

Article

Retrievals of Biomass Burning Aerosol and Liquid Cloud Properties from Polarimetric Observations Using Deep Learning Techniques

Michal Segal Rozenhaimer ^{1,2,*}, Kirk Knobelspiesse ³, Daniel Miller ^{3,4} and Dmitry Batenkov ^{5,6} 

¹ Bay Area Environmental Research Institute, Mountain View, CA 94035, USA

² NASA Ames Research Center, Mountain View, CA 94035, USA

³ NASA Goddard Space Flight Center, Earth Science Division, Greenbelt, MD 20771, USA; kirk.d.knobelspiesse@nasa.gov (K.K.); daniel.j.miller@nasa.gov (D.M.)

⁴ GESTAR, University of Maryland-Baltimore County, Baltimore, MD 21250, USA

⁵ Basis Research Institute, New York City, NY 10026, USA; dima@basis.ai

⁶ Department of Applied Mathematics, Tel Aviv University, P.O. Box 39040, Tel Aviv 6997801, Israel

* Correspondence: michal.segalrozenhaimer@gmail.com

Abstract: Biomass burning (BB) aerosols are the largest source of absorbing aerosols on Earth. Coupled with marine stratocumulus clouds (MSC), their radiative effects are enhanced and can cause cloud property changes (first indirect effect) or cloud burn-off and warm up the atmospheric column (semi-direct effect). Nevertheless, the derivation of their quantity and optical properties in the presence of MSC clouds is confounded by the uncertainties in the retrieval of the underlying cloud properties. Therefore, a robust methodology is needed for the coupled retrievals of absorbing aerosol above clouds. Here, we present a new retrieval approach implemented for a Spectro radiometric multi-angle polarimetric airborne platform, the research scanning polarimeter (RSP), during the ORACLES campaign over the Southeast Atlantic Ocean. Our approach transforms the 1D measurements over multiple angles and wavelengths into a 3D image-like input, which is then processed using various deep learning (DL) schemes to yield aerosol single scattering albedos (SSAs), aerosol optical depths (AODs), aerosol effective radii, and aerosol complex refractive indices, together with cloud optical depths (CODs), cloud effective radii and variances. We present a comparison between the different DL approaches, as well as their comparison to existing algorithms. We discover that the Vision Transformer (ViT) scheme, traditionally used by natural language models, is superior to the ResNet convolutional Neural-Network (CNN) approach. We show good validation statistics on synthetic and real airborne data and discuss paths forward for making this approach flexible and readily applicable over multiple platforms.

Keywords: biomass burning aerosol; convolutional neural networks; vision transformers; polarimetry; ORACLES



Academic Editor: Xiaoyang Zhang

Received: 12 February 2025

Revised: 5 May 2025

Accepted: 9 May 2025

Published: 12 May 2025

Citation: Segal Rozenhaimer, M.; Knobelspiesse, K.; Miller, D.; Batenkov, D. Retrievals of Biomass Burning Aerosol and Liquid Cloud Properties from Polarimetric Observations Using Deep Learning Techniques. *Remote Sens.* **2025**, *17*, 1693. <https://doi.org/10.3390/rs17101693>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Southeast Atlantic (SEA) ocean encompasses some of the highest biomass burning (BB) optical depths on earth [1]. In conjunction with the semi-permanent marine stratocumulus (MSC) cloud deck, it forms one of the world's largest regions of above cloud aerosol (ACA) [2–4]. This unique formation is responsible for the high uncertainty in the predictions of the earth's radiative budget over this region [1,4–7]. From a remote sensing perspective, retrievals of absorbing aerosol properties overlaying MSC clouds from passive

spectro-radiometers, like MODIS or SEVIRI, often rely on the assumption of the correct underlying cloud albedo [8,9], which is a function of the MSC cellular convection cloud type [10]. Closed cell formations have higher albedo values than open cell formations [10], but this is not taken into account in the retrieval process [8]. Moreover, the aerosol single scattering albedo (which is a proxy for aerosol type) needs to be assumed a priori for the derivation of ACA and cloud properties. As this variable is not well constrained over oceans, it can generate uncertainties in the retrieved AOD values between 15 and 40% [11], sometimes even up to 100% [12]. Cloud optical thickness (COT) and effective radii (CER) retrieval uncertainty in high AOD scenes range between 5 and 20%, with higher uncertainties found at higher AODs [12]. This effect is partially caused by the reduced sensitivity of the standard two-wavelength lookup table (LUT) retrieval scheme, over a domain with increasing AOD amounts. Another significant factor is the need to assume an a priori aerosol model [11] in such scenes.

Polarimetric observations, on the other hand, have more merit in obtaining coupled retrievals of ACA and liquid cloud properties, as the information content about the aerosol and the cloud properties lies within different scattering angles of the measured polarized light. These measurements are used in addition to the information content available from the non-polarized measurements, which is obtained by instruments like MODIS. Knobelspiesse et al. [13] used multi-angle, multi-wavelength polarized measurements from the airborne Research Scanning Polarimeter (RSP) to derive absorbing aerosol properties overlying liquid clouds over the Gulf of Mexico. They used the information contained in the polarized reflectance at scattering angles around 142° to obtain the cloud droplet size distribution, assuming high (opaque to the surface) cloud optical depth. Then, the cloud properties were used as an input to an optimization scheme utilizing a radiative transfer model to determine the aerosol optical properties such as AOD, the fine mode of the aerosol size distribution, and its complex refractive index. This iterative procedure was also used to refine the retrieved cloud property (effective radius and variance of the cloud droplet size distribution). Initial guess values needed for the iterative process were taken from other instruments, either co-flying with RSP or from ground-based measurements. A recent extension to the former and other approaches, such as [14], was developed by [15] for the airborne AirMSPI (Airborne Multiangle Spectro-Polarimetric Imager) instrument flown over the SEA in 2016. In their algorithm, the initial cloud droplet effective distribution is derived over the entire image scene (80–100 km by 10–25 km), followed by a coupled retrieval of image-scale cloud and above-cloud aerosol properties fitting the polarimetric data at all observation angles. That algorithm differs from the former ones by its ability to retrieve pixel scale (25 m) cloud droplet size distribution parameters via establishing an image-specific relationship between COT and CER and refining, iteratively, the retrieved values. Nevertheless, aerosol properties in this case are assumed to be constant over the entire image scene (~100 km), and the pixel-scale cloud properties rely on the assumption that there is a smooth relationship between COT and CER over the cloud observed domain.

To date, several machine learning-based algorithms are available for retrievals of geophysical variables from polarimetric observations. Among them, multilayer perceptron (MLP) neural-networks (NNs) were used to retrieve global liquid cloud properties from POLDER-3 (Polarization and Directionality of Earth's Reflectances-3) [16] and from the RSP instrument during the ORACLES (ObseRvations of Aerosols above CLouds and their intEractionS) campaign over the SEA [17,18]. In addition, a combined NN and an iterative regularization scheme was used by DiNoia [19] to retrieve aerosol properties from the RSP instrument. In the latter, they used the NN result as an initial guess to an iterative Phillips–Tikhonov algorithm. One of the main drawbacks in running an iterative optimal estimation-based algorithm is the computation cost of the forward radiative transfer (RT)

simulation that is needed at each iteration. Therefore, recent work has been conducted to speed up the retrieval process of aerosol properties using optimal estimation. For example, Gao et al. [20] utilized an MLP NN as an emulator for generating the forward RT simulations that are used in the iterative process for the HARP-2 (Hyper Angular Rainbow Polarimeter-2) instrument on-board the PACE satellite. This allowed a large speed-up of the retrieval process in their coupled retrieval algorithm for aerosol and ocean color properties (FastMAPOL).

However, despite the expanding usage of machine learning approaches in polarimetric retrieval schemes, existing algorithms still rely on the optimal estimation (OE) approach [19,20], where ML is used as a first guess or as a radiative transfer emulator to speed up the OE process. Moreover, even the latest retrieval schemes have challenges in the retrievals of aerosol optical properties, especially single scattering albedos (SSAs) and aerosol effective radii [20]. NN-based cloud retrieval algorithms [16,18] which performed an end-to-end retrieval process, suffered convergence issues when trying to utilize the entire polarized measurement vector as an input. This vector consists of the total reflectance, polarized reflectance, and DoLP (Degree of Linear Polarization) signals measured at multiple angles and wavelengths, resulting in a very high dimensionality, which substantially affected the number of network parameters to be optimized. A possible solution to this issue was to perform a dimensionality reduction of the measurement vector before its introduction to the network, but such an approach has not always yielded optimal results [18]. The dimensionality reduction can reduce the network capability to generate higher order interactions, thus affecting its performance.

In this work, we take advantage of the development in deep learning (DL) approaches for image processing (e.g., [21–25]) and utilize their inherent capabilities to handle high multi-dimensionality inputs of multi-angle, multi-wavelength polarimetric measurements. Instead of ingesting the measurements as a vector input, as was conducted in previous retrieval works mentioned above, we treat the multi-dimensional measurements as an image, where the data at different angles and wavelengths, and the different measurement states (reflectance, polarized reflectance, and DoLP) represent a 3D image array with its different channels. This approach enabled us to utilize deep learning algorithms that are based on convolutional operations [21,26] for feature extraction, for the preservation of both the angular and the spectral relationships in the polarimetric measurement signal. Such an approach is popular when processing satellite imagery for classification or semantic segmentation tasks (e.g., [24,26,27]) but is less established for regression tasks of geophysical variable retrievals [28].

To achieve the goal of a coupled ACA and liquid cloud property retrievals, we utilize the multi-dimensional signal from the airborne research scanning polarimeter (RSP) during the ORACLES campaign, preserving the spectro-angular structure in the measurements. We implement two deep-learning techniques that utilize (a) the recently introduced Vision Transformer architecture, and (b) the convolutional approach, comparing the results with a standard MLP NN, benchmarking recent work on cloud retrievals from RSP [17,18]. The main contributions from this work include the following:

- Developing a new coupled retrieval approach for aerosol and clouds from polarimetric measurements that takes into account multi-dimensional inputs and their spectral and angular spatial relations.
- Achieving better or similar accuracy compared to existing algorithms in retrieving aerosol optical properties, while maintaining fast computation time.
- Providing a general algorithm concept that can be easily extended to platforms other than RSP and can be utilized for real-time retrievals from existing space-borne platforms such as HARP2 [29,30] on PACE [31,32].

This manuscript is organized in the following manner: Section 2 describes the construction of the synthetic dataset that was used for the development of the algorithm including network training and testing, and the RSP data from ORACLES that was used for our validation. Section 3 describes the three machine learning architectures that were used in this investigation. Section 4 provides the results from our synthetic simulations, comparing the three ML algorithms, and validation of the results with collocated airborne data from ORACLES. Lastly, in Section 5, we discuss the strengths and weaknesses of the new methodology, also in the context of other works and conclude with steps forward.

2. Data

2.1. Simulations

We generated a synthetic dataset of aerosol and cloud scenes, which was used to train our machine learning algorithms. Specifically, our simulations produced synthetic signals that represent measurements from the NASA Research Scanning Polarimeter (RSP [33]), focusing on conditions of absorbing (BB) aerosol over low-level liquid clouds, which were prevalent over SEA Ocean during the ORACLES field campaign in 2016–2018 [1]. The training set was simulated by a vector radiative transfer software using the doubling-adding method [34]. This method is ideal for optically thick scene simulations and can produce all potential sun and viewing geometries with almost no added computational expense. We generated the reflectance from the three Stokes vector elements describing linearly polarized light (I, Q, U), with respect to the incident solar irradiance as follows:

$$R_{I,Q,U} = (I, Q, U) * \frac{\pi r_0^2}{F_0 \cos(\theta_0)} \quad (1)$$

where R_I represents the total reflectance (including unpolarized and polarized light), and R_Q and R_U represent the Q and U Stokes' vector parameters to describe linearly polarized reflectance. F_0 is the top of atmosphere solar irradiance, r_0 is the Sun-Earth distance in astronomical units, and θ_0 is the solar zenith angle (SZA). Since the simulations are defined in the solar principal plane (the plane containing both the incident solar and observation viewing direction vectors), measurements of R_U are expected to be near zero. Therefore, this element was not saved as an output from the simulations. In addition to the total and the polarized reflectance, we generated the degree of linear polarization as an output from our simulations, defined by the positive ratio of polarized reflectance $\sqrt{R_Q^2 + R_U^2}$ to total reflectance, which is reduced to the following equation:

$$DoLP = \frac{R_Q}{R_I} \quad (2)$$

Each individual simulation is executed with a different combination of twelve scene relevant (and retrievable) parameters. These parameters are randomly chosen from a defined numerical distribution (Figure 1), whereas the distributions were chosen based on the best available information about the nature of those parameters during ORACLES [1,4,5,35], including data from observations by other instruments that participated in the campaign. Simulations were generated for the seven non-gas absorption RSP instrument channels (spectrally centered at 0.410, 0.470, 0.555, 0.670, 0.865, 1.59, and 2.26 μm) at 152 view zenith angles (VZA) between $\pm 60.5^\circ$ along the flight track.

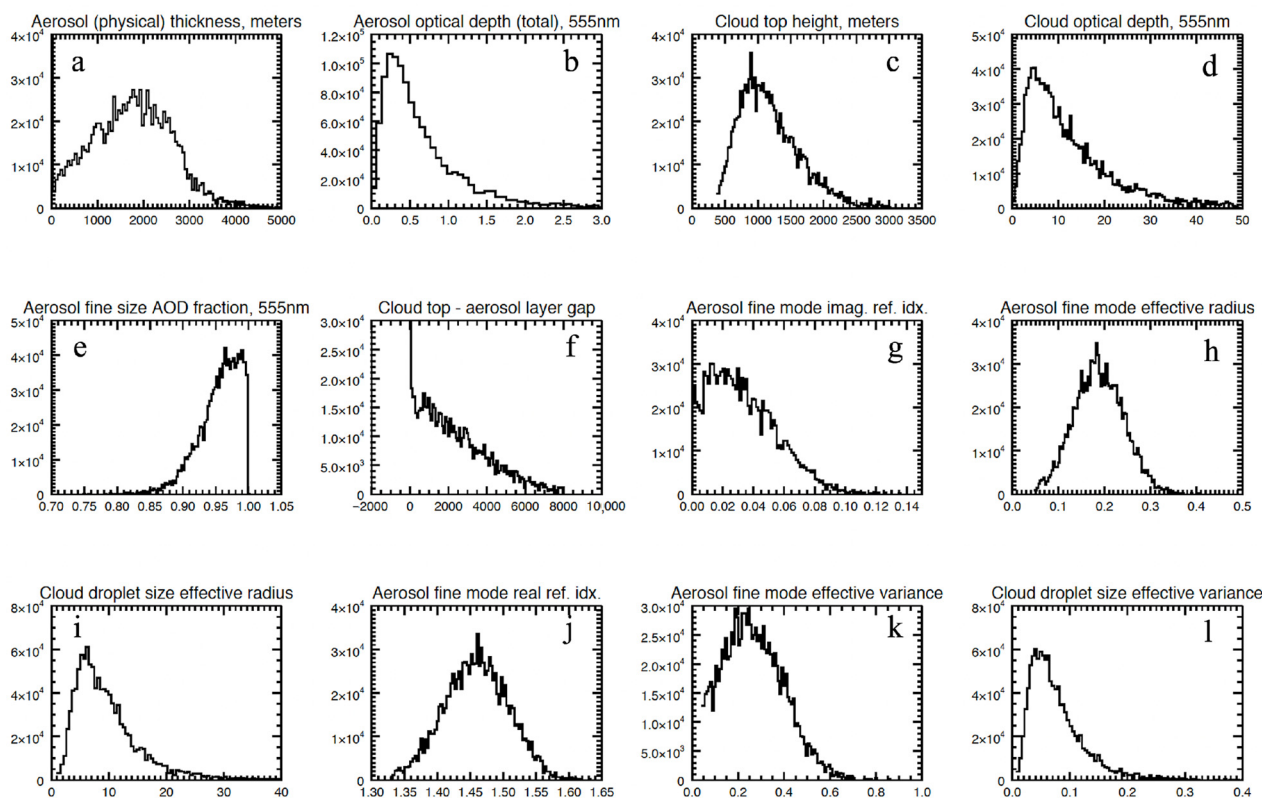


Figure 1. Training set histograms of the relevant parameters and their distributions for scenes of absorbing aerosol above liquid clouds: (a) aerosol thickness, (b) AOD, (c) cloud top height, (d) COD, (e) fine mode AOD fraction, (f) cloud-aerosol gap, (g) aerosol fine mode imaginary refractive index, Im_{IND} , (h) aerosol fine mode effective radius, R_{effAER} (i) cloud droplet size effective radius, R_{effCLD} , (j) aerosol fine mode real refractive index, Re_{IND} , (k) aerosol fine mode effective variance, V_{effAER} , and (l) cloud droplet size effective variance, V_{effCLD} .

To generate the training set, we randomly selected a combination of variables from the twelve scene relevant parameters, which were run for each of the solar zenith angles (SZAs) and the relative azimuth angles (AZIs) combinations. The AZI is defined as the difference between the instrument viewing azimuth to the solar azimuth angle. The random choice training grid was performed with the intent to reduce dependency upon arbitrarily defined ‘nodes’ in a regularly gridded training set, which was shown to affect prediction results [17]. The simulations were performed for 22 SZAs (2–86, every 4 degrees) and for 8 AZIs (0–84, every 12 degrees), totaling 1,139,600 scenarios for the training set. Simulated results were produced for R_I , R_Q , and DoLP, all of which are measured by RSP. There are different ways of expressing polarimetric information, each with its own means to represent measurement uncertainty. Because of these differences, we test in this work several expressions of polarization (e.g., R_I and R_Q vs. R_I and DoLP) because the impact of their measurement uncertainty varies.

Due to limitations of computational resources, file size and retrieval sensitivity, some of the descriptive scene parameters were kept constant, as detailed in Table 1.

Cloud top height, the physical thickness of the aerosol layer, and the gap between cloud top and the aerosol layer are all randomly generated parameters, based on the realistic value range observed during ORACLES. Aircraft height was held constant at 6100 m (the ORACLES average), so cloud top height is really an expression of the distance of the cloud top to the aircraft. We constructed 36% of the simulations to be with no cloud–aerosol gap, in order to simulate cases with touching clouds and aerosols, similarly to what was observed during ORACLES [1,36]. In some cases (25%), the combination of

the three variables (cloud top height, aerosol thickness, and gap) produced an aerosol layer whose top was close to the (fixed) aircraft altitude of 6100 m (~20,000 ft). In those cases, the aerosol layer thickness parameter was then reduced to constrain it by this aircraft height. The latter explains the somewhat non-Gaussian appearance of that parameter's histogram in Figure 1a.

Table 1. Fixed parameters of the RT simulations used for the training set.

| Parameter Description | Value (Unit) |
|---|---|
| Aircraft Altitude | 6100 (m) |
| Cloud droplet size distribution | Monomodal 2 parameter modified Gamma distribution (Hansen and Travis, 1974 [34], Equation 2.56) |
| Aerosol size distribution | Bimodal 2 parameter lognormal distribution (Hansen and Travis, 1974 [34], Equation 2.60) |
| Aerosol coarse size mode refractive index | 1.47-i0.01 |
| Aerosol coarse size mode effective radius | 6.91 (μm) |
| Aerosol coarse size mode effective variance | 0.867 |
| Trace gas absorption | Neglected (corrected in observational data) |
| Atmospheric surface pressure | 1013.25 (mbar) |
| Surface temperature | 288.15° (K) |
| Ocean surface reflectance | ^{&} None |
| Simulation geometry | Slab, plane parallel |

[&] due to the opaque nature of the low-level clouds simulated here ($\text{COD} > 3$) we assumed that ocean surface reflectance is not observed by the down looking instrument.

The aerosol distribution was defined by two size modes: fine and coarse. The fine mode fraction (FMF), Figure 1e, is determined by taking the ratio of the fine mode aerosol optical depth (AOD_f), to the total AOD at 555 nm. For ORACLES, the primary biomass burning (smoke) aerosols were dominated by the aerosol fine size mode [37,38], which is the reason to vary the fine mode parameters in the training set. Conversely, only minimal quantities of coarse mode aerosols were observed during ORACLES [1,15], and so the ability to uniquely retrieve their optical properties was not feasible. Therefore, the coarse mode AOD (AOD_c) is varied in the training set, but only in an indirect manner, as follows:

$$\text{AOD}(555\text{nm})_c = \text{AOD}(555\text{nm})_{\text{total}}(1 - \text{FMF}) \quad (3)$$

The single scattering albedo (SSA) is defined as the ratio between the scattered to total extinction, and as such, defines aerosol absorption, which is an important parameter that describes the state of the BB aerosol [39]. In our RT simulations, aerosol absorption is calculated from the aerosol imaginary refractive index, the real refractive index, and the aerosol size distribution. Thus, a single parameter such as the SSA may provide a more appropriate parameter to retrieve, as it is presumably more orthogonal to other parameters in the space of observations. The SSA is calculated as an intermediate output during the RT simulations and is saved per each simulation scenario (see Figure 2 for its values distribution). Then, in the training process, it is used as one of the retrieved variables. It is important to note that SSA is notoriously hard to retrieve and usually has high uncertainty

and a large span of values among different instrument and retrieval approaches [15,20,39]. Furthermore, this variable is often used as an input for ACA AOD from passive radiometers such as MODIS and SEVIRI, where small differences in its assumed values can result in large retrieval uncertainties [11,12].

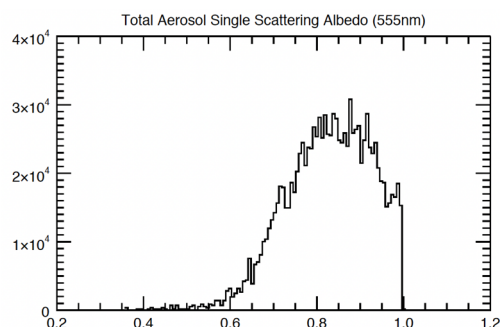


Figure 2. Single scattering albedo (SSA) at 555 nm histogram values.

To account for the measured uncertainty in our training set, and the fact that the total (and polarized) reflectance has significantly different uncertainties ($\sim 3\%$ relative for reflectance, and ~ 0.002 for DoLP), we standardized the reflectance vectors (R_i , R_Q , DoLP) as follows [17]:

$$\hat{x}_i = \frac{x_i - \bar{x}}{\sigma(\bar{x})} \quad (4)$$

where \hat{x}_i represents the standardized reflectance or DoLP vector (over all VZA and wavelengths, for a given measurement geometry), x_i is the actual measurement vector, \bar{x} is the mean signal calculated over a certain SZA, AZI, VZA and wavelength band, and $\sigma(\bar{x})$ is the instrument uncertainty calculated over the signal mean for all simulations at that wavelength and view zenith angle, using the derived uncertainty model for the RSP instrument [40]. By performing the standardization of the training set based on the instrument uncertainty model, we properly scale the input signals by their weight and improve the machine learning inference, as was shown in previous work [17,20]. Figure 3 shows some examples from the standardized training set, for two ranges of AOD, COD, SZA and AZI, for each of the measured vectors. It is apparent from the figure that each vector shows unique features, and that the spatial patterns created among the different wavelengths and VZA angles create distinct patterns for each of the variable combination. We also see that at high SZAs, the total reflectance signal feature and intensity is diminishing. This points to the importance of the polarized reflectance and DoLP data, which can still resolve the cloud, and aerosol features over the training domain.

2.2. Observations

In order to validate the machine learning algorithms utilized in this work, we used observational data from the ORACLES campaign in 2016–2018 over the SEA ocean [1]. The campaign was focused on in situ and remote-sensing measurements of BB aerosols above marine stratocumulus clouds, as detailed in Redemann et al. [1]. Observations in 2016 were taken by the RSP instrument on board the NASA ER-2 (a high-altitude aircraft that operates at 19–20 km), and on the NASA P-3 aircraft, during 2017–2018, through “remote-sensing” straight legs between 5 and 7 km. Most of the observations were taken above low-level liquid clouds, which were opaque enough to mimic our training simulations. AOD levels in the region averaged between 0.05 and 0.4, with BB aerosol representing the major component [1,15,41]. RSP made continuous along track scans over an area of clouds and aerosol along the flight trajectory, with 152 forward and aft angles between $+60$ and -60 degrees, respectively. However, during measurements, not all angles between $+60$

and -60 contained valid data due to field of regard vignetting in the aircraft installation. Therefore, our training set spanned the angle range between $+40$ and -40 . At each scan, RSP views a footprint of about 323 m.

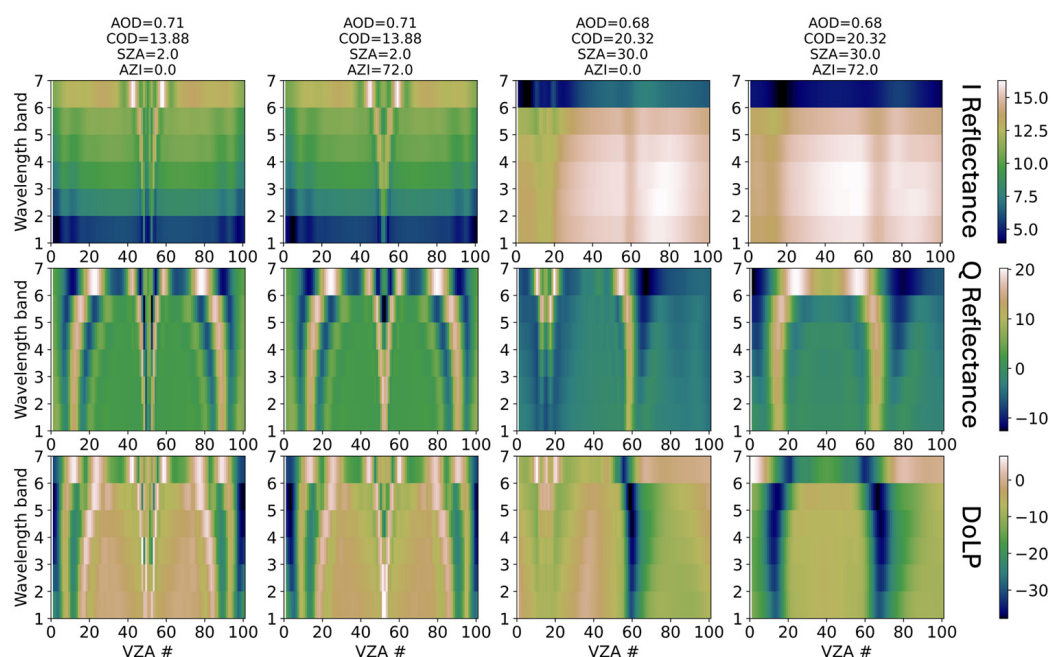


Figure 3. Standardized I reflectance (upper row), Q reflectance (middle row), and DoLP (bottom row) values over the seven RSP wavelength bands (410, 470, 555, 670, 865, 1590, and 2250 nm) and 101 viewing angles (between -40 and 40°), shown here as VZA index number (#). These matrices were used in the training of the machine learning algorithms, as described in Section 3.

To utilize the RSP scans from ORACLES as an input to our trained machine learning network, we processed the raw data following our algorithm theoretical database document used in our previous NN-based cloud retrieval algorithm [17]. In short, we proceed with the following steps:

- (1) Level 1 radiometric and polarimetric data were organized so they represent a multi-angle view of cloud top.
- (2) For each level 1 observation, if they represent clouds, then the above cloud water vapor pressure from model and reanalysis (MERRA-2 and standard atmosphere vertical profiles) are used to correct for trace gas absorption. This correction is performed since the training set did not include trace gas in the simulations.
- (3) The corrected data are then standardized according to Equation (4).
- (4) Finally, we use either the entire VZA values for each of the wavelengths as a vector input to the MLP-NN model, as was performed in [17] or as a 3D input, as showcased in Figure 3.

3. Methodology

The overall methodology in this work follows a standard data-driven research template, depicted in Figure 4. Namely, we first prepare the data for training as described in Section 2.1. For each candidate network architecture (described in detail in Sections 3.1–3.3 below) we train the network weights using the training data (Section 3.4), measure the resulting predictive performance on a hold-out test set (Section 4.1), and finally validate the trained models on real-world measurements (Sections 2.2 and 4.2).

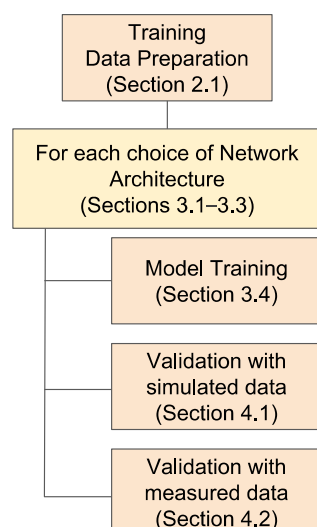


Figure 4. Flow chart describing algorithm test and validation process.

Our goal in this work was to develop a workflow that will allow coupled retrievals of aerosol above clouds and liquid cloud property using machine learning techniques. Specifically, we focused on techniques that utilize the spatial connections of a multi-dimensional dataset such as the one from polarimetric measurements. We implemented two methodologies that can ingest spatially structured high dimensional input data (i.e., ResNet and ViT, as detailed below), and compared their performance to a benchmark MLP-NN technique that was previously used by us to retrieve liquid clouds [17]. Below, we describe each of the algorithms and give detail on their training process and parameters. The input structures to each of the different models differ, as detailed in Table 2, but the output variables (8) are similar, as follows: AOD, COD, cloud droplet effective radii (R_{effCLD}), cloud droplet effective variance (V_{effCLD}), SSA, aerosol droplet effective radii (R_{effAER}), fine fraction aerosol imaginary refractive index (Im_{IND}), and real refractive index (Re_{IND}). Notably, ResNet and ViT are convolution and attention based, respectively, and operate on an image-like input (3D), while MLP-NN operates on a 1D input array.

Table 2. Input parameters for each of the network architectures.

| | MLP | ResNet | ViT |
|---------------|--|--|--|
| Layers | 4 hidden layers (1024 neurons) | 50 convolutional layers (ResNet50) | 10 transformer blocks, 8 multi-head attention each |
| Batch Size | 4000 | 4500 | 500 |
| Epochs | 500 | 500 | 500 |
| Learning Rate | 0.001 | 0.001 | 0.001 |
| Inputs | 1414 (2 channels) or 2121 (3 channels) + 2 | $2 \times 7 \times 101$ or $3 \times 7 \times 101 + 2$ | $2 \times 7 \times 101$ or $3 \times 7 \times 101 + 2$ |

3.1. ResNet

CNN (convolutional neural network) architectures (e.g., [23]) are considered to belong to the class of deep neural networks (DNNs). Benchmark CNN architectures, such as VGG [42], are constructed from multiple blocks of convolution and pooling layers. These have been found to perform well on image classification and many other tasks, compared to past benchmark architectures such as multi-layer perceptron (MLP) neural networks (NN). However, the parameter count in these deep networks increases exponentially with

depth, which ultimately impedes their performance, increasing both their training and train errors, and reducing their accuracy [21]. This issue is caused by the “vanishing gradient” phenomenon, which decreases the backpropagation gradient to near zero values during the training of the network. To overcome these challenges, a residual network architecture (ResNet) was suggested by [21] and is widely adopted by the computational community as superior to the “plain” deep-networks approaches [43]. Here, we adopt the ResNet architecture as a representative CNN deep-network candidate for the retrieval of aerosol and cloud properties from the image-like multi-dimensional polarimetry data. Specifically, and as shown in Figure 5, we adopt the ResNet50 architecture, implemented using the PyTorch 1.12 open-source software. The figure details the number of blocks and layers within each of the blocks, with our Input and Output layers that are different from the original ResNet50. The main difference between the ResNet and VGG or similar networks is that there is a “shortcut” connection between the input and the output of a network block (the expanded block in Figure 5 insert shows the residual short-connection structure), that combines the block output and the identity input. By doing so, the network learns the layer mapping plus the residual mapping instead of the underlying mapping of the entire function, which was proven to perform better in deep networks [21,43].

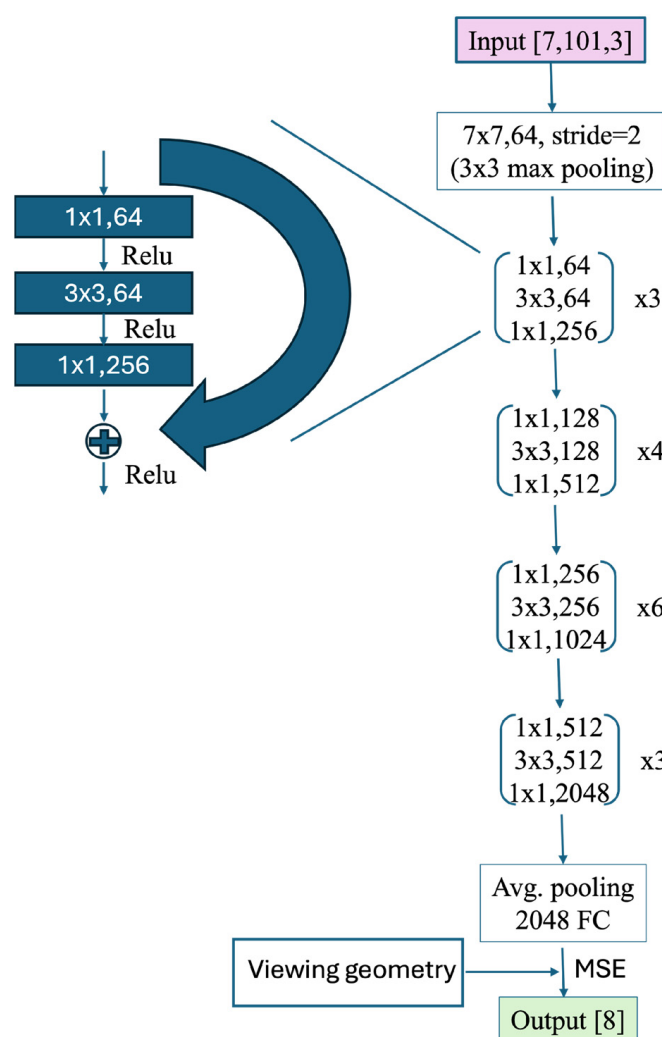


Figure 5. The ResNet50 architecture was used in this study. Relu (ReLU) stands for rectified linear unit activation function used between the layers. MSE stands for the mean square error loss function used in this work. The Figure is following He et al. [21] and modified per our changes, as detailed in the text.

The changes made to the original ResNet50 architecture are the input size, which is a 3D array in the size of (wavelengths, VZA, and channel) a fully connected layer of 2048 nodes at the end of all convolution blocks, a mean square error (MSE) loss function (instead of a SOFTMAX) to account for the regression character of our problem, and a layer embedding the viewing geometry inputs before the loss calculation is processed. The channels represent the combination of measurement inputs (total reflectance, polarized reflectance, and DoLP), and can take the size of 2 or 3, depending on input combination.

3.2. Vision Transformers (ViT)

Vision Transformers (ViT) emerged from transformer architectures used in large natural language processing (NLP) models [44,45]. ViT adapts the transformer architecture to computer vision by first dividing input images into fixed-size patches, which are linearly embedded and treated as tokens. This approach enables the model to process image data using the same self-attention mechanisms that have proven effective in NLP tasks, learning relationships between image patches to perform visual recognition tasks [44]. When compared to state-of-the-art convolutional architectures like ResNet, ViT achieves comparable or superior results in image classification tasks [44,46].

The transformer uses a self-attention block (Figure 6), where each sequence element is transformed into queries (Q), keys (K), and values (V) through learned linear projections. The attention weights are computed as scaled dot products between Q and K, which are then applied to V [45]. Since a single attention pattern is insufficient for capturing different types of relationships, transformers implement multi-head attention (see Figure 6 insert, following [45]). While the multi-head attention block is permutation invariant, ViT adds positional embeddings to the patch embeddings to enable the model to consider spatial relationships in the input image.

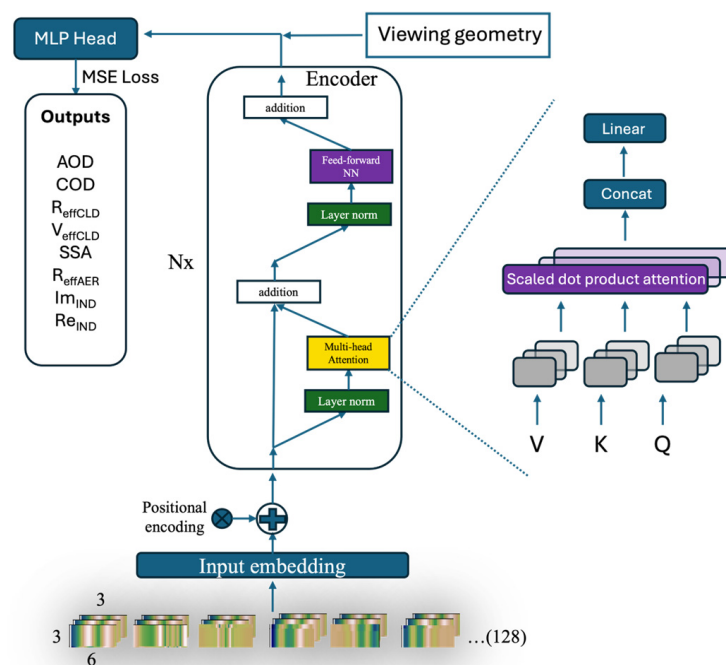


Figure 6. ViT architecture schematics that were used in our retrieval (see details in the text).

The ViT architecture used here is similar to [44], with geometry variables (SZA and AZI) inputs added before the fully connected operation. As shown in Figure 6 (adopted from [44]), each “image-like” input (wavelength, VZA, and channel), is divided into 128 patches with size (3×6) , with MSE loss function.

3.3. MLP-NN

To compare with previous works that were conducted with NN retrievals from polarized reflectance measurements [17–19], we constructed an architecture similar to the one used in [17], using a feed-forward multi-layer perceptron (MLP) NN architecture with four hidden layers, each with 1024 neurons, and a ReLU activation function (Figure 7). Inputs are used as a vector array, with a length of number of VZA \times number of measurement vectors (Table 2). The viewing geometry is added to the network by multiplying the encoded viewing zenith and azimuth inputs with the output of the hidden layers, before passing through the output layer.

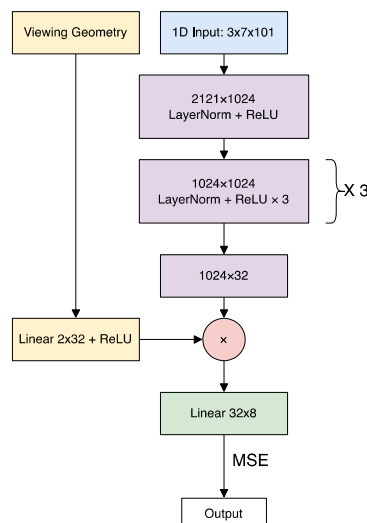


Figure 7. MLP architecture used in this study.

3.4. Training and Model Setup

We used a 70:30 percent ratio of randomly selected samples from our simulated dataset, for training and test, respectively. For reproducibility, we used the same fixed random seed for dataset splitting, to assure that each algorithm will use the same samples for training and test. All models were constructed using the PyTorch 1.12 open-source framework and trained on a NVIDIA A40 GPU processing card. The k-fold validation method was partially tested, without significant changes in the results, while also demanding extremely long computational time resources.

Table 2 summarizes the parameters and inputs used for each of the models. Notably, and as inherent by the model' architectures, ResNet and ViT receive inputs in the form of 3D arrays, while MLP-NN receives inputs in a form of a 1D vector. For example, as seen in Table 2, the MLP input of 1414 (2 channels) has 7 wavelengths, 101 viewing zenith angles, from each of the two channels selected (I, Q, or DoLP), and additional geometry input (SZA and AZI). The output variable values used in the training process were scaled using standard scaling procedures (mean subtracted and divided by the standard deviation over the variable range). We used an Adam stochastic gradient-based optimizer [47] for all models. Additionally, for training the ViT model, we used the Xavier weights initialization method [48] and a dropout of 0.1 to reduce overfitting.

4. Results

This section summarizes our results as follows: (1) comparing the different models' performance with the simulated test data, and (2) validating and comparing the different models using data from the RSP instrument, and other collocated instruments during ORACLES.

4.1. Comparing the Different Models on Test Data

We trained each of the deep networks with various input vector combinations as follows: (I,P), (I,Q), (I,Q,P), and (Q,P), where I, Q, P stand for the total reflectance, polarized reflectance, and degree of linear polarization vectors, respectively. We note that the AOD, COD, R_{eff} , and V_{eff} distributions shapes resembles more closely a log-normal distribution. Therefore, for each of the input combinations, we performed two sets of training simulations. The first set of training simulations was performed on the original distributions (i.e., log-normal experiments). The second set was performed on AOD, COD, R_{eff} , and V_{eff} distributions that were converted to normal distributions (notated as normal experiments). We found that using normal distributions slightly improves our test results, and so the majority of the results provided hereafter are the normal experiments. The total number of simulated scenes was 1,140,000, where we used 70/30% as our training/test ratio, with ~800,000 different scene combination for training, (as detailed in Section 2.1) and ~340,000 out-of-sample simulated data that were reserved for testing.

Figure 8 shows an example of the test results for a simulation using the ViT model with two input channels (I,Q) for the normal experiments. For each variable, we calculated the linear goodness of fit (R^2) between the predicted test set and the true values of the test set, the RMSE (root mean square error), and the normalized RMSE, which was calculated as follows: $(\text{RMSE}/\sigma) \times 100$, where σ is the training set standard deviation. In the figure, scatter points are color-coded by the density of samples. We see excellent retrieval capability for all the variables tested, measured by both high $R^2 > 90\%$ and low RMSE with values well within the requested accuracy for each of the variables. The real refractive index predicted values resulted in the lowest correlation and highest normalized RMSE values. Similar results were also obtained for the ResNet and MLP models, as seen in Figure 9. However, as will be shown later on, in Figures 10–15, the bias patterns between the models are different.

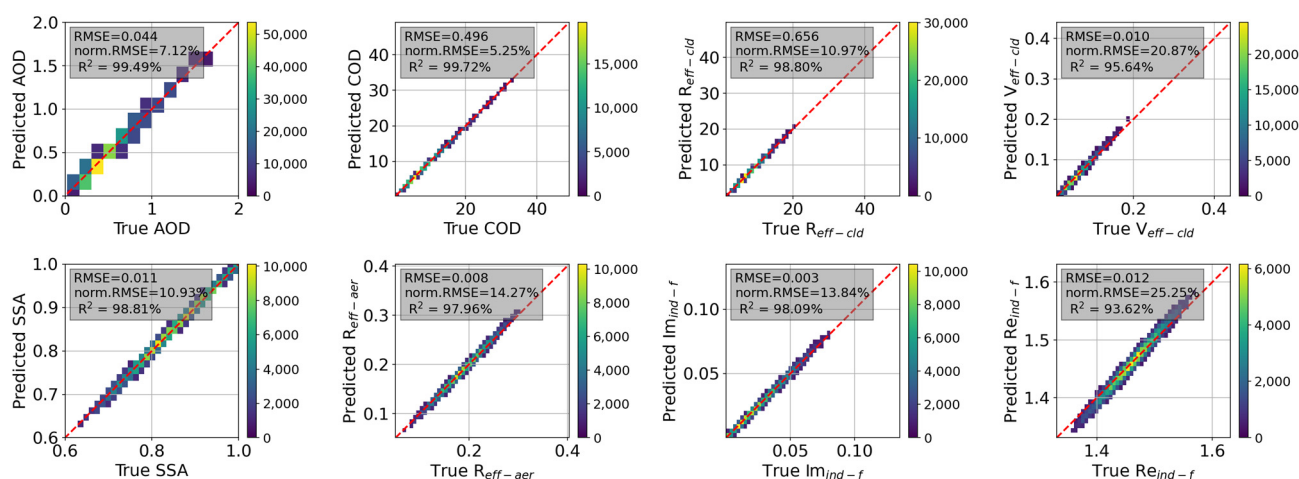


Figure 8. Predicted versus true values for the test set, using a trained ViT network with two input vectors (I,Q) and normal experiments for all variables. R^2 , RMSE (root mean square error), and normalized RMSE (calculated as RMSE divided by the data standard deviation and shown as percentage) are shown in grey text box. Red dashed line represents the 1:1 linear relationship.

Figure 9 shows very little difference between the three models used, with very high correlation coefficient values ($R^2 > 90\%$) for all the eight variables tested. A closer look at the RMSE and normalized RMSE provides a better perspective on the prediction capability of these variables. The most challenging variable to predict is the real refractive index, with about 20–30% of normalized RMSE for all models. This is probably the case due to its very narrow range, and therefore, changes in the real refractive index affect only slightly the

measured signal. The second most challenging variable to retrieve is the cloud effective variance (V_{eff}), with a relatively high normalized RMSE. This is an inherent characteristic of the polarized signal, where a large V_{eff} change ($\sim 30\%$) is needed for observing a noticeable change in the Q signal [15], meaning less sensitivity for retrieving this variable.

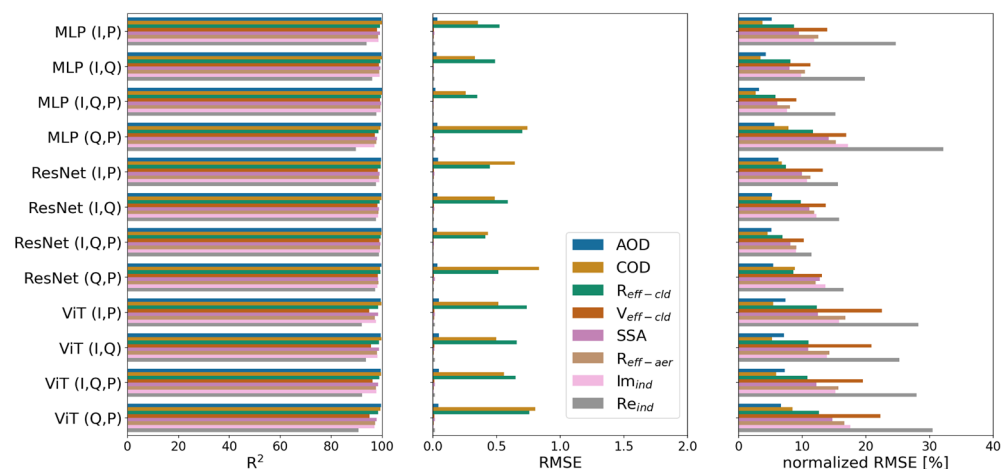


Figure 9. Overall predictions statistics: (left) correlation coefficient R^2 , (middle) RMSE, and (right) normalized RMSE, on test set simulated data generated with normal distributions of all variables for each of the models tested, and each input variation, color-coded as in the legend.



Figure 10. The difference between the true test and the predicted values versus SZA (upper panel) and AZI (lower panel). Test results are based on training a ViT network using two input vectors (I,Q) and normal distributions of all variables. Mean (dashed red line) represent the mean difference values with 1 standard deviation lines shown around the mean (dashed green lines). Colormap represents the number density of test points.

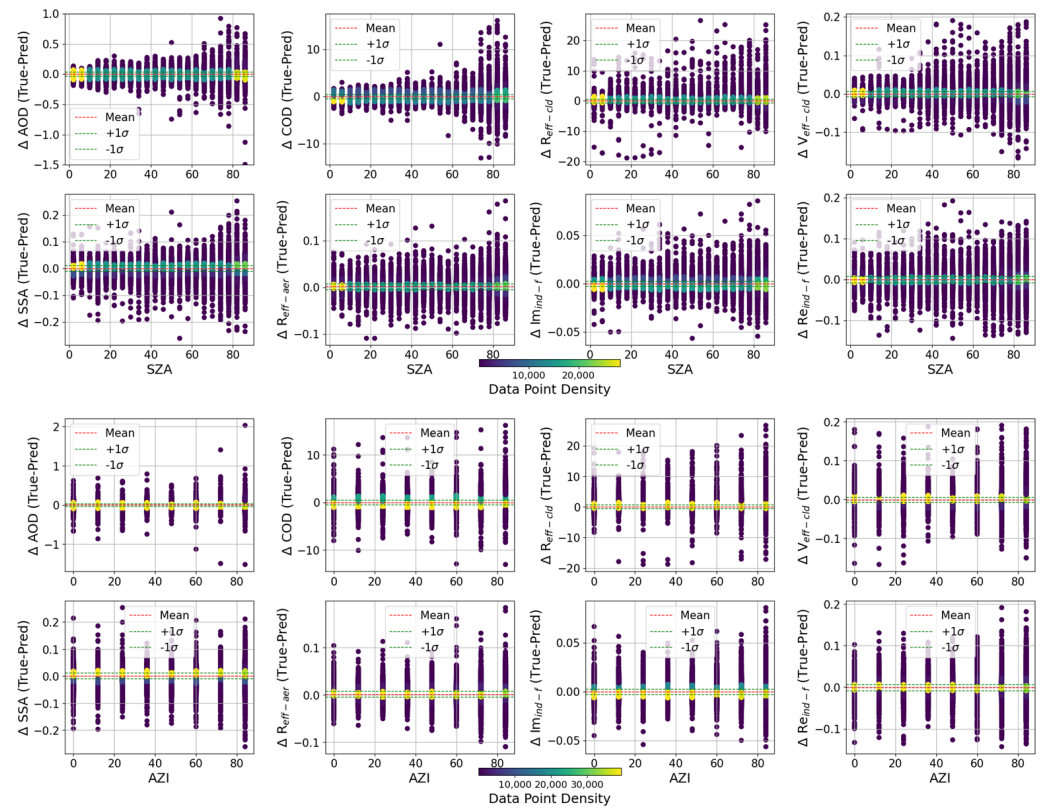


Figure 11. Same as Figure 10 but for the ResNet model.

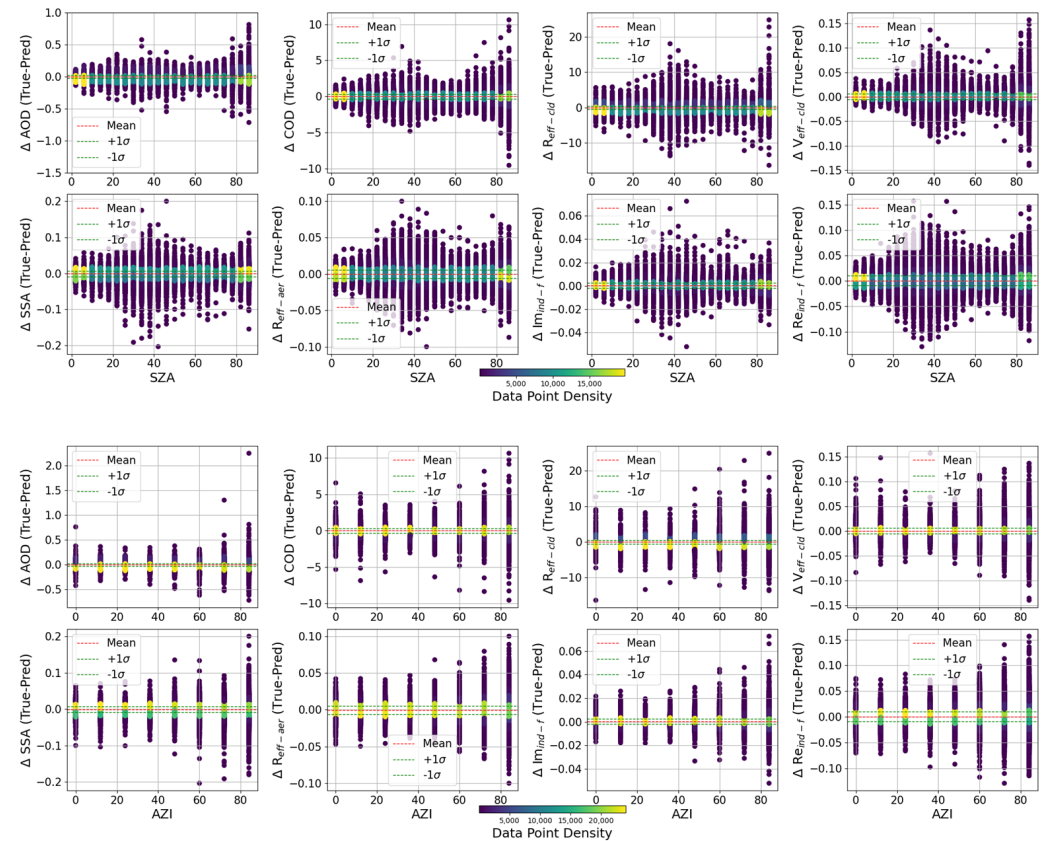


Figure 12. Same as Figure 10 but for the MLP model.

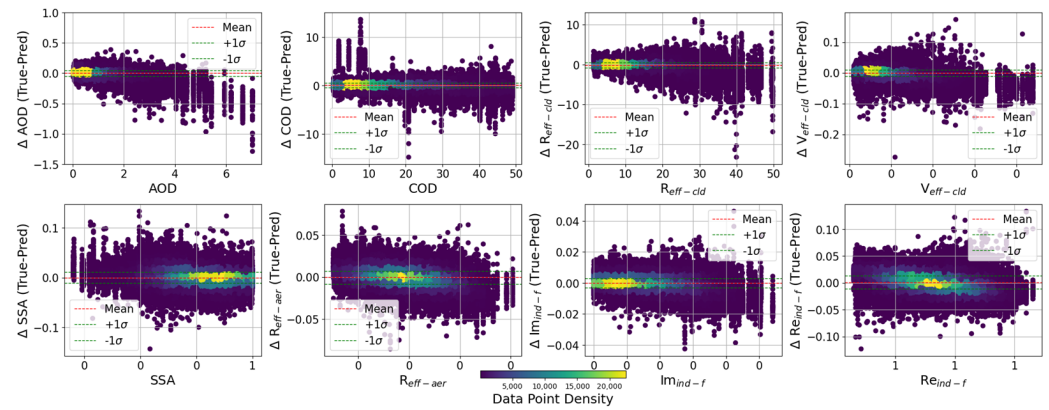


Figure 13. The difference between the true test and the predicted values versus the true variable values. Test results are based on training a ViT network using two input vectors (I,Q) and normal distributions of all variables. Mean (dashed red line) represent the mean difference values with 1 standard deviation lines shown around the mean (dashed green lines). Colormap represents the number density of test points.

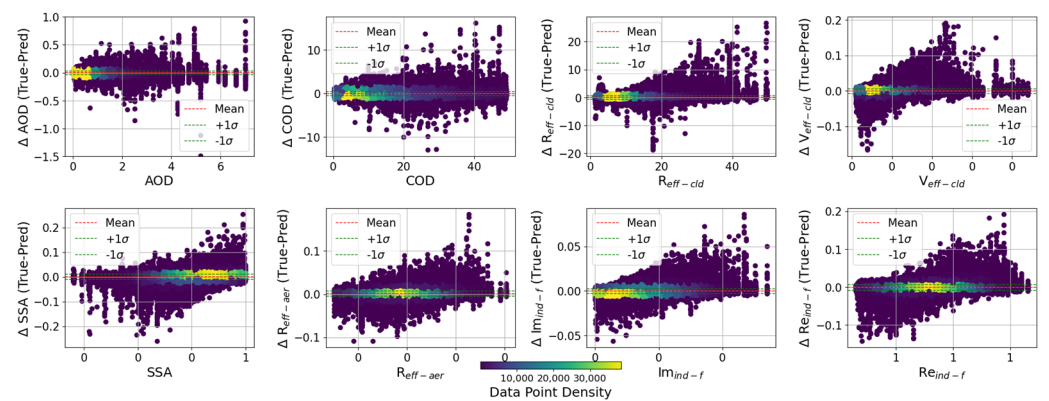


Figure 14. Same as in Figure 13 but for ResNet model.

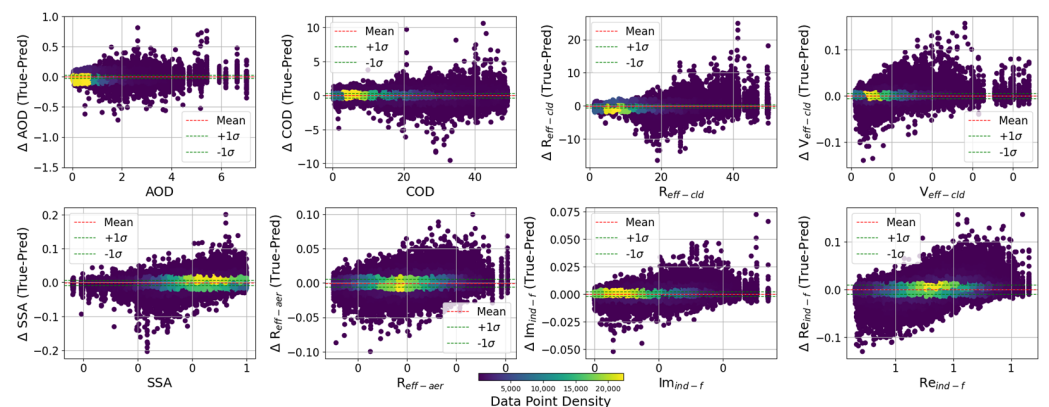


Figure 15. Same as in Figure 13 but for MLP model.

In Figures 10–12, we examine the behavior of the retrieved variable's bias with respect to the viewing geometry for the ViT, ResNet, and MLP models, respectively. We calculate the difference between the true and predicted values for the range of solar zenith angle SZA (upper plots panel) and azimuth angle AZI (lower plots panel). In general, we observed almost no or a very weak relationship between the goodness of prediction and the viewing geometries, except for the cloud effective variance variable, which showed increasing bias for larger SZA angles for all models but was most pronounced for the ViT model. Also, in the ViT model, it is interesting to note that although the majority of AOD points fall within

the zero-bias line, many predicted values are underestimated by the model, with growing deviation towards higher SZA. Among all three models, the ViT has the lowest bias values, where ResNet shows the least dependency on the viewing geometry.

Figures 13–15 show the prediction bias by variable value for each of the models. Similarly to the bias by viewing geometry, ResNet and MLP are slightly different than ViT. The latter shows similar bias values that do not depend on the variable value, except for AOD, where higher values lead to higher positive bias (prediction is much higher than the truth). The ResNet and the MLP models do not show an AOD dependent bias, but do show dependency with increasing variable values, especially for the cloud effective variance ($V_{\text{eff-cld}}$), and the real refractive index ($Re_{\text{ind-f}}$). These features will be further examined in our validation with other instrument and retrieval algorithms below.

4.2. Validation

4.2.1. Clouds

We validated the three models by testing them on RSP measurements obtained during the ORACLES campaigns. The input files to our retrieval were georeferenced L1B RSP files that were corrected for atmospheric water, as detailed in Section 2.2.

Figure 16 summarizes the similarity (coefficient of variation) values for all three model experiments, using models trained with normal distributions, and various input permutations. We note here that the models trained with the normal distributions yielded slightly better results for COD, R_{eff} , and V_{eff} . While R^2 evaluates how much of the variability in the actual values is explained by the model, and can suffer biases due to the “truth” (RSP) model erroneous values, the similarity measure evaluates the relative closeness of the predictions to the actual values, and was calculated as the dot product between the two arrays (RSP retrieved by either standard or Deep learning-DL models), divided by the product of their norm, as follows:

$$\text{Similarity} = \langle RSP_{\text{std}}^{\text{var}}, RSP_{\text{DL}}^{\text{var}} \rangle / (\|RSP_{\text{std}}^{\text{var}}\| \cdot \|RSP_{\text{DL}}^{\text{var}}\|) \quad (5)$$

where $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ stand for the standard Euclidean inner product and the norm, respectively, std represents RSP standard retrieval by the parametric polarimetric (PP) method [49] for the R_{eff} and V_{eff} variables, and the bi-spectral Nakajima King (NJK) method to retrieve COD, as detailed in Miller et al. [17,50]. Values shown in Figure 16 were calculated for each flight date, and averaged per campaign year, for each of the cloud variables retrieved. The NJK method uses R_{eff} and V_{eff} retrievals by the parametric polarimetric (PP) method, which is less sensitive to aerosols above clouds, and then uses the intensity measurement to determine COD. Therefore, it is more sensitive to large number of aerosols above cloud [50].

As shown in Figure 16, ViT and ResNet models produce higher similarity values for cloud optical depth (COD), compared with MLP predictions for all ORACLES years. The prediction similarity values for R_{eff} is high for all models and all years (between 0.85 and 0.98), but with ResNet model showing the lowest similarity values among the three models. This is in contrast to the relatively low normalized RMSE values for ResNet for R_{eff} on the test set (Figure 9). As R_{eff} retrieval depends on the cloud bow signal (between scattering angles of 135–160), it might be that the ResNet model representation is not capable of capturing the full variation of the signal in cases where partial cloud bow signal is available. With cloud V_{eff} predictions, we see lower similarity (~0.60–0.80), increasing in similarity from MLP, to ResNet to ViT. Although these values are relatively low compared with the COD and R_{eff} similarity values, they are much higher than comparable V_{eff} correlations obtained for the AirMSPI and RSP instruments during ORACLES 2016 (~0.2) using the joint aerosol–cloud retrieval by Xu et al. [15]. While COD and cloud effective radii (R_{eff}) similarity values seem to decrease from 2016 to 2018, the effective variance (V_{eff}) values

are almost constant. One explanation for the decreasing similarity from 2016 to 2018 is the fact that the absorbing aerosols overlaying the clouds were the highest in 2016 (September), and decreasing towards 2018 (October), corresponding better to the simulated scenes that the models were trained upon.

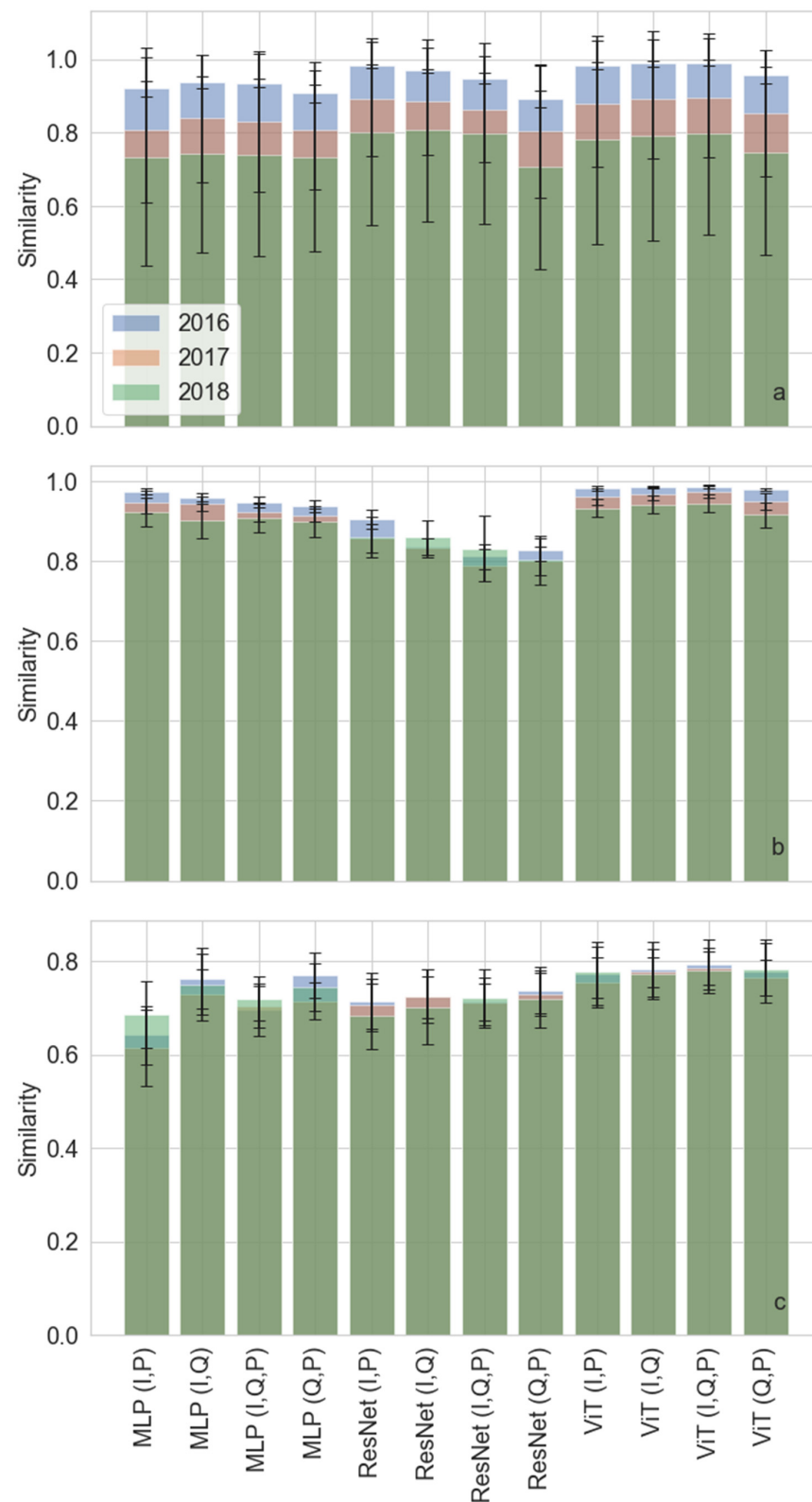


Figure 16. Similarity values comparing RSP polarimetry-based cloud only products: (a) COD, (b) R_{eff} , and (c) V_{eff} , with deep-learning ACA model predictions averaged over each of the ORACLES campaign years.

Figure 17 shows an example prediction for 20 September 2016, comparing PP cloud retrievals with the different model predictions of cloud properties. We compare MLP, ResNet, ViT, and the original RSP PP retrievals, where for each architecture, we calculated the corresponding similarity values, for each of the cloud variables (COD, R_{eff} , and V_{eff}). We see that ViT (green dots) provides the most stable retrieval, with the highest similarity values among the three models tested, for all cloud variables. As mentioned in Section 4, the training using the normal distributions of the cloud property yielded slightly more comparable results between the three models and RSP cloud products. Furthermore, using the normal distribution during the training improved the retrieval similarity of the cloud effective variance variable, which has been difficult to achieve previously [15,18].

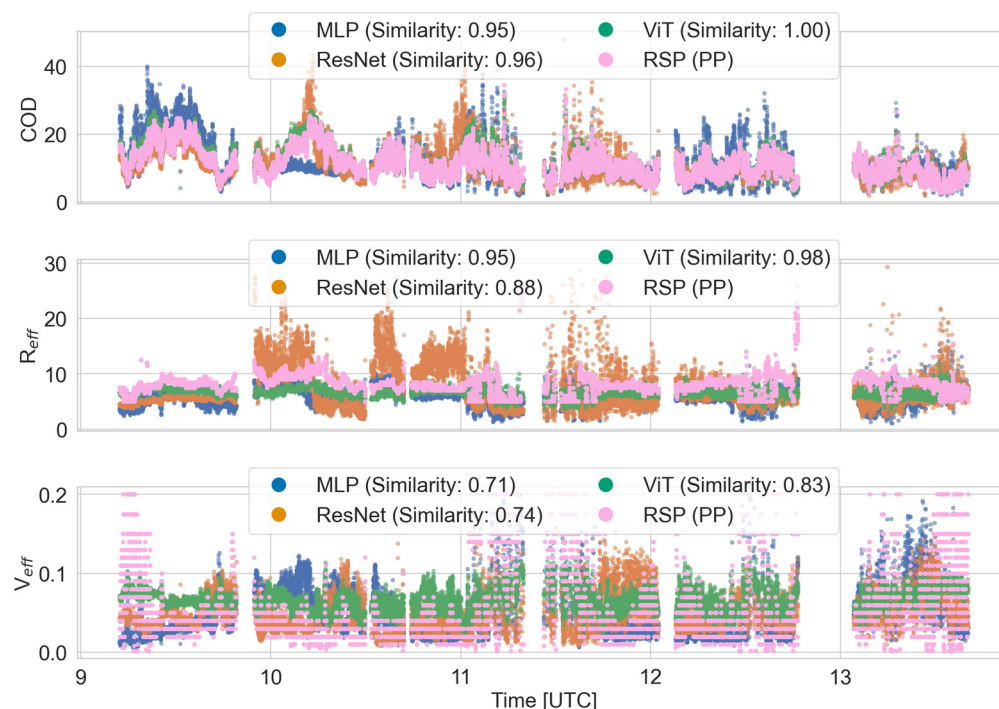


Figure 17. Time-series example for 20 September 2016 flight with RSP PP predicted values (pink dots) ViT (green dots), ResNet (orange dots), and MLP (blue dots) ACA prediction (log-normal experiment, (I,Q) input).

The summary of cloud retrieval comparison between the deep-learning models and RSP for all ORACLES years is shown in Figure 18. The upper panel corroborates the similarity plots, where COD variable shows the best match, with decreasing correlation among the R_{eff} and V_{eff} values compared between the two models. It is important to note that even the comparison between the two most common RSP retrieval algorithms (PP and NJK) yielded varying correlations, as shown previously by Miller et al. [17]. Specifically, they found that the correlation coefficient (R) between those methods was 0.75 and 0.2 for ORACLES 2016 and 2017, respectively. Consequently, computing R would not be informative of the actual performance of the algorithm; therefore, we introduce the similarity measure as a more useful quantity for comparison. In addition, when comparing their NN approach with RSP PP and NJK retrievals for R_{eff} , they found a bias of $\sim 3 \mu\text{m}$ for ORACLES 2016 data. With this regard, we note that the ResNet model shows a smaller R_{eff} bias ($\sim 1.5 \mu\text{m}$) with relatively higher correlation (for all ORACLES years), while the ViT model shows almost no bias ($\sim 0.25 \mu\text{m}$). The differences between the different retrieval methodologies and our deep-learning models are expected due to multiple reasons. First, the PP method makes use of a single wavelength for the retrieval (usually $0.865 \mu\text{m}$) where our method utilizes the interaction and spatial link between all available aerosol

wavelengths. Second, our retrieval is explicitly taking into account the amount of aerosol that exists in the scene, while the PP or NJK methods do not. In fact, the NJK method is known to be systematically high biased in R_{eff} in the presence of absorbing aerosol over clouds [12,17], as shown in Figure 18, in the lower panel.

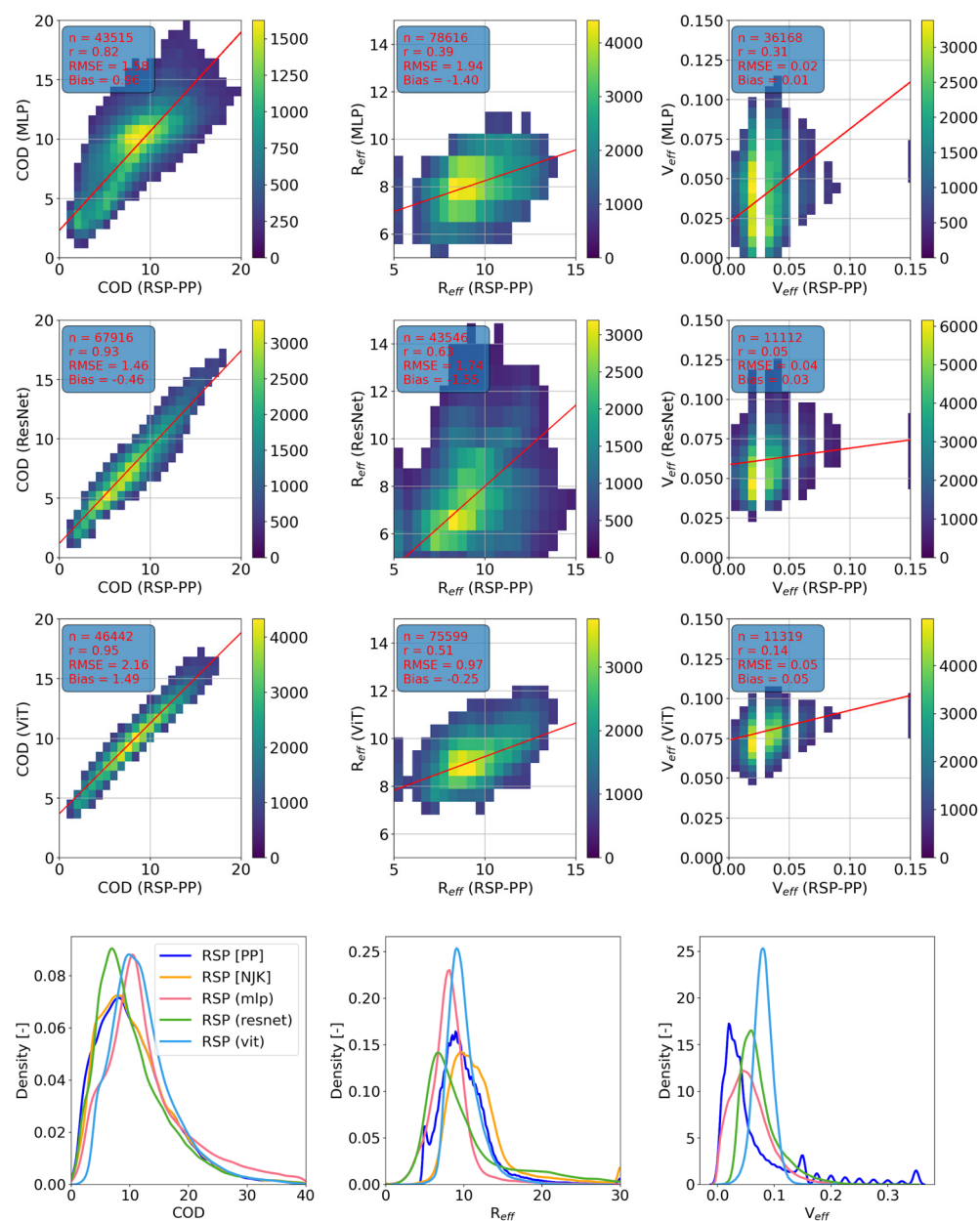


Figure 18. 2D histogram plots showing all ORACLES years RSP data, comparing between RSP parametric retrieval (PP) and each of the deep-learning models: (**top row**) MLP, (**middle row**) ResNet, and (**bottom row**) ViT. Subplot below shows the cloud microphysical variable distributions of the data among the different retrieval algorithms.

4.2.2. Aerosols

Comparing with AirMSPI ACA Algorithm

During ORACLES 2016, the Airborne Multiangle Spectro-Polarimetric Imager (AirMSPI) flew on the ER-2 high-flyer (~20 km) aircraft together with the RSP instrument. The AirMSPI is an imaging polarimeter, with swath size of 80–100 km along track and 10–25 km across track. AirMSPI acquires multi-angular observations over a $\pm 66^\circ$ along-track range in its sweep operation [51]. The data is mapped to a 25 m spatial grid.

In addition to radiance measurements in eight spectral bands covering the ultraviolet/visible/near-infrared bands, Stokes components Q and U are measured in the 470, 660, and 865 nm spectral bands. Although the two instruments have substantially different data acquisition modes and spatial, angular, and spectral resolutions, the polarimetric measurement capabilities and the available ACA retrieval algorithm by AirMSPI [15] makes this instrument an obvious candidate for validating our aerosol algorithm results for ORACLES 2016.

Figures 19–23 show the time-series comparison between RSP observations, converted to aerosol properties by our deep-learning schemes, and AirMSPI ACA retrievals [15], for SSA, AOD, Aerosol effective radii, Imaginary refractive index, and real refractive index, respectively, for all three models tested. AirMSPI AOD and SSA values are reported at 470, 660, and 865 nm bands, and were interpolated to 555 nm. All RSP retrieved variables were smoothed over a ~200 s time window.

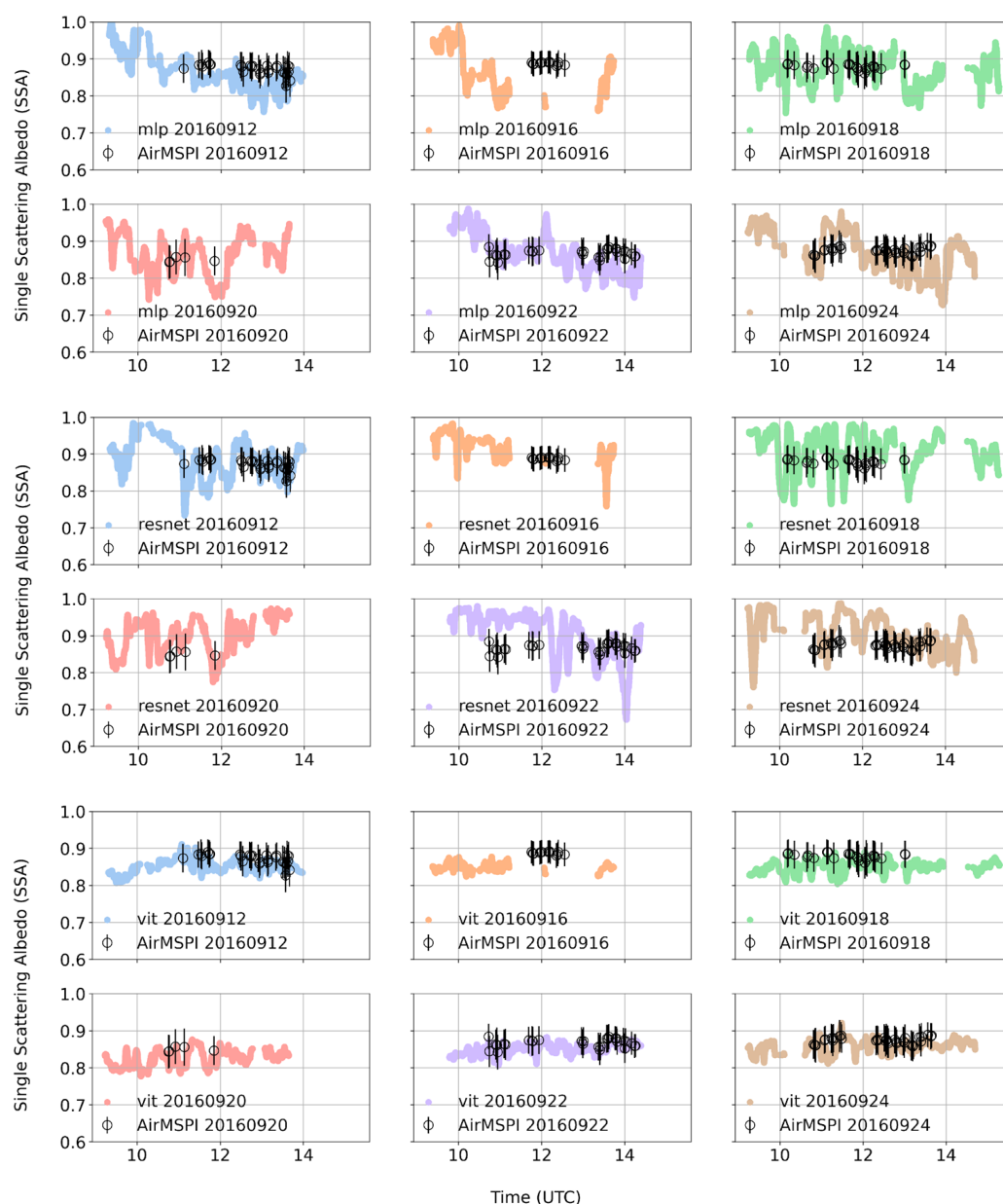


Figure 19. SSA (555 nm) comparisons using deep-learning model retrievals (color-coded by flight date), and AirMSPI (open circles with vertical lines as error bars) aerosol above cloud retrievals in 2016 for (**upper panel**) MLP, (**middle panel**) ResNet, and (**lower panel**) ViT models.

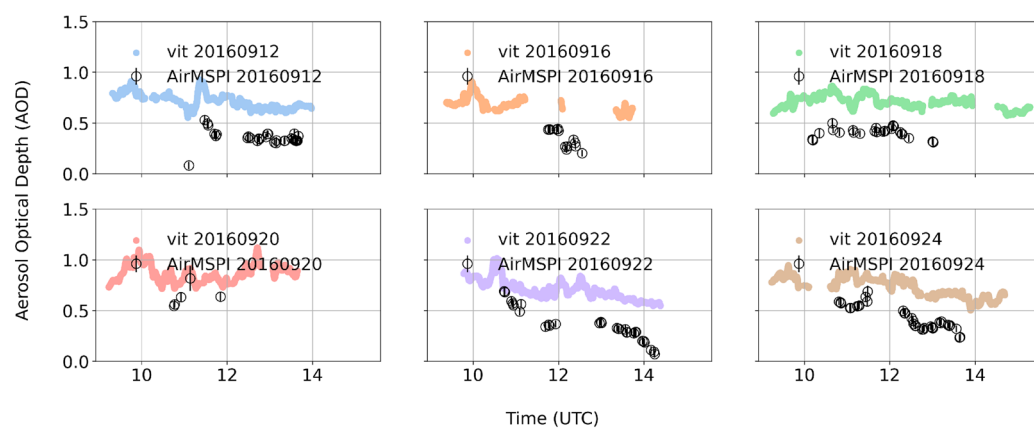


Figure 20. Similar to Figure 19 but for AOD comparisons (ViT model only).

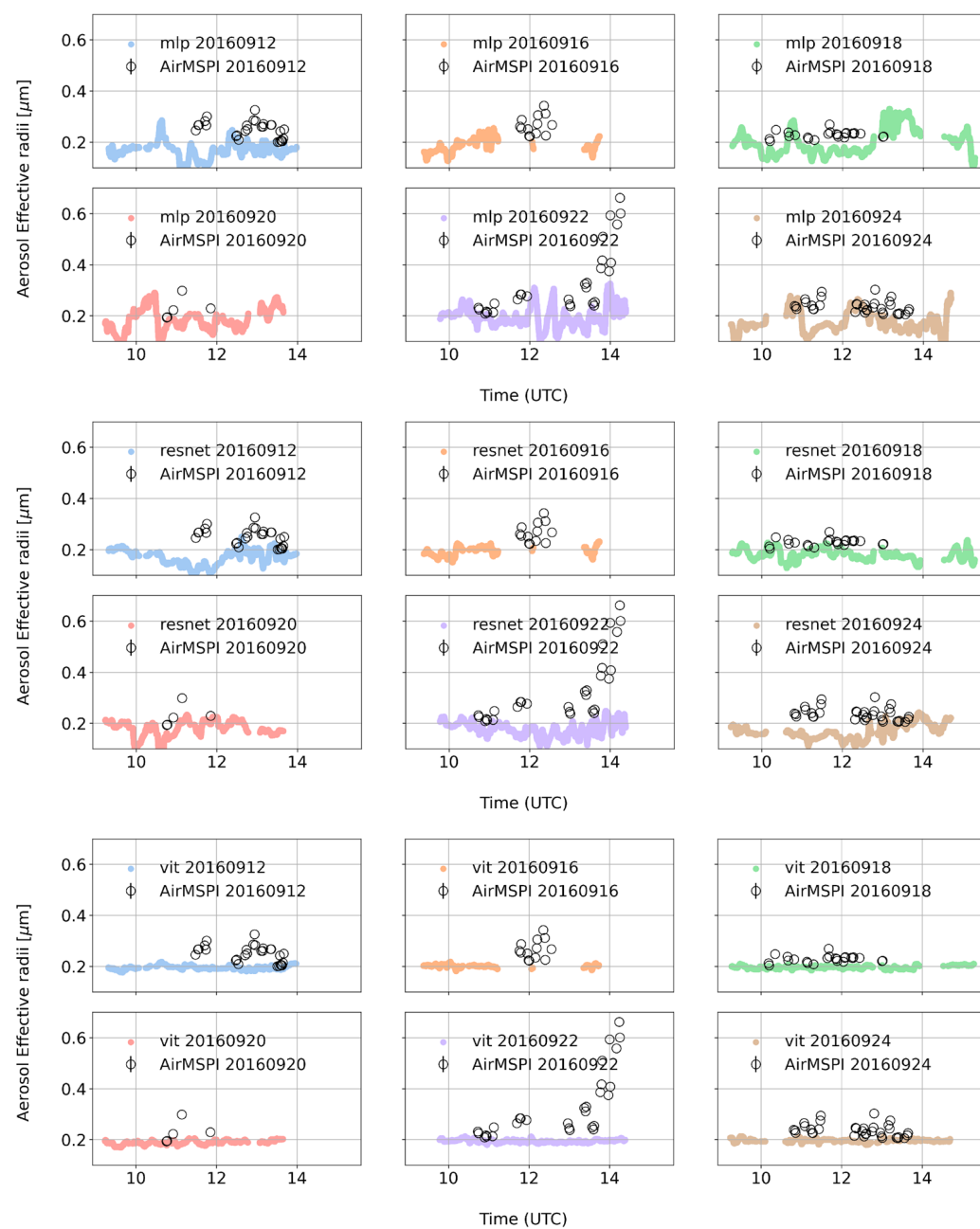


Figure 21. Similar to Figure 19 but for fine Aerosol effective radii comparisons.

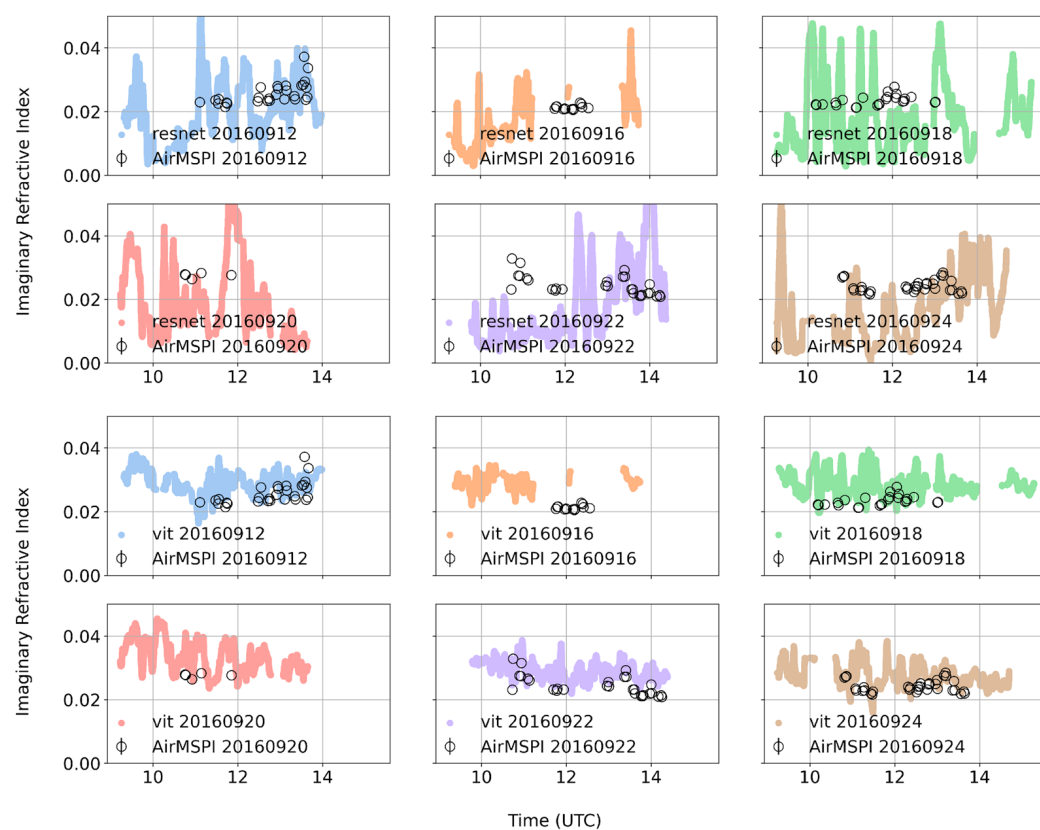


Figure 22. Similar to Figure 19 but for imaginary refractive index comparisons.

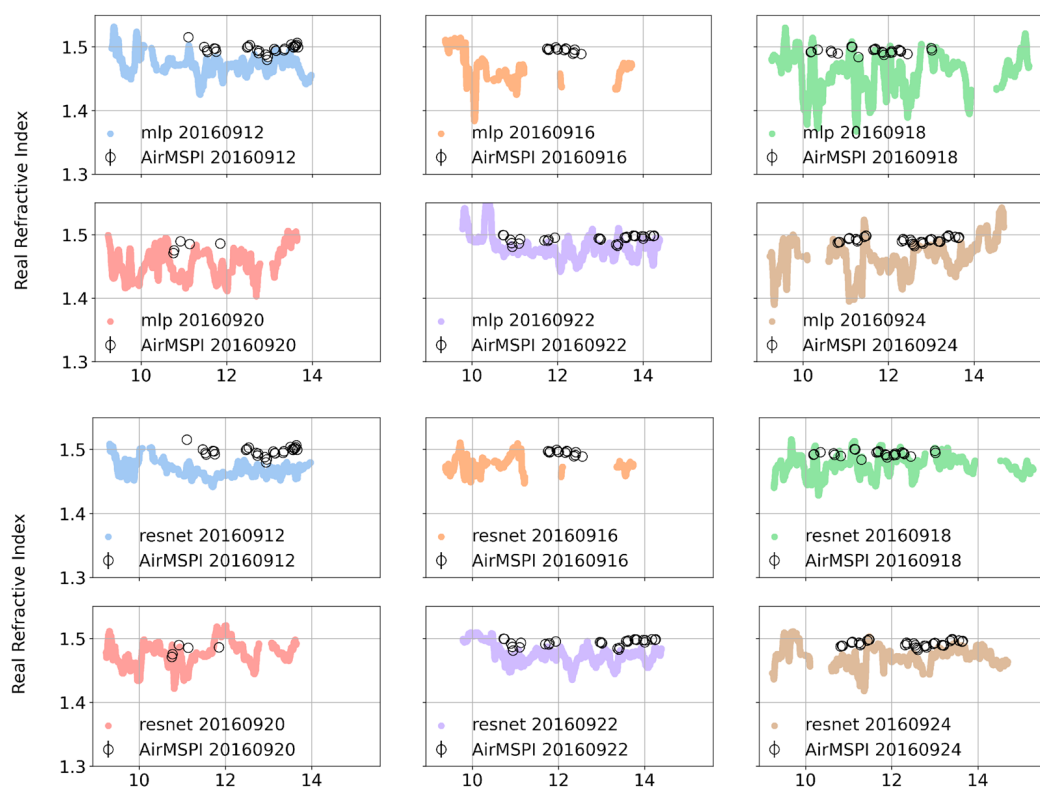


Figure 23. Cont.

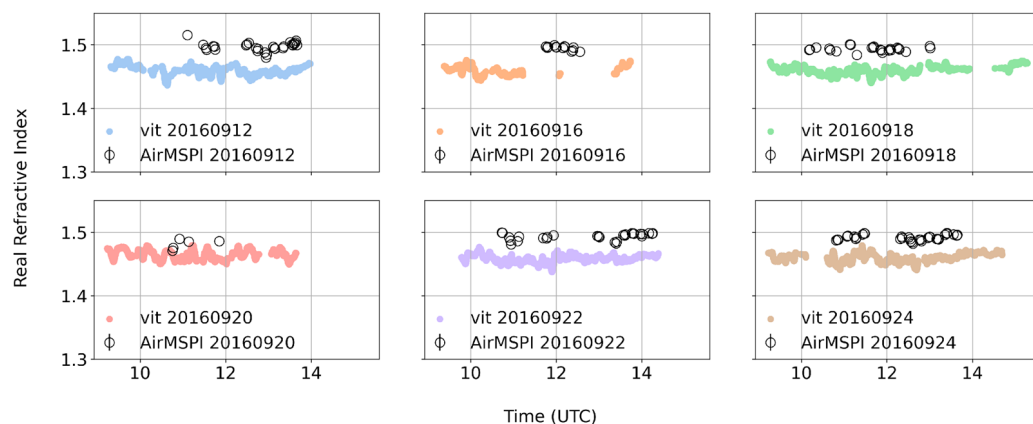


Figure 23. Similar to Figure 19 but for real refractive index comparisons.

We note that while RSP deep-learning retrievals yield a continuous output, AirMSPI retrieved values are much sparser. This is probably the result of their larger scene aggregation due to the larger image swath (100×25 km) compared with the 323 m scanning line of RSP.

For SSA, the two instruments show relatively similar values, with the ResNET and ViT schemes giving the closest values, and ViT showing less variability. Furthermore, our ViT results are aligned with campaign-wide SSA values for ORACLES 2016, as reported by Pistone et al. [39]. Such a good agreement (see Figure 24) between these two instruments is encouraging, since SSA is often used as an a priori assumption in passive instrument ACA retrievals, e.g., [12], which can affect AOD and cloud property values greatly (~ 10 – 50%) [52]. Therefore, lowering the uncertainty of the SSA values retrieved by polarimetric instruments will lower ACA retrieval uncertainties when combining them into radiometric only passive sensors retrieval schemes. Figure 20 is similar to Figure 19, but only showing the ViT retrieved AOD values, compared with those of AirMSPI. MLP and ResNet did not show any skill in predicting ACA AOD from RSP.

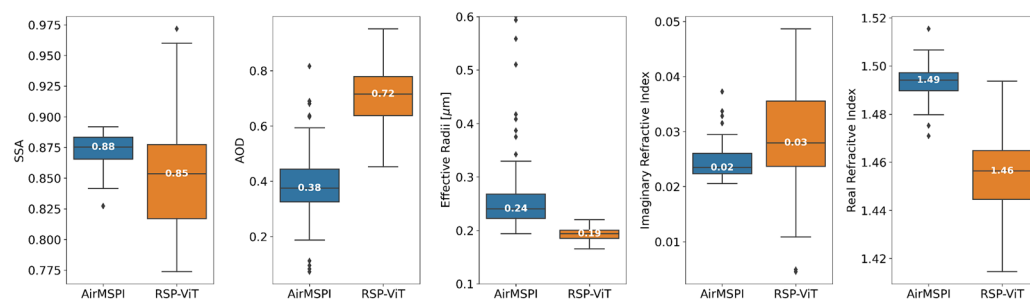


Figure 24. Boxplot (solid lines represent median values, with white text, and 5th and 95th percentiles) comparing AirMSPI (blue) with Vit-RSP retrievals (orange) above cloud aerosol microphysical property retrievals. Data points represent values beyond 5 and 95 percentiles.

The AOD values retrieved by our deep learning models is biased high in all tested models, with MLP and ResNet models producing very noisy and inaccurate results with no predictive power, compared to the ViT model. Indeed, Figures 13–15 show a retrieval bias for AOD, especially for ViT at higher AOD values. One reason for this bias is some inherent differences between the training dataset distribution and the real observational-based distribution. While other aerosol property values are better constrained (they have a well-defined range), AOD values are varied and do not have a rigid upper bound. Other reasons for this bias can result from the difference in the aerosol model treatment of each of the methodologies. Here, we trained our networks with one fine and one coarse mode

aerosol distributions, while AirMSPI retrievals are constructed to fit to a five-mode aerosol model (with three fine and two coarse mode aerosol models).

Figure 21 compares the fine aerosol effective radii retrievals between the two instruments. Such a comparison between the two instruments was not performed for ORACLES. Here, ResNet and ViT models produce effective radii values that are closest to AirMSPI, with less variability than AirMSPI. Still, the ViT model generates suboptimal results in terms of the variation in the aerosol effective radii values. Similar results were obtained when comparing the real refractive indices retrievals, in Figure 23 below, where the MLP and ResNet model produced closer values than the ViT model. In Figure 22, on the other hand, which compares the imaginary refractive index, ViT showed closer and more stable retrievals, mirroring the results obtained for SSA.

Figure 24 shows the aggregated boxplot statistics comparing the ViT RSP retrievals and AirMSPI retrievals for the variables shown in Figures 19–23. Values were taken only for time collocations between available AirMSPI retrievals and RSP retrievals for the ViT experiment, as in most cases MLP and ResNet resulted in larger variation range in the retrieved variables. The SSA median difference is 0.03, which is within the SSA accuracy requirement for space observations, as defined by the recent NASA decadal survey [53]. RSP ACA retrievals resulted in values that represent slightly more absorbing aerosol (lower SSA and higher imaginary index).

Despite the high bias in AOD for RSP deep-learning retrieval models, the correlation between the predicted RSP AOD and the AirMSPI retrieved AOD is relatively high (Figure 25), strengthening the assumption that the real-values AOD distribution is shifted toward lower AOD values relative to the training set. As future work, corrections for such biases should include an adversarial training component to allow for the model to adapt its results to slightly different distributions obtained by the real measurements versus the synthetic dataset, as demonstrated in [27].

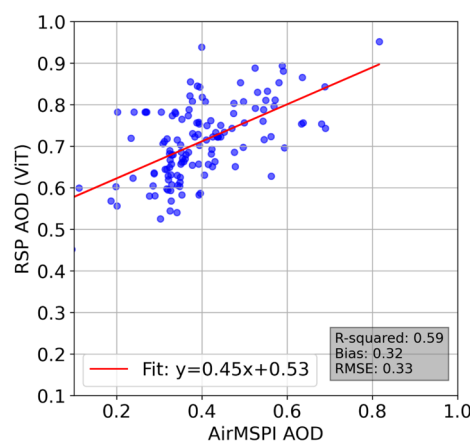


Figure 25. Scatter plot comparing RSP AOD values retrieved using our ViT model vs. AirMSPI ACA retrieved AOD for closest time collocations for ORACLES 2016 flights.

Comparing with HSRL

As AOD values retrieved by our models showed high bias compared with AirMSPI, we re-trained our ViT network to use, as inputs, a constrained AOD range of 0–1, instead of 0–3 as the original training distribution (see Figure 1), but kept the distribution shape similar. In Figure 26, we compare our results with AirMSPI and with the NASA Langley airborne HSRL (High Spectral Resolution Lidar) [54], flown concurrently on the ER-2 aircraft during the ORACLES 2016 campaign. The HSRL flown in ORACLES is the second generation HSRL, and it uses an independent measure of aerosol backscatter and extinction. We used the archived above cloud AOD product at a wavelength of 532 nm. As seen in

Figure 26, the new retrieved values are closer now to HSRL (black), and AirMSPI values (red), compared to the comparison shown in Figure 20 for AirMSPI. The overall bias is reduced by AOD mean values of 0.06 (0.66 versus 0.72 previously), with a smaller spread than before.

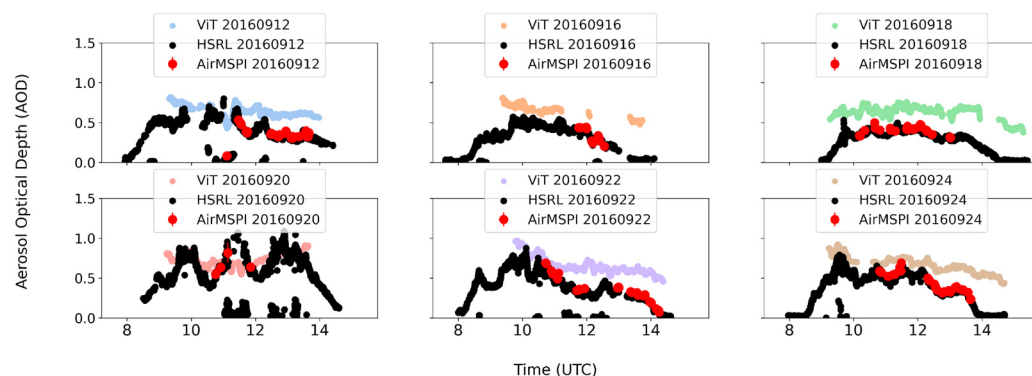


Figure 26. AOD time-series comparison between HSRL (black circles), AirMSPI (red circles), and RSP ViT retrievals (color coded by date).

5. Discussion

In this paper, we constructed two deep learning algorithms to retrieve aerosol and cloud properties from polarimetric measurements of above cloud absorbing aerosol scenes. These scenes are prevalent over the SEA ocean, where BB from the continent transport and overlay low level liquid marine clouds. Optical property retrievals for such scenes are important for the assessment of the radiative budget and cloud processes (formation and burn-off) of the region.

Standard retrieval algorithms (such as AirMSPI ACA or RSP standard algorithms) have several intrinsic limitations that restrict their applicability and accuracy. These limitations include:

1. Limited spectral information: retrievals often rely on only one or two wavelengths, missing valuable information encoded across broader spectral ranges.
2. Separate aerosol and cloud retrieval schemes: aerosols and clouds are typically retrieved independently, each utilizing different regions of spectral and angular measurements. This separation inherently reduces the accuracy and consistency of combined aerosol-cloud retrievals and hinders accurate quantification of aerosol-cloud interactions.
3. Strongly simplified assumptions: many traditional retrieval algorithms assume horizontally homogeneous aerosol layers extending over large spatial scales, as well as fixed empirical relationships between cloud optical depth (COD) and effective radius (R_{eff}). These simplifying assumptions often fail under realistic atmospheric conditions, reducing retrieval accuracy and limiting the valid retrievals across heterogeneous cloud-aerosol scenes, as observed during ORACLES. Furthermore, algorithms like the RSP MAPP (Microphysical Aerosol Properties from Polarimetry) retrievals can only provide aerosol properties under clear-sky conditions, severely limiting the quantity and representativeness of the retrieval data in cloudy regions.

In contrast, our proposed deep learning (DL) methods overcome these critical limitations by simultaneously retrieving aerosol and cloud properties using richer and more comprehensive scene information (full spectral, angular, and polarization signals). The more sophisticated neural architectures (especially Vision Transformers) explicitly exploit complex global dependencies in multi-angle polarized reflectances. As a result, the deep learning approach significantly improves spatial resolution and retrieval accuracy, enabling

the derivation of aerosol-cloud relationships at much finer scales compared to AirMSPI and RSP standard retrievals.

Additionally, the DL approach provides substantial computational advantages. Traditional algorithms rely on iterative online radiative transfer computations, making retrievals extremely computationally demanding, particularly in aerosol–cloud scenes due to the strongly forward-scattering nature of clouds and the large parameter space involved. In contrast, DL approaches, although computationally intensive during the initial training stage, perform inference (retrieval) almost instantaneously regardless of scene complexity. Thus, the computational cost of the DL approach does not scale significantly with complexity or number of retrieved parameters, making it particularly well-suited for real-time applications and operational scenarios involving large datasets.

We tested two different DL algorithm types: transformer (ViT) and CNN (ResNet), where a MLP model was used as the current benchmark. Overall, as shown above, the ViT model, traditionally used by natural language models is superior to the ResNet and MLP models, both in its capability to retrieve values with low RMSE and high correlation compared with our validation datasets (from RSP standard algorithms and AirMSPI ACA algorithm) and in its retrieval stability (low dispersion of retrieved values). Our results corroborate previous investigations that compared ViT and CNN for image classification tasks and found that ViTs are superior to CNNs (e.g., [47]). We interpret the ViT's superior performance as resulting primarily from its inherent ability to model long-range, global dependencies across multi-angular and spectral features in the polarized reflectance data, coupled with stable training. In contrast, the MLP architecture likely suffers from overfitting or insufficient representational capacity when dealing with the complex structure of polarimetric signals (especially for challenging retrievals like V_{eff}). Meanwhile, the convolutional inductive biases of ResNet—originally optimized for spatially structured natural images—do not effectively align with the inherently spectral-angular structure of multi-angle polarized reflectance measurements. This misalignment in inductive biases reduces the effectiveness of ResNet for these specific retrieval tasks.

In general, we found that our test set validation performs very well, with $R^2 > 93\%$ for all eight variables. We found that our retrieved SSA gives excellent results, with RMSE of 0.01 for our test set validation, and a RMSE of 0.03 when validating our results with AirMSPI for ORACLES 2016. To compare, the FastMAPOL algorithm, which combines OE iterative estimation procedure and NN emulators, achieved RMSE of 0.054 for their SSA test set validation.

We compared cloud property retrievals to the standard RSP algorithms (PP and NJK) and received very good correlation for COD and R_{eff} , comparable to the results by Miller et al. [17]. For V_{eff} , comparisons and correlations were generally lower, but still much better than the comparisons in Miller et al., and Xu et al. [15,17] (~ 0.6 – 0.7 vs. ~ 0.2). We note that cloud property similarity values are better for ORACLES 2016, which can be explained by the fact that in 2016 the aerosol above cloud scenes were the most similar to the scenes used for training our models. Later ORACLES years (2017–2018) had less BB aerosol amounts and more spatially inhomogeneous cloud scenes that would be less consistent with the training model assumptions.

Although we are using a training set that is defined for the low flyer (~ 6 km), which flew in 2017–2018, the 2016 comparisons between our retrievals and the RSP standard retrievals are better correlated despite the fact that in 2016 the RSP was flown on a high flyer platform (~ 20 km). We postulate that the aircraft altitude did not affect the trained model much, as the major difference would have been the atmospheric columnar gas content. However, since we used the non-gaseous absorption wavelengths of the instrument, the aircraft altitude effect has a minor effect, if at all.

In fact, we noticed that the scene specifications matter more. For example, 2016 saw the highest BB AOD amount and more low-level clouds, since the campaign concentrated over the southern domain of the SEA. Years 2017–2018 saw lower AOD levels and a more complex cloud scenes, including also mid-level clouds, as the flights concentrated over the middle-to-northern SEA domain. In this latter region, aerosol retrievals have shown higher noise and variability compared with the 2016 aerosol retrievals.

When comparing aerosol properties between our retrievals and AirMSPI ACA algorithm, we receive excellent agreement for corresponding flight times for all variables, except for AOD values, which are biased high in our retrievals. This bias can be the result of several reasons, among them a distributional shift between the training and the observational datasets, i.e., a possible domain gap which exists between the training set and the real measurements. We note that this is an inherent DL caveat, where if the training and real-world distributions do not align, the network will produce biased results [55].

One solution to solving this caveat is to apply domain adaptation methodologies during the training process. Several options exist, among them joint training, fine-tuning, and data augmentation [27,56,57] that are adequate for our supervised training case. As detailed above, we tested the fine-tuning methodology, where we re-trained the network on a smaller AOD distribution (0–1 instead of 0–3) and achieved a lower AOD value (average values of AOD were lowered by 0.06) and better comparison with HSRL. Other available options are adding training samples of the “target” domain, i.e., real RSP measurements. However, since the “truth” values of these measurements are actually retrieved values by the RSP standard algorithms, we risk that the predicted values by the network will converge to the RSP retrieved values. Another option is Test-time adaptation [58], which is an unsupervised method for model adaptation, and will be explored in future work to allow a more generalized retrieval training approach.

Another reason for the high AOD bias may be the fact that our simulations did not include gaseous absorption, and therefore, the real observations had to be corrected for any gaseous effect, including water vapor. As detailed in Section 2.2, we used the above cloud water vapor pressure from the MERRA-2 model reanalysis to correct for trace gas absorption, which might have resulted in a correction that leads to lower reflectance (higher absorption) values, thereby artificially increasing the AOD values as retrieved by our model. We note here that MERRA-2 water vapor amount above the aircraft was shown to be either high or low biased compared to real measurements during ORACLES 2016.

Finally, we note the difference between the underlying aerosol models in our training set versus AirMSPI [15]. While we used one fine and one coarse mode aerosol models in our simulations, AirMSPI integrated five models (three fine mode and two coarse modes, with coarse mode fraction between 0 and 0.5) into their optimal estimation retrieval scheme. Although our simulated set was based on real measurements from ORACLES, we postulate that the simplified two component aerosol model, and the lower limit constraint on the fine fraction (above 0.85) resulted in a less flexible domain fit, which forced the results to be biased high.

Nevertheless, despite the high absolute bias, the retrieved AOD values did not have a significant effect on other retrieved variables. For example, when comparing retrieved cloud and aerosol variables values for high AOD (>2) and low AOD (<2) cases, we observed that SSA was not affected, whereas COD was only slightly (~ 0.25 OD) underestimated for higher AOD, and both COD and R_{eff} had a higher spread.

Our retrieval can be further generalized as an end-to-end procedure for retrieving aerosol above clouds from different sets of instruments. To be compatible with instruments other than RSP, the training should take into account the different measurements (wavelengths and viewing angles) obtained by the different instrument. To be conducted in a

more generic way, simulations can span a large wavelength and viewing angle options (similar to the simulated scenes here), but the training should only utilize the available measurables (i.e., specific wavelengths and viewing angles). In addition, accounting for more scenarios during the training phase, such as large dust loads, various amounts of water vapor conditions, etc., would make the scheme more generalized. The key to the success of the DL methodology is to provide simulations as close to real conditions as possible, or to include domain gap solution methodologies to avoid retrieval discrepancies when measurements from modified instruments are used. In future work, we will consider utilizing simulations augmented by real observations, to balance the trained model and its generalization.

Author Contributions: M.S.R., Paper conceptualization, funding acquisition, methodology, data curation, formal analysis, visualization, software, and writing original draft; K.K., funding acquisition, methodology, simulations, and review and editing; D.M., simulations, RSP data corrections and review and editing; D.B., methodology, formal analysis, data analysis, software and review and editing; M.S.R., K.K., D.B. and D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the NASA ORACLES Earth Venture Suborbital campaign, NASA NNH13ZDA001N-EVS2 managed by Dr. Hal Maring. M.S.R has been funded by the ORACLES grant NASA NNH13ZDA001N-EVS2, Tel-Aviv University, and funds from NASA AOS science team through NASA Ames Research Center. The computational resources were partially funded by the Israel Science foundation, Grant 2036/20, and the training set was generated on infrastructure at the NASA Goddard Space Flight Center. KK was funded by the aforementioned ORACLES campaign, while DM was funded by a NASA Postdoctoral Program fellowship.

Data Availability Statement: The ORACLES data (RSP, AirMSPI, HSRL) used in this manuscript is publicly available at: <https://espoarchive.nasa.gov/archive/browse/oracles>. URL (accessed on 15 September 2024). The simulations and model results are available per request from the authors.

Acknowledgments: We thank the AirMSPI, RSP, and HSRL teams for the data that was used here. We also thank the ORACLES campaign and science team.

Conflicts of Interest: The corresponding author is a guest-editor at the Remote Sensing Journal. Other authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Redemann, J.; Wood, R.; Zuidema, P.; Doherty, S.J.; Luna, B.; LeBlanc, S.E.; Diamond, M.S.; Shinozuka, Y.; Chang, I.Y.; Ueyama, R.; et al. An overview of the ORACLES (ObseRVations of Aerosols above CLouds and their intERactionS) project: Aerosol–cloud–radiation interactions in the southeast Atlantic basin. *Atmos. Meas. Tech.* **2021**, *21*, 1507–1563. [\[CrossRef\]](#)
2. Muhlbauer, A.; McCoy, I.L.; Wood, R. Climatology of stratocumulus cloud morphologies: Microphysical properties and radiative effects. *Atmos. Meas. Tech.* **2014**, *14*, 6695–6716. [\[CrossRef\]](#)
3. Wood, R. Stratocumulus Clouds. *Mon. Weather. Rev.* **2012**, *140*, 2373–2423. [\[CrossRef\]](#)
4. Zuidema, P.; Redemann, J.; Haywood, J.; Wood, R.; Piketh, S.; Hipondoka, M.; Formenti, P. Smoke and Clouds above the Southeast Atlantic: Upcoming Field Campaigns Probe Absorbing Aerosol’s Impact on Climate. *Bull. Am. Meteorol. Soc.* **2016**, *97*, 1131–1135. [\[CrossRef\]](#)
5. Haywood, J.M.; Abel, S.J.; Barrett, P.A.; Bellouin, N.; Blyth, A.; Bower, K.N.; Brooks, M.; Carslaw, K.; Che, H.; Coe, H.; et al. The CLoud–Aerosol–Radiation Interaction and Forcing: Year 2017 (CLARIFY-2017) measurement campaign. *Atmos. Chem. Phys.* **2021**, *21*, 1049–1084. [\[CrossRef\]](#)
6. Painemal, D.; Kato, S.; Minnis, P. Boundary layer regulation in the southeast Atlantic cloud microphysics during the biomass burning season as seen by the A-train satellite constellation. *J. Geophys. Res. Atmos.* **2014**, *119*, 11288–11302. [\[CrossRef\]](#)
7. Sakaeda, N.; Wood, R.; Rasch, P.J. Direct and semidirect aerosol effects of southern African biomass burning aerosol. *J. Geophys. Res.* **2011**, *116*, D12205-1-19. [\[CrossRef\]](#)

8. Chang, I.; Christopher, S.A. Identifying Absorbing Aerosols Above Clouds From the Spinning Enhanced Visible and Infrared Imager Coupled With NASA A-Train Multiple Sensors. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3163–3173. [\[CrossRef\]](#)
9. Waquet, F.; P  r  , J.; Peers, F.; Goloub, P.; Ducos, F.; Thieuleux, F.; Tanr  , D. Global detection of absorbing aerosols over the ocean in the red and near-infrared spectral region. *J. Geophys. Res. Atmos.* **2016**, *121*, 10902–10918. [\[CrossRef\]](#)
10. McCoy, I.L.; Wood, R.; Fletcher, J.K. Identifying Meteorological Controls on Open and Closed Mesoscale Cellular Convection Associated with Marine Cold Air Outbreaks. *J. Geophys. Res. Atmos.* **2017**, *122*, 11678–11702. [\[CrossRef\]](#)
11. Peers, F.; Francis, P.; Fox, C.; Abel, S.J.; Szpek, K.; Cotterell, M.I.; Davies, N.W.; Langridge, J.M.; Meyer, K.G.; Platnick, S.E.; et al. Observation of absorbing aerosols above clouds over the south-east Atlantic Ocean from the geostationary satellite SEVIRI—Part 1: Method description and sensitivity. *Atmos. Meas. Tech.* **2019**, *19*, 9595–9611. [\[CrossRef\]](#)
12. Meyer, K.; Platnick, S.; Zhang, Z. Simultaneously inferring above-cloud absorbing aerosol optical thickness and underlying liquid phase cloud optical and microphysical properties using MODIS. *J. Geophys. Res. Atmos.* **2015**, *120*, 5524–5547. [\[CrossRef\]](#)
13. Knobelspiesse, K.; Cairns, B.; Redemann, J.; Bergstrom, R.W.; Stohl, A. Simultaneous retrieval of aerosol and cloud properties during the MILAGRO field campaign. *Atmos. Meas. Tech.* **2011**, *11*, 6245–6263. [\[CrossRef\]](#)
14. Waquet, F.; Cornet, C.; Deuz  , J.-L.; Dubovik, O.; Ducos, F.; Goloub, P.; Herman, M.; Lapyonok, T.; Labonnote, L.C.; Riedi, J.; et al. Retrieval of aerosol microphysical and optical properties above liquid clouds from POLDER/PARASOL polarization measurements. *Atmos. Meas. Tech.* **2013**, *6*, 991–1016. [\[CrossRef\]](#)
15. Xu, F.; van Harten, G.; Diner, D.J.; Davis, A.B.; Seidel, F.C.; Rheingans, B.; Tosca, M.; Alexandrov, M.D.; Cairns, B.; Ferrare, R.A.; et al. Coupled Retrieval of Liquid Water Cloud and Above-Cloud Aerosol Properties Using the Airborne Multiangle SpectroPolarimetric Imager (AirMSPI). *J. Geophys. Res. Atmos.* **2018**, *123*, 3175–3204. [\[CrossRef\]](#)
16. Di Noia, A.; Hasekamp, O.P.; van Diedenhoven, B.; Zhang, Z. Retrieval of liquid water cloud properties from POLDER-3 measurements using a neural network ensemble approach. *Atmos. Meas. Tech.* **2019**, *12*, 1697–1716. [\[CrossRef\]](#)
17. Miller, D.J.; Segal-Rozenhaimer, M.; Knobelspiesse, K.; Redemann, J.; Cairns, B.; Alexandrov, M.; van Diedenhoven, B.; Wasilewski, A. Low-level liquid cloud properties during ORACLES retrieved using airborne polarimetric measurements and a neural network algorithm. *Atmos. Meas. Tech.* **2020**, *13*, 3447–3470. [\[CrossRef\]](#)
18. Segal-Rozenhaimer, M.; Miller, D.J.; Knobelspiesse, K.; Redemann, J.; Cairns, B.; Alexandrov, M.D. Development of neural network retrievals of liquid cloud properties from multi-angle polarimetric observations. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, *220*, 39–51. [\[CrossRef\]](#)
19. Di Noia, A.; Hasekamp, O.P.; Wu, L.; van Diedenhoven, B.; Cairns, B.; Yorks, J.E. Combined neural network/Phillips–Tikhonov approach to aerosol retrievals over land from the NASA Research Scanning Polarimeter. *Atmos. Meas. Tech.* **2017**, *10*, 4235–4252. [\[CrossRef\]](#)
20. Gao, M.; Franz, B.A.; Zhai, P.-W.; Knobelspiesse, K.; Sayer, A.M.; Xu, X.; Martins, J.V.; Cairns, B.; Castellanos, P.; Fu, G.; et al. Simultaneous retrieval of aerosol and ocean properties from PACE HARP2 with uncertainty assessment using cascading neural network radiative transfer models. *Atmos. Meas. Tech.* **2023**, *16*, 5863–5881. [\[CrossRef\]](#)
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)
22. Koonce, B. ResNet 50. In *Convolutional Neural Networks with Swift for Tensorflow*; Apress: Berkeley, CA, USA, 2021. [\[CrossRef\]](#)
23. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Rozenhaimer, M.S.; Nukrai, D.; Che, H.; Wood, R.; Zhang, Z. Cloud Mesoscale Cellular Classification and Diurnal Cycle Using a Convolutional Neural Network (CNN). *Remote Sens.* **2023**, *15*, 1607. [\[CrossRef\]](#)
25. Zhong, Y.; Fei, F.; Liu, Y.; Zhao, B.; Jiao, H.; Zhang, L. SatCNN: Satellite image dataset classification using agile convolutional neural networks. *Remote Sens. Lett.* **2016**, *8*, 136–145. [\[CrossRef\]](#)
26. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [\[CrossRef\]](#)
27. Segal-Rozenhaimer, M.; Li, A.; Das, K.; Chirayath, V. Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (CNN). *Remote Sens. Environ.* **2020**, *237*, 111446. [\[CrossRef\]](#)
28. Malmgren-Hansen, D.; Laparra, V.; Nielsen, A.A.; Camps-Valls, G. Statistical retrieval of atmospheric profiles with deep convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 231–240. [\[CrossRef\]](#)
29. Martins, J.V.; Fernandez-Borda, R.; McBride, B.; Remer, L.; Barbosa, H.M.J. The Harp Hype Ran Gular Imaging Polarimeter and the Need for Small Satellite Payloads with High Science Payoff for Earth Science Remote Sensing. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6304–6307.
30. McBride, B.A.; Martins, J.; Puthukuddy, A.; Xu, X.; Borda, R.F.; Barbosa, H.M.J.; Hasekamp, O.; Remer, L.A. The Hyper-Angular Rainbow Polarimeter-2 (HARP2): A wide FOV polarimetric imager for high-resolution spatial and angular characterization of aerosol and cloud microphysics. In Proceedings of the 70th International Astronautical Congress (IAC), Washington, DC, USA, 21–25 October 2019.

31. Werdell, P.J.; Behrenfeld, M.J.; Bontempi, P.S.; Boss, E.; Cairns, B.; Davis, G.T.; Franz, B.A.; Gliese, U.B.; Gorman, E.T.; Hasekamp, O.; et al. The Plankton, Aerosol, Cloud, Ocean Ecosystem Mission: Status, Science, Advances. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 1775–1794. [\[CrossRef\]](#)
32. Werdell, P.J.; Franz, B.; Poulin, C.; Allen, J.; Cairns, B.; Caplan, S.; Cetinić, I.; Craig, S.; Gao, M.; Hasekamp, O.; et al. Life after launch: A snapshot of the first six months of NASA's plankton, aerosol, cloud, ocean ecosystem (PACE) mission. In *Proceedings of the Sensors, Systems, and Next-Generation Satellites XXVIII, Edinburgh, UK, 16–18 September 2024*; SPIE: Pune, India, 2024; Volume 13192, pp. 70–84.
33. Cairns, B.; Russell, E.E.; LaVeigne, J.D.; Tennant, P.M. Research scanning polarimeter and airborne usage for remote sensing of aerosols. *Polariz. Sci. Remote Sens.* **2003**, *5158*, 33–44.
34. Hansen, J.E.; Travis, L.D. Light scattering in planetary atmospheres. *Space Sci. Rev.* **1974**, *16*, 527–610. [\[CrossRef\]](#)
35. Zuidema, P.; Chiu, C.; Fairall, C.W.; Ghan, S.J.; Kollias, P.; McFarquhar, G.M.; Mechem, D.B.; Romps, D.M.; Wong, H.; Yuter, S.E.; et al. Layered Atlantic Smoke Interactions with Clouds (LASIC) Science Plan. (No. DOE/SC-ARM-14-037). DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States). 2015. Available online: <https://www.arm.gov/research/campaigns/amf2016lasic> (accessed on 8 May 2025).
36. Gupta, S.; McFarquhar, G.M.; O'Brien, J.R.; Delene, D.J.; Poellot, M.R.; Dobracki, A.; Podolske, J.R.; Redemann, J.; LeBlanc, S.E.; Segal-Rozenhaimer, M.; et al. Impact of the variability in vertical separation between biomass burning aerosols and marine stratocumulus on cloud microphysical properties over the Southeast Atlantic. *Atmos. Meas. Tech.* **2021**, *21*, 4615–4635. [\[CrossRef\]](#)
37. Zhang, L.; Segal-Rozenhaimer, M.; Che, H.; Dang, C.; Iii, A.J.S.; Lewis, E.R.; Dobracki, A.; Wong, J.P.S.; Formenti, P.; Howell, S.G.; et al. Light absorption by brown carbon over the South-East Atlantic Ocean. *Atmos. Meas. Tech.* **2022**, *22*, 9199–9213. [\[CrossRef\]](#)
38. Zuidema, P.; Sedlacek, A.J.; Flynn, C.; Springston, S.; Delgadillo, R.; Zhang, J.; Aiken, A.C.; Koontz, A.; Muradyan, P. The Ascension Island Boundary Layer in the Remote Southeast Atlantic is Often Smoky. *Geophys. Res. Lett.* **2018**, *45*, 4456–4465. [\[CrossRef\]](#)
39. Pistone, K.; Redemann, J.; Doherty, S.; Zuidema, P.; Burton, S.; Cairns, B.; Cochrane, S.; Ferrare, R.; Flynn, C.; Freitag, S.; et al. Intercomparison of biomass burning aerosol optical properties from in situ and remote-sensing instruments in ORACLES-2016. *Atmos. Meas. Tech.* **2019**, *19*, 9181–9208. [\[CrossRef\]](#)
40. Knobelspiesse, K.; Tan, Q.; Bruegge, C.; Cairns, B.; Chowdhary, J.; van Diedenhoven, B.; Diner, D.; Ferrare, R.; van Harten, G.; Jovanovic, V.; et al. Intercomparison of airborne multi-angle polarimeter observations from the Polarimeter Definition Experiment. *Appl. Opt.* **2019**, *58*, 650–669. [\[CrossRef\]](#)
41. LeBlanc, S.E.; Redemann, J.; Flynn, C.; Pistone, K.; Kacenelenbogen, M.; Segal-Rosenheimer, M.; Shinozuka, Y.; Dunagan, S.; Dahlgren, R.P.; Meyer, K.; et al. Above-cloud aerosol optical depth from airborne observations in the southeast Atlantic. *Atmos. Chem. Phys.* **2020**, *20*, 1565–1590. [\[CrossRef\]](#)
42. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.
43. Mascarenhas, S.; Agarwal, M. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In *Proceedings of the 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, Bengaluru, India, 19–21 November 2021; pp. 96–99.
44. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
45. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
46. Chen, X.; Hsieh, C.-J.; Gong, B. When Vision Transformers Outperform ResNets without Pre-training or Strong Data Augmentations. *arXiv* **2022**, arXiv:2106.01548.
47. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
48. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, Sardinia, Italy, 13–15 May 2010*; pp. 249–256.
49. Alexandrov, M.D.; Cairns, B.; Emde, C.; Ackerman, A.S.; van Diedenhoven, B. Accuracy assessments of cloud droplet size retrievals from polarized reflectance measurements by the research scanning polarimeter. *Remote Sens. Environ.* **2012**, *125*, 92–111. [\[CrossRef\]](#)
50. Miller, D.J.; Zhang, Z.; Platnick, S.; Ackerman, A.S.; Werner, F.; Cornet, C.; Knobelspiesse, K. Comparisons of bispectral and polarimetric retrievals of marine boundary layer cloud microphysics: Case studies using a LES–satellite retrieval simulator. *Atmos. Meas. Tech.* **2018**, *11*, 3689–3715. [\[CrossRef\]](#)
51. Diner, D.J.; Xu, F.; Garay, M.J.; Martonchik, J.V.; Rheingans, B.E.; Geier, S.; Davis, A.; Hancock, B.R.; Jovanovic, V.M.; Bull, M.A.; et al. The Airborne Multiangle SpectroPolarimetric Imager (AirMSPI): A new tool for aerosol and cloud remote sensing. *Atmos. Meas. Tech.* **2013**, *6*, 2007–2025. [\[CrossRef\]](#)

52. Jethva, H.T.; Torres, O.; Ferrare, R.A.; Burton, S.P.; Cook, A.L.; Harper, D.B.; Hostetler, C.A.; Redemann, J.; Kayetha, V.; LeBlanc, S.; et al. Retrieving UV–Vis spectral single-scattering albedo of absorbing aerosols above clouds from synergy of ORACLES airborne and A-train sensors. *Atmos. Meas. Tech.* **2024**, *17*, 2335–2366. [[CrossRef](#)]
53. *Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space: An Overview for Decision Makers and the Public*; The National Academies Press: Washington, DC, USA, 2018. [[CrossRef](#)]
54. Hair, J.W.; Hostetler, C.A.; Cook, A.L.; Harper, D.B.; Ferrare, R.A.; Mack, T.L.; Welch, W.; Izquierdo, L.R.; Hovis, F.E. Airborne High Spectral Resolution Lidar for profiling aerosol optical properties. *Appl. Opt.* **2008**, *47*, 6734–6752. [[CrossRef](#)] [[PubMed](#)]
55. Wang, J.; Lan, C.; Liu, C.; Ouyang, Y.; Qin, T.; Lu, W.; Chen, Y.; Zeng, W.; Yu, P. Generalizing to Unseen Domains: A Survey on Domain Generalization. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 8052–8072. [[CrossRef](#)]
56. Kellenberger, B.; Tasar, O.; Damodaran, B.B.; Courty, N.; Tuia, D. Deep domain adaptation in earth observation. In *Deep Learning for the Earth Sciences*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2021; Chapter 7; pp. 90–104. [[CrossRef](#)]
57. Mateo-García, G.; Laparra, V.; López-Puigdollers, D.; Gómez-Chova, L. Transferring deep learning models for cloud detection between Landsat-8 and Proba-V. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 1–17. [[CrossRef](#)]
58. Fang, Y.; Yap, P.-T.; Lin, W.; Zhu, H.; Liu, M. Source-free unsupervised domain adaptation: A survey. *Neural Netw.* **2024**, *174*, 106230. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.