelet transform. The information of aerosol types based on the HSRL aerosol classification scheme (Burton et al., 2012). The algorithm identifies aerosols based on the lidar ratio, depolarization, backscatter color ratio,
110   and spectral depolarization ratio from HSRL-2 measurements. Eight types of aerosols are classified, including dust, polluted marine, fresh smoke, smoke, urban, marine, dusty mix, and ice. The HSRL aerosol classification scheme has been shown to improve a similar scheme for the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) instrument (Burton et al., 2013; Ferrare et al., 2023).

   To expedite comparison with gridded global reanalysis data, selected instantaneous meteorological fields from the Modern-
115   Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al., 2017) are interpolated to the HSRL-2 curtains along the flight tracks. These 3 hourly MERRA-2 variables have a horizontal resolution of $0.5° \times 0.625°$ (latitude × longitude) on 42 pressure levels from 1000 hPa to 0.1 hPa.

2.4   Other ACTIVATE datasets

The King Air navigational data for RSP and HSRL-2 are provided by an Applanix 610 system at 2 Hz, including flight date,
120   time, longitude, latitude, altitude, and other aircraft motion variables (Sorooshian et al., 2023). Two cameras were mounted on the King Air: one facing nadir and the other facing forward (Sorooshian et al., 2023). The images collected from the airborne cameras are used to identify the presence of clouds above and below the aircraft. Recently, a cloud detection neural network (CDNN) algorithm was developed to speed up the cloud identification process by training and testing the camera imageries via convolutional neural network, which resulted in above-aircraft and below-cloud cloud mask products at 1 Hz (Nied et al.,
125   2023). The accuracy of the forward-facing CDNN is 96% by validating against human-labeled testing data.

2.5   Moderate Resolution Imaging Spectroradiometer

The Moderate Resolution Imaging Spectroradiometer (MODIS) is an imaging instrument with 36 spectral channels ranging from 0.41 μm to 14.5 μm. The instrument collects data at three nadir spatial resolutions (250 m, 500 m, and 1 km) (Remer et al., 2005). As part of the NASA Earth Observing System, two almost identical MODIS sensors were launched on the Terra
130   and Aqua satellites in December 1999 and May 2002, respectively. Both satellites observe the Earth from a polar-orbiting, sun-synchronous orbit at an altitude of ∼700 km. The equatorial overpass times for the descending mode of Terra and the

Atmospheric
Measurement
Techniques
Discussions
Open Access
EGU

ascending mode of Aqua are 1030 and 1330 local solar time, respectively. The swath of each scan is 2330 km × 10 km (across track × along track at nadir) (Levy et al., 2013). Seven bands are used for aerosol retrievals.

135   In this study, we use the following level 2 aerosol products from the MODIS Collection 6.1: MOD04_3K from Terra and MYD04_3K from Aqua. The horizontal resolution is 3 km × 3 km at nadir. The Dark Target (DT) aerosol algorithm is used to retrieve the 3 km product (Levy et al., 2013; Remer et al., 2013). To compare with the RSP and HSRL-2 AOD at 532 nm, the MODIS AOD at 550 nm is converted to 532 nm using the Ångström exponent from the MODIS AOD at 470 nm and 550 nm.

2.6   Cloud-Aerosol Lidar with Orthogonal Polarization

The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) is the primary instrument that was launched on the Cloud-
140   Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite in April 2006 (Winker et al., 2009). We use the CALIOP level 3 stratospheric aerosol profile products which are based on the version 4 of CALIOP level 1B and level 2 datasets (Kar et al., 2019). The horizontal resolution of the monthly stratospheric AOD data is 5°×20° (latitude×longitude). To maximize the length of available data, we stitched the datasets from version 1.00 and 1.01. The version 1.01 begins in July 2020.

145   3   Evaluation methods

3.1   Traditional metrics

To summarize the distribution of an AOD dataset, it is common to use mean and standard deviation (SD) to measure its central tendency and dispersion, which are known to be affected by outliers. Median and interquartile range (IQR = third quartile minus first quartile) are also used as they are more robust against outliers.

150   To evaluate AOD between two collocated datasets X and Y, each having n elements, we use the following metrics,

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}, \tag{1}$$

$$\text{MB} = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \overline{X} - \overline{Y}, \tag{2}$$

$$\text{RMSD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2}, \tag{3}$$

where r, MB, and RMSD are the Pearson's correlation coefficient, mean bias, and root mean square deviation, respectively; overbar denotes the mean value of a dataset.

3.2   Triple collocation

Triple collocation (TC) is a statistical method that estimates error statistics of a target variable from three independent and
155   collocated datasets (Stoffelen, 1998). The method does not require any one of the triplet items being ground truth. It has been

Atmospheric
Measurement
Techniques
Open Access
EGU
Discussions

used for characterizing errors in a wide range of geophysical variables such as sea surface temperature (O'Carroll et al., 2008), soil moisture (Draper et al., 2013), and ocean wave height (Caires and Sterl, 2003; Janssen et al., 2007).

In this study, we take advantage of three available AOD retrievals from RSP, HSRL-2, and MODIS. To start with, each dataset is considered as an estimate of the (unknown) ground truth. Therefore, a functional relationship (or an error model) is assumed to exist between each dataset and the ground truth to account for the potential deviation from the truth. Linear relationship is one of the simplest error models. Some linear error models are discussed in detail in Zwieback et al. (2012). We adopt an affine (also linear) model for the AOD datasets,

$$\tau_i = a_i + b_i\Theta + \varepsilon_i, \tag{4}$$

where τ_i ($i \in \{\text{RSP, HSRL-2, MODIS}\}$) represents the AOD from dataset i; Θ is the true AOD; a_i, b_i, and ε_i are additive bias, multiplicative bias, and random error from dataset i, respectively. To simplify the notation, we assign RSP $= 1$, HSRL-2 $= 2$, and MODIS $= 3$.

Together with the error model being linear, the following assumptions are made (Gruber et al., 2016). First, the random error has zero mean ($E[\varepsilon_i] = 0$). Second, the random error is uncorrelated with the true AOD ($\text{Cov}[\varepsilon_i, \Theta] = 0$). Third, the random error of different products is uncorrelated with each other ($\text{Cov}[\varepsilon_i, \varepsilon_j] = 0, i \neq j$). Lastly, the additive and multiplicative biases are time invariant.

The equations of variance and covariance between the collocated datasets can then be obtained (McColl et al., 2014; Stoffelen, 1998) by

$$\text{Cov}[\tau_i, \tau_j] = b_i b_j \text{Var}[\Theta] + b_i \text{Cov}[\Theta, \varepsilon_j] + b_j \text{Cov}[\Theta, \varepsilon_i] + \text{Cov}[\varepsilon_i, \varepsilon_j], \tag{5}$$

$$\text{Var}[\tau_i] = b_i^2 \text{Var}[\Theta] + 2b_i \text{Cov}[\Theta, \varepsilon_i] + \text{Var}[\varepsilon_i], \tag{6}$$

where $\text{Var}[\Theta] = \sigma_\Theta^2$ and $\text{Var}[\varepsilon_i] = \sigma_{\varepsilon_i}^2$ is the error variance. Note that $\text{Cov}[\tau_i, \tau_j] = \text{Cov}[\tau_j, \tau_i]$.

The error variances are then solved with the help of above assumptions (McColl et al., 2014),

$$\sigma^2_{\varepsilon_{\text{RSP}}} = \text{Var}[\tau_1] - \frac{\text{Cov}[\tau_1, \tau_2]\text{Cov}[\tau_1, \tau_3]}{\text{Cov}[\tau_2, \tau_3]}, \tag{7}$$

$$\sigma^2_{\varepsilon_{\text{HSRL-2}}} = \text{Var}[\tau_2] - \frac{\text{Cov}[\tau_1, \tau_2]\text{Cov}[\tau_2, \tau_3]}{\text{Cov}[\tau_1, \tau_3]}, \tag{8}$$

$$\sigma^2_{\varepsilon_{\text{MODIS}}} = \text{Var}[\tau_3] - \frac{\text{Cov}[\tau_1, \tau_3]\text{Cov}[\tau_2, \tau_3]}{\text{Cov}[\tau_1, \tau_2]}. \tag{9}$$

McColl et al. (2014) extended the theory by solving the correlation coefficient of each dataset with respect to the *ground truth*, resulting in

$$r_{\text{RSP}} = \pm\sqrt{\frac{\text{Cov}[\tau_1, \tau_2]\text{Cov}[\tau_3, \tau_j]}{\text{Var}[\tau_1]\text{Cov}[\tau_2, \tau_3]}}, \tag{10}$$

$$r_{\text{HSRL-2}} = \pm\text{sgn}(\text{Cov}[\tau_1, \tau_3]\text{Cov}[\tau_2, \tau_3])\sqrt{\frac{\text{Cov}[\tau_1, \tau_2]\text{Cov}[\tau_2, \tau_3]}{\text{Var}[\tau_2]\text{Cov}[\tau_1, \tau_3]}}, \tag{11}$$

$$r_{\text{MODIS}} = \pm\text{sgn}(\text{Cov}[\tau_1, \tau_2]\text{Cov}[\tau_2, \tau_3])\sqrt{\frac{\text{Cov}[\tau_1, \tau_3]\text{Cov}[\tau_2, \tau_3]}{\text{Var}[\tau_3]\text{Cov}[\tau_1, \tau_2]}}, \tag{12}$$

where sgn is the signum function, and there is a sign ambiguity for each r_i but practically all the AOD datasets are expected to be positively correlated with the unknown true value. If the ground truth is known, the correlation coefficient can be easily computed from (1). However, if the ground truth is *unknown*, the TC method enables the computation of correlation coefficient from (10)–(12).

3.3 Data alignment

To ensure the robustness of triple collocation analysis, we attempt to minimize the representation error due to the different support scales between the AOD datasets. The 10-s HSRL-2 data are coarsened to 30-s resolution by horizontally averaging three consecutive data points, which results in a 3 km product. Then all RSP data points within the same 30-s window are averaged. To minimize cloud impact, all HSRL-2 and RSP data points must be valid within the same window. The data alignment and aggregation result in 6988 pairs of data from 134 flights.

4 Results and discussion

4.1 Comparison of RSP and HSRL-2

The design of ACTIVATE to take place over the northwest Atlantic across different seasons was partly to characterize aerosol-cloud interactions across a wide range of aerosol conditions (Sorooshian et al., 2019). The overall distribution of RSP and HSRL-2 AOD is shown in Figure 2, which demonstrates that a broad range of values was observed. The mean values of RSP and HSRL-2 AOD from 6988 pairs for the whole field campaign period are 0.173 and 0.120, respectively. RSP has a higher standard deviation than HSRL-2 (0.109 vs 0.074). Both mean and standard deviation are sensitive to outliers so median and
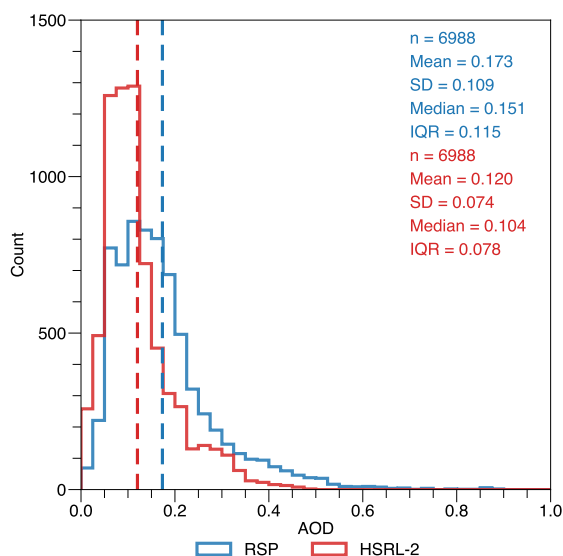
Figure 2. Distribution of RSP (blue) and HSRL-2 (red) AOD for the whole field campaign period (2020–2022). Each bin is 0.025. Mean values (color dashed lines) and selected statistics (mean, median, SD, and IQR) are also shown.

interquartile range are also computed. The mean and median differences between RSP and HSRL-2 are about the same (0.053 vs 0.047) and the interquartile range of RSP is larger than HSRL-2.

195   Table 1 shows the monthly breakdown of the AOD statistics. More than 80% of data points come from March, May, and June. The largest mean AOD happens in May, August, and September for both instruments and the median AOD corroborates the corresponding maxima. The same months also exhibit the largest variations as confirmed from the values of both standard deviation and interquartile range. The mean AOD is smallest in January and November although the number of samples in winter is somewhat inadequate. The seasonal difference is largely consistent with Corral et al. (2021) who used coastal

200   monitoring sites and a reanalysis product to study the regional AOD properties. During summer, there are more African dust events (Aldhaif et al., 2020) and more biomass burning events originated from the western United States (Mardi et al., 2021).

The discrepancy between RSP and HSRL-2 AOD is shown in Figure 3. Around 80% of data points are within ±0.1 intervals (Figure 3a). The correlation coefficient r, RMSD, and MB are 0.727, 0.092, and 0.053, respectively. Most of the extreme outliers occur during the summer deployments (Figures 3b–c).

205   A complementary approach of analyzing the deviation between RSP and HSRL-2 is to examine their mean squared deviation (MSD = RMSD^2) decomposition. If two datasets x and y are identical to each other, all data points fall onto the 1:1 line ($y = x$, thick gray dashed lines in Figure 3) and MSD equals zero. If there are discrepancies between two datasets (MSD > 0), a linear fit $y = a + bx$ (black dashed lines in Figure 3), where a is intercept and b is slope, quantifies the deviations from 1:1 line with

Atmospheric
Measurement
Techniques
Discussions

Table 1. Monthly RSP and HSRL-2 AOD statistics. SD and IQR denote standard deviation and interquartile range, respectively.

Month	Count	RSP				HSRL-2			
		Mean	Median	SD	IQR	Mean	Median	SD	IQR
Jan	44	0.026	0.020	0.026	0.019	0.010	0.004	0.011	0.020
Feb	230	0.080	0.067	0.051	0.057	0.049	0.047	0.029	0.046
Mar	1610	0.128	0.124	0.058	0.085	0.084	0.083	0.040	0.059
Apr	0	–	–	–	–	–	–	–	–
May	1932	0.220	0.182	0.146	0.204	0.153	0.115	0.098	0.151
Jun	2331	0.158	0.146	0.075	0.089	0.110	0.106	0.039	0.045
Jul	0	–	–	–	–	–	–	–	–
Aug	496	0.236	0.208	0.105	0.096	0.161	0.145	0.074	0.098
Sep	327	0.224	0.189	0.102	0.116	0.189	0.173	0.090	0.112
Oct	0	–	–	–	–	–	–	–	–
Nov	4	0.014	0.015	0.004	0.007	0.037	0.036	0.002	0.002
Dec	14	0.064	0.058	0.029	0.013	0.026	0.023	0.013	0.018
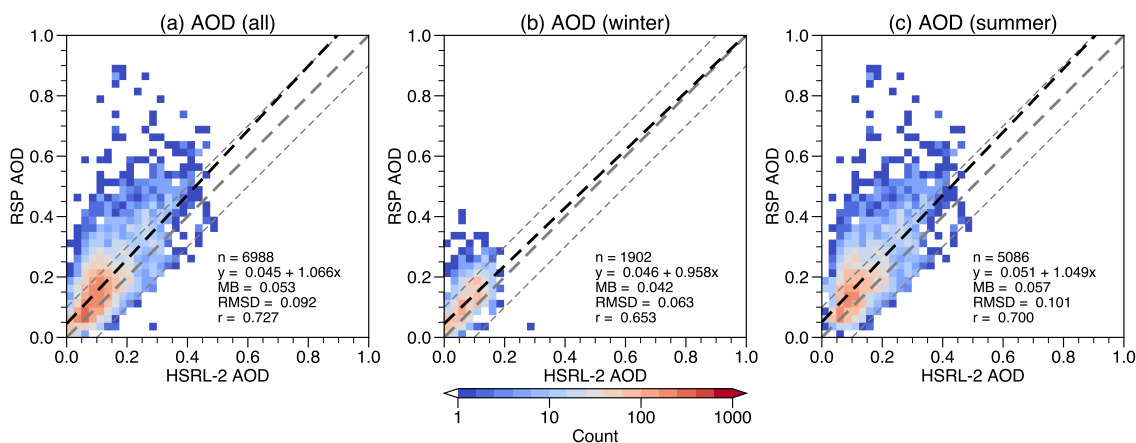
Figure 3. Two-dimensional histograms of RSP and HSRL-2 AOD. (a) All data. (b) Winter deployments only. (c) Summer deployments only. Bin sizes are 0.025 for both instruments. In each plot, the linear fit (black dashed), 1:1 line (thick gray dashed), ±0.1 intervals (thin gray dashed), and selected error metrics are shown.

Table 2. Mean square deviation (MSD) decomposition of RSP and HSRL-2 AOD.

Period	MSD	SB (%)	NU (%)	LC (%)
All	0.00847	0.00283 (33.3)	2.45×10^{-5} (0.29)	0.00562 (66.4)
Winter	0.00392	0.00179 (45.7)	2.99×10^{-6} (0.08)	0.00213 (54.3)
Summer	0.01018	0.00327 (32.2)	1.43×10^{-5} (0.14)	0.00689 (67.7)

$a \neq 0$ and/or $b \neq 1$. Here we use the MSD decomposition proposed by Gauch et al. (2003),

$$MSD = SB + NU + LC, \tag{13}$$
$$= (\bar{y} - \bar{x})^2 + (1-b)^2\sigma_x^2 + (1-r^2)\sigma_y^2, \tag{14}$$

210 where SB is squared bias, NU is nonunity slope, LC is lack of correlation, overlines denote the mean values, σ is standard deviation, r is correlation coefficient between x and y, and b is the slope of regression line of y on x.

Such partitioning has a straightforward geometric meaning for all three components. SB represents the translation of the linear fit from 1:1 line when there is a systematic bias, NU represents the rotation of linear fit when the slope b becomes different from 1, and LC represents the scattering of data points. The MSD decomposition in Table 2 shows that data scattering

215 is the main contribution to the MSD for both winter and summer deployments. Systematic data bias is more severe in winter than in summer although the magnitude of bias is smaller in winter.

A question naturally arises from the above analysis as to which dataset is more accurate. Traditional metrics cannot answer this question because there is no available reference dataset over the western Northern Atlantic Ocean. To overcome this issue, we are motivated to carry out the triple collocation analysis.

220 ## 4.2 Triple collocation analysis

We introduce MODIS AOD as the third dataset. Since the Terra and Aqua satellites only provide two measurements during daytime because of their sun-synchronous orbits, we combine both datasets to maximize the number of files for collocation. For each pair of RSP and HSRL-2 data, we collocate the combined MODIS data point within ±60 minutes and 25 km that has the closest distance which deems suitable for mesoscale aerosol variabilities (Anderson et al., 2003). The collocation process

225 results in 2344 triplets, roughly 34% of the original data points. Zwieback et al. (2012) suggested that at least 500 triplets are needed for robust error estimates.

Table 3 summarizes the error metrics of the three instruments. HSRL-2 has the lowest error standard deviation ($\sigma_{\varepsilon_{\text{HSRL-2}}} = 0.0273$) and highest correlation coefficient ($r_{\text{HSRL-2}} = 0.926$) with respect to the true underlying values. After triple collocation, data scattering not only prevails between RSP and HSRL-2 (Figure 4a), but also occurs between RSP and MODIS (Figure 4c).

230 HSRL-2 also maintains relatively higher correlation coefficients with both RSP and MODIS (Figures 4a–b). The results suggest

Table 3. Triple collocation analysis of RSP, HSRL-2, and MODIS AOD.

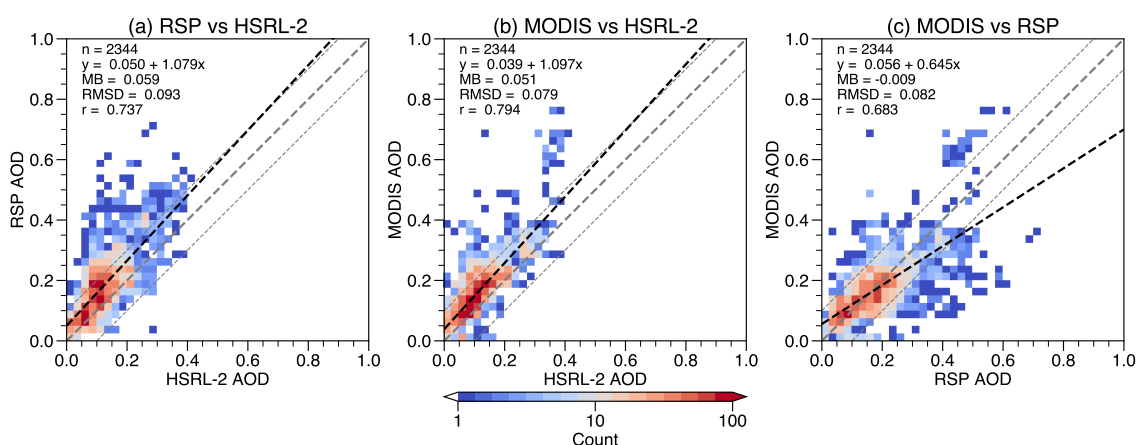| RSP		HSRL-2		MODIS	
$\sigma_{\varepsilon_{RSP}}$	r_{RSP}	$\sigma_{\varepsilon_{HSRL-2}}$	r_{HSRL-2}	$\sigma_{\varepsilon_{MODIS}}$	r_{MODIS}
0.0637	0.796	0.0273	0.926	0.0511	0.858

Figure 4. Two-dimensional histograms of the triple collocated AOD datasets. Bin sizes are 0.02 for both instruments. (a) RSP vs HSRL-2. (b) MODIS vs HSRL-2. (c) MODIS vs RSP. In each plot, the linear fit (black dashed), 1:1 line (thick gray dashed), ±0.1 intervals (thin gray dashed), and selected performance metrics are also shown.

that HSRL-2 is the most accurate dataset among the three. Note also that MODIS AOD corresponds to the whole atmospheric column whereas RSP and HSRL-2 AOD correspond to the column below the aircraft. Therefore, MODIS AOD should be slightly larger than RSP and HSRL-2 AOD due to the contribution above the aircraft. The mean bias shows that MODIS is larger than HSRL-2 (Figure 4b) but RSP is larger than MODIS (Figure 4c). This suggests that it is more likely that RSP has a high bias over HSRL-2, not that HSRL-2 has a low bias over RSP (Figure 3).

4.3 RSP filter

The triple collocation analysis indicates that the most accurate AOD dataset over the western North Atlantic Ocean is HSRL-2 and paves the way for improving the data quality of RSP: design a simple RSP filter to reduce disagreement between the two instruments. There are three guiding principles. First, the filter should effectively get rid of outliers as suggested from the result

Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

Table 4. Same as Table 2 but for HSRL-2 and filtered RSP AOD.

Period	MSD	SB (%)	NU (%)	LC (%)
All	0.00374	0.00091 (24.3)	2.33×10^{-5} (0.62)	0.00281 (75.1)
Winter	0.00335	0.00153 (45.7)	2.44×10^{-5} (0.73)	0.00179 (53.5)
Summer	0.00388	0.00071 (18.4)	9.65×10^{-6} (0.25)	0.00316 (81.4)

240   of MSD decomposition (Table 2). Second, at the same time, the filter should keep as many data points as possible. Third, the criterion should be independent of HSRL-2 properties.

Three filter candidates are sought, including RSP normalized cost function χ', coarse-mode AOD, and fine-mode AOD. Although it is tempting to use the total RSP AOD to filter out outliers, it is anticipated to be only effective for summer data because on average AOD is higher and fluctuates more during summer deployments (Table 1). Therefore, instead we choose

245   coarse-mode and fine-mode AOD. Normalized cost function χ' quantifies the performance of the vector radiative transfer model in the RSP-MAPP algorithm (Stamnes et al., 2018).

The candidates are tested by iteratively decreasing the maximum value allowed in the dataset and three performance metrics (r, MB, RMSD) are plotted in Figure 5. For a balanced candidate, it is expected that when the criterion gets more stringent, r should increase and MB and RMSD should decrease at the same time. It is clear that the fine-mode AOD does not qualify

250   because r generally *decreases* while the maximum allowed fine-mode AOD is reduced (Figure 5c). In other words, constraining the fine-mode AOD will lead to more data outliers, which goes against one of our principles. The remaining cost function and coarse-mode AOD are balanced candidates. A relatively fair way to compare the two balanced candidates is to focus on their performance metrics when 50% of data points are left (intersection of red dashed lines in Figures 5a–b). That means the ranges of cost function and coarse-mode AOD are 0 to 0.049 and 0 to 0.014, respectively. Compared to the coarse-mode AOD, the

255   cost function performs better on the whole. Therefore, the cost function is regarded as the best simple filtering criterion. It has two additional benefits. First, it is globally applicable. Second, it also allows offline post-processing. For simplicity, the cost function threshold is set to be 0.05 thereupon.

To evaluate the cost function filter, we perform again the comparison between RSP and HSRL-2 (Figure 6). All metrics generally improve. Compared to 80% in Figure 3a, more than 89.8% of data points are now within ±0.1 intervals after filtering.

260   By analyzing MSD decomposition before and after filtering (Tables 2 and 4), it is found that the filter is effective in reducing MSD during summer deployments by cutting down around 80% of the systematic bias.

4.4   Why do RSP and HSRL-2 disagree with each other?

Stamnes et al. (2018) suggested some possible sources of the discrepancies between two instruments: (i) aerosols above the aircraft; (ii) clouds above the aircraft; (iii) aerosols near the aircraft; (iv) three-dimensional effects due to clouds surrounding
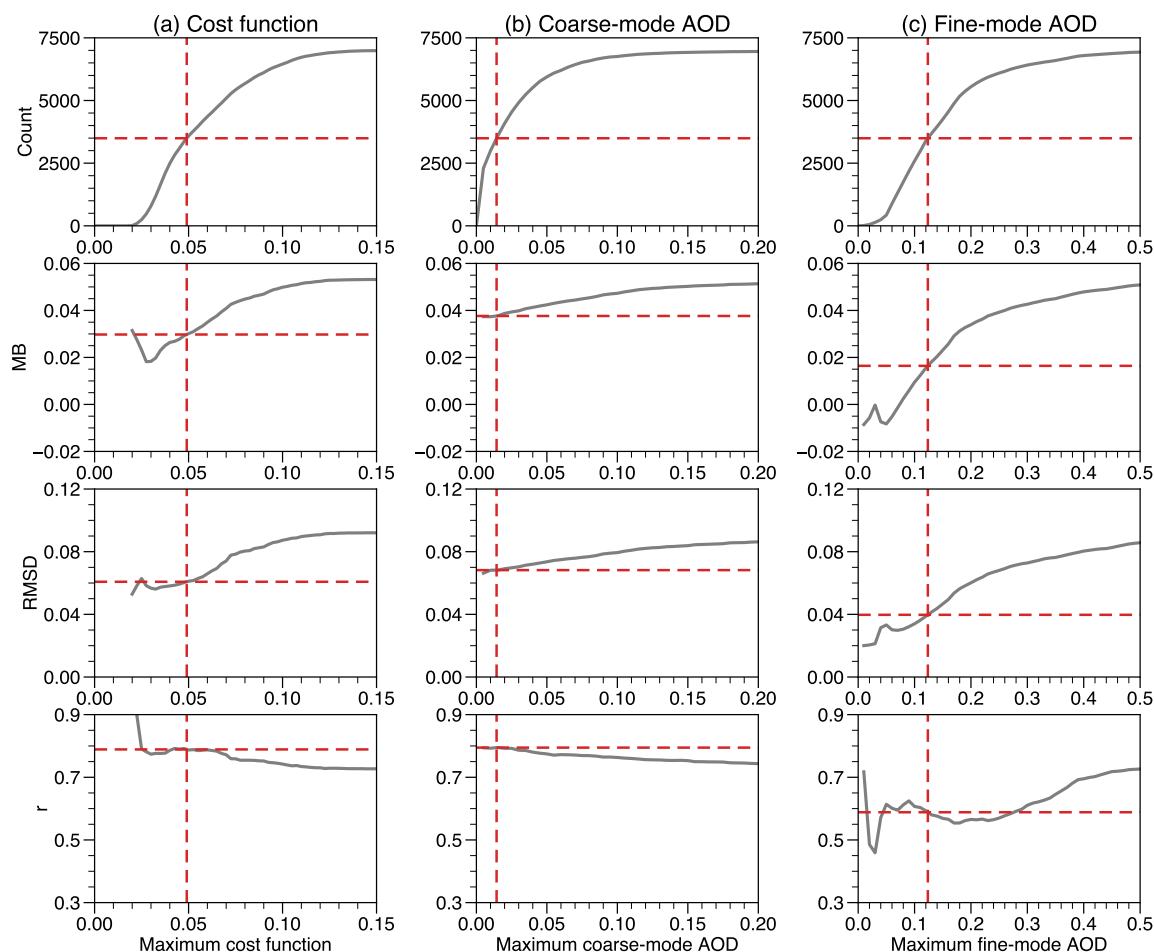
Figure 5. Sensitivity of the performance of RSP filter criterion candidates. (a) Maximum normalized cost function (left column). (b) Maximum coarse-mode AOD (middle column). (c) Maximum fine-mode AOD (right column). In each plot, we calculate a performance metric as a function of criterion candidate threshold. As a result, the number of data points decreases when the criterion gets more stringent (smaller value). Intersecting red dashed lines indicate 50% of the data points (y-axis) meet the criterion threshold (x-axis). First to fourth rows show, in order, the number of valid data points, mean bias, root-mean-squared deviation, and correlation coefficient.

265 the aircraft; (v) aerosols close to the surface (within 100 m); (vi) multiple aerosol layers from different types of aerosol; (vii) presence of absorbing aerosols; (viii) presence of non-spherical aerosols; (ix) intrinsic retrieval and measurement uncertainties. These factors are examined using the available field campaign data.

To streamline the analysis, the aforementioned factors can be broadly organized into two groups: cloud and aerosol impacts. Because the temporal resolution of cloud mask is 1 Hz, we compute the fraction of cloud mask being one (i.e., cloud is present)
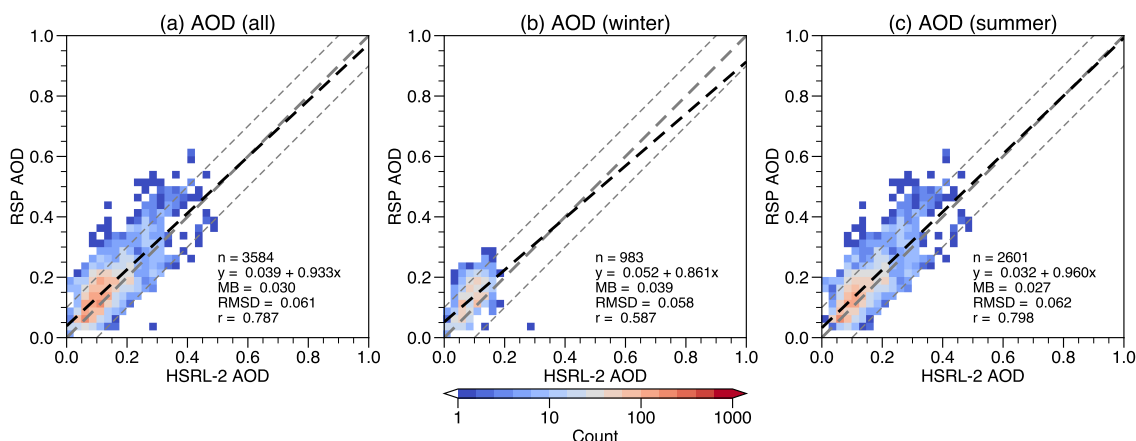
Figure 6. Same as Figure 3 but for HSRL-2 and filtered RSP data.

during each 30-s window of an HSRL-2 data point. Above the aircraft, over 65% (4546 out of 6988) of cloud mask fraction are greater than zero (Figure 7a). It is interesting that once the fraction is above zero the mean and median AOD biases are almost invariant, which indicates that the bias is possibly lagging behind the cloud influence. Most of the outliers occur when the cloud mask is on for over 80% of time. There is only one year of below-aircraft cloud mask data but it includes other properties such as dynamic cloud fraction which also considers the influence from glint. Below the aircraft, around 34% (724 out of 2156) of cloud mask fraction are greater than zero (Figure 7b). The mean and median biases are again similar across a wide range of cloud mask fraction groups. Higher cloud mask fraction is associated with higher below-aircraft cloud cover (Figure 7c). This implies that there is not a linear relationship between the cloud cover and bias.

For aerosol impact, first it is helpful to be aware of the diversity of aerosols encountered during the field campaign because all eight aerosol types are detected using the HSRL aerosol classification scheme (Figure 8a). Pure dust and ice are scarcely identified but all other types occur around 50–60% of time. The total bin thickness for each aerosol type in each column is estimated by the product of the number of bin associated with that aerosol type and the thickness of each bin ($\sim$15 m). The bin thickness among different aerosol types varies a lot. The whiskers of urban and dusty mix are more than 2500 m. The median of other aerosols is usually below 200 m. The vertical aerosol variability can be very complicated so we computed the number of aerosol types in each column, which can be viewed as the *minimum* number of aerosol layers in each column. Over 73% of time there are more than two aerosol types (Figure 8b), which can potentially invalidate one of the RSP aerosol model assumptions. Note that zero type occurs because the classification scheme occasionally fails to distinguish and identifies it as ambiguous type which is not counted. The median bias of all groups with different number of aerosol type is similar but the bias gets more dispersed when the number of aerosol type increases.
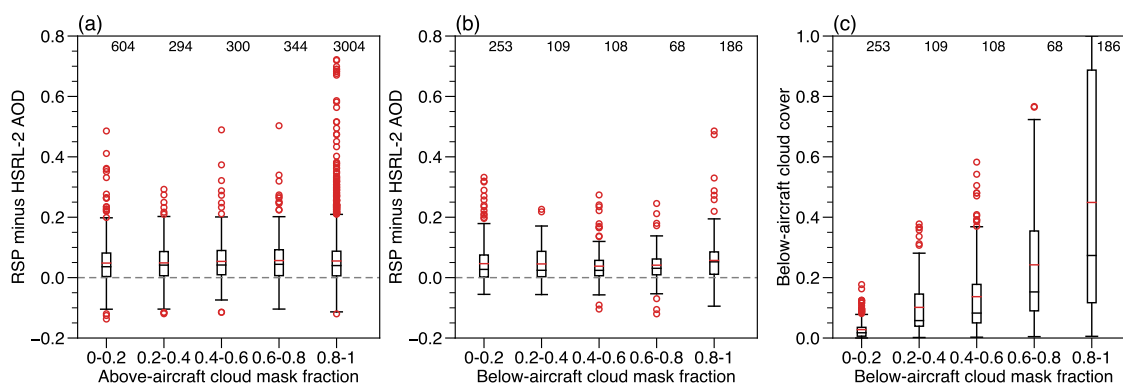
Figure 7. Cloud impact on the bias between RSP and HSRL-2. (a) Bias grouped by above-aircraft cloud mask fraction. (b) Bias grouped by below-aircraft cloud mask fraction. (c) Cloud cover grouped by below-aircraft cloud mask fraction. Numbers at the top indicate the number of each data group. The box and whiskers indicates the variability of each group (box, first and third quartiles; black line in box, median; red line in box, mean; whiskers, a distance of 1.5×IQR beyond the first and third quartiles; red circles, outliers).

The presence of absorbing aerosols may be indicated from the presence of four HSRL-2 aerosol types: smoke, fresh smoke, pure dust, and dusty mix (Figure 8a). Practically dusty mix is a mixture of maritime and pollution aerosols (Burton et al., 2012). Another way is to examine the RSP fine-mode single-scattering albedo (SSA) at 532 nm, which is defined as the ratio of scattering coefficient and extinction coefficient. The larger the absorption efficiency, the more SSA departs from one. An SSA value from 0.90 to almost 1 represents a weakly to moderately absorbing aerosol (Stamnes et al., 2018). It is clear that the minimum mean bias occurs when SSA is close to 1 and the mean bias gradually increases when SSA decreases (Figure 8c). Around 47% of SSA is below 0.90, which gives a similar result compared to HSRL-2 absorbing aerosols (dusty mix accounts for around 53% of time).

The presence of non-spherical aerosols may be indicated from the presence of three HSRL-2 aerosol types: ice, pure dust, and dusty mix (Figure 8a). All these aerosol types are characterized by a high degree of aerosol depolarization although there is not a clear-cut threshold for all aerosol types. Schlosser et al. (2022) deemed 0.13 a suitable threshold for the maritime study region of ACTIVATE. Burton et al. (2013) found that the first quartiles of depolarization for ice, pure dust, dusty mix are 0.23, 0.31, and 0.13, respectively, from measurements collected in 109 flights during 2006 and 2012. The HSRL-2 column-maximum depolarization is more than 0.1 for over 70% of measurements (Figure 8d). However, unless the column-maximum depolarization is less than 0.05, all other data groups show a large bias range.

As part of marine aerosols, sea salt is another potential source for contributing to high depolarization in the marine boundary layer but it is complicated by the fact that the shape of a sea salt depends not only on the magnitude but also the history of ambient relative humidity (Haarig et al., 2017). Ferrare et al. (2023) reported that sea salt may cause high depolarization within
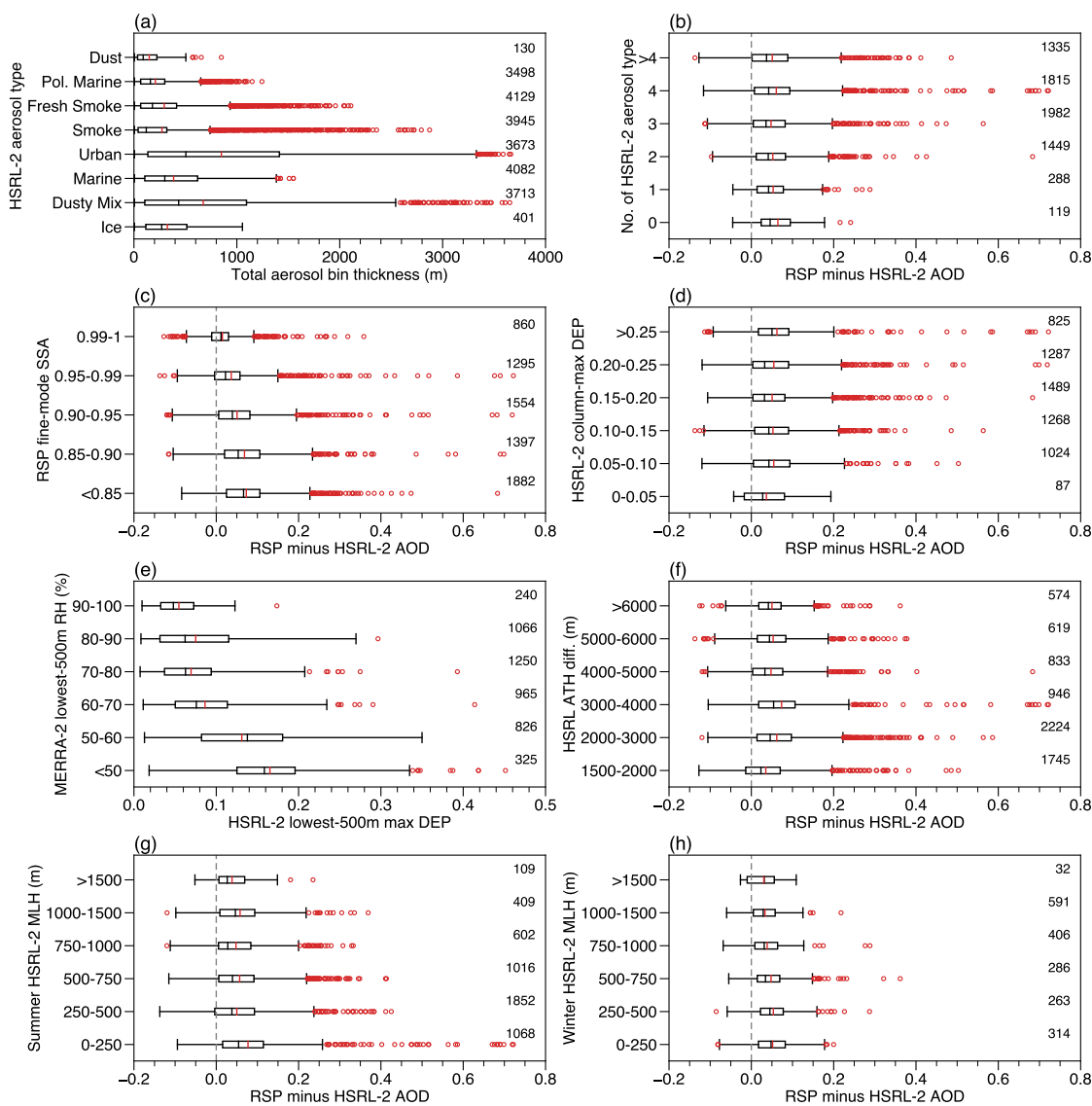
Figure 8. Same as Figure 7 but for aerosol impact. (a) Average HSRL-2 aerosol column total thickness grouped by HSRL-2 aerosol type. (b) Bias grouped by the number of HSRL-2 aerosol types. (c) Bias grouped by RSP fine-mode single-scattering albedo at 532 nm. (d) The HSRL-2 maximum depolarization ratio at 532 nm at the lowest 500 m grouped by the collocated MERRA-2 relative humidity (RH). (e) Bias grouped by the difference of aircraft altitude and HSRL-2 (effective) aerosol top height. (f) Bias grouped by HSRL-2 column-maximum depolarization ratio at 532 nm. (g) Bias grouped by HSRL-2 mixed layer height during summer deployments. (h) Bias grouped by HSRL-2 mixed layer height during winter deployments. Numbers on the right indicate the number of each data group.

the boundary layer for around one third of flights during the first two years of ACTIVATE based on HSRL-2 and dropsonde measurements. Here we extend the result by collocating the maximum HSRL-2 depolarization at the lowest 500 m with the interpolated MERRA-2 relative humidity for the whole campaign period (Figure 8e). Seethala et al. (2021) showed that despite

310   having a rather coarse spatial resolution, MERRA-2 adequately reconstructs the observed lower tropospheric thermodynamic structure during ACITVATE, especially the relative humidity and specific humidity fields. Median depolarization gradually increases when relative humidity decreases, suggesting that sea salt may change its shape over a wide range of relative humidity due to both deliquescence and efflorescence.

The high depolarization associated with sea salt provides evidence of the existence of aerosols close to the surface because

315   sea surface is the major source of sea salt in marine boundary layer. Conversely, it is more difficult to detect aerosols near and above the aircraft. The near-range limit for all ACTIVATE HSRL-2 extinction profiles is 1500 m so that the transmitter-to-receiver overlap function is close to one (Burton et al., 2012; Hair et al., 2008). To detect the possibility of aerosols within the near range of aircraft, we compute an effective HSRL-2 aerosol top height h_t, which is analogous to the RSP MAPP aerosol top height (Stamnes et al., 2018):

$$\int_0^{h_t} \alpha dz = 0.95 \int_0^h \alpha dz, \tag{15}$$

320   where α is extinction coefficient and h is aircraft altitude.

The aerosol top height difference can then be defined as the difference between the aircraft altitude and effective aerosol top height. Obviously, the difference is at least 1500 m due to the near-range limit (Figure 8f). A small difference indicates that the aerosol profile is top heavy, but the result for 1500–2000 m does not show larger bias compared to other groups. Note also that the dispersion of bias is much smaller when the effective top height is very low.

325   Aerosols above the aircraft can be estimated from the CALIPSO stratospheric AOD (Figure 9). Climatologically, the mean stratospheric AOD is around 0.011 and swings between 0.008 and 0.013 throughout a year. During ACTIVATE, the first winter deployment in 2020 experienced the record-high AOD for February and March and the second summer deployment in 2021 experienced the record-high AOD for May. A major source of stratospheric AOD perturbation is volcanic eruption (Kremser et al., 2016). The peak AOD in 2020 is very likely due to the stratospheric eruption at Raikoke (48° N, 153° E) in June 2019

330   (Kloss et al., 2021). Subsequently the stratospheric AOD increased up to 0.027 in October 2019 (not shown) and gradually decreased in the ensuing months. The spike in May 2021 may be due to the eruption at La Soufrière (13° N, 61° W) in April.

Other than the single homogeneous aerosol layer assumption, the assumption of coarse-mode sea salt aerosol being located between the surface and 1 km may also lead to retrieval uncertainties. During summer deployments, the mean HSRL-2 mixed layer height (MLH) is around 534 m. The bias is much larger when the RSP coarse-mode aerosol layer height (assumed to be 1

335   km) is much higher than the HSRL-2 MLH (Figure 8g). During winter deployments, the mean HSRL-2 MLH (around 764 m) is much closer to the RSP coarse-mode aerosol layer height but the large bias prevails when there is big mismatch between two heights (Figure 8h). Note that on average the coarse-mode RSP AOD accounts around 14% of total RSP AOD so the overall influence is likely to be tiny.
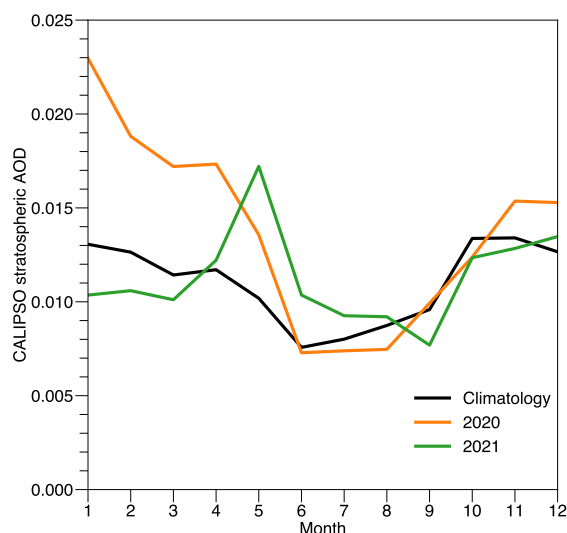
Figure 9. Annual cycle of CALIPSO stratospheric AOD centered at 37.5°N and 70°W. The 2006-2021 climatology is drawn in black and all available ACTIVATE years are drawn in other colors.

It is beneficial to consolidate the result by revisiting all factors: (i) There is no direct evidence for the presence of above-aircraft aerosols but elevated stratospheric AOD is seen at least in 2020 and 2021, but the overall impact should be small. (iii) There is no direct evidence for aerosols near aircraft. (ii,iv–viii) There is evidence for the presence of above-aircraft clouds, clouds surrounding the aircraft, sea salt near the sea surface, more than two aerosol layers, absorbing aerosols, and non-spherical aerosols. (ix) There is evidence for simplified and unwarranted assumptions, but the impact may be small.

5 Conclusions

In this paper we analyzed 3 years of AOD data over the western North Atlantic Ocean and attempted to estimate and understand their uncertainties, using two retrieval products collected during ACTIVATE and one satellite retrieval product. The key findings are as follows:

1. RSP and HSRL-2 AOD have large seasonal variations but also exhibit considerable deviations between the two retrievals. The absence of a reference dataset over the western North Atlantic Ocean, and arguably all global marine areas, prompts us to use triple collocation to determine which dataset is more accurate.

2. Triple collocation analysis indicates that HSRL-2 is the most accurate dataset over the study region. The correlation coefficient and error standard deviation of AOD with respect to the expected ground truth are 0.93 and 0.027, respectively.

3. The RSP retrieval quality can be improved considerably by applying a simple, yet efficient, filtering criterion: a more stringent normalized cost function of 0.05. There are two additional benefits. First, it can be easily applied globally. Second, it is suitable for offline post-processing.

4. The reasons of disagreement between RSP and HSRL-2 are diagnosed using RSP, HSRL-2 and other ACTIVATE datasets. There is a basket of potential factors, including cloud contamination, aerosols near surface, multiple aerosol layers, absorbing aerosols, non-spherical aerosols, and simplified retrieval assumptions. It proves that the quality of AOD retrievals can be optimally improved by combining information from both lidar and polarimeter.

Motivated by the last point, future efforts are needed along two lines: multi-angle polarimeter and HSRL (or HSRL-2) data from other field campaigns should be analyzed to test the robustness of the conclusions here; and a more efficient retrieval of HSRL-2 MLH data should be developed potentially through two-dimensional image processing (rather than the current one-dimensional retrieval at each horizontal point along the flight path). Earlier recommendation such as combining HSRL-2 elevated aerosol layer altitudes to RSP retrievals should be reconsidered (Knobelspiesse et al., 2011). Above results will provide a better constraint on aerosol properties and help prepare concurrent deployments of lidar and polarimeter for future spaceborne missions.

References

Aldhaif, A. M., Lopez, D. H., Dadashazar, H., and Sorooshian, A.: Sources, frequency, and chemical nature of dust events impacting the United States East Coast, Atmospheric Environment, 231, 117 456, https://doi.org/10.1016/j.atmosenv.2020.117456, 2020.

Alexandrov, M. D., Cairns, B., Emde, C., Ackerman, A. S., and van Diedenhoven, B.: Accuracy assessments of cloud droplet size retrievals from polarized reflectance measurements by the Research Scanning Polarimeter, Remote Sensing of Environment, 125, 92–111, https://doi.org/10.1016/j.rse.2012.07.012, 2012.

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., and Holmén, K.: Mesoscale variations of tropospheric aerosols, Journal of the Atmospheric Sciences, 60, 119–136, https://doi.org/10.1175/1520-0469(2003)060<0119:mvota>2.0.co;2, 2003.

Andreae, M. O.: Correlation between cloud condensation nuclei concentration and aerosol optical thickness in remote and polluted regions, Atmospheric Chemistry and Physics, 9, 543–556, https://doi.org/10.5194/acp-9-543-2009, 2009.

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.-L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T., McCoy, D. T., Myhre, G., Mülmenstädt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, M., Schwartz, S. E., Sourdeval, O., Storelvmo, T., Toll, V., Winker, D., and Stevens, B.: Bounding global aerosol radiative forcing of climate change, Reviews of Geophysics, 58, https://doi.org/10.1029/2019rg000660, 2020.

Burton, S. P., Ferrare, R. A., Hostetler, C. A., Hair, J. W., Rogers, R. R., Obland, M. D., Butler, C. F., Cook, A. L., Harper, D. B., and Froyd, K. D.: Aerosol classification using airborne High Spectral Resolution Lidar measurements – methodology and examples, Atmospheric Measurement Techniques, 5, 73–98, https://doi.org/10.5194/amt-5-73-2012, 2012.

Burton, S. P., Ferrare, R. A., Vaughan, M. A., Omar, A. H., Rogers, R. R., Hostetler, C. A., and Hair, J. W.: Aerosol classification from airborne HSRL and comparisons with the CALIPSO vertical feature mask, Atmospheric Measurement Techniques, 6, 1397–1412, https://doi.org/10.5194/amt-6-1397-2013, 2013.

Burton, S. P., Chemyakin, E., Liu, X., Knobelspiesse, K., Stamnes, S., Sawamura, P., Moore, R. H., Hostetler, C. A., and Ferrare, R. A.: Information content and sensitivity of the $3\beta + 2\alpha$ lidar measurement system for aerosol microphysical retrievals, Atmospheric Measurement Techniques, 9, 5555–5574, https://doi.org/10.5194/amt-9-5555-2016, 2016.

Burton, S. P., Hostetler, C. A., Cook, A. L., Hair, J. W., Seaman, S. T., Scola, S., Harper, D. B., Smith, J. A., Fenn, M. A., Ferrare, R. A., Saide, P. E., Chemyakin, E. V., and Müller, D.: Calibration of a high spectral resolution lidar using a Michelson interferometer, with data examples from ORACLES, Applied Optics, 57, 6061, https://doi.org/10.1364/ao.57.006061, 2018.

Caires, S. and Sterl, A.: Validation of ocean wind and wave data using triple collocation, Journal of Geophysical Research, 108, https://doi.org/10.1029/2002jc001491, 2003.

Cairns, B., Russell, E. E., and Travis, L. D.: Research Scanning Polarimeter: calibration and ground-based measurements, in: Proceedings of SPIE, edited by Goldstein, D. H. and Chenault, D. B., vol. 3754, pp. 186–196, SPIE, https://doi.org/10.1117/12.366329, 1999.

Cairns, B., Russell, E. E., LaVeigne, J. D., and Tennant, P. M. W.: Research Scanning Polarimeter and airborne usage for remote sensing of aerosols, in: Proceedings of SPIE, edited by Shaw, J. A. and Tyo, J. S., vol. 5158, pp. 33–44, SPIE, https://doi.org/10.1117/12.518320, 2003.

Chen, Y.-C., Christensen, M. W., Stephens, G. L., and Seinfeld, J. H.: Satellite-based estimate of global aerosol–cloud radiative forcing by marine warm clouds, Nature Geoscience, 7, 643–646, https://doi.org/10.1038/ngeo2214, 2014.

Chylek, P., Henderson, B., and Mishchenko, M.: Aerosol radiative forcing and the accuracy of satellite aerosol optical depth retrieval, Journal of Geophysical Research: Atmospheres, 108, 1–8, https://doi.org/10.1029/2003jd004044, 2003.

Corral, A. F., Braun, R. A., Cairns, B., Gorooh, V. A., Liu, H., Ma, L., Mardi, A. H., Painemal, D., Stamnes, S., van Diedenhoven, B., Wang, H., Yang, Y., Zhang, B., and Sorooshian, A.: An overview of atmospheric features over the Western North Atlantic Ocean and North American East Coast – Part 1: Analysis of aerosols, gases, and wet deposition chemistry, Journal of Geophysical Research: Atmospheres, 126, e2020JD32 592, https://doi.org/10.1029/2020jd032592, 2021.

Draper, C., Reichle, R., de Jeu, R., Naeimi, V., Parinussa, R., and Wagner, W.: Estimating root mean square errors in remotely sensed soil moisture over continental scale domains, Remote Sensing of Environment, 137, 288–298, https://doi.org/10.1016/j.rse.2013.06.013, 2013.

Ferrare, R., Hair, J., Hostetler, C., Shingler, T., Burton, S. P., Fenn, M., Clayton, M., Scarino, A. J., Harper, D., Seaman, S., Cook, A., Crosbie, E., Winstead, E., Ziemba, L., Thornhill, L., Robinson, C., Moore, R., Vaughan, M., Sorooshian, A., Schlosser, J. S., Liu, H., Zhang, B., Diskin, G., DiGangi, J., Nowak, J., Choi, Y., Zuidema, P., and Chellappan, S.: Airborne HSRL-2 measurements of elevated aerosol depolarization associated with non-spherical sea salt, Frontiers in Remote Sensing, 4, 1–18, https://doi.org/10.3389/frsen.2023.1143944, 2023.

Gao, M., Knobelspiesse, K., Franz, B. A., Zhai, P.-W., Sayer, A. M., Ibrahim, A., Cairns, B., Hasekamp, O., Hu, Y., Martins, V., Werdell, P. J., and Xu, X.: Effective uncertainty quantification for multi-angle polarimetric aerosol remote sensing over ocean, Atmospheric Measurement Techniques, 15, 4859–4879, https://doi.org/10.5194/amt-15-4859-2022, 2022.

Gauch, H. G., Hwang, J. T. G., and Fick, G. W.: Model evaluation by comparison of model-based predictions and measured values, Agronomy Journal, 95, 1442–1446, https://doi.org/10.2134/agronj2003.1442, 2003.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, International Journal of Applied Earth Observation and Geoinformation, 45, 200–211, https://doi.org/10.1016/j.jag.2015.09.002, 2016.

Haarig, M., Ansmann, A., Gasteiger, J., Kandler, K., Althausen, D., Baars, H., Radenz, M., and Farrell, D. A.: Dry versus wet marine particle optical properties: RH dependence of depolarization ratio, backscatter, and extinction from multiwavelength lidar measurements during SALTRACE, Atmospheric Chemistry and Physics, 17, 14 199–14 217, https://doi.org/10.5194/acp-17-14199-2017, 2017.

Hair, J. W., Hostetler, C. A., Cook, A. L., Harper, D. B., Ferrare, R. A., Mack, T. L., Welch, W., Izquierdo, L. R., and Hovis, F. E.: Airborne High Spectral Resolution Lidar for profiling aerosol optical properties, Applied Optics, 47, 6734, https://doi.org/10.1364/ao.47.006734, 2008.

Hansen, J., Rossow, W., Carlson, B., Lacis, A., Travis, L., Del Genio, A., Fung, I., Cairns, B., Mishchenko, M., and Sato, M.: Low-cost long-term monitoring of global climate forcings and feedbacks, Climatic Change, 31, 247–271, https://doi.org/10.1007/BF01095149, 1995.

Holben, B., Eck, T., Slutsker, I., Tanré, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET–A federated instrument network and data archive for aerosol characterization, Remote Sensing of Environment, 66, 1–16, https://doi.org/10.1016/S0034-4257(98)00031-5, 1998.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom, https://doi.org/10.1017/9781009157896, 2023.

Janssen, P. A. E. M., Abdalla, S., Hersbach, H., and Bidlot, J.-R.: Error estimation of buoy, satellite, and model wave height data, Journal of Atmospheric and Oceanic Technology, 24, 1665–1677, https://doi.org/10.1175/jtech2069.1, 2007.

Kar, J., Lee, K.-P., Vaughan, M. A., Tackett, J. L., Trepte, C. R., Winker, D. M., Lucker, P. L., and Getzewich, B. J.: CALIPSO level 3 stratospheric aerosol profile product: Version 1.00 algorithm description and initial assessment, Atmospheric Measurement Techniques, 12, 6173–6191, https://doi.org/10.5194/amt-12-6173-2019, 2019.

Kloss, C., Berthet, G., Sellitto, P., Ploeger, F., Taha, G., Tidiga, M., Eremenko, M., Bossolasco, A., Jégou, F., Renard, J.-B., and Legras, B.: Stratospheric aerosol layer perturbation caused by the 2019 Raikoke and Ulawun eruptions and their radiative forcing, Atmospheric Chemistry and Physics, 21, 535–560, https://doi.org/10.5194/acp-21-535-2021, 2021.

Knobelspiesse, K., Cairns, B., Ottaviani, M., Ferrare, R., Hair, J., Hostetler, C., Obland, M., Rogers, R., Redemann, J., Shinozuka, Y., Clarke, A., Freitag, S., Howell, S., Kapustin, V., and McNaughton, C.: Combined retrievals of boreal forest fire aerosol properties with a polarimeter and lidar, Atmospheric Chemistry and Physics, 11, 7045–7067, https://doi.org/10.5194/acp-11-7045-2011, 2011.

Kremser, S., Thomason, L. W., von Hobe, M., Hermann, M., Deshler, T., Timmreck, C., Toohey, M., Stenke, A., Schwarz, J. P., Weigel, R., Fueglistaler, S., Prata, F. J., Vernier, J.-P., Schlager, H., Barnes, J. E., Antuña-Marrero, J.-C., Fairlie, D., Palm, M., Mahieu, E., Notholt, J., Rex, M., Bingen, C., Vanhellemont, F., Bourassa, A., Plane, J. M. C., Klocke, D., Carn, S. A., Clarisse, L., Trickl, T., Neely, R., James, A. D., Rieger, L., Wilson, J. C., and Meland, B.: Stratospheric aerosol–Observations, processes, and impact on climate, Reviews of Geophysics, 54, 278–335, https://doi.org/10.1002/2015rg000511, 2016.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989–3034, https://doi.org/10.5194/amt-6-2989-2013, 2013.

Lohmann, U. and Feichter, J.: Global indirect aerosol effects: A review, Atmospheric Chemistry and Physics, 5, 715–737, https://doi.org/10.5194/acp-5-715-2005, 2005.

Mardi, A. H., Dadashazar, H., Painemal, D., Shingler, T., Seaman, S. T., Fenn, M. A., Hostetler, C. A., and Sorooshian, A.: Biomass burning over the United States East Coast and western North Atlantic Ocean: Implications for clouds and air quality, Journal of Geophysical Research: Atmospheres, 126, https://doi.org/10.1029/2021jd034916, 2021.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophysical Research Letters, 41, 6229–6236, https://doi.org/10.1002/2014gl061322, 2014.

Mishchenko, M. I., Cairns, B., Hansen, J. E., Travis, L. D., Burg, R., Kaufman, Y. J., Vanderlei Martins, J., and Shettle, E. P.: Monitoring of aerosol forcing of climate from space: Analysis of measurement requirements, Journal of Quantitative Spectroscopy and Radiative Transfer, 88, 149–161, https://doi.org/https://doi.org/10.1016/j.jqsrt.2004.03.030, 2004.

MODIS Science Team: MODIS/Terra Aerosol 5-Min L2 Swath 3km, https://doi.org/10.5067/MODIS/MOD04_3K.061, 2017a.

MODIS Science Team: MODIS/Aqua Aerosol 5-Min L2 Swath 3km, https://doi.org/10.5067/MODIS/MYD04_3K.061, 2017b.

NASA/LARC/SD/ASDC: CALIPSO Lidar Level 3 Stratospheric Aerosol Profiles Standard V1-00, https://doi.org/10.5067/CALIOP/CALIPSO/LID_L3_Stratospheric_APro-Standard-V1-00, 2018.

NASA/LARC/SD/ASDC: CALIPSO Lidar Level 3 Stratospheric Aerosol Profiles Standard V1-01, https://doi.org/10.5067/CALIOP/CALIPSO/CAL_LID_L3_Stratospheric_APro-Standard-V1-01, 2020.

NASA/LARC/SD/ASDC: Aerosol Cloud meTeorology Interactions oVer the western ATlantic Experiment (ACTIVATE), https://doi.org/10.5067/SUBORBITAL/ACTIVATE/DATA001, 2021.

Nied, J., Jones, M., Seaman, S., Shingler, T., Hair, J., Cairns, B., Gilst, D. V., Bucholtz, A., Schmidt, S., Chellappan, S., Zuidema, P., Diedenhoven, B. V., Sorooshian, A., and Stamnes, S.: A cloud detection neural network for above-aircraft clouds using airborne cameras, Frontiers in Remote Sensing, 4, https://doi.org/10.3389/frsen.2023.1118745, 2023.

O'Carroll, A. G., Eyre, J. R., and Saunders, R. W.: Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations, Journal of Atmospheric and Oceanic Technology, 25, 1197–1207, https://doi.org/10.1175/2007jtecho542.1, 2008.

Pöschl, U.: Atmospheric aerosols: Composition, transformation, climate and health effects, Angewandte Chemie International Edition, 44, 7520–7540, https://doi.org/10.1002/anie.200501122, 2005.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS aerosol algorithm, products, and validation, Journal of the Atmospheric Sciences, 62, 947–973, https://doi.org/10.1175/jas3385.1, 2005.

Remer, L. A., Mattoo, S., Levy, R. C., and Munchak, L. A.: MODIS 3 km aerosol product: Algorithm and global perspective, Atmospheric Measurement Techniques, 6, 1829–1844, https://doi.org/10.5194/amt-6-1829-2013, 2013.

Sawamura, P., Moore, R. H., Burton, S. P., Chemyakin, E., Müller, D., Kolgotin, A., Ferrare, R. A., Hostetler, C. A., Ziemba, L. D., Beyersdorf, A. J., and Anderson, B. E.: HSRL-2 aerosol optical measurements and microphysical retrievals vs. airborne in situ measurements during DISCOVER-AQ 2013: An intercomparison study, Atmospheric Chemistry and Physics, 17, 7229–7243, https://doi.org/10.5194/acp-17-7229-2017, 2017.

Sayer, A. M., Govaerts, Y., Kolmonen, P., Lipponen, A., Luffarelli, M., Mielonen, T., Patadia, F., Popp, T., Povey, A. C., Stebel, K., and Witek, M. L.: A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing, Atmospheric Measurement Techniques, 13, 373–404, https://doi.org/10.5194/amt-13-373-2020, 2020.

Scarino, A. J., Obland, M. D., Fast, J. D., Burton, S. P., Ferrare, R. A., Hostetler, C. A., Berg, L. K., Lefer, B., Haman, C., Hair, J. W., Rogers, R. R., Butler, C., Cook, A. L., and Harper, D. B.: Comparison of mixed layer heights from airborne high spectral resolution lidar, ground-based measurements, and the WRF-Chem model during CalNex and CARES, Atmospheric Chemistry and Physics, 14, 5547–5560, https://doi.org/10.5194/acp-14-5547-2014, 2014.

Schlosser, J. S., Stamnes, S., Burton, S. P., Cairns, B., Crosbie, E., Van Diedenhoven, B., Diskin, G., Dmitrovic, S., Ferrare, R., Hair, J. W., Hostetler, C. A., Hu, Y., Liu, X., Moore, R. H., Shingler, T., Shook, M. A., Thornhill, K. L., Winstead, E., Ziemba, L., and Sorooshian, A.: Polarimeter + lidar–derived aerosol particle number concentration, Frontiers in Remote Sensing, 3, 1–13, https://doi.org/10.3389/frsen.2022.885332, 2022.

Seethala, C., Zuidema, P., Edson, J., Brunke, M., Chen, G., Li, X.-Y., Painemal, D., Robinson, C., Shingler, T., Shook, M., Sorooshian, A., Thornhill, L., Tornow, F., Wang, H., Zeng, X., and Ziemba, L.: On assessing ERA5 and MERRA2 representations of cold-air outbreaks across the Gulf Stream, Geophysical Research Letters, 48, https://doi.org/10.1029/2021gl094364, 2021.

Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., Feingold, G., Ghan, S., Guenther, A. B., Kahn, R., Kraucunas, I., Kreidenweis, S. M., Molina, M. J., Nenes, A., Penner, J. E., Prather, K. A., Ramanathan, V., Ramaswamy, V., Rasch, P. J., Ravishankara, A. R., Rosenfeld, D., Stephens, G., and Wood, R.: Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system, Proceedings of the National Academy of Sciences, 113, 5781–5790, https://doi.org/10.1073/pnas.1514043113, 2016.

Shinozuka, Y., Johnson, R. R., Flynn, C. J., Russell, P. B., Schmid, B., Redemann, J., Dunagan, S. E., Kluzek, C. D., Hubbe, J. M., Segal-
530 Rosenheimer, M., Livingston, J. M., Eck, T. F., Wagener, R., Gregory, L., Chand, D., Berg, L. K., Rogers, R. R., Ferrare, R. A., Hair,
 J. W., Hostetler, C. A., and Burton, S. P.: Hyperspectral aerosol optical depths from TCAP flights, Journal of Geophysical Research:
 Atmospheres, 118, 12,180–12,194, https://doi.org/10.1002/2013jd020596, 2013.

Sorooshian, A., Anderson, B., Bauer, S. E., Braun, R. A., Cairns, B., Crosbie, E., Dadashazar, H., Diskin, G., Ferrare, R., Flagan, R. C.,
 Hair, J., Hostetler, C., Jonsson, H. H., Kleb, M. M., Liu, H., MacDonald, A. B., McComiskey, A., Moore, R., Painemal, D., Russell,
535 L. M., Seinfeld, J. H., Shook, M., Smith, W. L., Thornhill, K., Tselioudis, G., Wang, H., Zeng, X., Zhang, B., Ziemba, L., and Zuidema,
 P.: Aerosol–cloud–meteorology interaction airborne field investigations: Using lessons learned from the U.S. West Coast in the design of
 ACTIVATE off the U.S. East Coast, Bulletin of the American Meteorological Society, 100, 1511–1528, https://doi.org/10.1175/bams-d-
 18-0100.1, 2019.

Sorooshian, A., Alexandrov, M. D., Bell, A. D., Bennett, R., Betito, G., Burton, S. P., Buzanowicz, M. E., Cairns, B., Chemyakin, E. V., Chen,
540 G., Choi, Y., Collister, B. L., Cook, A. L., Corral, A. F., Crosbie, E. C., van Diedenhoven, B., DiGangi, J. P., Diskin, G. S., Dmitrovic,
 S., Edwards, E.-L., Fenn, M. A., Ferrare, R. A., van Gilst, D., Hair, J. W., Harper, D. B., Hilario, M. R. A., Hostetler, C. A., Jester, N.,
 Jones, M., Kirschler, S., Kleb, M. M., Kusterer, J. M., Leavor, S., Lee, J. W., Liu, H., McCauley, K., Moore, R. H., Nied, J., Notari, A.,
 Nowak, J. B., Painemal, D., Phillips, K. E., Robinson, C. E., Scarino, A. J., Schlosser, J. S., Seaman, S. T., Seethala, C., Shingler, T. J.,
 Shook, M. A., Sinclair, K. A., Smith Jr., W. L., Spangenberg, D. A., Stamnes, S. A., Thornhill, K. L., Voigt, C., Vömel, H., Wasilewski,
545 A. P., Wang, H., Winstead, E. L., Zeider, K., Zeng, X., Zhang, B., Ziemba, L. D., and Zuidema, P.: Spatially coordinated airborne data and
 complementary products for aerosol, gas, cloud, and meteorological studies: the NASA ACTIVATE dataset, Earth System Science Data,
 15, 3419–3472, https://doi.org/10.5194/essd-15-3419-2023, 2023.

Stamnes, S., Hostetler, C., Ferrare, R., Burton, S., Liu, X., Hair, J., Hu, Y., Wasilewski, A., Martin, W., van Diedenhoven, B., Chowdhary,
 J., Cetinić, I., Berg, L. K., Stamnes, K., and Cairns, B.: Simultaneous polarimeter retrievals of microphysical aerosol and ocean color
550 parameters from the "MAPP" algorithm with comparison to high-spectral-resolution lidar aerosol and ocean products, Applied Optics,
 57, 2394, https://doi.org/10.1364/ao.57.002394, 2018.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, Journal of Geophysical
 Research: Oceans, 103, 7755–7766, https://doi.org/10.1029/97jc03180, 1998.

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H., and Young, S. A.: Overview of the
555 CALIPSO mission and CALIOP data processing algorithms, Journal of Atmospheric and Oceanic Technology, 26, 2310–2323,
 https://doi.org/10.1175/2009jtecha1281.1, 2009.

Zwieback, S., Scipal, K., Dorigo, W., and Wagner, W.: Structural and statistical properties of the collocation technique for error characteri-
 zation, Nonlinear Processes in Geophysics, 19, 69–80, https://doi.org/10.5194/npg-19-69-2012, 2012.