https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

3-D Cloud Masking Across a Broad Swath using Multi-angle Polarimetry and Deep Learning

Sean R. Foley[1,2], Kirk D. Knobelspiesse[1], Andrew M. Sayer[1,3], Meng Gao[1,4], James Hays[5], and Judy Hoffman[5]

[1]NASA Goddard Space Flight Center, Code 616, Greenbelt, MD 20771, USA
[2]Goddard Earth Sciences Technology and Research (GESTAR) II, Morgan State University, Baltimore, MD, USA
[3]Goddard Earth Sciences Technology and Research (GESTAR) II, University of Maryland, Baltimore County, Baltimore, MD, USA
[4]Science Systems and Applications, Inc., Greenbelt, MD, USA
[5]Georgia Institute of Technology, Atlanta, GA, USA

Correspondence: Sean Foley (sean.r.foley@nasa.gov)

Abstract. Understanding the 3-dimensional structure of clouds is of crucial importance to modeling our changing climate. Active sensors, such as radar and lidar, provide accurate vertical cloud profiles, but are mostly restricted to along-track sampling. Passive sensors can capture a wide swath, but struggle to see beneath cloud tops. In essence, both types of products are restricted to two dimensions: as a cross-section in the active case, and an image in the passive case. However, multi-angle sensor configurations contain implicit information about 3D structure, due to parallax and atmospheric path differences. Extracting that implicit information can be challenging, requiring computationally expensive radiative transfer techniques that must make limiting assumptions. Machine learning, as an alternative, may be able to capture some of the complexity of a full 3D radiative transfer solution with significantly less computational expense. In this work, we make three contributions towards understanding 3D cloud structure from multi-angle polarimetry. First, we introduce a large-scale, open-source dataset that fuses existing cloud products into a format more amenable to machine learning. This dataset treats multi-angle polarimetry as an input, and radar-based vertical cloud profiles as an output. Second, we describe and evaluate strong baseline machine learning models based that predict these profiles from the passive imagery. Notably, these models are trained only on center-swath labels, but can predict cloud profiles over the entire passive imagery swath. Third, we leverage the information-theoretic nature of machine learning to draw conclusions about the relative utility of various sensor configurations, including spectral channels, viewing angles, and polarimetry. These findings have implications for Earth-observing missions such as NASA's Plankton, Aerosol, Cloud-ocean Ecosystem (PACE) and Atmosphere Observing System (AOS) missions, as well as in informing future applications of computer vision to atmospheric remote sensing.

1 Introduction

Clouds regulate the global climate system in many important ways. As the climate changes, the global cloud distribution will be affected by numerous feedback cycles, some of which are not fully understood. As a result, clouds and cloud feedbacks are among the greatest sources of uncertainty in climate sensitivity models (Pörtner et al., 2022; Meehl et al., 2020; Bony et al.,

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

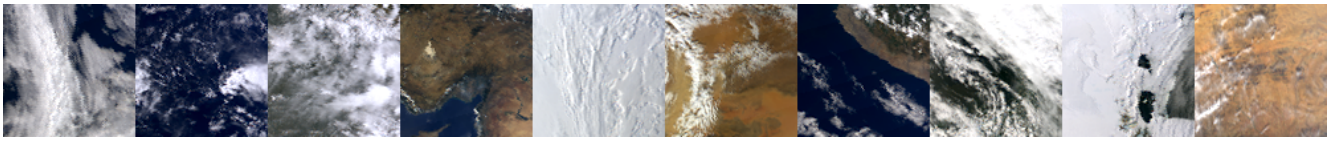

Figure 1. True color POLDER-3 images from the test set, illustrating a diversity of cloudy scenes.

2015). Monitoring the response of global cloud distributions to a changing climate will be of utmost importance in the coming decades. Understanding the 3D structure of clouds, as well as the vertical distribution of cloud phase and thickness, is of great importance to climate modeling studies, such as the characterization of positive and negative cloud feedbacks (Rossow et al., 2022; Marchand et al., 2010).

Three missions are of particular relevance to this work. One of these is Polarization and Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar (PARASOL), a French Centre National D'études Spatiales (CNES) mission which launched in 2004 and remained operational until 2013. POLarization and Directionality of the Earth's Reflectances (POLDER) was a multi-angle polarimeter, the third of which (POLDER-3) was mounted on PARASOL. It was designed to improve understanding of clouds, aerosols, and their climate interactions (Deschamps et al., 1994; Buriez et al., 1997). POLDER data continues to provide diverse insights on aerosol properties near clouds (Waquet et al., 2013), constraints on global emissions (Chen et al., 2019), multi-layer cloud identification (Desmons et al., 2017), and more. The second relevant mission is CloudSat - a NASA mission which observed the vertical distribution of clouds, aerosols, and precipitation using backscatter from its Cloud-Profiling Radar (CPR) (Stephens et al., 2002; Im et al., 2005). From its launch in 2006 to its end of life in 2023, CloudSat provided diverse insights on clouds, aerosols, and precipitation (L'Ecuyer et al., 2015; Smalley et al., 2014; Haynes et al., 2009; Liu, 2008). Finally, Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO), which launched in 2006, carried both an active lidar sensor and passive imagers, enabling the study of the vertical distribution of clouds and aerosols (Winker et al., 2009). All three of these satellites at one point shared an orbital constellation known as the A-Train, yielding nearly-simultaneous observations of the surface (Stephens et al., 2018). CloudSat left the A-Train in 2018, followed by CALIPSO, but remained in orbit for several more years. These missions have reached end of life now, but remain an invaluable resource for preparing for the next generation of atmospheric satellites, like PACE (Werdell et al., 2019).

The characterization of the vertical structure of clouds can be done from both passive and active sensors. Cloud-top height is often derived from passive thermal infrared measurements (Baum et al., 2012). These tend to provide only the top altitude of clouds in a column and struggle in multi-layered cloud systems, but have the advantage of broad spatial coverage. Active sensors, like CloudSat's CPR, have the advantage of providing a more complete vertical profile, at the expense of coverage. For example, CloudSat observed a narrow cross-section of clouds at nadir, whereas POLDER's maximum swath width (across view angles) was approximately 2200 kilometers (Buriez et al., 1997). In addition, passive sensors can be multi-angle, allowing the use of stereoscopic methods to retrieve cloud structure. The introduction of uncertainty as a part of stereo algorithms must be weighed against the benefit of a wider swath. Prior work has explored the usage of POLDER data to understand the structure of clouds, using a decision tree to predict whether each pixel contains single-layer or multi-layer clouds (Desmons et al.,

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

2017). Another work uses the Multi-angle Imaging SpectroRadiometer (MISR) in tandem with Moderate resolution Imaging Spectroradiometer (MODIS) to estimate two-layer cloud properties (Mitra et al., 2023). Airborne multi-angle platforms like RSP have been used to study multi-layer clouds (Sinclair et al., 2017). The estimation of vertical cloud profiles in geostationary data using machine learning has been studied (Brüning et al., 2023). This concurrent work employs a similar strategy to ours, but on geostationary, single-view imagery. To our knowledge, the estimation of full vertical cloud profiles from POLDER data has not been attempted.

The problem of reconstructing 3D geometry has a rich history in both remote sensing and computer vision, but the techniques used differ greatly between those fields. Whereas computer vision techniques nearly always assume a pinhole camera model, remote sensing uses a rational polynomial camera model (Gupta and Hartley, 1997; Zhang et al., 2019). Additionally, remote sensing reconstruction typically operates on individual stereo pairs, rather than simultaneously reconstructing more than two views (Schonberger and Frahm, 2016). Even within the remote sensing space, most 3D reconstruction approaches operate on relatively high resolution imagery (on the order of several meters), and not on wide-swath imagery like POLDER, whose pixels are several kilometers wide (Buriez et al., 1997). Another scale-related difficulty is the difference between horizontal and vertical resolution. Many vertical cloud profiles are reported at sub-kilometer vertical resolution (Stephens et al., 2002), which may be difficult or impossible to accurately retrieve from low-resolution passive imagery like POLDER. One example of 3D reconstruction applied to a coarser spatial resolution is the MISR Interactive eXplorer project (MINX), which performs stereo height retrieval of aerosol plumes in MISR data (Nelson et al., 2010). However, the smaller pixel size and larger parallax simplifies the determination of height. Cloud-top height retrievals were performed with the Along-Track Scanning Radiometer (ATSR) sensor series (Muller et al., 2007). Both MISR and ATSR have sharper resolutions than the 6×7 km^2 pixels of POLDER (Buriez et al., 1997), with MISR at 275-1100 m (Diner et al., 1998), and with ATSR at 1 km (Muller et al., 2007).

As an alternative to stereoscopic 3D reconstruction, we elect to directly estimate the vertical cloud distribution of each location. This is accomplished with a deep learning method, which essentially takes passive imagery as input and estimates an active sensor product as output, but on the same spatial grid as the input. The model takes a flat representation of the multi-angle imagery as an input and produces a dense 3D binary cloud / no-cloud grid. The 3D masking task is known in the computer vision literature as volumetric segmentation (as opposed to image segmentation). Deep learning for volumetric segmentation is well-studied, particularly within medical applications like computed tomography (Çiçek et al., 2016; Soffer et al., 2021; Ardila et al., 2019; Jnawali et al., 2018). However, these methods take a 3D input. By contrast, our method stacks the various viewpoints of the surface (as seen from different angles), as is done in 3DeepCT (Sde-Chen et al., 2021), in which deep learning is used to volumetrically segment clouds in simulated multi-angular satellite data. The motivation for stacking multiple viewpoints is "depth-from-disparity", where the disparity (in appearance between viewpoints) is an indicator of the depth (distance from sensor).

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

Table 1. POLDER-3 spectral bands.

Band (nm)	443	490	565	670	763	765	865	910	1020
Bandwidth (nm)	20	20	20	20	10	40	40	40	20
Polarization	no	yes	no	yes	no	no	yes	no	no

2 Data

In order to satisfy the data requirements of a deep neural network, we synthesize POLDER data with a CloudSat product called 2B-CLDCLASS (Sassen and Wang, 2008). Both the original POLDER and CloudSat data are made available by the AERIS/ICARE Data and Services Center (see Code and Data Availability). The original POLDER data comes from the level-1B product, whereas for CloudSat we use a product called 2B-CLDCLASS as provided by the Calxtract application from ICARE. We call the fused and re-formatted dataset the A-Train Cloud Segmentation (ATCS) Dataset, after the orbital constellation these satellites shared. The ATCS dataset and related API are available as an archive in the SeaWiFS Bio-optical Archive and Storage System (SeaBASS).

2.1 POLDER data

The POLDER level-1B data used in this study are in HDF5 format, organized by date. Each file corresponds to one PARASOL half-orbit (the daytime side). POLDER level 1 and higher products are generated on a gridded equal-area sinusoidal projection, with each pixel in this grid corresponding to approximately 6×7 km on the surface (Hagolle et al., 1996). Each pixel contains the co-registered quasi-simultaneous multi-angular viewpoints. There are up to 16 viewing angles (from PARASOL's perspective) from which a pixel can be observed. The maximum delay between these quasi-simultaneous observations is on the order of several minutes. See Table 1 for a description of the POLDER-3 spectral bands.

2.2 2B-CLDCLASS

These files are merged with 2B-CLDCLASS data, a CloudSat product which contains CPR data, classified into eight cloud types (Sassen and Wang, 2008). While the eight cloud types preserved in the ATCS dataset, this study treats cloud presence as a binary label. The CPR range extends from beneath the surface up to 25 km, the vertical resolution is 240 m, and there are 125 height bins in the data. There are few to no clouds in many of the higher altitude bins (above 14 km) for the sampled data, which are discarded. Sub-surface bins are discarded as well. There are 59 bins in the valid range.

There are related products which combine the CloudSat radar and the CALIOP lidar, such as 2B-GEOPROF-LIDAR (Mace and Zhang, 2014). When including lidar, the data contain many optically thin clouds, to which POLDER and many other passive sensors are often not sensitive. This caused issues when applying supervised learning; the loss function was dominated

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

by optically thin clouds and performance on the radar-observed clouds suffered as a result. This course of study was therefore discontinued, but could prove interesting for future work.

Generally, POLDER-3 and CPR are sensitive to different physics. It should be impossible for a method with access to only the shortwave information in POLDER-3 to fully capture a radar product. Some loss of information is expected, but this study aims to provide a reasonable lower bound on the degree to which POLDER-3 contains the 2B-CLDCLASS information.

2.3 Sampling Strategy

As the quantity of available data is more than sufficient to train a deep network, we make use of uniform sampling to compile the dataset. We randomly sampled half-orbit files for both PARASOL / POLDER and CloudSat / 2B-CLDCLASS from every day in a predetermined date range, and discarded invalid data. The date range was determined by the availability of valid data: before November 27, 2007, our data extraction process was unable to recover valid data. In December 2009, PARASOL lowered its orbit to exit the A-Train (Stephens et al., 2018), meaning POLDER-3 data were no longer quasi-simultaneous with CloudSat. Data were considered invalid under any of the following conditions: 1) any of the POLDER, 2B-CLDCLASS, or CALIPSO/CALIOP half-orbit files were missing, 2) the half-orbit files start-times were off by more than ten minutes (potentially indicating an incomplete record), 3) the produced records did not contain valid data for 13 or more angles in the POLDER imagery, or 4) the produced records did not contain 50 or more labeled pixels.

Standard practice in machine learning involves splitting datasets into separate sets for training, validation, and testing. Labels from the training set are used to optimize the model's parameters. The validation set is used to examine the model's performance, and its labels are held out so they cannot directly affect the model's parameters. However, hyperparameters (e.g. learning rate, how many layers in a model) are typically optimized with respect to the validation set. Neither parameters nor hyperparameters should be optimized with respect to the test set, which is the final measure of a method's efficacy. Commonly, the training set is the largest. We separately generate a training + validation (trainval) set and a test set, with the test set being approximately one quarter the size of the trainval set. The trainval set is then split with 80% of half-orbit files assigned to training and 20% assigned to validation.

The trainval set has two files per day, and the test set has one file per day. For each 2B-CLDCLASS half-orbit file, we uniformly sampled random locations along the CloudSat track, attempting the alignment procedure described in the next section until acquiring enough patches: 16 for the trainval set, 8 for the test set. The patches are 100x100 pixels, as these have sufficient spatial context without yielding overly large file sizes. Since each pixel contains hundreds of values corresponding to various angles and spectra, increasing patch size quickly becomes costly to storage space. Conversely, smaller patch sizes reduce spatial context, which contains useful information for the segmentation algorithm.

2.4 Alignment Strategy

Alignment between POLDER and 2B-CLDCLASS data is performed with sub-pixel accuracy, with respect to POLDER. The POLDER grid is a sinusoidal (equal-area) projection of the globe (Hagolle et al., 1996). Each pixel in this grid contains multi-angular sensor data, geography (latitude, longitude, altitude), and geometry (viewing angle, solar zenith angle, relative

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

azimuth). The 2B-CLDCLASS data is provided temporally, with each timestamp associated to a vertical cloud profile, a latitude/longitude point on the Earth, and some metadata, such as quality flags. Although we could quantize the 2B-CLDCLASS locations to the POLDER grid, doing so would cause substantial quantization error. Instead, we compute ground distances to find nearest neighbors in the POLDER data for each 2B-CLDCLASS observation. As it would be computationally prohibitive to compute these distances between all point pairs, we first use a KD-Tree (Bentley, 1975) to find the top 20 POLDER matches in latitude / longitude space for each CloudSat observation. Although the latitude / longitude grid is far from equal-area, at sufficiently small distances, the effect of Earth's curvature on the distance-ranking approaches zero. We then compute approximate ground distances from the CloudSat observation to these 20 points using the WGS-84 ellipsoid model of the Earth. From these 20 points, we select the closest ones that are found to the northeast, southeast, southwest, and northwest of the Cloudsat observation. These four points are used as interpolation corners, with weights defined by their ground distances. Standard bilinear interpolation leverages the separability of the x- and y-components of the interpolation equation, but these corners are neither on a flat plane nor are they guaranteed to be rectangular. For these reasons, the normalized inverse distance is a more appropriate choice for the corner weighting function.

For a pair of half-orbit files, the above defines a mapping from CloudSat observations to their corners in the POLDER grid. In order to sample patches, we compute all of the latitude intervals which would yield a 100 pixel north-south distance in the POLDER grid. We also compute the latitude-dependent longitude intervals which would yield a 100 pixel east-west distance in the POLDER grid. We uniformly sample latitude intervals, and use our longitude intervals to compute an index from a 100x100 patch into the POLDER grid. As an additional step, we uniformly shift patches east or west by one quarter of the patch window so that labels aren't always found in the center of the patch. All patches are north-aligned; we do not perform any rotation. Only Cloudsat observations which have 4 corners within the patch are kept. If there are fewer than 100 Cloudsat observations in a patch, that patch is discarded. Additionally, occasional quantization or out-of-bounds errors can cause the patch to contain fewer than 100x100 pixels or to contain 101x100 pixels. In these cases, we discard the patch. This process loops until sufficient patches are found for a file. Sufficient valid patches were found for every file.

We validate our alignment accuracy by computing (approximate) ground distance between the latitude and longitude values of the projected Cloudsat observations and the weighted interpolation of the latitude and longitude values of the corresponding corners in the POLDER grid. The mean projection error across both the training and validation datasets is 881 m, with a maximum error of 2,216 m. This is significantly less than a single pixel in the POLDER grid, which are approximately 6×7 km^2. Sub-pixel alignment accuracy ensures that each pixel in the POLDER grid contains sensor values from the correct location.

2.5 Dataset Statistics

ATCS is a large-scale dataset, enabling the training of deep neural networks. The 2B-CLDCLASS vertical cloud profiles are used as labels for the PARASOL imagery, which is used as input to the neural network. The ATCS dataset contains 20,352 labeled instances, and 5,032 instances in the test set, whose labels are withheld. The labeled instances are divided, with an

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

80/20% training/validation split. On average, in the labeled set, there are 116 labeled locations per 100x100 patch, with a standard deviation of 18, a minimum of 55, and a maximum of 171.

Data were sampled uniformly in the geographic sense, except for latitudes greater than $\pm 80°$, which cause complications with geometric processing. There is great value in this global coverage, as cloud dynamics and appearance vary greatly by region. The difficulty of the cloud segmentation task is also strongly regionally dependent. For example, near the poles, there are high-albedo icy and snowy surfaces, which are more visually cloud-like than typical terrestrial surfaces, complicating the segmentation task (Stillinger et al., 2019).

Cloudiness also correlates strongly with altitude. There is a strong 'class imbalance' between the clouds at various altitudes – a randomly sampled pixel is much more likely to contain clouds at a lower altitude than to contain clouds at the upper altitude range, i.e. near 14 km.

3 Approach

Our approach is to learn a functional mapping from multi-angle imagery to per-pixel vertical profiles of cloud occurrence using a deep convolutional neural network (CNN). This had two primary motivations. First, by training a model on labeled nadir pixels and applying it to unlabeled off-nadir pixels, we produce a wide-swath vertical cloud product, which has independent value. Second, the supervised learning setup allows for ablation experiments; various sensor properties (angle, spectral band, polarization) can be omitted from the input data. The resulting change in model skill provides insights into the utility of these properties for estimation of cloud structure. The use of a neural network allows the characterization of highly non-linear relationships between these input fields and the cloud profile outputs, and deep learning outperforms shallow learning, as is shown in Table 2.

We used convolutional neural networks to incorporate spatial context, necessary in order to account for the parallax present in multi-angle data. Convolutions are useful for data which exhibit some form of translational invariance, meaning that some elements of appearance do not change due to translations in image space. Interestingly, convolutions are particularly suitable for satellite data, since it exhibits lower translational variance than ground-level imagery, due to less perspective variation in depth. The spatial context provided by convolutional architectures can capture the apparent shift in location due to parallax present in a cloudy scene, which is useful for the network to predict depth and vertical structure.

3.1 Representation

Instances are defined as input / output pairs, where input contains a 100x100 patch of multi-angle, polarimetric imagery from POLDER, as well as the pre-computed weights used for interpolation, and output contains the vertical cloud profiles from 2B-CLDCLASS. Each pixel in the input has data for up to 16 viewing angles. When data is missing for one or more viewing angles, a standard missing value of -1 is used. The median number of available viewing angles over the training and validation sets is 14, and 98% of pixels have at least 13 available viewing angles. At each of the viewing angles, there are nine spectral bands, three of which measure linear polarization (Table 1). Linear polarization is represented as the Q and U components

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

of the normalized Stokes parameters (also known as S_2, S_3). Therefore, the channels without polarization have one value: I, while the polarization channels have three: I, Q, and U.

Raw sensor inputs must be converted into features which can be used by a neural network. Most inputs can simply be normalized before use, but some inputs must be more heavily modified. In addition to the nine intensity values I, and the three pairs of Q, U values, there are four geometry fields: viewing angle, solar zenith angle, relative azimuth, and solar azimuth. We found there to be little difference from using view azimuth instead of relative azimuth. Rotations are discontinuous at $360°$, which poses issues for a neural network as a small change in rotation can cause a large change in the output. We use a technique from the object recognition literature which encodes angles into binary membership in two of eight overlapping bins, as well as two in-bin floating point offsets (Mousavian et al., 2017), resulting in a length-ten vector. The eight overlapping intervals are $[0°, 90°], [45°, 135°], \ldots, [315°, 45°]$. The in-bin regression values are relative to the bin centers. As an example, consider the angle $100°$. This occupies the second bin $[45°, 135°]$ and the third bin $[90°, 180°]$, giving it a bin-membership encoding of 01100000. Its differences with the centers of the second and third bins are $10°$ and $-35°$, respectively. This yields a length-ten feature vector: $\{0, 1, 1, 0, 0, 0, 0, 0, 10°, -35°\}$, although angles are converted to radians in practice. This transformation is only necessary for the azimuth angles, not the zenith angles, as zenith angles are bounded in the range $[0°, 90°]$.

There are 27 input values per view: six non-polarized channels, three polarized channels with three values each (I,Q,U), a length-10 view azimuth feature, one view zenith angle, and one solar zenith angle. The length-10 solar azimuth feature does not vary with view angle. Therefore, when using all 16 available angles, the channel depth (features per pixel) is 442. In addition, there are (non-angular) fields included as a convenience for other researchers. These values are not provided as an input to any model trained in this study, and include latitude, longitude, surface altitude, and a land-or-sea flag.

Each input patch is represented as an image cube, created by stacking the multi-spectral, multi-angular features, as well as the geometry features described above. In line with standard practice in computer vision, each of the spectral features are normalized by subtracting their mean and dividing by their standard deviation, with respect to the training set.

3.2 Architectures

3.2.1 Single-Pixel Network

A simple baseline approach to cloud segmentation is to independently estimate the vertical profile of cloud occurrence at each pixel. A multi-layer perceptron (i.e. a simple neural network) predicts this vertical profile of a single pixel at a time from the sensor observations of that pixel. The single pixel model is implemented using 2D convolutions with kernel size 1×1. Since the kernel size is always 1×1, the independence of pixels is preserved (i.e. the network has no spatial context), but has the important property of keeping the same number of pixels per batch as the later experiments using 2D convolutions. Batch size would otherwise be a potentially confounding variable in any comparison between the single-pixel model and models which ingest an entire image at a time. We test a single-pixel model with three linear layers, with all but the last layer followed by batch normalization (features averaged across each batch) and a rectified linear unit (ReLU), which is defined as $f(x) = x$ for $x \geq 0$, $f(x) = 0$ for $x < 0$.

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

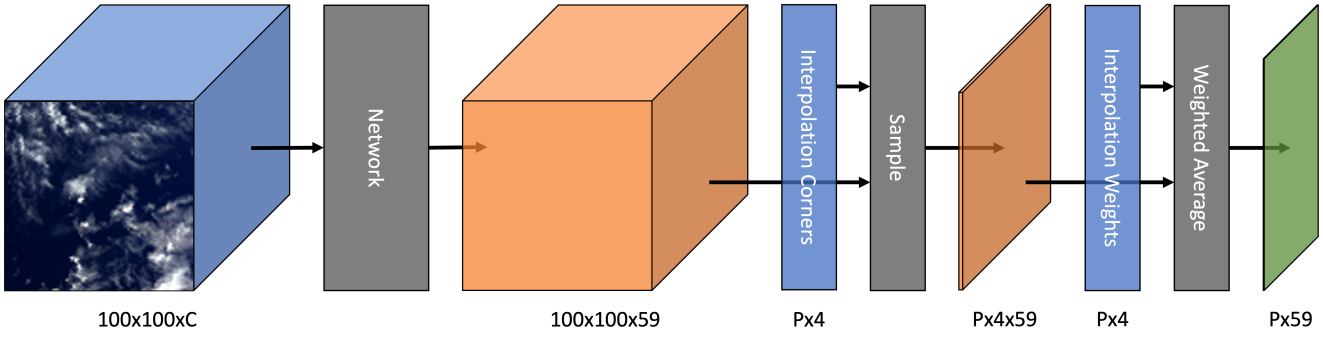

Figure 2. Diagram illustrating the forward pass. The color blue denotes inputs, orange denotes intermediate tensors, and green denotes output. C denotes channel depth and P denotes the number of labeled pixels in an instance. The model maps the input imagery to predictions for each of the 59 height bins. These predictions are then interpolated, represented by the Sample and Weighted Average blocks. Simply skipping the interpolation step allows the generation of wide-swath predictions.

3.2.2 Simple CNN

We also test a simple multi-layer convolutional neural network. This has five 3×3 convolutional layers. As with the single-pixel model, all but the last convolutional layer are followed by batch normalization and ReLU. Each convolutional layer includes one pixel of padding, so the image resolution stays constant during the forward pass.

While the above are relatively simple choices, the choice of per-layer channel depths is more complicated. A common approach is to use increasing powers of two. As our data has a much higher number of input channels than RGB imagery does, increasing the channel depth by a factor of two would result in an unreasonably high feature depth after even a few convolutional layers. Instead, we compute a scaling factor which yields a depth of 1024 after five layers. This avoids bottlenecking the feature depth at the start of the network, while keeping the feature depth in a reasonable range. For example, an input channel depth of 226 (as in the eight-angle experiments) yields feature depths of 306, 414, 560, 757, and 1024. These layer-wise feature depths c_i are given by the following, where c_{input} is the input channel depth, and $\lfloor \cdot \rfloor$ is the floor operator:

$$c_i = \lfloor c_{\text{input}} \wedge (1 + (\frac{\log(1024)}{\log(c_{\text{input}})} - 1)\frac{i}{5}) + 0.5 \rfloor \quad (1)$$

This scaling ensures a consistent rate of change in the feature depth, rounded to the nearest integer. Note that the base of the log does not matter as long as it is consistent.

3.2.3 U-Net

A U-Net (Ronneberger et al., 2015) consists of a fixed number of down-sampling 'blocks', followed by the same number of up-sampling blocks, as well as skip connections between blocks at the same level of the spatial pyramid. The down-sampling blocks capture the typical structure of a convolutional neural network for image classification, where each subsequent layer represents a gradual trade-off of spatial resolution for increasing feature depth. U-Net also adds up-sampling blocks, which

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

effectively do the opposite: decreasing feature depth while increasing spatial resolution. This architecture is related to the commonly used encoder-decoder network. Unlike encoder-decoder networks, U-Net has skip connections, which allow the preservation of spatially located features in the up-sampling path. We use five blocks in our U-Net, with the per-block feature depths using the same scheme as described in Equation 1. It is worth noting that increasing the channel depth in this way substantially affects both U-Net's parameter density and the size of the intermediate features during the forward pass. We experimented with several different schemes to decide channel depth, and found this approach to be the most stable configuration for U-Net.

3.3 Interpolation

One novel element of our network architecture is the use of interpolation during the forward pass of the model. The CloudSat labels are only available for some locations in the input images, and these locations are not quantized to the POLDER grid. Therefore, after the up-sampling blocks, the model has an interpolation layer, using the corners and corner weights described in 2.4. As a weighted average is differentiable, back-propagation can still be used to train the network. After the interpolation layer, the network has two fully-connected layers. Figure 2 illustrates the forward pass, including a depiction of the interpolation process. Note the difference between this and bilinear interpolation: bilinear interpolation is linearly separable. For accuracy, the interpolation model utilized here uses a latitude/longitude grid rather than a Cartesian grid, so the standard separability does not apply.

The networks can be applied to data *without* interpolating. This is as simple as omitting the interpolation module. In Figure 2, the $100 \times 100 \times 59$ tensor can simply be used as the wide-swath prediction. The reduction down to a $P \times 59$ tensor is only necessary for the application of the loss function. Interpolation is slightly more complicated for U-Net, which is discussed in Appendix A.

3.4 Training Procedure

In all experiments, the model is trained for 30 epochs using the Adam optimizer (Kingma and Ba, 2017) and the binary cross-entropy loss. Let $\sigma(\cdot)$ be the sigmoid operator, y the binary labels, and x the network's unbounded outputs (also referred to as logits). The binary cross-entropy loss is given by:

$$\ell_n = y_n \cdot \log\left(\sigma(x_n)\right) + (1 - y_n) \cdot \log\left(1 - \sigma(x_n)\right) \tag{2}$$

We experimented with data augmentation, including random flips and rotations, but found these to significantly worsen performance, even when accounting for these changes in the geometry features. Therefore, we omitted data augmentation from our final experiments.

Each experiment is repeated three times, for reasons discussed further in 4.2. Model checkpoints are saved every 5 epochs. Our implementation uses the Pytorch framework (Paszke et al., 2019). Training is performed using a single NVIDIA Tesla V100 GPU with 32GB of memory. Training plus validation typically takes between two and six hours per experiment.

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

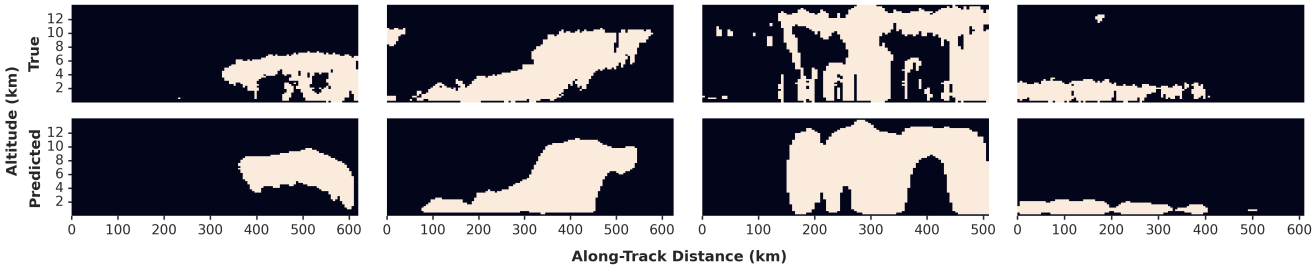

Figure 3. Results for four instances in the test set, using the default experiment. The true cloud profiles are shown in the top row, and predictions are in the bottom row.

4 Results

We perform a series of ablation experiments. The first experiment compares the skill of the three aforementioned architectures on the test data. In subsequent experiments, we deprive the model of various aspects of the available input data and measure the resulting effect on skill. These ablation studies reveal the impact of the viewing angles, the spectral configuration, polarimetry, and the scene geometry. The baseline 'default' experiment consists of the simple convolutional model with 8 of 16 viewing angles, all available spectral channels (with polarimetry when applicable), and all geometry fields. Other experiments vary from the default experiment only in the exact ways specified in the tables and figures.

As the network is trained using the cross-entropy loss, its unbounded outputs (also known as logits) cannot directly be treated as probabilities. Within the loss function, the logits are passed into a sigmoid function to bound them within the $[0, 1]$ range. A threshold of 0.5 after the sigmoid function corresponds to a threshold of 0 before. Therefore, to evaluate the output of the network, we simply threshold the logits at 0 to convert them to a binary value, which we compare with the ground-truth cloud profiles.

4.1 Dice Score

The Dice score, originally used in an ecological context (Dice, 1945), was adopted in the medical segmentation literature as a validation metric for segmentation of MRI imagery (Zijdenbos et al., 1994). Its use has since become standard, due to its several advantages over pixel-wise accuracy metrics, including its tolerance to infrequent positive samples and that it penalizes differences in location more than differences in size. These and other qualities mean that the Dice score is a better measure of perceptual quality than pixel-wise metrics. Let A and B be two sets representing all of the discretized locations of two respective objects. The Dice score between A and B is twice the intersection divided by the sums of the sizes of A and B. It can be contextualized in terms of true positives (TP), false positives (FP), and false negatives (FN):

$$\text{Dice}(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3}$$

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

The Dice score is related to intersection-over-union, also known as the Jaccard index, another metric commonly used in the segmentation literature. Both metrics have a range of $[0, 1]$, but the Dice score is strictly greater than the Jaccard index.

Notably, there is significant evidence that for many applications, the loss function, which is used to optimize the network, should be metric-specific. In segmentation, Jaccard-like loss functions often outperform their pixel-wise counterparts (such as cross-entropy) (Eelbode et al., 2020; Mohajerani and Saeedi, 2021; Wang and Blaschko, 2023). We experimented with Jaccard-like losses early in project development, but observed no apparent improvement over the standard cross-entropy loss.

The Dice score generalizes to an arbitrary number of dimensions. While it is typically used for 2D data, we make use of it in both 1D and 2D contexts. The overall Dice scores presented in the tables are 2D, while the altitude-dependent Dice scores in the figures are 1D, as they describe a single row of the time, altitude cross-section given in the 2B-CLDCLASS labels.

Alongside the Dice score, we report the bin-wise accuracy, which is the rate at which the model correctly assigns a pixel, altitude-bin pair as cloudy or not cloudy. This metric is less strict than the Dice score, and is less suited to labels with a strong imbalance between positives and negatives, as in our data. Our findings are therefore mostly based on the Dice score, but the inclusion of accuracy allows for easier comparison with other works.

$$\text{Acc}(A, B) = \frac{|A \cap B| + |\neg A \cap \neg B|}{|A| + |\neg A|} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

4.2 Inter-run Variability

Due to the stochastic nature of machine learning, the same experiment will yield variable results given different initializations of the network, as well as differences in the shuffled order of the training set. In this work, the inter-run variability is an important factor to consider. Repeating the U-Net experiment 3 times, for example, yielded validation set Dice scores of 72.8, 73.1, and 73.4, although the simpler architectures experience less variance. We report the maximum test set accuracy over three runs; for all runs of an experiment, for all saved model checkpoints, we use the model which yields the highest Dice score on the validation set, and report its Dice score on the test set. There is little variation between the validation set and the test set metrics: U-Net, for example, has a max validation set Dice score of 73.4%, which drops to 73.2% on the test set.

4.3 Architecture Complexity

The three architectures described in 3.2 are evaluated, with results shown in Table 2. There is a large increase in skill from the single pixel model to the simple convolutional model, and a negligible increase from the simple convolutional model to the U-Net.

As the POLDER-3 data are geo-referenced using the surface elevation (and not the cloud-top height), there is a parallax-induced shift when clouds are present, particularly if those clouds occur at a higher altitude. The single-pixel model, which lacks spatial context, does not have access to adjacent pixels. The simple convolutional network and the U-Net, by contrast, can leverage surrounding pixels, making use of the information contained in the parallax. Another potentially useful quality of spatial context is that it captures more information about the scene dynamics, which may be used by the model.

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

Table 2. Comparison of three architectures of increasing complexity. Results are reported for the test set. U-Net achieves the highest performance but has more parameters.

Architecture	# Params.	Dice Score (%)	Accuracy (%)
Single-Pixel	1.2E+05	68.6	93.7
Simple ConvNet	1.2E+06	73.0	**94.3**
U-Net	9.5E+07	**73.2**	**94.3**

Table 3. Comparison of various viewing geometries. The 2-angle view contains only the pair of angles closest to nadir. The 4-angle view adds the next-innermost pair of angles, proceeding outward until all 16 angles are included.

# View Angles	Dice Score (%)	Accuracy (%)
2	68.9	93.6
4	71.6	94.1
6	72.2	94.2
8	73.0	94.3
10	73.2	94.3
12	73.8	**94.5**
14	**74.1**	**94.5**
16	73.8	94.4

The results suggest that there are diminishing returns from increasing model capacity. It is unlikely, in our estimation, that larger models will achieve significantly higher performance on this dataset, without changes in other aspects of the data processing pipeline or optimization procedure.

4.4 Viewing Angles

Multi-angle POLDER measurements are particularly sensitive to parallax, providing useful information on vertical cloud distribution. The impact of parallax on model skill was studied via an ablation experiment in which various viewing geometries were provided for the model. First, we started with an experiment using only the 2 central (closest to nadir) viewing angles. Next, we progressively added subsequent pairs of angles (in increasing zenith angle order) until all angles were included, running independent experiments on each configuration. The results are shown in Table 3, as well as Figure 4, which captures the altitude-dependent Dice score of experiments using various numbers of angles. There is a significant improvement between 2 and 4 angles, with diminishing returns from the inclusion of more angles. There is a clear increasing trend, with 14 viewing angles achieving the best performance. The drop in performance from including the outermost pair of angles is unsurprising,

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

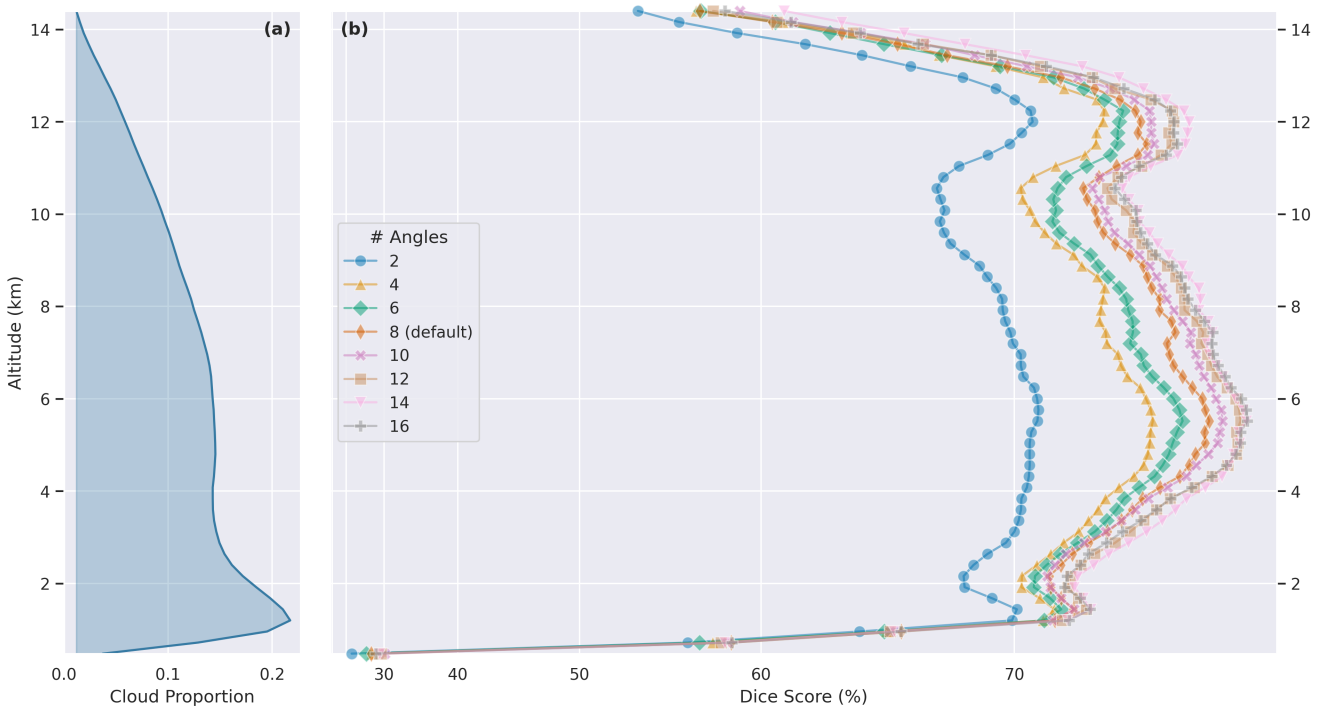

Figure 4. Panel (a) shows the altitude-dependent proportion of pixels labeled as clouds in the dataset. Panel (b) shows the altitude-dependent Dice Score of experiments with different viewing geometries. Note the significant difference between the 2- and 4-angle experiments, with diminishing returns with more angles.

as these angles contain little to no valid data in most scenes, due to the viewing geometry of POLDER-3. We elected to use 8 angles as the default for other experiments, as it marked a good trade-off between performance and training time.

An important factor to consider in the evaluation of these results is the difference between the 2B-CLDCLASS vertical resolution of 240 m with the 6×7 km^2 horizontal resolution of POLDER-3. Even an extreme difference in viewing angles may not be enough to overcome the stark resolution difference. The multiple angles may aid the model to constrain the possible reflectance distribution functions attributable to a surface, helping to identify surface type and altitude, rather than being useful in the stereoscopic sense. This offers one potential explanation for the diminishing return from including more angles, but other explanations remain possible.

4.5 Spectra

We perform two ablation studies to understand the impact the nine spectral bands have on model skill. In the first, shown in Figure 5, one or two bands are omitted at a time, and the resulting decrease in model skill provides a measure of the unique (non-redundant) information content present in that band. In the second, shown in Figure 6, only one or two bands are provided at a time, and the model skill represents the bulk information content of that single band. In both studies, we include the 763

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

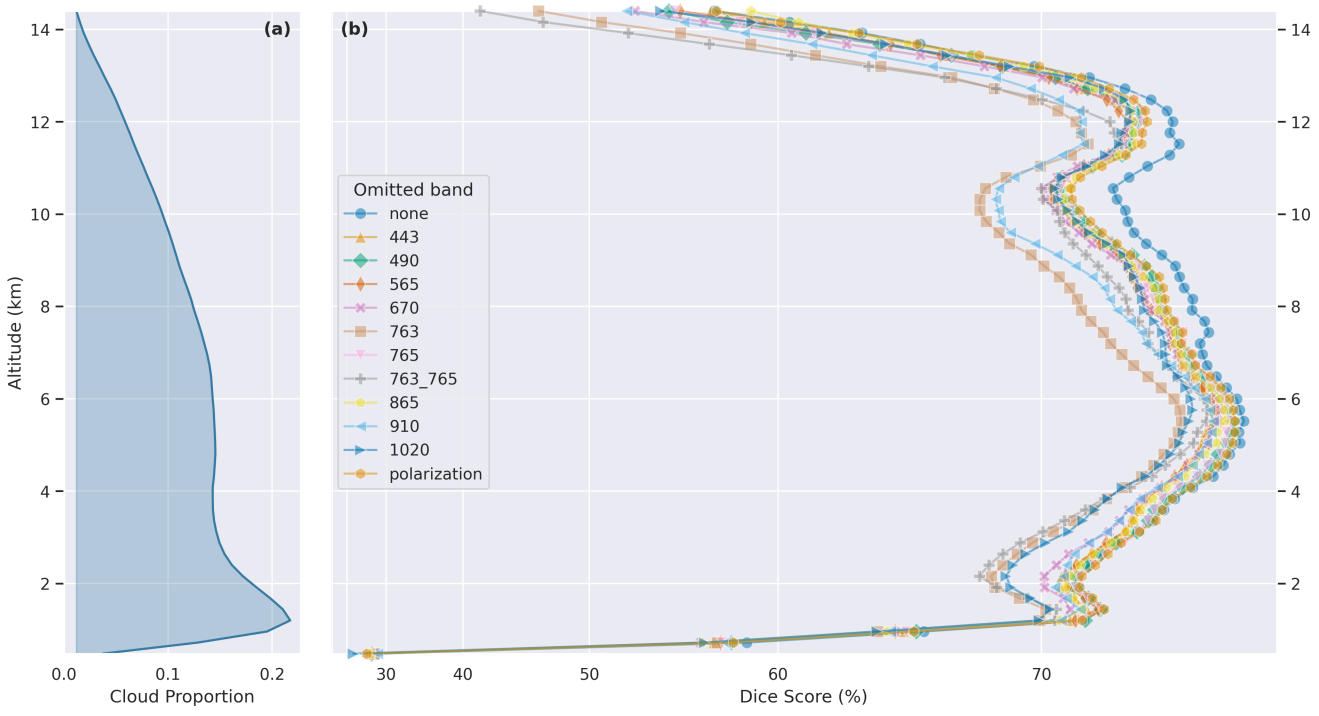

Figure 5. Panel (a) as in Figure 4. Panel (b) shows the altitude-dependent Dice Score of experiments with various channels omitted, as compared to the default, labeled 'none'.

nm (oxygen A-band) and 765 nm bands, as well as a combination of both 763 nm and 765 nm band, as they are often jointly used to derive the oxygen pressure within an atmospheric column, which can be used to infer cloud structure (Ferlay et al., 2010). A table of all the spectral experiments can be found in Appendix B, Table B1.

Figure 5 demonstrates the utility of various bands, with larger drops in performance suggesting that the omitted band contains uniquely useful information content. The oxygen absorption feature at 763 nm is highly correlated with model skill, consistent with the known relationship between this feature and cloud structure. The next greatest drop in skill is captured in the 910 nm and 1020 nm band omission experiments. 910 nm is a water vapor band, which provides information on scattering and absorption interactions between clouds and vapor (Dubuisson et al., 2004). The near-infrared 1020 nm band might be useful for low-altitude clouds over the ocean, as clouds are quite bright in this wavelength, while oceans are quite dark.

4.6 Polarimetry

The importance of polarization to model skill is evaluated by comparing the default experiment to one without the polarization parameters. This is achieved by omitting the Q and U channels from the I,Q,U Stokes parameterization. This caused a reduction in Dice Score from 73.0% to 72.6%. This difference is smaller than we hypothesized. One possibility is that the Stokes vector is not the ideal parametrization for this task. Another possibility is that the particular polarization channels

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

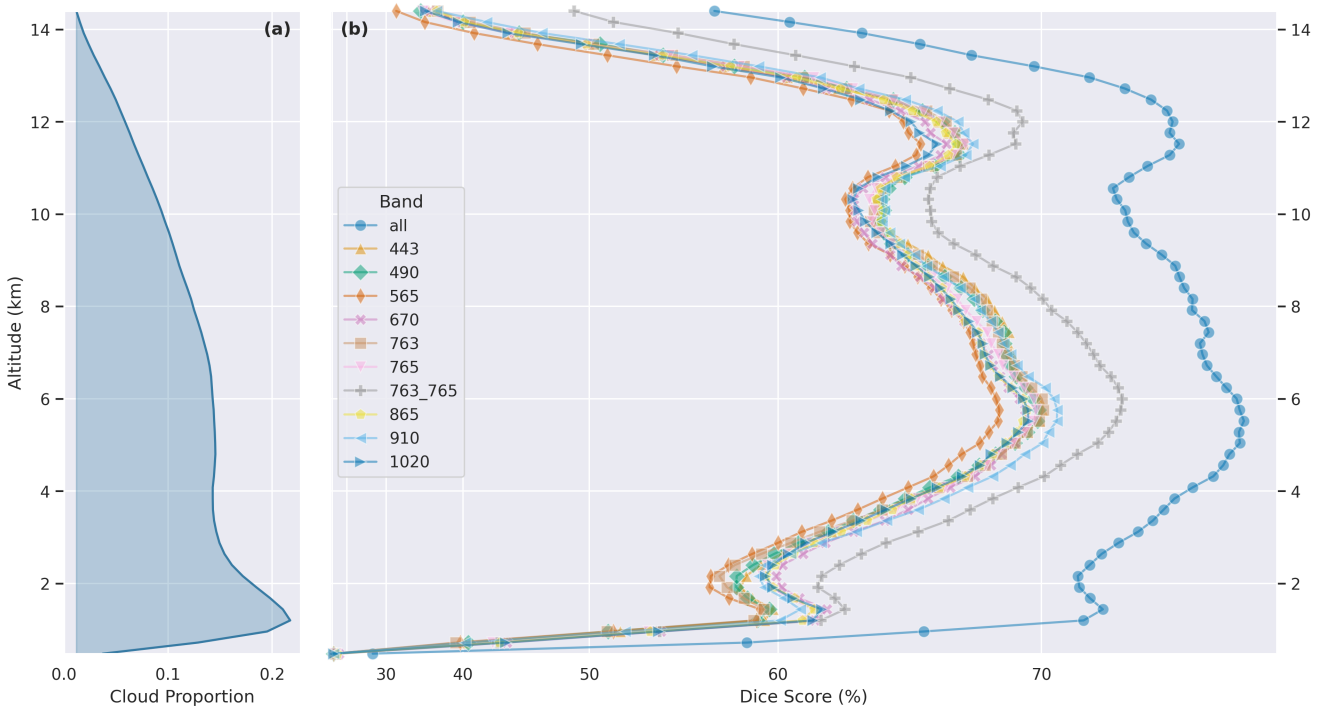

Figure 6. Panel (a) as in Figure 4. Panel (b) shows the altitude-dependent Dice Score of experiments using only one or two channels at a time, as compared to the default, labeled 'all'.

in POLDER-3 have only limited unique utility for the derivation of cloud structure. Prior work has found that blue-light polarization is disproportionately useful for the retrieval of aerosol layer height, especially with wavelengths of 410 nm or lower (Wu et al., 2016), but POLDER-3's shortest-wavelength polarization band is 490 nm. The shorter-wavelength polarization information from future missions, like the Hyper-Angular Rainbow Polarimeter (HARP2) aboard PACE (Werdell et al., 2019), might prove more useful. For example, it has been shown that the polarized 440 nm band in HARP2 provides some sensitivity to aerosol altitude (Gao et al., 2023).

5 Discussion

The results demonstrate the feasibility of estimating 3D cloud structure from passive, multi-angle imagery. The inclusion of spatial context significantly improved results, while the improvement from using a more complex model was modest (Table 2). Increasingly extreme view angles exhibit diminishing returns (Fig. 4). It is possible that stacking the view angles in the channel dimension might not be the most effective feature representation for retrieving 3D structure. The spectral results (Figs. 5, 6, Table B1) confirm the known utility of the oxygen-A band, water vapor band, and the near infra-red for the study of clouds.

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

Polarization only proved modestly useful for the cloud profiling task (Fig. 5, Table B1), but this may be a limitation of the specific polarization bands available in POLDER-3 data.

Even the best experimental configuration only reaches a Dice score of 74.1%, which is slightly lower than average results in the segmentation literature (Eelbode et al., 2020). Seemingly, this dataset is challenging, and it is likely that there exists a limit to a model's skill. This should be unsurprising, given the nature of the labeled data: in most segmentation results, the labels come from human annotators. The human annotation process guarantees that the labels are predictable from the imagery, up to the skill of the human annotator. However, our dataset is not annotated by humans, but involves the fusion of two different sources of satellite data, which are sensitive to inherently different physics. This manifests in several ways. For example, one notable characteristic of the qualitative results is that all models struggle to mask the lower layer of multi-layer clouds, or to predict the extent of optically thick clouds. The penetration depth of clouds is quite low in the visible and near-visible spectrum. It is possible the models may not have any inputs that allow correct classification beneath cloud tops. Related to this problem is the performance drop across all models at low altitudes. Some of this skill decrease is likely due to penetration depth, but may also relate to lower parallax near the surface, making it harder for the model to vertically resolve cloud location. Another phenomenon this approach may struggle to capture is optically thin clouds. As the labels are based on a cloud fraction product and not an optical depth product, there are extremely optically thin clouds present in the data. The information contained in the visible and near-visible wavelengths may not be enough to detect such clouds.

6 Conclusions

We designed a supervised machine learning method to perform 3D cloud masking over a wide swath from multi-angle polarimetry, and introduced a dataset to support this method. The dataset should be useful for future research in this area, as it is designed for ease-of-use in machine learning applications. The code accompanying the dataset includes everything necessary to reproduce the results in this paper: to train and validate models, generate figures, and even to create custom datasets.

By performing extensive ablations with various model inputs and hyperparameters, we analyzed the qualities of both POLDER-3 and CloudSat CPR data, as well as their relationship. Our conclusions both confirm existing knowledge and offer new insights. The results are promising, and suggest the continued use of machine learning as a means to understand the relationships between various sensor modalities. Additionally, machine learning might, with further refinement, offer a useful way to retrieve 3D cloud masks from satellite data at an unprecedented scale, which would be an invaluable source of data for climate modeling.

This study constitutes an initial foray into the combined use of machine learning and multi-angle polarimetry for 3D cloud masking. Whereas its use here provided insights on the POLDER-3 sensor, other sensors have yet to be studied in such a way. Upcoming missions such as PACE and AOS will carry multi-angle polarimeters, providing a useful testbed for this approach.

The model's supervision in this study was provided by an active radar, constrained to nadir locations. The off-nadir performance of our approach has yet to be validated, and would likely require cross-referencing with ground-based radar. The validation of its 2D (flattened) accuracy could be performed with other wide-swath cloud products, like the MODIS cloud

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

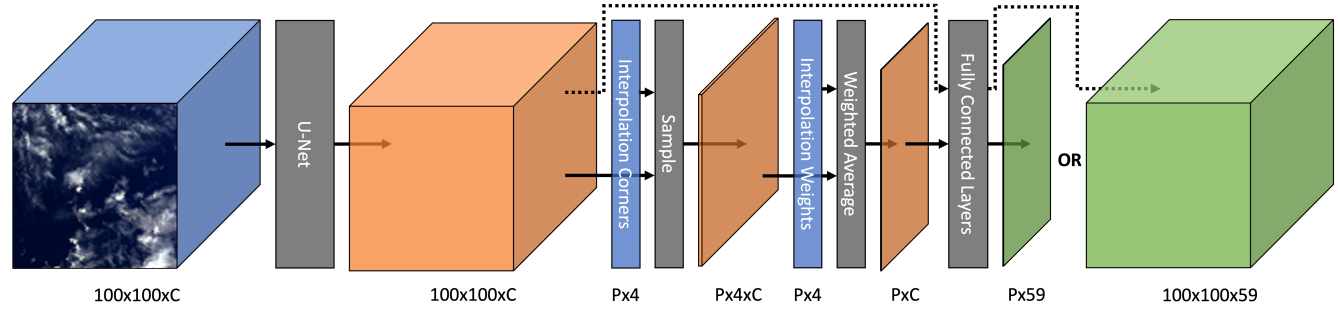

Figure A1. Diagram illustrating the forward pass. The color blue denotes inputs, orange denotes intermediate tensors, and green denotes output. C denotes channel depth and P denotes the number of labeled pixels in an instance. The dotted line represents an alternate pathway which skips the interpolation module, used to get wide-swath results.

mask (Ackerman et al., 2015). Having such a wide-swath 3D cloud mask product, were its accuracy sufficient, could prove useful for climate modeling.

Alternative approaches from the computer vision literature may be better-suited to the stereoscopic nature of this data. 3D reconstruction pipelines such as COLMAP (Schönberger and Frahm, 2016; Schönberger et al., 2016) might be adapted for wide-swath multi-angle imagery, as has been done for high-resolution satellite imagery (Zhang et al., 2019). These would allow the estimation of 3D cloud structure without the need for radar-based supervision.

Code and data availability. Both code and data will be made available prior to publication, and can be retrieved using SeaBASS, found at the following link: https://seabass.gsfc.nasa.gov/search, by searching for "ATCS".

Appendix A: U-Net Forward Pass

Interpolation during the forward pass is more complicated for the U-Net model, as illustrated in Figure A1. During development, we found that the U-Net model performed better if it included fully connected layers after the interpolation module. However, these fully connected layers can still operate on the full (not interpolated) features, of shape $100 \times 100 \times C$. By flattening this to a $10000 \times C$ tensor, passing it into the fully connected layers to get a 10000×59 tensor, and unflattening it to $100 \times 100 \times 59$, the network can be applied to wide-swath data. At training time, the interpolation module is applied to the $100 \times 100 \times C$ tensor to retrieve a $P \times C$ feature, which is then passed through the fully-connected layers to get a $P \times 59$ tensor, which is then compared with the ground-truth labels in the loss function.

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

Appendix B: Spectral Results

The full table of spectral results is included here for reference. The experiments here are the same ones illustrated in Figure 5 and Figure 6.

Table B1. Results for a variety of experimental configurations. The first section shows with and without polarization, the second section shows the effect of omitting each spectral band, and the third section shows the results when only one band is included at a time.

Wavelength (nanometers)									Polarization?	Dice Score (%)	Accuracy (%)
443	490*	565	670*	763	765	865*	910	1020			
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	73.0	94.3
✓	✓	✓	✓	✓	✓	✓	✓	✓		72.6	94.2
	✓	✓	✓	✓	✓	✓	✓	✓		72.1	94.2
✓		✓	✓	✓	✓	✓	✓	✓		72.4	94.2
✓	✓		✓	✓	✓	✓	✓	✓		72.0	94.2
✓	✓	✓		✓	✓	✓	✓	✓		71.8	94.2
✓	✓	✓	✓		✓	✓	✓	✓		70.3	93.8
✓	✓	✓	✓	✓		✓	✓	✓		72.4	94.2
✓	✓	✓	✓			✓	✓	✓		71.0	94.0
✓	✓	✓	✓	✓	✓		✓	✓		72.2	94.2
✓	✓	✓	✓	✓	✓	✓		✓		71.3	94.0
✓	✓	✓	✓	✓	✓	✓	✓			71.3	94.0
✓										64.1	92.9
	✓									64.0	92.9
		✓								62.9	92.7
			✓							64.2	92.8
				✓						63.9	92.8
					✓					64.3	92.9
				✓	✓					66.9	93.4
						✓				64.2	92.9
							✓			64.6	92.8
								✓		64.0	92.8

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

A1

Author contributions. SRF developed the code and dataset, and wrote the draft. KDK, AMS, and MG guided the experiments and revised the paper. JaH and JuH advised the machine learning-related aspect of the project.

Competing interests. One of the authors (AMS) is an Associate Editor for AMT. The other authors declare no competing interests.

Acknowledgements. We'd like to acknowledge the team at ICARE for maintaining the POLDER and 2B-CLDCLASS data, as well as the CloudSat team for providing such an invaluable source of vertical cloud data. We'd also like to acknowledge Jie Gong (NASA Goddard Space Flight Center) for useful discussions about the relevant data products. A significant portion of this work was performed during an internship through Science Applications International Corporation (SAIC). That internship was managed by Fred Patt (NASA Goddard Space Flight Center, SAIC), who provided valuable advice.

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

References

Ackerman, S., Frey, R., Kathleen Strabala, K., Liu, Y., Gumley, L., Baum, B., and Menzel, P.: MODIS Atmosphere L2 Cloud Mask Product, NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA, https://doi.org/http://dx.doi.org/10.5067/MODIS/MOD35_L2.006, 2015.

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D. P., and Shetty, S.: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, Nature Medicine, 25, 954–961, https://doi.org/10.1038/s41591-019-0447-x, 2019.

Baum, B. A., Menzel, W. P., Frey, R. A., Tobin, D. C., Holz, R. E., Ackerman, S. A., Heidinger, A. K., and Yang, P.: MODIS Cloud-Top Property Refinements for Collection 6, Journal of Applied Meteorology and Climatology, 51, 1145 – 1163, https://doi.org/https://doi.org/10.1175/JAMC-D-11-0203.1, 2012.

Bentley, J. L.: Multidimensional Binary Search Trees Used for Associative Searching, Commun. ACM, 18, 509–517, https://doi.org/10.1145/361002.361007, 1975.

Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A. P., Sobel, A. H., Watanabe, M., and Webb, M. J.: Clouds, circulation and climate sensitivity, Nature Geoscience, 8, 261–268, https://doi.org/10.1038/ngeo2398, 2015.

Brüning, S., Niebler, S., and Tost, H.: AI-derived 3D cloud tomography from geostationary 2D satellite data, EGUsphere, 2023, 1–24, https://doi.org/10.5194/egusphere-2023-1834, 2023.

Buriez, J. C., Vanbauce, C., Parol, F., Goloub, P., Herman, M., Bonnel, B., Fouquart, Y., Couvert, P., and Seze, G.: Cloud Detection and Derivation of Cloud Properties from POLDER, International Journal of Remote Sensing, 18, 2785–2813, https://doi.org/10.1080/014311697217332, 1997.

Chen, C., Dubovik, O., Henze, D. K., Chin, M., Lapyonok, T., Schuster, G. L., Ducos, F., Fuertes, D., Litvinov, P., Li, L., Lopatin, A., Hu, Q., and Torres, B.: Constraining global aerosol emissions using POLDER/PARASOL satellite remote sensing observations, Atmospheric Chemistry and Physics, 19, 14 585–14 606, https://doi.org/10.5194/acp-19-14585-2019, 2019.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O.: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, 2016.

Deschamps, P.-Y., Breon, F.-M., Leroy, M., Podaire, A., Bricaud, A., Buriez, J.-C., and Seze, G.: The POLDER mission: instrument characteristics and scientific objectives, IEEE Transactions on Geoscience and Remote Sensing, 32, 598–615, https://doi.org/10.1109/36.297978, 1994.

Desmons, M., Ferlay, N., Parol, F., Riédi, J., and Thieuleux, F.: A Global Multilayer Cloud Identification with POLDER/PARASOL, Journal of Applied Meteorology and Climatology, 56, 1121–1139, https://doi.org/10.1175/JAMC-D-16-0159.1, 2017.

Dice, L. R.: Measures of the Amount of Ecologic Association Between Species, Ecology, 26, 297–302, https://doi.org/https://doi.org/10.2307/1932409, 1945.

Diner, D., Beckert, J., Reilly, T., Bruegge, C., Conel, J., Kahn, R., Martonchik, J., Ackerman, T., Davies, R., Gerstl, S., Gordon, H., Muller, J.-P., Myneni, R., Sellers, P., Pinty, B., and Verstraete, M.: Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview, IEEE Transactions on Geoscience and Remote Sensing, 36, 1072–1087, https://doi.org/10.1109/36.700992, 1998.

Dubuisson, P., Dessailly, D., Vesperini, M., and Frouin, R.: Water vapor retrieval over ocean using near-infrared radiometry, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/https://doi.org/10.1029/2004JD004516, 2004.

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

Eelbode, T., Bertels, J., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., and Blaschko, M. B.: Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index, IEEE Transactions on Medical Imaging, 39, 3679–3690, https://doi.org/10.1109/TMI.2020.3002417, 2020.

Ferlay, N., Thieuleux, F., Cornet, C., Davis, A. B., Dubuisson, P., Ducos, F., Parol, F., Riédi, J., and Vanbauce, C.: Toward New Inferences about Cloud Structures from Multidirectional Measurements in the Oxygen A Band: Middle-of-Cloud Pressure and Cloud Geometrical Thickness from POLDER-3/PARASOL, Journal of Applied Meteorology and Climatology, 49, 2492 – 2507, https://doi.org/https://doi.org/10.1175/2010JAMC2550.1, 2010.

Gao, M., Franz, B. A., Zhai, P.-W., Knobelspiesse, K., Sayer, A., Xu, X., Martins, V., Cairns, B., Castellanos, P., Fu, G., Hannadige, N., Hasekamp, O., Hu, Y., Ibrahim, A., Patt, F., Puthukkudy, A., and Werdell, P. J.: Simultaneous retrieval of aerosol and ocean properties from PACE HARP2 with uncertainty assessment using cascading neural network radiative transfer models, EGUsphere, 2023, 1–31, https://doi.org/10.5194/egusphere-2023-1843, 2023.

Gupta, R. and Hartley, R.: Linear pushbroom cameras, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19, 963–975, https://doi.org/10.1109/34.615446, 1997.

Hagolle, O., Guerry, A., Cunin, L., Millet, B., Perbos, J., Laherrere, J.-M., Bret-Dibat, T., and Poutier, L.: POLDER level-1 processing algorithms, Proceedings of SPIE - The International Society for Optical Engineering, 2758, 308–319, https://doi.org/10.1117/12.243226, 1996.

Haynes, J. M., L'Ecuyer, T. S., Stephens, G. L., Miller, S. D., Mitrescu, C., Wood, N. B., and Tanelli, S.: Rainfall retrieval over the ocean with spaceborne W-band radar, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/https://doi.org/10.1029/2008JD009973, 2009.

Im, E., Wu, C., and Durden, S.: Cloud profiling radar for the CloudSat mission, in: IEEE International Radar Conference, 2005., pp. 483–486, https://doi.org/10.1109/RADAR.2005.1435874, 2005.

Jnawali, K., Arbabshirani, M. R., Rao, N., and M.D., A. A. P.: Deep 3D convolution neural network for CT brain hemorrhage classification, in: Medical Imaging 2018: Computer-Aided Diagnosis, edited by Petrick, N. and Mori, K., vol. 10575, p. 105751C, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.2293725, 2018.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, 2017.

Liu, G.: Deriving snow cloud characteristics from CloudSat observations, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/https://doi.org/10.1029/2007JD009766, 2008.

L'Ecuyer, T. S., Beaudoing, H. K., Rodell, M., Olson, W., Lin, B., Kato, S., Clayson, C. A., Wood, E., Sheffield, J., Adler, R., Huffman, G., Bosilovich, M., Gu, G., Robertson, F., Houser, P. R., Chambers, D., Famiglietti, J. S., Fetzer, E., Liu, W. T., Gao, X., Schlosser, C. A., Clark, E., Lettenmaier, D. P., and Hilburn, K.: The Observed State of the Energy Budget in the Early Twenty-First Century, Journal of Climate, 28, 8319 – 8346, https://doi.org/https://doi.org/10.1175/JCLI-D-14-00556.1, 2015.

Mace, G. G. and Zhang, Q.: The CloudSat radar-lidar geometrical profile product (RL-GeoProf): Updates, improvements, and selected results, Journal of Geophysical Research: Atmospheres, 119, 9441–9462, https://doi.org/https://doi.org/10.1002/2013JD021374, 2014.

Marchand, R., Ackerman, T., Smyth, M., and Rossow, W. B.: A review of cloud top height and optical depth histograms from MISR, ISCCP, and MODIS, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/https://doi.org/10.1029/2009JD013422, 2010.

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., Taylor, K. E., and Schlund, M.: Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models, Science Advances, 6, eaba1981, https://doi.org/10.1126/sciadv.aba1981, 2020.

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

Mitra, A., Loveridge, J. R., and Di Girolamo, L.: Fusion of MISR Stereo Cloud Heights and Terra-MODIS Thermal Infrared Radiances to Estimate Two-Layered Cloud Properties, Journal of Geophysical Research: Atmospheres, 128, e2022JD038 135, https://doi.org/https://doi.org/10.1029/2022JD038135, e2022JD038135 2022JD038135, 2023.

Mohajerani, S. and Saeedi, P.: Cloud and Cloud Shadow Segmentation for Remote Sensing Imagery Via Filtered Jaccard Loss Function and Parametric Augmentation, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 4254–4266, https://doi.org/10.1109/JSTARS.2021.3070786, 2021.

Mousavian, A., Anguelov, D., Flynn, J., and Kosecka, J.: 3D Bounding Box Estimation Using Deep Learning and Geometry, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

Muller, J., Denis, M., Dundas, R. D., Mitchell, K. L., Naud, C., and Mannstein, H.: Stereo cloud-top heights and cloud fraction retrieval from ATSR-2, International Journal of Remote Sensing, 28, 1921–1938, https://doi.org/10.1080/01431160601030975, 2007.

Nelson, D., Garay, M., Diner, D., and Kahn, R.: The MISR Wildfire Smoke Plume Height Project, AGU Fall Meeting Abstracts, 2010.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E. Z., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, CoRR, abs/1912.01703, http://arxiv.org/abs/1912.01703, 2019.

Pörtner, H.-O., Roberts, D., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., and Rama, B.: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, IPCC Sixth Assessment Report, 2022.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597 [cs], 2015.

Rossow, W. B., Knapp, K. R., and Young, A. H.: International Satellite Cloud Climatology Project: Extending the Record, Journal of Climate, 35, 141 – 158, https://doi.org/10.1175/JCLI-D-21-0157.1, 2022.

Sassen, K. and Wang, Z.: Classifying Clouds around the Globe with the CloudSat Radar: 1-Year of Results, Geophysical Research Letters, 35, https://doi.org/10.1029/2007GL032591, 2008.

Schönberger, J. L. and Frahm, J.-M.: Structure-from-Motion Revisited, in: Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

Schonberger, J. L. and Frahm, J.-M.: Structure-From-Motion Revisited, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

Schönberger, J. L., Zheng, E., Pollefeys, M., and Frahm, J.-M.: Pixelwise View Selection for Unstructured Multi-View Stereo, in: European Conference on Computer Vision (ECCV), 2016.

Sde-Chen, Y., Schechner, Y. Y., Holodovsky, V., and Eytan, E.: 3DeepCT: Learning Volumetric Scattering Tomography of Clouds, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5651–5662, https://doi.org/10.1109/ICCV48922.2021.00562, 2021.

Sinclair, K., van Diedenhoven, B., Cairns, B., Yorks, J., Wasilewski, A., and McGill, M.: Remote sensing of multiple cloud layer heights using multi-angular measurements, Atmospheric Measurement Techniques, 10, 2361–2375, https://doi.org/10.5194/amt-10-2361-2017, 2017.

Smalley, M., L'Ecuyer, T., Lebsock, M., and Haynes, J.: A Comparison of Precipitation Occurrence from the NCEP Stage IV QPE Product and the CloudSat Cloud Profiling Radar, Journal of Hydrometeorology, 15, 444 – 458, https://doi.org/https://doi.org/10.1175/JHM-D-13-048.1, 2014.

https://doi.org/10.5194/egusphere-2023-2392
Preprint. Discussion started: 19 January 2024

Soffer, S., Klang, E., Shimon, O., Barash, Y., Cahan, N., Greenspana, H., and Konen, E.: Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis, Scientific Reports, 11, 15 814, https://doi.org/10.1038/s41598-021-95249-3, 2021.

Stephens, G., Winker, D., Pelon, J., Trepte, C., Vane, D., Yuhas, C., L'Ecuyer, T., and Lebsock, M.: CloudSat and CALIPSO within the A-Train: Ten Years of Actively Observing the Earth System, Bulletin of the American Meteorological Society, 99, 569 – 581, https://doi.org/https://doi.org/10.1175/BAMS-D-16-0324.1, 2018.

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., Illingworth, A. J., O'connor, E. J., Rossow, W. B., Durden, S. L., Miller, S. D., Austin, R. T., Benedetti, A., and Mitrescu, C.: THE CLOUDSAT MISSION AND THE A-TRAIN: A New Dimension of Space-Based Observations of Clouds and Precipitation, Bulletin of the American Meteorological Society, 83, 1771 – 1790, https://doi.org/https://doi.org/10.1175/BAMS-83-12-1771, 2002.

Stillinger, T., Roberts, D. A., Collar, N. M., and Dozier, J.: Cloud Masking for Landsat 8 and MODIS Terra Over Snow-Covered Terrain: Error Analysis and Spectral Similarity Between Snow and Cloud, Water Resources Research, 55, 6169–6184, https://doi.org/https://doi.org/10.1029/2019WR024932, 2019.

Wang, Z. and Blaschko, M. B.: Jaccard Metric Losses: Optimizing the Jaccard Index with Soft Labels, 2023.

Waquet, F., Cornet, C., Deuzé, J.-L., Dubovik, O., Ducos, F., Goloub, P., Herman, M., Lapyonok, T., Labonnote, L. C., Riedi, J., Tanré, D., Thieuleux, F., and Vanbauce, C.: Retrieval of Aerosol Microphysical and Optical Properties above Liquid Clouds from POLDER/-PARASOL Polarization Measurements, Atmospheric Measurement Techniques, 6, 991–1016, https://doi.org/10.5194/amt-6-991-2013, 2013.

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., Knobelspiesse, K. D., Mannino, A., Martins, J. V., McClain, C. R., Meister, G., and Remer, L. A.: The Plankton, Aerosol, Cloud, Ocean Ecosystem Mission: Status, Science, Advances, Bulletin of the American Meteorological Society, 100, 1775 – 1794, https://doi.org/https://doi.org/10.1175/BAMS-D-18-0056.1, 2019.

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 CALIPSO Mission and CALIOP Data Processing Algorithms, Journal of Atmospheric and Oceanic Technology, 26, 2310 – 2323, https://doi.org/https://doi.org/10.1175/2009JTECHA1281.1, 2009.

Wu, L., Hasekamp, O., van Diedenhoven, B., Cairns, B., Yorks, J. E., and Chowdhary, J.: Passive remote sensing of aerosol layer height using near-UV multiangle polarization measurements, Geophysical Research Letters, 43, 8783–8790, https://doi.org/https://doi.org/10.1002/2016GL069848, 2016.

Zhang, K., Sun, J., and Snavely, N.: Leveraging Vision Reconstruction Pipelines for Satellite Imagery, 2019.

Zijdenbos, A., Dawant, B., Margolin, R., and Palmer, A.: Morphometric analysis of white matter lesions in MR images: method and validation, IEEE Transactions on Medical Imaging, 13, 716–724, https://doi.org/10.1109/42.363096, 1994.