Cloud Fields and Aerosol Classification with Lidar using Advanced AI Approach

Yonatan Peleg[1], Lior Zeida-Cohen[1], Imri Tzror[1], Johannes Bühl[2], Albert Ansmann[3], Alexandra Chudnovsky[4], and Zohar Yakhini[1]

[1]Efi Arazi School of Computer Science, Reichman University, Herzliya, Israel
[2]Harz University of Applied Sciences, Wernigerode, Germany
[3]Leibniz Institute for Tropospheirc Research (TROPOS), Leibniz, Germany
[4]Faculty of Exact Sciences, Department of Geophysics, Tel Aviv University, Tel Aviv, Israel

Correspondence: Yonatan Peleg (yoni.peleg@post.runi.ac.il)

Abstract. Understanding the vertical distribution of aerosol and clouds is critical for climate modeling, weather forecasting, and air quality monitoring. Lidar observations are central to profiling atmospheric composition, yet signal attenuation in optically thick layers limits the effective retrieval of some important properties above those layers. More complex measurement approaches, using a combination of Lidar and cloud radar systems, can be taken to support more inclusive and accurate in-
5   ference. In this study, we develop a deep learning framework to address this trade-off and gap in cost of data acquisition by enabling full-column aerosol and cloud classification using only standard lidar inputs. The approach is based on a U-Net architecture trained to predict combined aerosol and cloud types from vertical profiles of backscatter and depolarization. Classification targets integrate established aerosol typing from PollyXT with cloud and precipitation categorization from Cloudnet, facilitating a unified scheme. The model achieves high precision, recall and F1-scores above 95%. By evaluating numerous
10   complex case studies, we establish the model's ability to exploit information embedded in the lidar signal below attenuating layers, including structural and contextual features, to infer atmospheric conditions at higher altitudes, offering a robust AI-based enhancement to lidar-based atmospheric profiling and target classification. The application of AI in this context closes the gap between the need for vertical cloud maps and the sparse availability of Cloudnet.

1   Introduction

15   Aerosols and clouds are fundamental components of the Earth's atmosphere, exerting profound influences on the planet's climate system. They modulate radiative energy balance through scattering and absorption of solar and terrestrial radiation, significantly impacting the hydrological cycle through precipitation processes (Ramanathan et al., 2001; Albrecht, 1989). They also play a role in atmospheric chemistry by interacting with gasses and providing surfaces for chemical reactions to occur (Intergovernmental Panel on Climate Change (IPCC), 2014; Jacob, 2000; Lelieveld and Crutzen, 1991). Clouds, in particular,
20   have large but opposing effects on short-wave and long-wave radiation, resulting in a significant net cooling effect globally, although the magnitude remains uncertain (Hartmann and Doelling, 1991). Aerosol-cloud interactions (ACI), represent one of the largest uncertainties in current climate projections (Intergovernmental Panel on Climate Change (IPCC), 2014; Rosenfeld

et al., 2014). Understanding the vertical distribution and properties of both aerosol and clouds is therefore critical to accurately understand and quantify their climatic impacts and to improve the representation of atmospheric processes in weather prediction

25 and climate models (Winker et al., 2010; Weitkamp, 2005; Rogozovsky et al., 2023).

Significant progress has been made in the development of remote sensing techniques capable of profiling atmospheric constituents. Lidar has emerged as a powerful tool for providing detailed vertical profiles of aerosol particles and clouds with high spatial and temporal resolution (Cairo et al., 2024). Advanced lidar systems, such as multiwavelength Raman and polarization lidars, can retrieve not only the vertical distribution of aerosol backscatter but also intensive optical properties. These properties

30 include the lidar ratio (extinction-to-backscatter ratio) and the particle linear depolarization ratio, which provide crucial information about particle size, shape, and absorption characteristics, enabling the classification of different aerosol types and its different vertical distribution (e.g., dust, smoke, marine, urban haze) (Rogozovsky et al., 2025; Baars et al., 2017). Networks like PollyNET, using standardized and automated PollyXT lidars, demonstrate the capability for continuous, near-real-time monitoring and characterization of aerosol profiles in diverse global locations (Baars et al., 2018, 2017).

35 Lidar is highly effective for detecting aerosols and thin cloud layers, yet it is constrained by signal attenuation (Winker et al., 2010; Weitkamp, 2005). The laser beam can be strongly scattered and absorbed by dense atmospheric constituents, particularly liquid water droplets (Haarig et al., 2023). In optically thick clouds, especially those containing liquid water, the lidar signal is usually fully attenuated within a few hundred meters above the cloud base, typically at optical depths (τ) around 3–5 (Winker et al., 2017). This attenuation prevents the lidar from probing the full vertical extent of the cloud, hindering the characterization

40 of the height of the cloud top, the internal structure, and the thermodynamic phase (Kalesse-Los et al., 2022). This physical limitation is the primary reason why synergistic approaches like Cloudnet rely on cloud radar, which can easily penetrate multiple cloud layers, to provide information above the lidar attenuation height (Bühl et al., 2017).

Cloudnet integrates measurements from a lidar ceilometer, a cloud radar (typically millimeter-wavelength), a microwave radiometer (for integrated liquid water path), and thermodynamic profiles (temperature, humidity) from numerical weather

45 prediction models. It provides a detailed target classification, distinguishing between clear sky, aerosol, various cloud phases (liquid droplets, supercooled liquid, ice), precipitation types (drizzle, rain, snow), and even non-meteorological targets like insects (Illingworth et al., 2007). Due to the complexity and high cost of operating and maintaining full Cloudnet setups, there is a need for new scalable, data driven methods that can extract comparable cloud information from single instrument observations such as lidar. While techniques exist to determine cloud base from lidar (Pal et al., 1992), accurately classifying

50 the entire cloud column using only lidar data remains a significant challenge, particularly for multi-layer, mixed-phase, or deep convective cloud systems (Kalesse-Los et al., 2022).

The rapid advancement of deep learning, particularly through architectures such as Convolutional Neural Networks (CNNs) and the U-Net, has introduced innovative approaches for analyzing complex datasets. These methods have achieved outstanding performance in pattern recognition and image segmentation across a wide range of disciplines (Krizhevsky et al., 2017; LeCun

55 et al., 2015; Ronneberger et al., 2015; Reichstein et al., 2019). In atmospheric science, deep learning has already demonstrated significant potential, for example in cloud phase classification from radar Doppler spectra (Schimmel et al., 2022), cloud detection and segmentation from satellite imagery, and the analysis of lidar point clouds (Biasutti et al., 2019). A particularly

compelling application lies in addressing the long-standing problem of signal attenuation in lidar observations. Conventional retrieval methods are limited once the backscatter signal is extinguished, leaving the atmospheric structure above the attenuation

60 height poorly constrained. Deep learning offers a new perspective: by leveraging the rich set of features, correlations, and contextual cues embedded in the portion of the lidar profile below the attenuation threshold, and potentially its spatial and temporal evolution (Bansal et al., 2022), it infer the atmospheric properties above. A U-Net model, with its strong capacity for hierarchical feature extraction and pattern recognition, is especially well suited to capture these subtle and non-intuitive relationships.

65 Machine learning has been increasingly applied to the detection and classification of aerosols and clouds from lidar observations. In most approaches, lidar measurements—commonly represented as time–height or along-track–height cross-sections—are treated as two-dimensional images. This formulation allows CNNs and U-Net architectures to learn spatial textures, morphological features, and contextual patterns directly from the data. For example, a CNN was developed for cloud–aerosol discrimination using only lidar measurements from NASA's Ice, Cloud, and Land Elevation Satellite (ICESat-2)

70 (Oladipo et al., 2024). Similarly, a U-Net model enhanced with self-attention mechanisms was constructed to classify cloud and aerosol layers in atmospheric vertical profiles using CALIPSO L1 data (Zhou et al., 2024). Both studies demonstrated the capacity of deep learning to reliably separate aerosols from clouds. However, their focus remained limited to binary cloud–aerosol discrimination, without further subdivision into specific categories.

Another recent study introduced a multitask machine learning framework for space-based lidar, capable of simultaneous

75 cloud-aerosol discrimination and aerosol typing (Fuller et al., 2025) . While their approach successfully improves the spatial resolution of retrievals compared to standard products, their model is trained on lidar-derived optical products and is therefore strictly bound by the physical signal limitations of the lidar instrument itself. Consequently, the model cannot infer or characterize atmospheric structures in regions where the lidar signal is fully attenuated.

Other direction of research has targeted aerosol sub-classification using lidar data. One study applied traditional algorithms to

80 first detect atmospheric layers and compute their integrated optical properties; these derived feature vectors were subsequently classified by a standard artificial neural network (ANN) (Nicolae et al., 2018). A more recent comparison of six machine learning models for aerosol typing identified LightGBM as the most effective (del Águila et al., 2025). While these efforts highlight the promise of machine learning for aerosol categorization, they do not extend to the joint classification of cloud and aerosol subtypes. Importantly, distinct aerosol and cloud categories often exhibit complex cross-category and cross-type

85 interactions. Capturing these interactions requires integrated datasets that explicitly combine both aerosol and cloud categories as classification targets.

This paper introduces a deep learning methodology aimed at achieving unified aerosol and cloud classification throughout the vertical atmospheric column using only standard lidar measurements as input. The approach utilizes a U-Net architecture, trained end-to-end to map vertical profiles of lidar backscatter and depolarization to a combined target classification derived

90 from PollyXT aerosol typing and Cloudnet cloud precipitation classification.

2 Data

The primary input to the deep learning model consists of vertically resolved profiles obtained from ground-based lidar measurements in Limassol, Cyprus, between November 2016 and April 2018. Data were formatted as two-dimensional (2D) time-height images representing a sequence of profiles. Using 2D inputs allows the CNN architecture to exploit spatio-temporal context, capturing dynamic features or advection patterns relevant to the classification task. Our data consists of a temporal resolution of 90 seconds and a vertical resolution of 37 meters, where each image spans over 24 hours and 22.5 kilometers in height. Data were provided by the Leibniz Institute for Tropospheric Research (TROPOS). The terms 'sample' and 'image' will be used intermittently to describe one 2D (time-height) input, where an 'image' is a multi-channel time-height dataset.

The input features used in this study include: attenuated backscatter coefficient at 532 nm and 1064 nm, aerosol backscatter coefficients at 532 nm and 1064 nm, particle depolarization ratio at 532 nm , volume depolarization ratio at 532 nm, the backscatter-related Ångström exponent between 532 nm and 1064 nm, model pressure and model temperature. For generation of the training dataset, all variables of the Cloudnet processing scheme were mapped to the PollyNET time-height grid.

The target variable represents a unified classification that combines aerosol and cloud/precipitation types for each vertical bin in the profile. It was constructed by synergistically combining outputs from PollyXT aerosol target classification and Cloudnet cloud and precipitation target classification. The label classes are numbered as detailed in Table 1, and Figure 1 shows an example target image that the model aims to learn.

Table 1. Classification Labels

| Class Number	Class Label
0	No class
1	Clean atmosphere
2	None-typed particles/low concentration
3	Aerosol small
4	Aerosol large spherical
5	Aerosol mixture partly non-spherical
6	Aerosol large non-spherical
7	Cloud non-typed
8	Cloud water droplets
9	Cloud likely water droplets
10	Cloud ice crystals
11	Cloud likely ice crystals
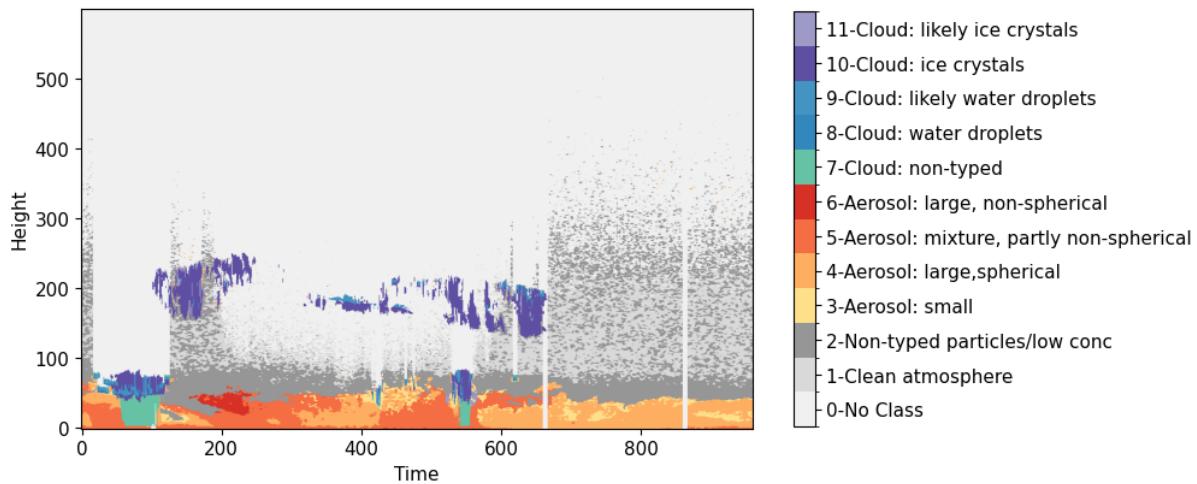
Figure 1. Unified atmospheric target classification mask. A 2D time-height training label for November 3, 2016, created by integrating Pol-lyXT aerosol typing and Cloudnet cloud/precipitation categorization. The vertical axis represents altitude (up to 22.5 km), and the horizontal axis represents 24 hours of observations at 90-second resolution. Classes range from clear atmosphere (Class 1) to specific aerosol types (Classes 3–6) and various cloud phases including water droplets and ice crystals (Classes 8–11).

3   Methodology

3.1   Data preprocessing

Prior to ingestion by the U-Net model, the raw lidar data were subjected to a series of preprocessing steps to ensure data quality
and suitability for the network architecture. Initially, each input sample was filtered to conform to the expected input dimensions
of 600 vertical bins by 960 time steps. Instances containing missing timestamps were identified; since they constituted only
3% of the entire dataset, these samples were excluded from the training, validation and testing sets.

Subsequently, the lidar data were clipped to zero, given that lidar-derived physical quantities such as backscatter and de-
polarization ratios are inherently non-negative. If negative values exist in lidar data, they typically resulted from instrumental
noise or background subtraction procedures, particularly in regions with low signal-to-noise ratios.

Next, we impute NaN entries which occurr in lidar profiles due to factors such as incomplete correction of signal attenuation
in dense clouds, instrument malfunction, or measurements beyond the reliable operational range, were replaced using the global
average of the respective feature.

To address the highly skewed distributions of lidar signals, especially the backscatter coefficients that often span several
orders of magnitude, a $log(1 + x)$ transformation was applied to the features. This transformation compresses the range of
high-intensity values and expands the range of low-intensity values, resulting in a more symmetrical, approximately Gaussian
distribution. An example of this transformation for the backscatter-related Ångström exponent between 532 nm and 1064 nm

from one image can be seen in Figure 2. This transformation is beneficial in stabilizing the variance, reducing the disproportionate influence of extreme outliers, and improving the learning process of the neural network (Box and Cox, 1964).
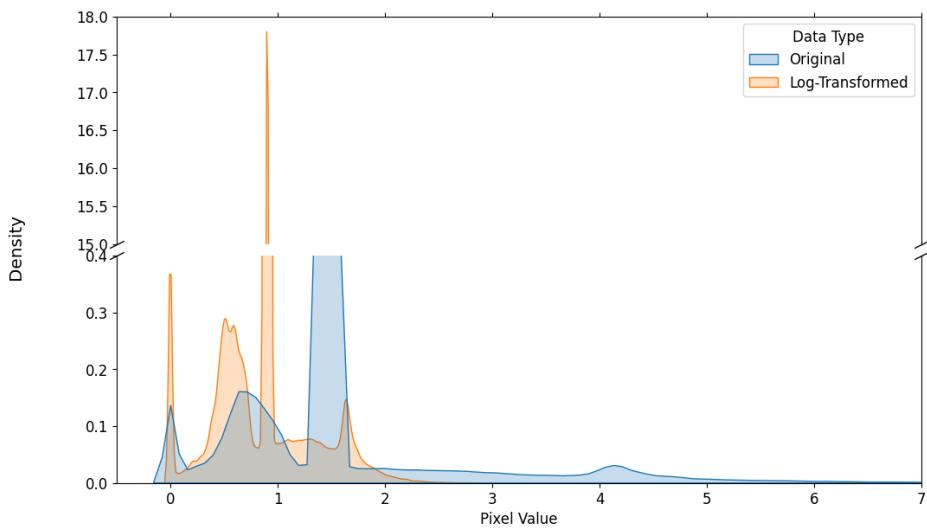

Figure 2. Statistical distribution of lidar input features. Density plot comparing the original skewed distribution (blue) of the backscatter-related Ångström exponent (532/1064 nm) against its log-transformed $log(1+x)$ counterpart (orange) for February 1, 2018. This preprocessing step stabilizes variance and compresses the dynamic range of high-intensity backscatter values to improve neural network convergence and reduce the impact of extreme outliers.

Following the logarithmic transformation, the features were standardized according to:

$$z_i = \frac{x_i - \mu}{\sigma}$$

where μ and σ were calculated from the log-transformed per-feature training data, and then applied to the validation and test datasets. Standardization transforms the features to have a mean of approximately zero and a standard deviation of approximately one, ensuring that all features contribute more equally to the learning process, preventing domination by features with larger numerical ranges, and promoting faster and more stable training (Ioffe and Szegedy, 2015).

Additionally, for each input feature, a corresponding binary indicator feature was created to indicate signal absence. This indicator took a value of 1 if the original data at that specific pixel had been a NaN (and subsequently imputed), and 0 otherwise. This step provided the model with explicit information about the original data quality at each point (Jones, 1996).

3.2 Learning

The dataset was partitioned into training, validation, and testing subsets in a 70:10:20 ratio, resulting in 284 samples allocated for training, 40 for validation, and 80 for testing. Care was taken to ensure that each subset maintained a similar variance in the distribution of the total distinct number of atmospheric classes present within each sample. For each subset, let I be the set of
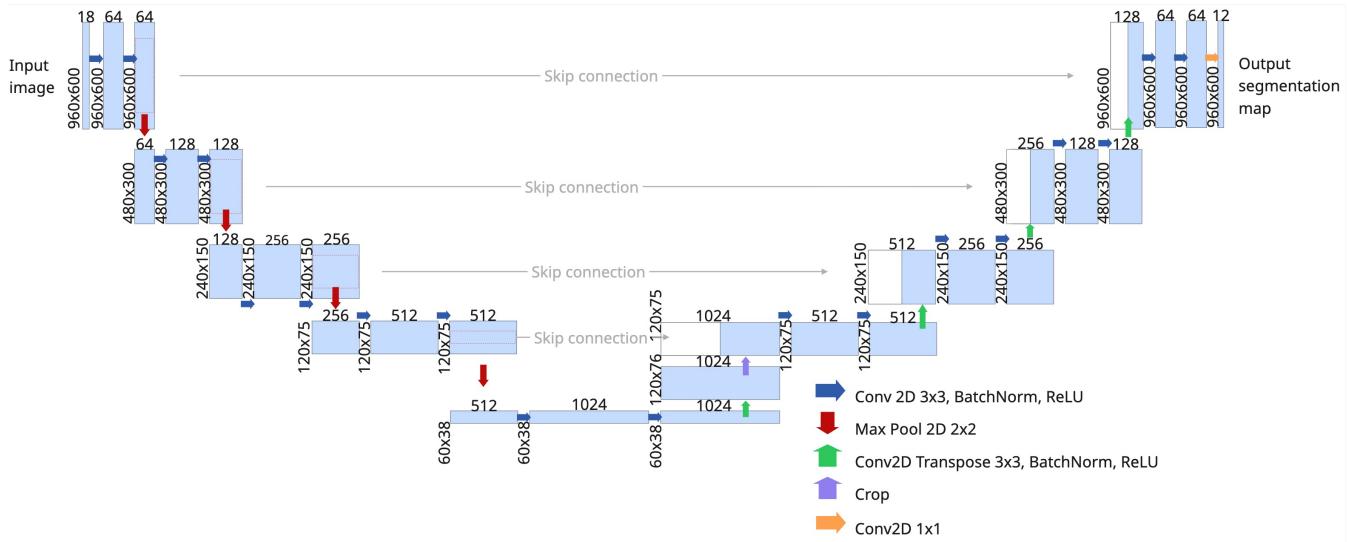
Figure 3. U-Net architecture for pixel-wise atmospheric segmentation. This U-Net-style architecture features an encoder, a bottleneck, and a decoder. The encoder captures context through four downsampling blocks, where each block applies two sequential 3×3 convolution layers (followed by batch normalization and ReLU) with an increasing number of filters (64, 128, 256, 512), a dropout layer with a rate of 0.1, and a final 2×2 max-pooling operation. Feature maps from the encoder are passed to the decoder via skip connections. The central bottleneck layer consists of two 3×3 convolutions with 1024 filters, batch norm, ReLU, and a dropout layer with a rate of 0.2. The decoder path enables exact localization by mirroring the encoder. Each of its four upsampling blocks uses a 2×2 transposed convolution, concatenates its output with the corresponding feature map from the skip connection, and applies two more 3×3 convolutions. The network terminates with a 1×1 convolution and a softmax activation function to generate the final pixel-wise segmentation mask.

all samples and c be a specific class. Thus:

$$I_c = \{i \in I \mid sample\ i\ contains\ class\ c\}$$

$$Count_c = |I_c| \tag{1}$$

140   We then plot the kernel density estimate (KDE) of each subset and compare the curves to make sure that the distributions behave similarly (Figure 4). This stratified approach helped ensure that each dataset split was representative of the overall class diversity, preventing potential biases, and ensuring robust model training and evaluation.

A U-Net architecture (Figure 3) was selected for this task due to its demonstrated success in image segmentation, including applications in earth and atmospheric sciences (Foley et al., 2024) (Rusyn et al., 2019) (Galea et al., 2023) and other applica-
145 tions (Levy-Jurgenson et al., 2020). Its structure is particularly well-suited for tasks requiring both contextual understanding and precise localization. The architecture follows a symmetric encoder-decoder structure with skip connections, tailored to capture both hierarchical features and fine-grained spatial details. The encoder decreases the dimensions of the input and increases the number of feature channels, and the decoder increases the dimensions of the bridge data and decreases the number of feature channels.
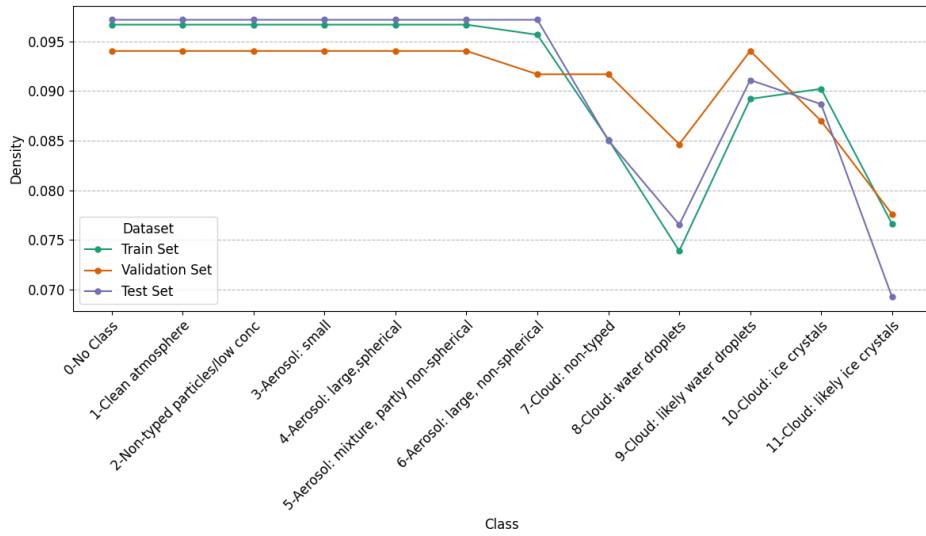
Figure 4. Dataset stratification analysis. Kernel Density Estimate (KDE) plot showing the frequency of the 12 atmospheric classes across the Training (70%), Validation (10%), and Test (20%) subsets. The alignment of the curves ensures that each split is representative of the overall atmospheric diversity, preventing class frequency bias during model evaluation.

To effectively train the U-Net model for our multi-class segmentation task, a composite loss function was employed. This tailored loss function was designed to address two key challenges: achieving accurate segmentation for each individual atmospheric class, many of which are imbalanced, and specifically penalizing confusion between aerosol and clouds. The total loss, L_{total}, is defined as:

$$L_{total} = L_{MWSD} + \lambda \cdot L_{GC} \tag{2}$$

Where L_{MWSD} is the multiclass weighted squared dice loss, L_{GC} (described below) is the group confusion loss (Equations 3 and 5 respectively). The penalty λ is a penalty factor that balances the contribution of the group confusion term (L_{GC} - described below). This penalty was a hyperparameter and was optimized based on the best results of the validation set.

Dice coefficient is a common metric for evaluating overlap in segmentation tasks, and its loss variant has proven to be effective, particularly for imbalanced classes (Sudre et al., 2017) (Milletari et al., 2016). To address the inherent imbalance in the frequency of different atmospheric aerosol and cloud types, class weights w_{class} are introduced (Bressan et al., 2022). These weights were chosen to be inversely proportional to the frequency or volume of each class in the training dataset, thereby giving more importance to underrepresented classes. A multiplicative factor was applied for training stability. A weight of 0 was applied to the first class which is defined as 'no-class' (Class 0). This was done to ensure that the model only learns physical features and assigns anything that doesn't fit a physical phenomenon to 'no class' (Class 0). The dice coefficient was

165 calculated as follows:

$$L_{MWSD} = 1 - 2 \cdot \left(\frac{\sum_{j=1}^{h} \sum_{i=1}^{t} y_t(i,j) \cdot y_p(i,j) \cdot w_c(y_t(i,j))}{\sum_{j=1}^{h} \sum_{i=1}^{t} (y_t^2(i,j) + y_p^2(i,j)) \cdot w_c(y_t(i,j))} \right) \tag{3}$$

$$w_c(i,j) = 200 \cdot \frac{1}{n_c} \tag{4}$$

Where h and t are the height and time indices, y_t are the ground truth labels, y_p are the prediction probabilities, w_c is the scaled

170 inverse frequency of the true training class and n_c is the number of times a specific true class appears in the training dataset. Note that w_c is the weight associated to each class and is a function of y_t.

Although L_{MWSD} focuses on individual class performance, a critical requirement for this application is to strongly discourage misclassifications between fundamentally different atmospheric categories, specifically between aerosol and clouds. To address this, a group confusion loss term, L_{GC} is introduced:

175 $$L_{GC} = \sum_{j=1}^{h} \sum_{i=1}^{t} y_{ta}(i,j) \cdot y_{pc}(i,j) + y_{tc}(i,j) \cdot y_{pa}(i,j) \tag{5}$$

Where y_{ta} are the ground truth aerosol labels, y_{pc} are the predicted cloud probabilities, y_{tc} are the ground truth cloud labels and y_{pa} are the predicted aerosol probabilities.

The composite loss function (Eq. 5) is designed to guide the U-Net model not only to accurately segment individual aerosol and cloud types (even rare ones, due to weighting in L_{MWSD}), but also to maintain a clear distinction between the broader

180 aerosol and cloud categories.

The U-Net model was developed and trained using the TensorFlow and Keras libraries. Adam optimization algorithm, with an initial learning rate of $5 \cdot 10^{-3}$, was employed to minimize the loss function (Eq. 2). To prevent overfitting and reduce unnecessary training time, early stopping was implemented as follows. Validation loss was monitored and training was halted if no improvement was observed for 20 consecutive epochs, starting from epoch 50. The best weights achieved during training

185 (post epoch 50) were restored upon stopping. The learning rate was adaptively adjusted during training. If the validation loss did not improve for 10 epochs, the learning rate was reduced by a factor of 0.2, down to a minimum learning rate of $5 \cdot 10^{-7}$. This allows for finer adjustments as the model approaches convergence. Training was implemented using Tensorflow on Google Cloud Vertex AI, Colab Enterprise notebook, using NVIDIA TESLA A100x4 GPUs.

4 Results and discussion

190 Model training stopped after 167 steps due to the early stopping mechanism, where epoch 148 was chosen as the best epoch with a training loss value of 0.1198 and a validation loss value of 0.1288.

Jaccard index at the chosen epoch was 0.9009 for the training set and 0.5955 for the validation set, and cloud-aerosol confusion loss was 0.0023.

Both loss values and Jaccard indices for the training and validation sets plateaued at around step 120, where the most

195 significant learning was done between steps 0 and 60. Results henceforth will be discussed solely regarding the test dataset, which comprises 80 samples that the model has not seen. The model's performance was quantitatively assessed using metrics such as precision, recall, and F1-score for each atmospheric class:

$$P(i) = \frac{TP(i)}{TP(i) + FP(i)} \tag{6}$$

$$R(i) = \frac{TP(i)}{TP(i) + FN(i)} \tag{7}$$

200 $$F1(i) = 2 \cdot \frac{P(i) \cdot R(i)}{P(i) + R(i)} = \frac{TP(i)}{TP(i) + \frac{1}{2}(FP(i) + FN(i))} \tag{8}$$

Where TP, FP and FN are the true positive, false positive and false negative results, and where i represents the class. To determine the overall performance of the model, the macro-averaged F1 scores, weighted macro-averaged F1 scores, and micro F1 scores were calculated:

$$F1_{macro} = \frac{\sum_{i=1}^{12} F1(i)}{12}, \quad P_{macro} = \frac{\sum_{i=1}^{12} P(i)}{12}, \quad R_{macro} = \frac{\sum_{i=1}^{12} R(i)}{12}, \tag{9}$$

205 $$F1_{weighted} = \sum_{i=1}^{12} w_i F1(i), \quad P_{weighted} = \sum_{i=1}^{12} w_i P(i), \quad R_{weighted} = \sum_{i=1}^{12} w_i R(i), \tag{10}$$

$$F1_{micro} = \frac{\sum_{i=1}^{12} TP(i)}{\sum_{i=1}^{12} TP(i) + \frac{1}{2}(\sum_{i=1}^{12} FP(i) + \sum_{i=1}^{12} FN(i))}$$
$$P_{micro} = \frac{\sum_{i=1}^{12} TP(i)}{\sum_{i=1}^{12} TP(i) + \sum_{i=1}^{12} FP(i)} \tag{11}$$
$$R_{micro} = \frac{\sum_{i=1}^{12} TP(i)}{\sum_{i=1}^{12} TP(i) + \sum_{i=1}^{12} FN(i)}$$

Where i runs on all classes, and w_i is the fraction of that class within the entire test dataset, and TP, FP, FN are the sum of the true positive, false positive and false negative results. The evaluation is based on the test dataset, which comprises 46,080,000 individual pixel classifications in all test samples, and is detailed in Table 2.
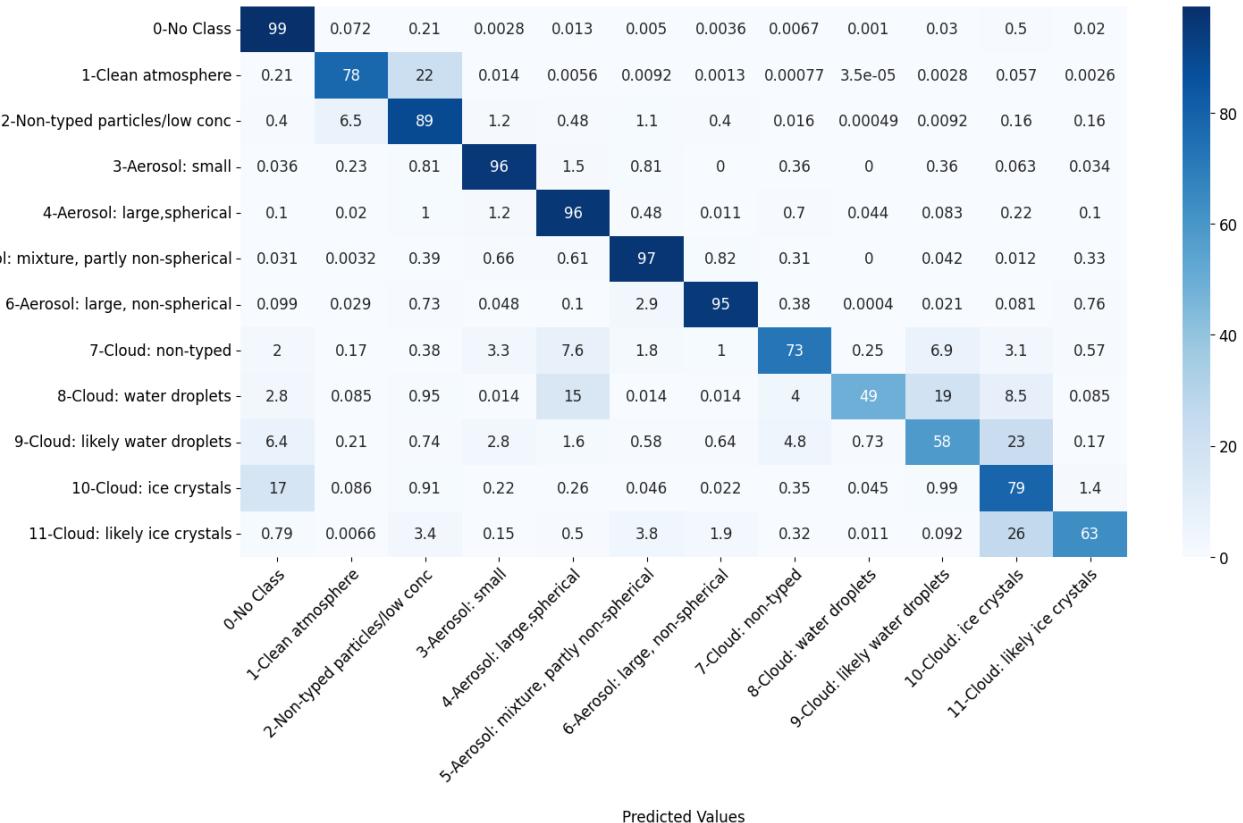
Figure 5. Comprehensive model performance matrix. Normalized confusion matrix for the test dataset. Diagonal elements represent the recall for each class, showing high accuracy for aerosol types (89%–95%). The matrix highlights physically plausible confusion between liquid cloud categories (Classes 8 and 9) and the model's high precision in identifying the "No Class" background.

4.1 Analysis

The model achieved consistent performance across classes, with a micro-average (Eq. 11) as well as weighted averages (Eq. 10 for accuracy, precision, recall, and F1-score of 0.96. These results indicate that, when class imbalances are taken into account, the classification is reliable and unbiased toward specific categories. However, the macro average (Eq. 9), which calculates the metric independently for each class and then averages them assuming equal weights, shows more moderate results: a precision of 0.82, recall of 0.81, and an F1-score of 0.81. The difference between the weighted and macro averages points to class imbalance, where the model performs well on the aerosol classes which are more common, but struggles more with cloud classes which are more rare.

Table 2. Classification Performance Metrics

Class description	Precision	Recall	F1-Score	Support
0-No Class	0.99	0.99	0.99	33197713
1-Clean atmosphere	0.88	0.78	0.83	2840232
2-Non-typed particles/ low concentration	0.84	0.89	0.87	4274238
3-Aerosol: small	0.89	0.96	0.92	751168
4-Aerosol: large, spherical	0.95	0.96	0.96	1323590
5-Aerosol: mixture, partly non-spherical	0.94	0.97	0.95	1301967
6-Aerosol: large, non-spherical	0.93	0.95	0.94	499600
7-Cloud: non-typed	0.80	0.73	0.76	178227
8-Cloud: water droplets	0.54	0.49	0.52	7061
9-Cloud: likely water droplets	0.60	0.58	0.59	113802
10-Cloud: ice crystals	0.83	0.79	0.81	1486952
11-Cloud: likely ice crystals	0.59	0.63	0.61	105450
Averages				
micro	0.96	0.96	0.96	46080000
macro	0.82	0.81	0.81	46080000
weighted	0.96	0.96	0.96	46080000

Support = number of pixels pertaining to the corresponding truth class.

The confusion matrix (Figure 5) provides a granular view of the model's classification accuracy and error patterns. The diagonal elements represent the percentage of correctly classified pixels (recall) for each class. Key aerosol types are classified

220 with high recall rates: "Aerosol: small" (96% recall), "Aerosol: large, spherical" (96% recall), "Aerosol: mixture, partly non-spherical" (97% recall), and "Aerosol: large, non-spherical" (95% recall). This indicates that the model successfully learned to distinguish the nuanced lidar signatures corresponding to different aerosol properties, like size and shape. There is some minor confusion between aerosol types, such as 1% of "Aerosol: large, spherical" being misclassified as "Non-typed particles," and 2.9% of "Aerosol: large, non-spherical" being misclassified as "Aerosol: mixture." This is expected, as atmospheric aerosol

225 populations are often complex mixtures rather than discrete types.

The classification of cloud types reveals a more complex challenge. While the model correctly identifies "Cloud: ice crystals" (Class 10) in 79% of cases, its performance on liquid water clouds is notably lower. Only 49% of "Cloud: water droplets" (Class 8) are correctly identified, and 58% of "Cloud: likely water droplets" (Class 9) are correctly identified. An important observation is the confusion between similar and physically adjacent classes. For "Cloud: water droplets" (Class 8), while the recall is low

230 (49%), a significant portion of the misclassifications go to neighboring liquid cloud classes: 19% are mislabeled as "Cloud: likely water droplets" (Class 9) and 4% as "Cloud: non-typed" (Class 7). Similarly, for "Cloud: likely ice crystals" (Class 11), the main source of error is misclassification as "Cloud: ice crystals" (26%). This pattern of confusion is physically plausible.

Distinguishing between definite and "likely" water droplets, or between droplets and small ice crystals near the freezing level, can be ambiguous even for synergistic algorithms, let alone for a model relying only on lidar.

235   "No Class" (class 0) is identified with 99% F1 score, which is very high given that the model was trained with a corresponding weight of 0 in the loss function. This is due to the fact that the U-Net's final layer is a convolution layer with 12 classes and a softmax activation function. The softmax function forces the model to output a probability distribution across all 12 classes for every pixel, and these probabilities must sum to 1. Thus, even if class 0 has zero weight in the loss, the model must still assign some probability to the channel corresponding to class 0 for every pixel. As the model improves in identifying and segmenting

240 classes 1-11, it learns the features and contexts associated with them. For pixels that do not exhibit strong features of any of classes 1-11, the probabilities assigned by the softmax to these classes will naturally be low. Thus, the "leftover" probability mass is often assigned to the remaining classes, which in this case includes class 0. If class 0 is a general "background" or "none of the above specific particle types," this can lead to it being correctly predicted for those pixels. In addition, class 0 is much more prevalent than other classes (Table 2).

245   To further investigate the model's capabilities and limitations, Jaccard index was calculated for each height index across the test dataset, excluding the dominant "No Class" category (Equation 12).

$$J = \frac{TP}{TP + FP + FN} \tag{12}$$

Figure 6 shows the resulting relationship between classification performance and altitude. The general trend aligns with the physical expectations of lidar performance: the highest Jaccard values (0.4-0.6) are concentrated in the lowest height indices

250 (approximately 0–100). Performance gradually degrades through indices 100–300 and becomes poorest at high altitudes (indices > 300), where Jaccard values drop below 0.2. This vertical decline in performance is primarily attributable to the smaller number of classes present at higher altitudes (portrayed by the darker colors). Given that the model trained on fewer examples at higher altitudes, we would expect to see lower accuracy at those heights. Furthermore, the decline in performance may also be caused by the degradation of the lidar's signal-to-noise ratio with increasing altitude and the significant signal attenuation

255 caused by intervening clouds and dense aerosol layers. An interesting and seemingly counterintuitive trend is observed in the lowest part of the atmosphere (height indices 0–50). Within this range, where the pixel count is highest (indicated by the lighter color of the points), the mean Jaccard value shows a slight decrease with decreasing height. This is contrary to the assumption that performance should be uniformly best where the lidar signal is strongest and where there are more examples to train on.

4.2   Case studies to test model performance

260 To provide a qualitative assessment of the performance of the model, and to better understand its behavior under specific atmospheric conditions, we analyze a series of case studies from the test set. Each case compares the ground truth classification with the model's prediction and examines a corresponding confusion heatmap to identify specific areas and types of misclassification. The results of the case studies are summarized in Table 3.
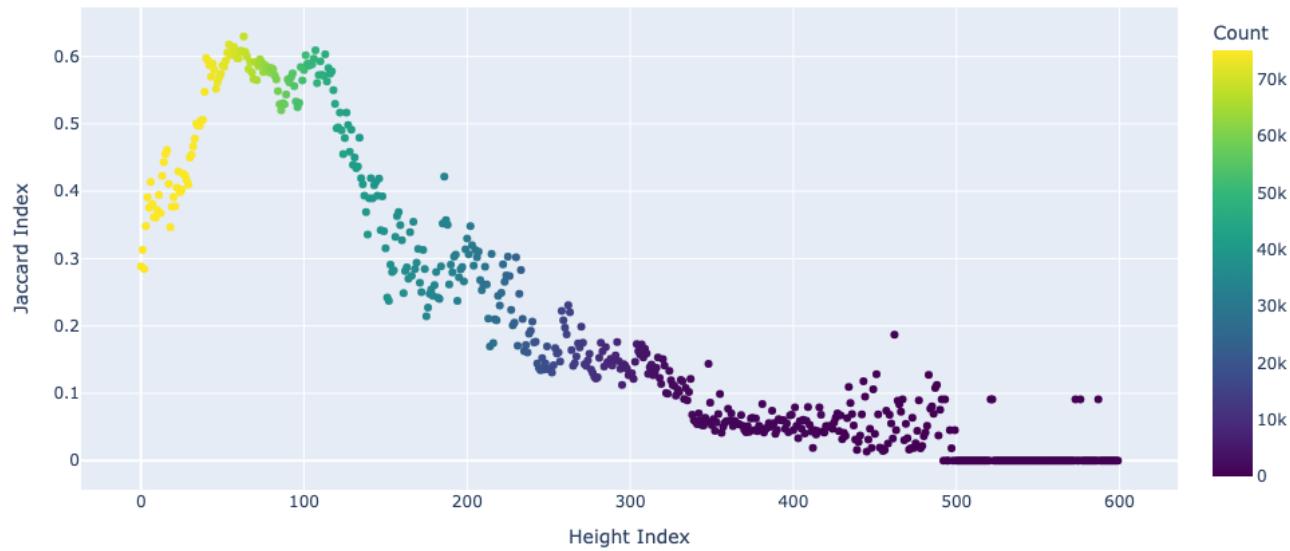
Figure 6. Vertical profiles of classification reliability. The Jaccard Index (Intersection over Union) plotted as a function of height index, with points colored by the total pixel count (Support) at that altitude. Peak performance occurs in the free troposphere (indices 0–100), with degradation at higher altitudes caused by reduced signal-to-noise ratios and signal attenuation from intervening cloud layers.

4.2.1 Case 1: Cloud-free multiple layering conditions

The first case study (7 December 2016) examines a complex, cloud-free scene with several distinct aerosol layers (Figure 7). This scenario serves as a validation of the model's ability to classify aerosols. Given that the target classification is done using only lidar data, we expect the model to perform well. The model successfully captures the vertical extent, boundaries, and temporal evolution of the different aerosol layers. The "Aerosol: mixture, partly non-spherical" (Class 5) layer is accurately reproduced by the model in both its location and classification. "Large, non-spherical" class is also correctly identified in the time steps between 200 to 450 and 800 to 960. Furthermore, the model also correctly identifies the overlying layer of "non-typed particles/low concentration" (Class 2) and the "clean atmosphere" (Class 1) above it. The per-image classification report confirms this strong performance with a weighted f1-score of 0.96. The most prevalent class, "Aerosol: mixture, partly non-spherical," achieves an F1-score of 0.97. Other aerosol classes also show high F1-scores, such as "Aerosol: large, spherical" (0.98) and "Aerosol: small" (0.95).

The confusion matrix (Figure A4) shows only minor, physically reasonable errors, such as overlap between "Clean atmosphere" and "Non-typed particles" and some confusion of "Aerosol: small" with "Aerosol: mixture." Notably, the group confusion heatmap (Figure A1) indicates no aerosol–cloud misclassifications in this case. These results indicate that the com-

Table 3. F1-Scores for Single Image Case Studies (1-3)

Class Description	c.s. 1	c.s. 2	c.s. 3
0-No Class	1.00	0.98	0.98
1-Clean atmosphere	0.85	0.83	0.84
2-Non-typed particles/ low concentration	0.74	0.80	0.87
3-Aerosol: small	0.95	0.90	0.92
4-Aerosol: large, spherical	0.98	0.95	0.97
5-Aerosol: mixture, partly non-spherical	0.97	0.89	0.95
6-Aerosol: large, non-spherical	0.82	0.00	0.78
7-Cloud: non-typed	n/a	0.81	0.42
8-Cloud: water droplets	n/a	0.51	0.68
9-Cloud: likely water droplets	n/a	0.63	0.45
10-Cloud: ice crystals	n/a	0.80	0.80
11-Cloud: likely ice crystals	n/a	0.50	0.66
Averages			
micro	0.96	0.95	0.93
macro	0.90	0.72	0.78
weighted	0.96	0.95	0.93

posite loss function improved discrimination, with the group confusion term reducing aerosol–cloud ambiguity and enabling clear separation between the two categories in this case. Furthermore, The generalized confusion heatmap (Figure A1) shows only minor ambiguity, mainly between closely related aerosol types (e.g., Class 5 within Class 4). This case study illustrates the model's ability to resolve aerosol subtypes in cloud-free conditions.

4.2.2 Case 2: Aerosol-cloud interaction study

The second case study investigates the interaction between a near-surface aerosol layer and an overlying liquid cloud (2016-11-13). This case directly tests the model's ability to delineate the boundary between aerosol and cloud and to correctly classify both in close proximity. As in the previous case, the model's prediction shows a very good structural agreement with the ground truth (Figure 8). It accurately identifies the general location and extent of the low-level aerosol layer (primarily "Aerosol: large, spherical") and the cloud system above it (a mix of "likely water droplets" and "ice crystals"). The temporal evolution of both features is also well replicated.

The Confusion Heatmap highlights that the most significant misclassifications are concentrated precisely at this boundary. This indicates that while the model can identify the core of the aerosol and cloud layers, it struggles to precisely delineate the transition zone between them. The Aerosol-Cloud Group Confusion Density Heatmap (Figure A2) reveals distinct, though localized, instances of confusion between aerosol and cloud groups. These areas are co-located with the aerosol-cloud interface
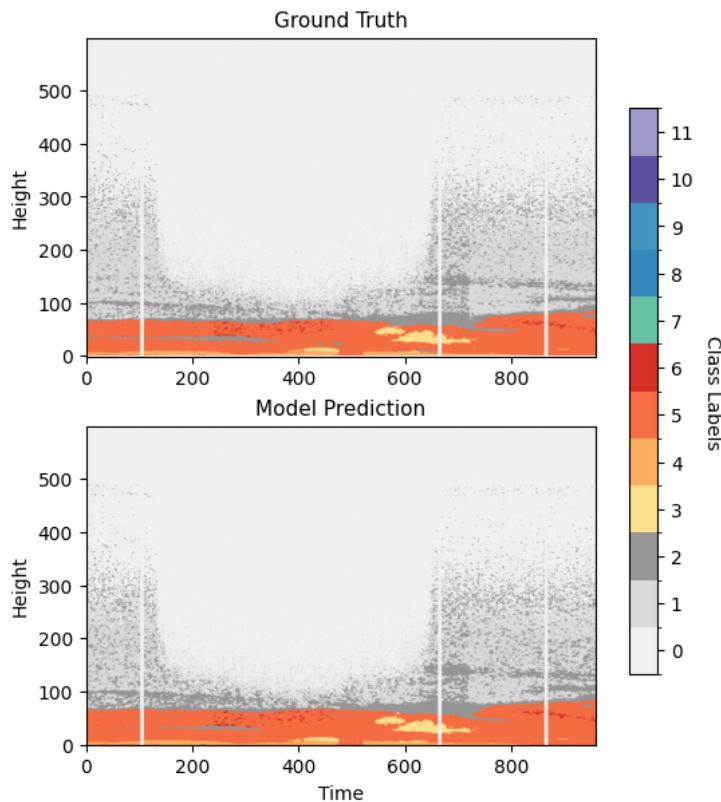
Figure 7. Case study 1: cloud free, distinct aerosol layers qualitative comparison. Side-by-side comparison of the ground truth labels (left) and the U-Net model's predictions (right) for 2016-12-07. Demonstrates the model's ability to resolve multiple distinct aerosol layers (mixture vs. non-spherical) in cloud-free conditions.

shown in the main plots. This indicates that the majority of the model's errors in this scene are aerosol-cloud misclassifications. The confusion matrix (Figure A5) quantifies this: for example, 7.5% of "Cloud: non-typed" (Class 7) is misclassified as "Aerosol: large, spherical" (Class 4). Most notably, the entire class of "Aerosol: large non-spherical" (Class 6) was misclassified and was primarily confused with Class 10 and Class 11 (a mix of "likely water droplets" and "ice crystals"). However, due to it's temporal-spacial singularity, where it only appears in one timestamp and at a specific height index, and due to its small support of only 54 pixels (table A2), it is safe to disregard these results as anomalous.

4.2.3 Case 3: Mid–High Clouds and Low-Level Dust Event

The last case study explores a complex multilayered cloud system that stretches from mid to high altitudes, with an air quality event (dust storm) at low altitudes, which took place on the date 2017-01-15. This case was chosen to test whether the model can predict clouds correctly at high altitudes at the very limit of lidar attenuation. The ground truth plot (Figure 9) shows a mid-level "Cloud: likely water droplets" (Class 9) layer at height index 200, where directly below and above it are classes
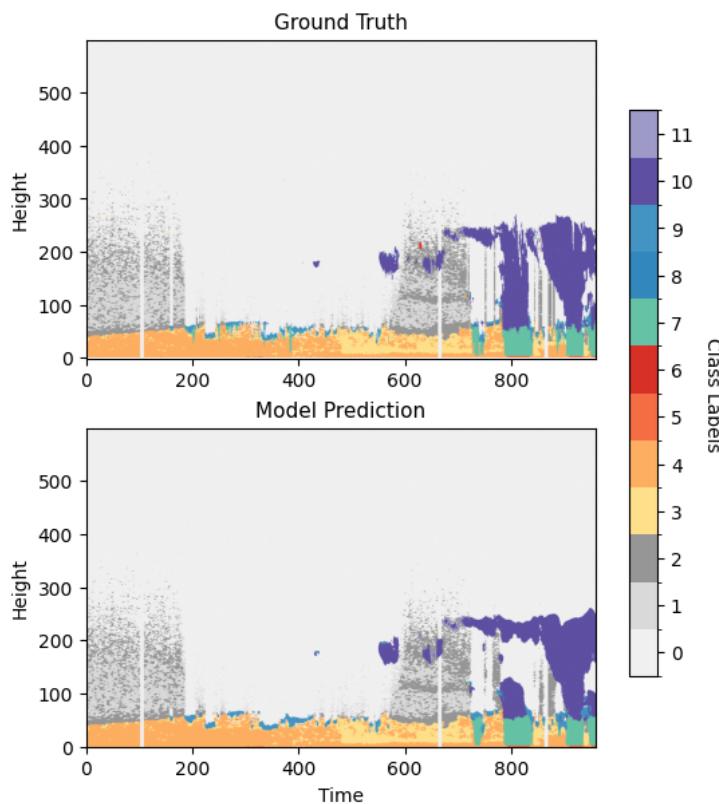
Figure 8. Case study 2: low level liquid cloud and aerosol layers qualitative comparison. Side-by-side comparison of the ground truth labels (left) and the U-Net model's predictions (right) for 2016-11-13. Illustrates the delineation of the boundary between a near-surface aerosol layer and an overlying liquid cloud.

10 and 11 (ice clouds). The model's prediction correctly identifies the low-level aerosol and liquid cloud layers with good
305 structural accuracy. Most importantly, it successfully infers the presence of the upper-level ice cloud in a position and with a structure that closely matches the ground truth. The per-image classification report shows a strong F1-score of 0.80 for "ice crystals" (Class 10), which make up the bulk of the upper cloud. This quantitatively confirms the model's successful inference. However, the confusion heatmap (Figure A3) reveals errors concentrated at the top and bottom boundaries of the inferred ice cloud. The per-image confusion matrix (Figure A6) shows that 23% of "likely ice crystals" are misclassified as "ice crystals,"
310 a plausible and minor error. Critically, the "Aerosol/Cloud Group Confusion Density Heatmap" remains predominantly dark, indicating that even in this complex multi-layer, multi-phase scene, the model rarely confuses the fundamental aerosol and cloud categories.
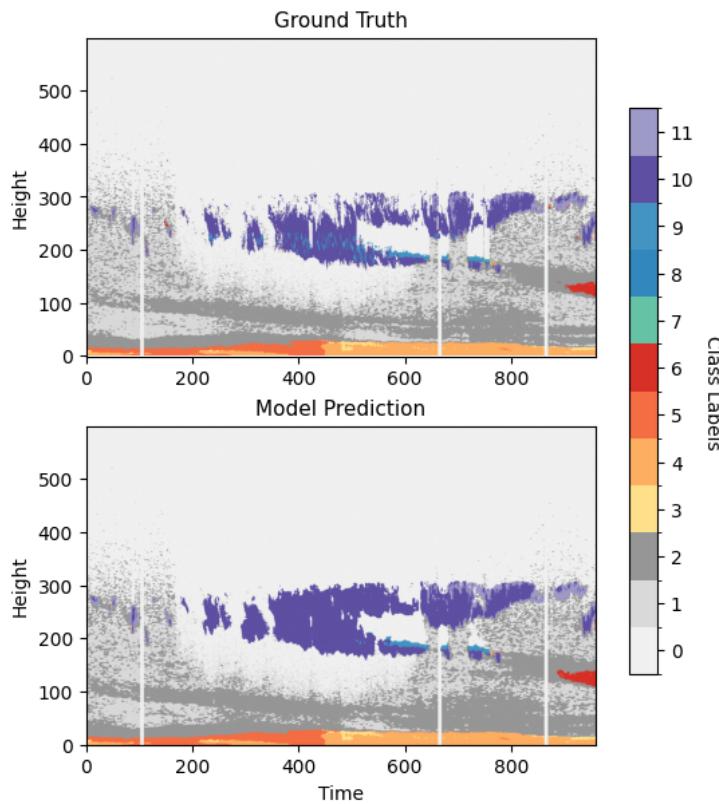
Figure 9. Case study 3: mid–high clouds and low-Level dust event qualitative comparison. Side-by-side comparison of the ground truth labels (left) and the U-Net model's predictions (right) for 2017-01-15. Highlights the model's capability of inferring upper-level ice clouds (Class 10) even at the limit of lidar signal attenuation.

5 Concluding Remarks

This study investigated a novel deep learning approach for unified aerosol and cloud classification using only ground-based lidar
315   data. The approach is motivated by the need for comprehensive atmospheric composition profiling and the inherent limitations
of lidar-only retrievals (primarily signal attenuation in clouds). A U-Net architecture was developed and trained. The model
takes standard lidar measurements (attenuated backscatter, depolarization ratio) as input and aims to predict a detailed vertical
classification encompassing both aerosol types (PollyXT outputs) and cloud/precipitation categories (Cloudnet outputs).

The deep learning model developed in this study demonstrated significant capability in classifying atmospheric constituents
320   from lidar data alone. The model achieves excellent overall performance (weighted F1 score of 0.96) and is particularly adept
at classifying diverse aerosol types. In cloud-free conditions, it successfully distinguishes between different aerosol categories
with high recall rates, validating its ability to learn the subtle signatures associated with particle size and shape based on
the input lidar features. The model shows proficiency in identifying ice clouds (F1-score of 0.81) but struggles more with

classifying specific liquid cloud categories, particularly the rare "water droplets" class. The observed confusion between similar

325 classes (e.g., "water droplets" and "likely water droplets") is physically reasonable and highlights the inherent ambiguity in defining discrete boundaries for continuous atmospheric processes. The use of a composite loss function, which features a group confusion penalty, proved effective in minimizing the most critical classification errors. Across all case studies, including complex multi-layer and multi-phase scenes, the model consistently and reliably discriminated between the broader categories of aerosol and clouds, with confusion being rare and localized to the most ambiguous interface regions. Model performance

330 varies predictably with altitude, achieving optimal results in the free troposphere, where signals are strong and targets are well-defined. Performance degrades at high altitudes due to decreasing signal-to-noise ratio and at very low altitudes in part due to instrumental effects and the high complexity of the planetary boundary layer.

The findings of our study have important implications for atmospheric remote sensing, suggesting that ground-based observational systems could be simplified. By applying advanced algorithms to relatively simple and cost-effective lidar data, it

335 may be possible to reduce the reliance on co-located, complex, and expensive instruments such as cloud radars and microwave radiometers, thereby facilitating the establishment of denser observational networks(Illingworth et al., 2007). Such networks could provide valuable data streams for improving weather forecasts, evaluating climate models (particularly concerning cloud feedback and ACI), and supporting air quality and environment monitoring as well as aviation safety.

Despite the potential, several limitations and challenges must be acknowledged. The performance of the deep learning

340 model is fundamentally dependent on the quality, accuracy, and representativeness of the complex training dataset. Any biases or errors inherent in the reference PollyXT and Cloudnet algorithms used to generate the target variable will likely be learned and propagated by the U-Net model. Further more, a model trained on data from one specific site or lidar instrument, as is the case in this study, may not perform equally well and may not generalize in different atmospheric regimes or with data from different lidar systems without either retraining or the application of domain adaptation techniques.

345 We are planning future work to focus on further exploring different deep learning architectures, including variants of the U-Net or transformer-based models. Exploring physics-informed neural networks, which incorporate physical constraints into the learning process, might also improve the physical consistency of the predictions. Validation efforts using independent datasets could further build confidence and establish generalization capabilities. This includes data from different geographical locations, seasons, and lidar instruments. Comparison with data from field campaigns involving airborne in-situ measurements

350 or overpasses of satellites with cloud-penetrating capabilities would provide valuable independent validation as well.

Code and data availability. The combined PollyXT and Cloudnet dataset can be accessed at https://doi.org/10.5281/zenodo.17424878. Python code files are available at https://doi.org/10.5281/zenodo.17422969.

Appendix A: Case Studies

Table A1. Classification Performance Metrics for case study 1 - Complex Multi-Layer Aerosol

Class Description	Precision	Recall	F1-Score	Support
0-No Class	1.00	1.00	1.00	392451
1-Clean atmosphere	0.91	0.81	0.85	79939
2-Non-typed particles/ low concentration	0.69	0.80	0.74	43936
3-Aerosol: small	0.94	0.97	0.95	2965
4-Aerosol: large, spherical	0.97	0.99	0.98	4146
5-Aerosol: mixture, partly non-spherical	0.95	0.99	0.97	51200
6-Aerosol: large, non-spherical	0.87	0.78	0.82	1363
7-Cloud: non-typed	na	na	na	na
8-Cloud: water droplets	na	na	na	na
9-Cloud: likely water droplets	na	na	na	na
10-Cloud: ice crystals	na	na	na	na
11-Cloud: likely ice crystals	na	na	na	na
Averages				
micro	0.96	0.96	0.96	576000
macro	0.90	0.90	0.90	576000
weighted	0.96	0.96	0.96	576000

Table A2. Classification Performance Metrics for case study 2 - Low Level Liquid Cloud and Aerosol

Class Description	Precision	Recall	F1-Score	Support
0-No Class	0.98	0.99	0.98	443893
1-Clean atmosphere	0.84	0.82	0.83	24579
2-Non-typed particles/ low concentration	0.77	0.83	0.80	28258
3-Aerosol: small	0.88	0.92	0.90	11605
4-Aerosol: large, spherical	0.94	0.96	0.95	26893
5-Aerosol: mixture, partly non-spherical	0.95	0.84	0.89	313
6-Aerosol: large, non-spherical	0.00	0.00	0.00	54
7-Cloud: non-typed	0.88	0.75	0.81	7530
8-Cloud: water droplets	0.48	0.55	0.51	173
9-Cloud: likely water droplets	0.56	0.71	0.63	3420
10-Cloud: ice crystals	0.86	0.75	0.80	28622
11-Cloud: likely ice crystals	0.50	0.50	0.50	660
Averages				
micro	0.95	0.95	0.95	576000
macro	0.72	0.72	0.72	576000
weighted	0.95	0.95	0.95	576000

Table A3. Classification Performance Metrics for case study 3 - Lidar Signal Attenuation

Class Description	Precision	Recall	F1-Score	Support
0-No Class	1.00	0.97	0.98	360174
1-Clean atmosphere	0.92	0.77	0.84	69998
2-Non-typed particles/ low concentration	0.82	0.92	0.84	85916
3-Aerosol: small	0.90	0.93	0.92	1266
4-Aerosol: large, spherical	0.98	0.97	0.97	12059
5-Aerosol: mixture, partly non-spherical	0.92	0.99	0.95	6543
6-Aerosol: large, non-spherical	0.69	0.90	0.78	1075
7-Cloud: non-typed	0.57	0.33	0.42	85
8-Cloud: water droplets	0.70	0.66	0.68	139
9-Cloud: likely water droplets	0.78	0.32	0.45	2889
10-Cloud: ice crystals	0.69	0.94	0.80	30870
11-Cloud: likely ice crystals	0.58	0.76	0.66	4986
Averages				
micro	0.93	0.93	0.93	576000
macro	0.80	0.79	0.78	576000
weighted	0.93	0.93	0.93	576000
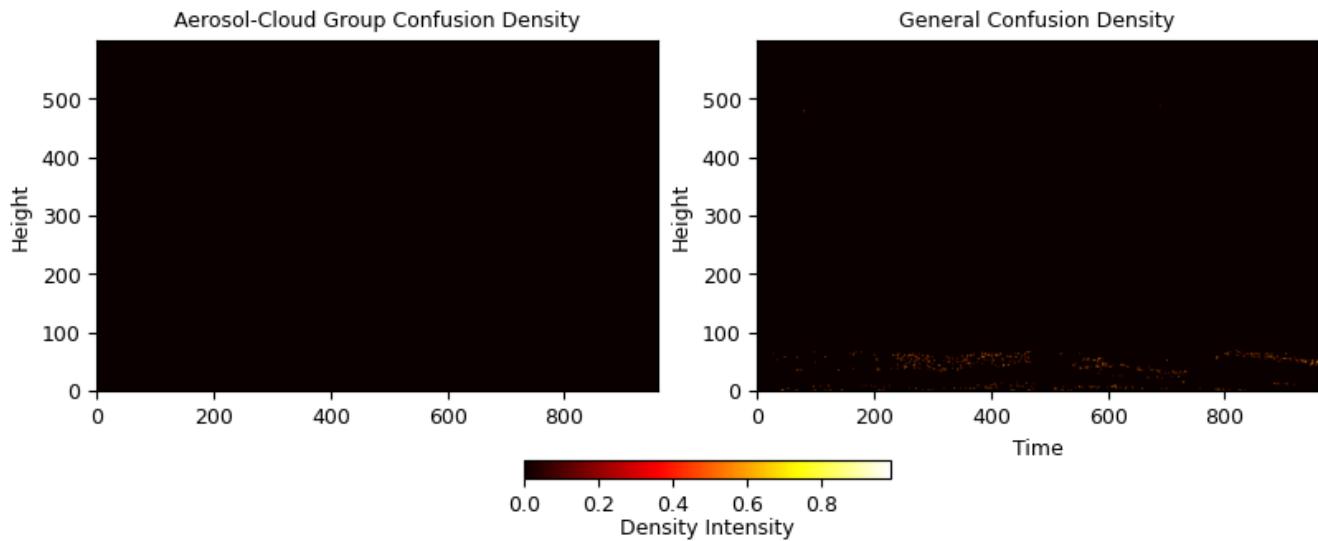
Figure A1. Case study 1 - spatial distribution of classification errors. (Left) Group confusion density heatmap highlighting specific regions where the model confuses the broad categories of aerosols and clouds. (Right) Generalized confusion heatmap excluding background and clean atmosphere classes (0–2) to visualize the nuances of subtype misclassification.
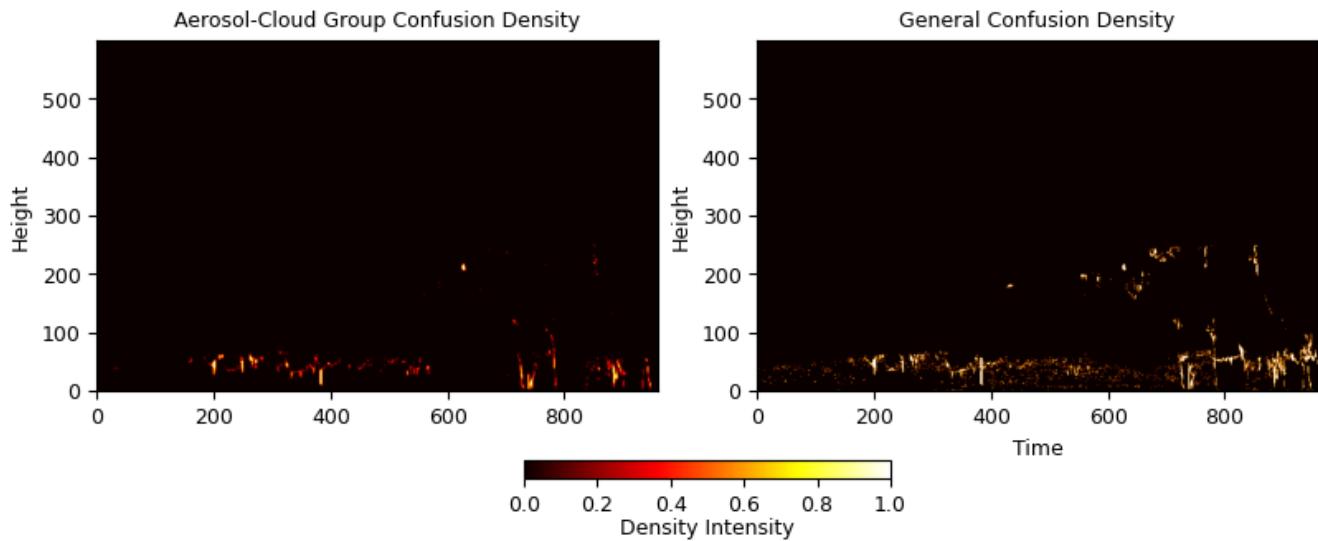
Figure A2. Case study 2 - spatial distribution of classification errors. (Left) Group confusion density heatmap highlighting specific regions where the model confuses the broad categories of aerosols and clouds. (Right) Generalized confusion heatmap excluding background and clean atmosphere classes (0–2) to visualize the nuances of subtype misclassification.
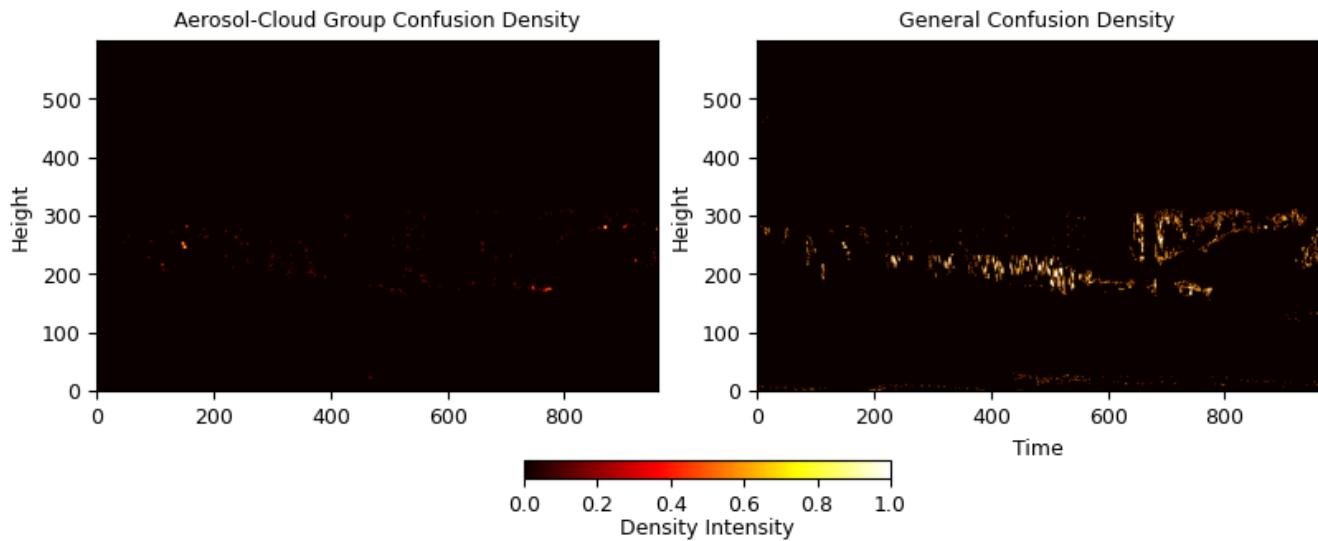
Figure A3. Case study 3 - spatial distribution of classification errors. (Left) Group confusion density heatmap highlighting specific regions where the model confuses the broad categories of aerosols and clouds. (Right) Generalized confusion heatmap excluding background and clean atmosphere classes (0–2) to visualize the nuances of subtype misclassification.
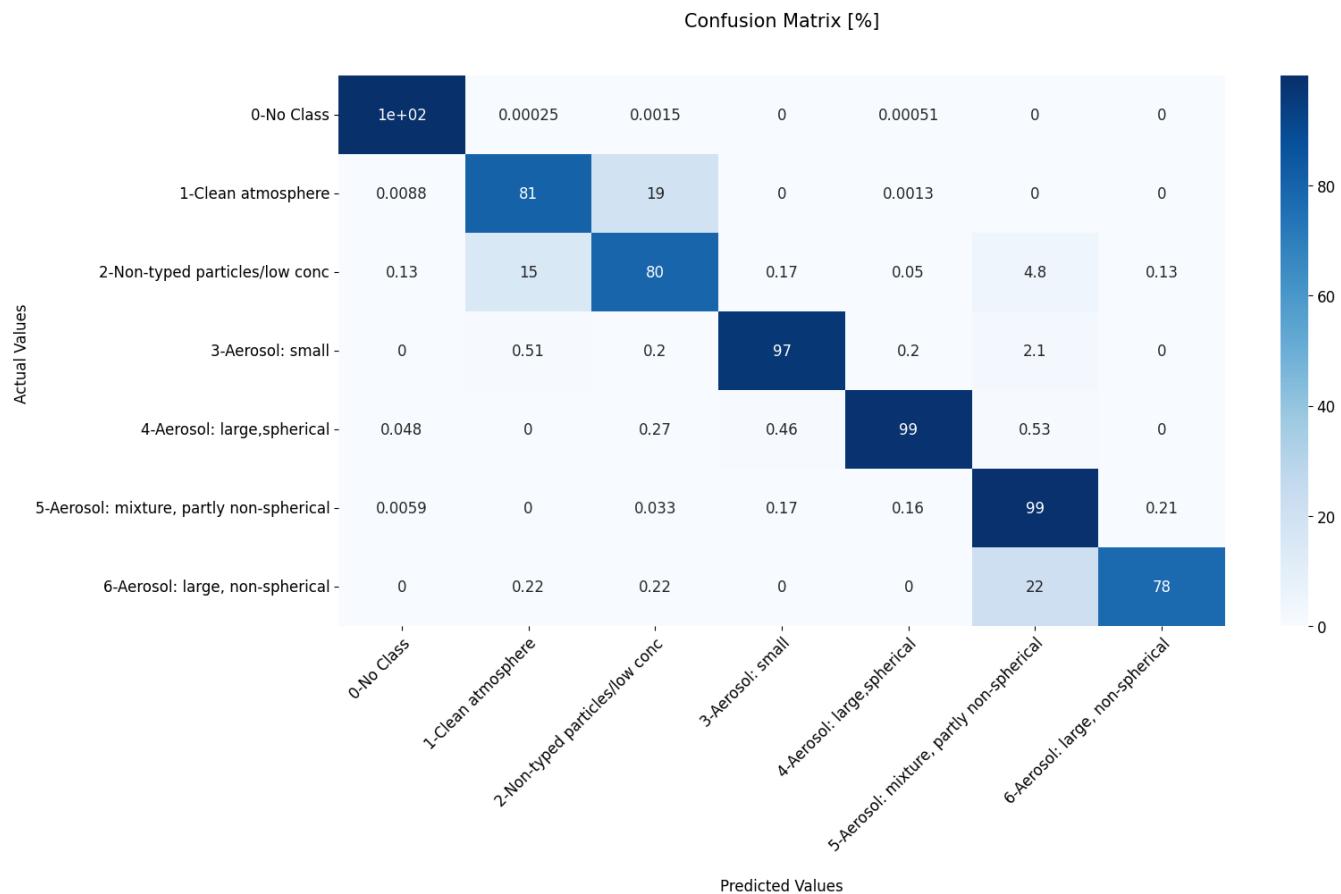
Figure A4. Confusion matrix for case study 1 - Complex Multi-Layer Aerosol
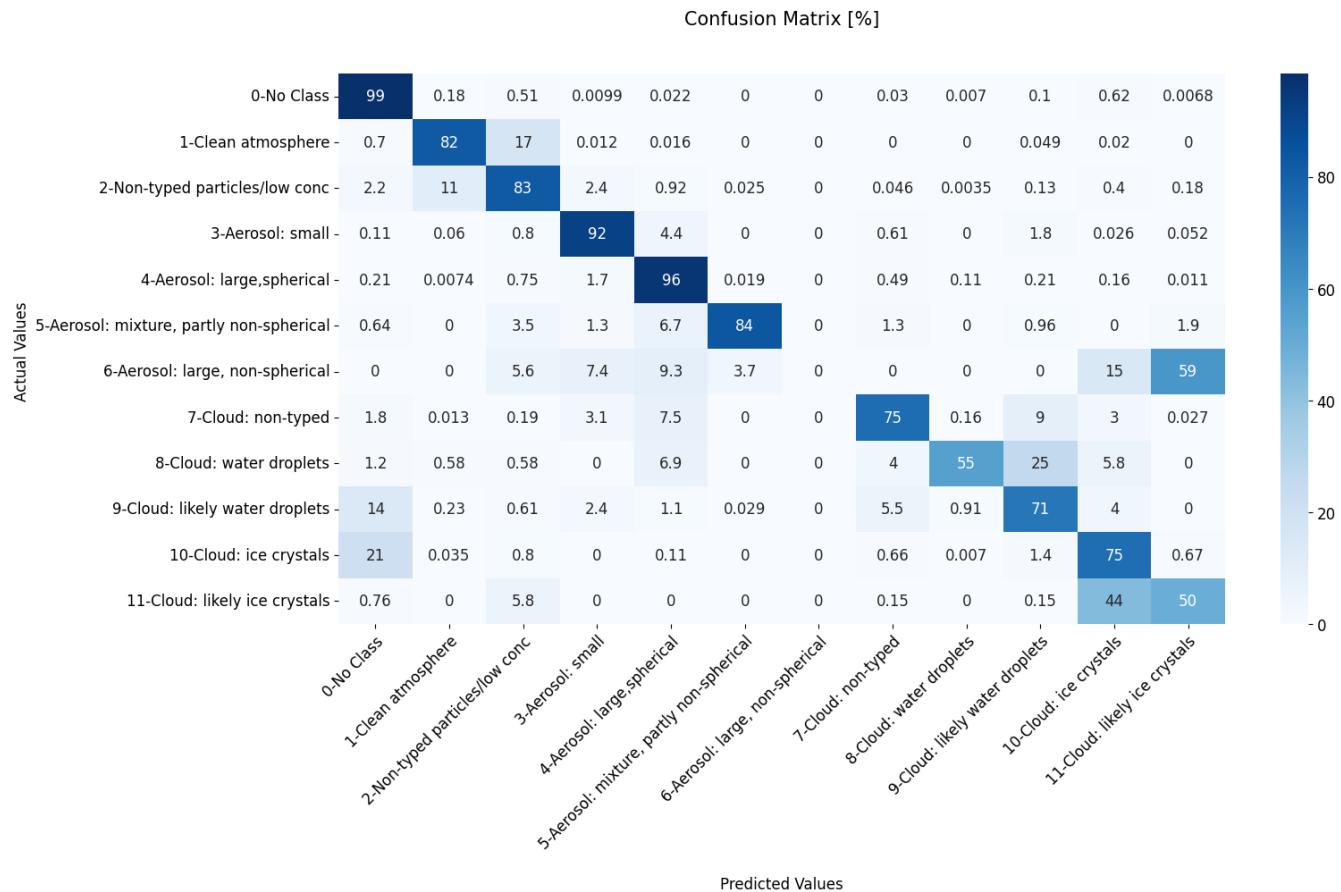
Confusion Matrix [%]

Figure A5. Confusion matrix for case study 2 - Low Level Liquid Cloud and Aerosol
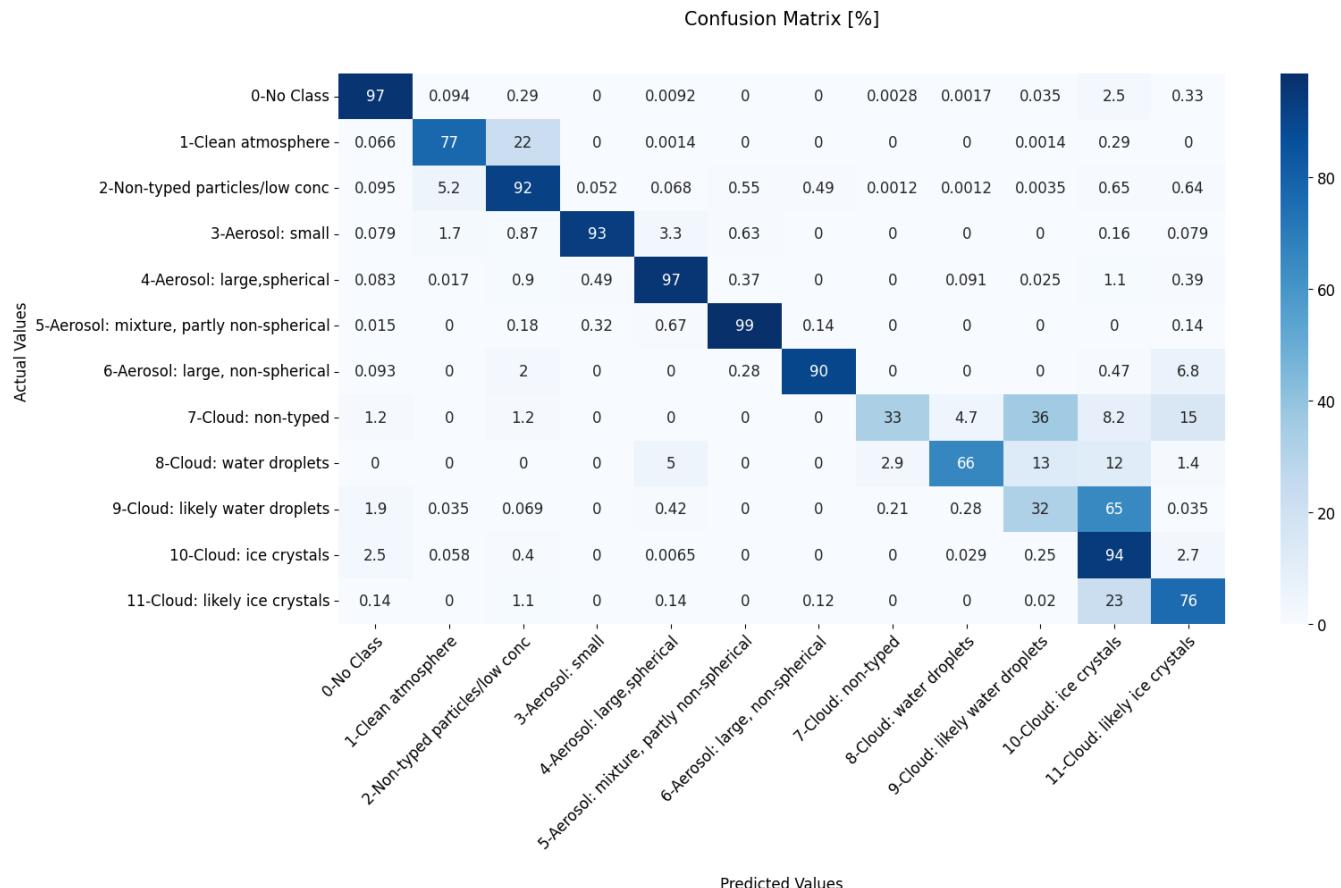
Confusion Matrix [%]

**Figure A6.** Confusion matrix for case study 3 - Lidar Signal Attenuation
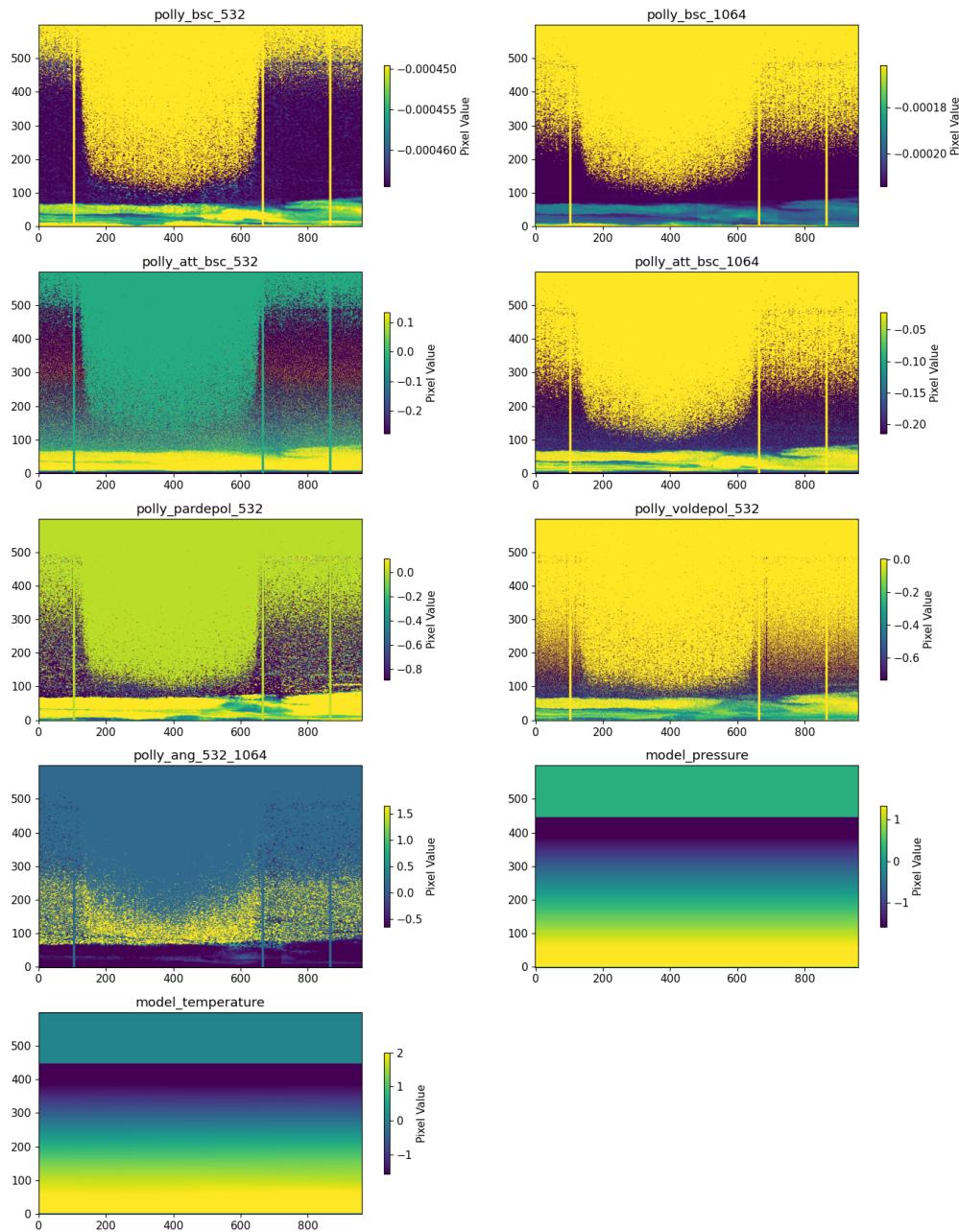
Figure A7. Raw signals case study 1 - Input lidar and meteorological feature suite. Visualization of the multi-channel input data used by the U-Net, including attenuated and aerosol backscatter (532/1064 nm), depolarization ratios, the backscatter-related Ångström exponent, and NWP-derived pressure/temperature profiles. These raw signals provide the hierarchical features necessary for the model to infer atmospheric state even in signal-attenuated regions.
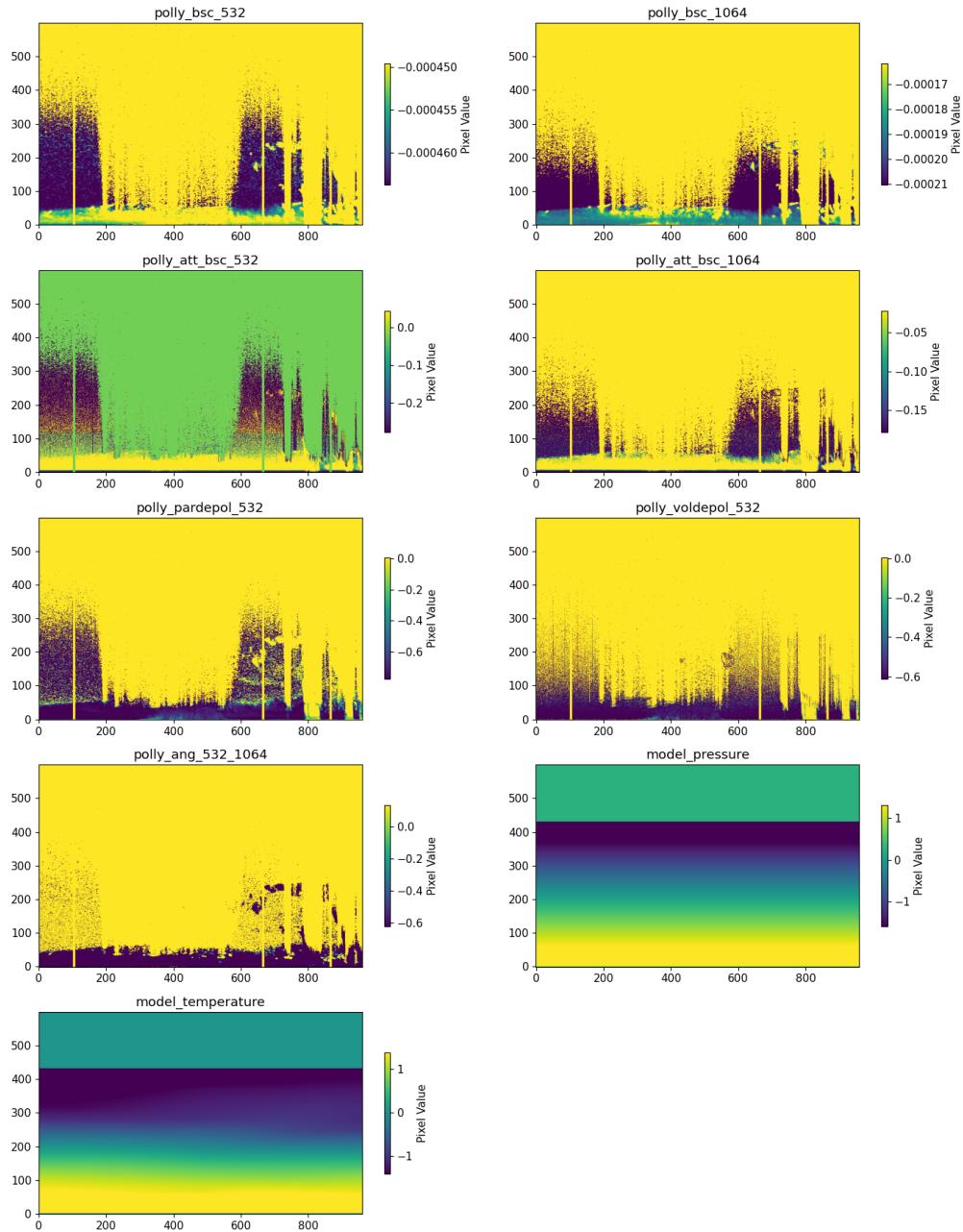
Figure A8. Raw signals case study 2 - Input lidar and meteorological feature suite. Visualization of the multi-channel input data used by the U-Net, including attenuated and aerosol backscatter (532/1064 nm), depolarization ratios, the backscatter-related Ångström exponent, and NWP-derived pressure/temperature profiles. These raw signals provide the hierarchical features necessary for the model to infer atmospheric state even in signal-attenuated regions.
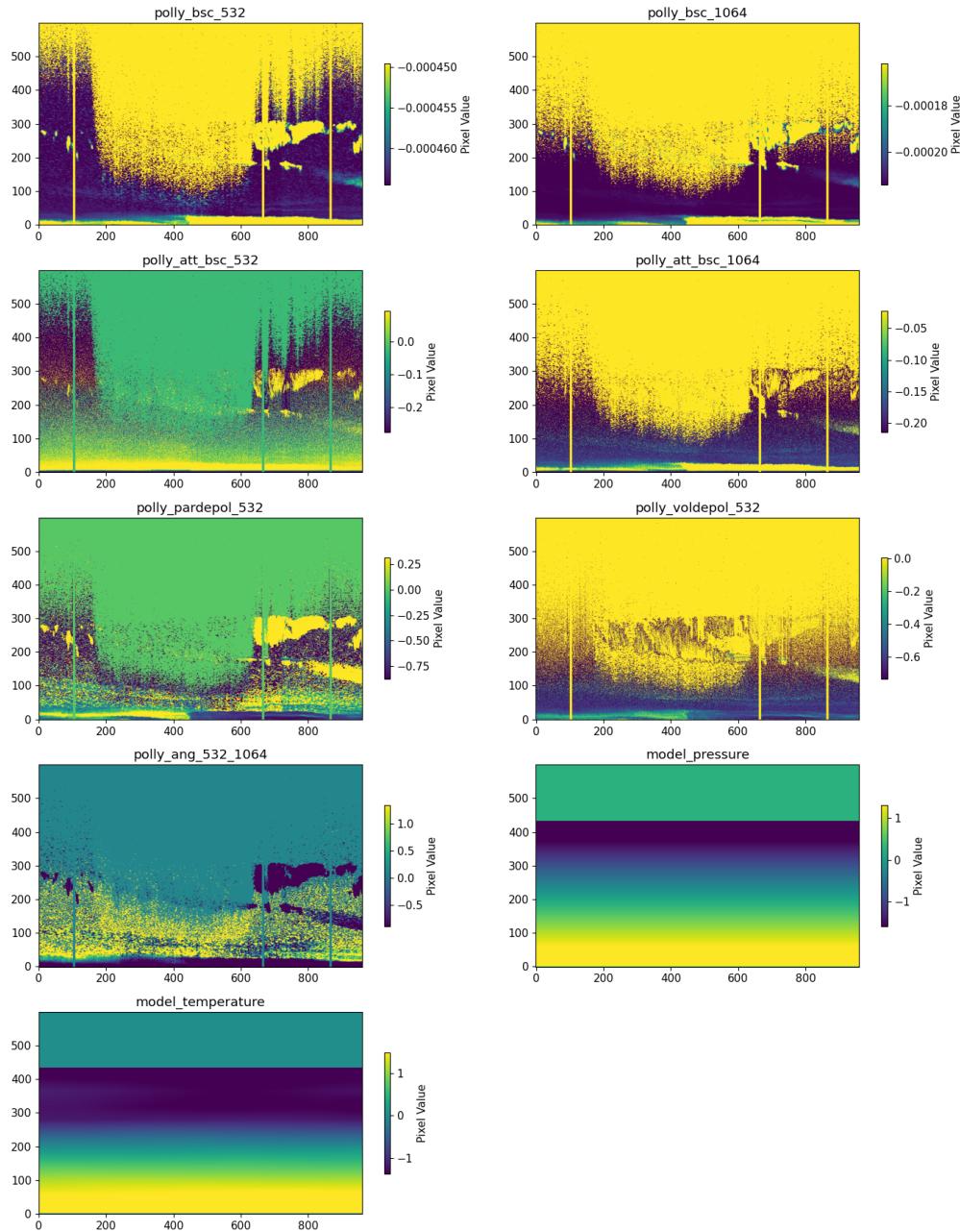
Figure A9. Raw signals case study 3 - Input lidar and meteorological feature suite. Visualization of the multi-channel input data used by the U-Net, including attenuated and aerosol backscatter (532/1064 nm), depolarization ratios, the backscatter-related Ångström exponent, and NWP-derived pressure/temperature profiles. These raw signals provide the hierarchical features necessary for the model to infer atmospheric state even in signal-attenuated regions.

References

Albrecht, B. A.: Aerosols, cloud microphysics, and fractional cloudiness, Science, 245, https://doi.org/10.1126/science.245.4923.1227, 1989.

Baars, H., Seifert, P., Engelmann, R., and Wandinger, U.: Target categorization of aerosol and clouds by continuous multiwavelength-

365   polarization lidar measurements, Atmospheric Measurement Techniques, 10, https://doi.org/10.5194/amt-10-3175-2017, 2017.

Baars, H., Althausen, D., Engelmann, R., Heese, B., Ansmann, A., Wandinger, U., Hofer, J., Skupin, A., Komppula, M., Giannakaki, E., Filioglou, M., Bortoli, D., Silva, A. M., Pereira, S., Stachlewska, I. S., Kumala, W., Szczepanik, D., Amiridis, V., Marinou, E., Kottas, M., Mattis, I., and Müller, G.: PollyNET - An emerging network of automated raman-polarizarion lidars for continuous aerosolprofiling, in: EPJ Web of Conferences, vol. 176, ISSN 2100014X, https://doi.org/10.1051/epjconf/201817609013, 2018.

370   Bansal, A., Lee, Y., Hilburn, K., and Ebert-Uphoff, I.: Tools for Extracting Spatio-Temporal Patterns in Meteorological Image Sequences: From Feature Engineering to Attention-Based Neural Networks, https://doi.org/10.48550/arXiv.2210.12310, 2022.

Biasutti, P., Lepetit, V., Aujol, J. F., Bredif, M., and Bugeau, A.: LU-net: An efficient network for 3D LiDAR point cloud semantic segmentation based on end-to-end-learned 3D features and U-net, in: Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019, https://doi.org/10.1109/ICCVW.2019.00123, 2019.

375   Box, G. E. P. and Cox, D. R.: An Analysis of Transformations, Journal of the Royal Statistical Society Series B: Statistical Methodology, 26, https://doi.org/10.1111/j.2517-6161.1964.tb00553.x, 1964.

Bressan, P. O., Junior, J. M., Correa Martins, J. A., de Melo, M. J., Gonçalves, D. N., Freitas, D. M., Marques Ramos, A. P., Garcia Furuya, M. T., Osco, L. P., de Andrade Silva, J., Luo, Z., Garcia, R. C., Ma, L., Li, J., and Gonçalves, W. N.: Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping, International Journal of Applied Earth Observation and Geoinformation,

380   108, https://doi.org/10.1016/j.jag.2022.102690, 2022.

Bühl, J., Alexander, S., Crewell, S., Heymsfield, A., Kalesse, H., Khain, A., Maahn, M., Van-Tricht, K., and Wendisch, M.: Chapter 10: Remote Sensing, Meteorological Monographs, 58, 2017.

Cairo, F., Di Liberto, L., Dionisi, D., and Snels, M.: Understanding Aerosol–Cloud Interactions through Lidar Techniques: A Review, Remote Sensing, 16, https://doi.org/10.3390/rs16152788, 2024.

385   del Águila, A., Ortiz-Amezcua, P., Tabik, S., Bravo-Aranda, J. A., Fernández-Carvelo, S., and Alados-Arboledas, L.: Aerosol type classification with machine learning techniques applied to multiwavelength lidar data from EARLINET, EGUsphere, 2025, 1–28, https://doi.org/10.5194/egusphere-2025-269, 2025.

Foley, S. R., Knobelspiesse, K. D., Sayer, A. M., Gao, M., Hays, J., and Hoffman, J.: 3D cloud masking across a broad swath using multi-angle polarimetry and deep learning, Atmospheric Measurement Techniques, 17, 7027–7047, https://doi.org/10.5194/amt-17-7027-2024,

390   2024.

Fuller, C. A., Selmer, P. A., Gomes, J., and McGill, M. J.: Using Multitask Machine Learning to Type Clouds and Aerosols from Space-Based Photon-Counting Lidar Measurements, Remote Sensing, 17, https://doi.org/10.3390/rs17162787, 2025.

Galea, D., Ma, H.-Y., Wu, W.-Y., and Kobayashi, D.: Deep Learning Image Segmentation for Atmospheric Rivers, Artificial Intelligence for the Earth Systems, 3, https://doi.org/10.1175/aies-d-23-0048.1, 2023.

395   Haarig, M., Hünerbein, A., Wandinger, U., Docter, N., Bley, S., Donovan, D., and Van Zadelhoff, G. J.: Cloud top heights and aerosol columnar properties from combined EarthCARE lidar and imager observations: The AM-CTH and AM-ACD products, Atmospheric Measurement Techniques, 16, https://doi.org/10.5194/amt-16-5953-2023, 2023.

Hartmann, D. L. and Doelling, D.: On the net radiative effectiveness of clouds, Journal of Geophysical Research, 96, https://doi.org/10.1029/90JD02065, 1991.

400 Illingworth, A. J., Hogan, R. J., O'Connor, E. J., Bouniol, D., Brooks, M. E., Delanoë, J., Donovan, D. P., Eastment, J. D., Gaussiat, N., Goddard, J. W., Haeffelin, M., Klein Baltinik, H., Krasnov, O. A., Pelon, J., Piriou, J. M., Protat, A., Russchenberg, H. W., Seifert, A., Tompkins, A. M., van Zadelhoff, G. J., Vinit, F., Willen, U., Wilson, D. R., and Wrench, C. L.: Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations, Bulletin of the American Meteorological Society, 88, https://doi.org/10.1175/BAMS-88-6-883, 2007.

405 Intergovernmental Panel on Climate Change (IPCC): Clouds and Aerosols, in: Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by (IPCC), I. P. o. C. C., pp. 571–658, Cambridge University Press, Cambridge, ISBN 9781107057999, https://doi.org/DOI: 10.1017/CBO9781107415324.016, 2014.

Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: 32nd International
410 Conference on Machine Learning, ICML 2015, vol. 1, 2015.

Jacob, D. J.: Heterogeneous chemistry and tropospheric ozone, Atmospheric Environment, 34, https://doi.org/10.1016/S1352-2310(99)00462-8, 2000.

Jones, M. P.: Indicator and stratification methods for missing explanatory variables in multiple linear regression, Journal of the American Statistical Association, 91, https://doi.org/10.1080/01621459.1996.10476680, 1996.

415 Kalesse-Los, H., Schimmel, W., Luke, E., and Seifert, P.: Evaluating cloud liquid detection against Cloudnet using cloud radar Doppler spectra in a pre-trained artificial neural network, Atmospheric Measurement Techniques, 15, https://doi.org/10.5194/amt-15-279-2022, 2022.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, Communications of the ACM, 60, https://doi.org/10.1145/3065386, 2017.

420 LeCun, Y., Hinton, G., and Bengio, Y.: Deep learning (2015), Y. LeCun, Y. Bengio and G. Hinton, Nature, 521, 2015.

Lelieveld, J. and Crutzen, P. J.: The role of clouds in tropospheric photochemistry, Journal of Atmospheric Chemistry, 12, https://doi.org/10.1007/BF00048075, 1991.

Levy-Jurgenson, A., Tekpli, X., Kristensen, V. N., and Yakhini, Z.: Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer, Scientific Reports, 10, https://doi.org/10.1038/s41598-020-75708-z, 2020.

425 Milletari, F., Navab, N., and Ahmadi, S. A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016, https://doi.org/10.1109/3DV.2016.79, 2016.

Nicolae, D., Vasilescu, J., Talianu, C., Binietoglou, I., Nicolae, V., Andrei, S., and Antonescu, B.: A neural network aerosol-typing algorithm based on lidar data, Atmospheric Chemistry and Physics, 18, https://doi.org/10.5194/acp-18-14511-2018, 2018.

Oladipo, B., Gomes, J., McGill, M., and Selmer, P.: Leveraging Deep Learning as a New Approach to Layer Detection and Cloud–Aerosol
430 Classification Using ICESat-2 Atmospheric Data, Remote Sensing, 16, 2344, https://doi.org/10.3390/rs16132344, 2024.

Pal, S. R., Steinbrecht, W., and Carswell, A. I.: Automated method for lidar determination of cloud-base height and vertical extent, Applied Optics, 31, https://doi.org/10.1364/ao.31.001488, 1992.

Ramanathan, V., Crutzen, P. J., Kiehl, J. T., and Rosenfeld, D.: Atmosphere: Aerosols, climate, and the hydrological cycle, https://doi.org/10.1126/science.1064034, 2001.

435 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Rogozovsky, I., Ohneiser, K., Lyapustin, A., Ansmann, A., and Chudnovsky, A.: The impact of different aerosol layering conditions on the high-resolution MODIS/MAIAC AOD retrieval bias: The uncertainty analysis, Atmospheric Environment, 309, https://doi.org/10.1016/j.atmosenv.2023.119930, 2023.

440 Rogozovsky, I., Ansmann, A., Hofer, J., and Chudnovsky, A.: Unveiling Atmospheric Layers: Vertical Pollution Patterns and Prospects for High-Resolution Aerosol Retrievals Using the Eastern Mediterranean as a Case Study, Environmental Science & Technology, 59, 12 181–12 195, https://doi.org/10.1021/acs.est.4c14556, 2025.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351, ISSN 16113349,
445 https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Rosenfeld, D., Andreae, M. O., Asmi, A., Chin, M., De Leeuw, G., Donovan, D. P., Kahn, R., Kinne, S., Kivekäs, N., Kulmala, M., Lau, W., Schmidt, K. S., Suni, T., Wagner, T., Wild, M., and Quaas, J.: Global observations of aerosol-cloud-precipitation-climate interactions, https://doi.org/10.1002/2013RG000441, 2014.

Rusyn, B., Korniy, V., Lutsyk, O., and Kosarevych, R.: Deep Learning for Atmospheric Cloud Image Segmentation, in: 2019
450 11th International Scientific and Practical Conference on Electronics and Information Technologies, ELIT 2019 - Proceedings, https://doi.org/10.1109/ELIT.2019.8892285, 2019.

Schimmel, W., Kalesse-Los, H., Maahn, M., Vogl, T., Foth, A., Garfias, P. S., and Seifert, P.: Identifying cloud droplets beyond lidar attenuation from vertically pointing cloud radar observations using artificial neural networks, Atmospheric Measurement Techniques, 15, https://doi.org/10.5194/amt-15-5343-2022, 2022.

455 Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10553 LNCS, ISSN 16113349, https://doi.org/10.1007/978-3-319-67558-9_28, 2017.

Weitkamp, C.: Lidar : range-resolved optical remote sensing of the atmosphere, 102, 2005.

Winker, D., Chepfer, H., Noel, V., and Cai, X.: Observational Constraints on Cloud Feedbacks: The Role of Active Satellite Sensors, Surveys
460 in Geophysics, 38, https://doi.org/10.1007/s10712-017-9452-0, 2017.

Winker, D. M., Pelon, J., Coakley, J. A., Ackerman, S. A., Charlson, R. J., Colarco, P. R., Flamant, P., Fu, Q., Hoff, R. M., Kittaka, C., Kubar, T. L., Le Treut, H., McCormick, M. P., Mégie, G., Poole, L., Powell, K., Trepte, K., Vaughan, M. A., and Wielicki, B. A.: The Calipso Mission: A Global 3D View of Aerosols and Clouds, Bulletin of the American Meteorological Society, 91, https://doi.org/10.1175/2010BAMS3009.1, 2010.

465 Zhou, X., Chen, B., Ye, Q., Zhao, L., Song, Z., Wang, Y., Hu, J., and Chen, R.: Cloud–Aerosol Classification Based on the U-Net Model and Automatic Denoising CALIOP Data, Remote Sensing, 16, https://doi.org/10.3390/rs16050904, 2024.