Evaluating cloud liquid detection against Cloudnet using cloud radar Doppler spectra in a pre-trained artificial neural network~~against Cloudnet liquid detection~~

Heike Kalesse-Los[1,2], Willi Schimmel[1], Edward Luke[3], and Patric Seifert[2]

[1]Institute for Meteorology, Universität Leipzig, Leipzig, Germany
[2]Leibniz Institute for Tropospheric Research, Leipzig, Germany
[3]Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, New York, USA

Correspondence: Heike Kalesse-Los (heike.kalesse@uni-leipzig.de)

Abstract. Detection of liquid-containing cloud layers in thick mixed-phase clouds or multi-layer cloud situations from ground-based remote sensing instruments still pose observational challenges yet improvements are crucial since the existence of multi-layer liquid layers in mixed-phase cloud situations influences cloud radiative effects, cloud life time, and precipitation formation processes. Hydrometeor target classifications such as from Cloudnet that require a lidar signal for the classification of liquid are limited to the maximum height of lidar signal penetration and thus often lead to underestimations of liquid-containing cloud layers. Here we evaluate the Cloudnet liquid detection against the approach of Luke et al. (2010) which extracts morphological features in cloud-penetrating cloud radar Doppler spectra measurements in a artificial neural network (ANN) approach to classify liquid beyond full lidar signal attenuation based on the simulation of the two lidar parameters particle backscatter coefficient and particle depolarization ratio. We show that the ANN of Luke et al. (2010) which was trained in Arctic conditions can successfully be applied to observations in the mid-latitudes obtained during the seven-week long ACCEPT field experiment in Cabauw, the Netherlands, 2014. In a sensitivity study covering the whole duration of the ACCEPT campaign, different liquid-detection thresholds for ANN-predicted lidar variables are applied and evaluated against the Cloudnet target classification. Independent validation of the liquid mask from the standard Cloudnet target classification against the ANN-based technique is realized by comparisons to observations of microwave radiometer liquid water path, ceilometer liquid-layer base altitude, and radiosonde relative humidity. ~~Four~~ In addition, a case-study comparison against the cloud feature mask detected by the spaceborne lidar aboard the CALIPSO satellite is presented. Three conclusions were drawn from the investigation: First, it was found that the threshold selection criteria of liquid-related lidar backscatter and depolarization alone control the liquid detection considerably. Second, ~~nevertheless,~~ all threshold values used in the ANN-framework were found to outperform the Cloudnet target classification for deep or multi-layer cloud situations where the lidar signal is fully attenuated within low liquid layers and the cloud ~~reflectivity~~ radar is able to detect the microphysical fingerprint of liquid in higher cloud layers~~was sufficiently high to be detectable by the cloud radar~~. Third, ~~in convective situations for which~~ if lidar data is available~~and for which the~~ , Cloudnet is at least as good as the ANN. The times when Cloudnet outperforms the ANN in liquid detections are often associated with situations where cloud dynamics smear the imprint of cloud microphysics on the radar Doppler ~~spectrum~~

~~is decreased, Cloudnet outperforms the ANN retrieval. Fourth, in high-level clouds both approaches (Cloudnet and the ANN technique), are limited.~~ spectra.

1 Introduction

In mixed-phase clouds the variable mass ratio between liquid water and ice as well as its spatial distribution within the cloud plays an important role for cloud life time, precipitation processes, and the radiation budget (Sun and Shine, 1994; Yong-Sang et al., 2014; Morrison et al., 2012). The complexity of interactions in mixed-phase clouds may result in parameterizations that are based on highly uncertain mixed-phase cloud classifications and thus lead to a misrepresentation of those clouds in models of all scales. Illingworth (2007) compared vertical ice water and liquid-water content as observed by a combination of ground-based radar, lidar, and microwave radiometer (MWR) comprised within the Cloudnet project with Global Climate Models (GCM). They showed that many GCMs underestimate the presence of mid-level clouds (As, Ac) by at least 30 % and that there is a large spread in the stated frequency of occurrence of liquid water in the models. This underestimation of the supercooled liquid fraction (SLF) in mixed-phase clouds in many GCM was e.g. also described in Komurcu et al. (2014). Tan et al. (2016) argued that a realistic representation of the SLF in GCM is needed to better constrain the equilibrium climate sensitivity. They stated that this can only be reached by more accurate observations of the distribution of supercooled liquid in mixed-phase clouds. This remains a challenge due to the difficulty of identifying the presence of supercooled liquid water layers embedded in cloud regions dominated by ice (Shupe et al., 2008; Luke et al., 2010; Silber et al., 2020). Besides single-layer mixed-phase clouds existing of a supercooled liquid top where ice particles are nucleated and precipitate out, ~~multilayer~~ multi-layer clouds (MLC) often exist (Vassel et al., 2019). MLC can interact microphysically via the seeder-feeder effect (e.g., ~~(Cotton and Anthes, 1989; Hobbs and Rangno, 1998)~~ (Cotton and Anthes, 1989; Hobbs and Rangno, 1998; Radenz et al., 2019; Ramelli et al., 2021), i.e. ice crystals nucleated in an upper liquid layer can fall into lower liquid layers, interact with its hydrometeors and influence cloud lifetime and precipitation efficiency. We thus argue that it is important to improve the detection of ~~multilayer~~ multi-layer liquid layer occurrences.

Synergistic measurements of cloud Doppler radar and polarization lidar can be used to identify cloud thermodynamic phase in mixed-phase clouds (e.g., Shupe (2007); Illingworth (2007); de Boer et al. (2009); Kalesse et al. (2016a) based on differences in the scattering mechanisms at the different wavelengths. While cloud radars are highly sensitive to large particles such as ice crystals (backscattering cross section is proportional to the particle size D^6 for the size range in which the Rayleigh approximation is valid), lidars are sensitive to higher concentrations of smaller particles such as cloud droplets and aerosol particles as the backscattering cross section is proportional to the projected surface area of the scatterers (O'Connor et al., 2005). As an additional variable, the state of polarization of the received lidar backscatter cross section gives information about particle shape. This is usually utilized by means of the detection of the circular or linear depolarization ratio (Sassen, 1991), hereafter referred to as lidar depolarization ratio. When multiple scattering is negligible, a low (high) lidar depolarization ratio indicates the presence of spherical (non-spherical) particles (Hu et al., 2006). Except for small quasi-spherical ice particles, ice is usually non-spherical so that the lidar depolarization ratio can also be used to infer cloud phase (Seifert et al., 2010). Concluding, liquid-dominated layers are characterized by high lidar backscattering cross section, low lidar depolarization ratio concurrent with small radar reflectivities and small mean radar Doppler velocities. Ice-dominated layers lead to a low lidar backscattering cross section, a high lidar depolarization ratio as well as higher radar reflectivities and higher mean Doppler velocities. Such synergistic lidar-radar retrievals are however only applicable up to the maximum lidar observation height determined by complete signal attenuation at a penetrated optical depth of about three and thus do not allow for the characterization of cloud liquid in the entire vertical column, e.g. in the presence of ~~mulilayered~~ multi-layered mixed-phase clouds.

Since cloud Doppler radars are able to penetrate multiple liquid layers, they can be used to detect warm and supercooled liquid layers (SCL) beyond the lidar measurement range via identification of morphological features in the cloud radar Doppler spectrum (Luke et al., 2010; Verlinde et al., 2013; Kalesse et al., 2016b) and thus have great potential to characterize the distribution of SCL in the entire vertical column. Specifically, if cloud ice and liquid are observed in the same radar sampling volume and if their populations are sufficiently separated by their respective terminal fall velocities, the cloud radar Doppler spectra may contain multiple peaks. Since the terminal velocity of small cloud droplets is negligible they cause a peak at about $0\,\mathrm{m\ s^{-1}}$ in the Doppler spectra; any deviation from this is caused by vertical motions (Shupe et al., 2004). Ice particles have larger and broader fall velocity distributions and thus cause a spectral peak at higher Doppler velocities. If the fall velocity difference between liquid and ice is small (for example when the ice population is comprised of smaller crystals), single-peak skewed (non-Gaussian) Doppler spectra are observed (Williams et al., 2018). Sub-volume turbulence does however induce spectrum broadening which can smear microphysically-induced morphological features in the Doppler spectrum (Kollias et al., 2007). The separation of both hydrometeor populations is thus only possible if the cloud radar settings are optimized to reduce spectrum broadening by a short dwell time, a small beam width, and a small resolution volume (Kollias et al., 2016). Sufficient range-dependent sensitivity of the cloud radar is also required as the reflectivity of the liquid peak comprised of small droplets can be as low as -40 dBz for convective situations Lamer et al. (2015).

As specific technical settings and cloud conditions are required in order to identify liquid water directly from cloud radar measurements, more sophisticated approaches are needed to make cloud radars applicable to a broader range of conditions. Artificial neural networks (ANN) are increasingly being used in atmospheric science to evaluate large datasets and/or to combine the advantages of different sensors. In short, ANNs are mathematical models trained to recognize patterns. Validation is often done by comparison to other (physical) retrievals. As emphasized in Liljegren (2009), ANN-based retrievals have been proven to be reliable statistical techniques that are preferable to computationally expensive variational retrievals for certain applications. Liljegren (2009) developed an ANN algorithm in which G-band vapor radiometer measurements are used to retrieve low amounts of liquid water and water vapor. Strandgren et al. (2017a) determine cirrus properties from the SEVIRI imager on Meteosat Second Generation satellites based on a set of ANN trained SEVIRI thermal observations and satellite-based lidar backscatter products among others. Andersen et al. (2017) use an ANN based on 15 years of monthly averaged Moderate Resolution Imaging Spectroradiometer (MODIS) liquid cloud products to determine the drivers of marine liquid-water cloud occurrence. All of the above studies employ ~~multilayer~~ multi-layer perceptrons (MLP, a specific type of feed-forward artificial neural network) that are commonly used in atmospheric sciences as they are able to model highly nonlinear functions (Andersen et al., 2017). Generally speaking, a vector of output data is estimated from an input data vector by modeling the relationship between the input- and output data. The training of the MLP is done for a variety of examples where the input- and corresponding output is known. The MLP structure consists of an input layer, a chosen number of hidden layers, and an output layer. Each of these layers is made of a certain number of neurons that exchange information in a way that the output of the previous layer is used to process the output for each connected neuron in the subsequent layer according to the corresponding numeric weights assigned to each neuron–neuron connection through an activation function (Strandgren et al., 2017b). By using error back-propagation introduced in Rumelhart et al. (1986), the numeric weights of the neurons are adjusted in an iterative process until the squared error between the predicted (estimated) output and the known reference output data reaches its minimum.

In the present study ~~a pre-trained ANN~~ an ANN pre-trained in Arctic conditions developed by Luke et al. (2010) for cloud radar-based liquid detection beyond full lidar signal attenuation ~~(pre-trained in Arctic conditions)~~ is applied to mid-latitude observations (Section 2). The objective of the study is to evaluate the ANN-based liquid classification against the Cloudnet target classification (Hogan and O'Connor, 2006) by using independent measurements of MWR liquid water path (LWP), first liquid-dominated cloud base height from ceilometer observations~~and~~ , relative humidities with respect to liquid as obtained from radio soundings and for one case study also space-borne lidar observations from a CALIPSO overpass (Section 3). A short conclusion summarizing the findings is provided in Section 4.

2 Methods

2.1 Observations

2.1.1 ACCEPT field experiment

Data used in this study were obtained during the Analysis of the Composition of Clouds with Extended Polarization Techniques (ACCEPT) field experiment which took place at the Cabauw Experimental Site for Atmospheric Research (CESAR, (51.971°N, 4.927°E)) in the Netherlands during 1 October- 18 November, 2014. During that field experiment, the remote-sensing instrumentation suite operated by the Royal Netherlands Meteorological Institute (KNMI) was complemented by the Leipzig Aerosol and Cloud Remote Observations System (LACROS; Büehl et al. (2013)) mainly consisting of a vertically-pointing 35 GHz MIRA-35 cloud radar (Görsdorf et al., 2015), a ceilometer, a multi-wavelength polarization Raman lidar (PollyXT; Engelmann et al. (2016)), and a HATPRO-MWR (Rose et al., 2005). Additionally, a new polarimetric hybrid-mode 35 GHz cloud radar (named hybrid MIRA-35) from METEK GmbH described in Myagkov et al. (2016a, b) and the Transportable Atmospheric Radar (TARA, S-band) operated by the TU-Delft were deployed (Pfitzenmaier et al., 2017).

2.1.2 MIRA-35 characteristics

In the present study, data from the vertically-pointing MIRA-35 was used as input to the ANN of Luke et al. (2010) to predict liquid beyond full lidar signal attenuation. The MIRA-35 was operated with a pulse length of 208 ns, resulting in a vertical range resolution of 31.18 m. Incoherent averages of 20 Doppler spectra produced from a series of 256 consecutive radar pulses with a pulse repetition frequency of 5000 Hz led to a temporal resolution of 1.024 s. The MIRA-35 Doppler spectra resolution was 8 cm s^{-1}.

2.1.3 Cloudnet target classification

The observations of the MIRA-35, the ceilometer and the MWR have been processed using the widely-used Cloudnet processing chain. One of the main products of Cloudnet is the target classification product (Hogan and O'Connor, 2006) which is illustrated in Fig. 1 and which we use to validate the ANN-predicted liquid detections. In order to classify a cloud volume to contain liquid, the Cloudnet target classification algorithm requires a valid lidar attenuated backscatter coefficient. For deep- or multiple liquid layers and situations with low-level fog the lidar signal can get fully attenuated, so the Cloudnet target classification thus underestimates the occurrence of liquid in the entire vertical atmospheric column and overestimates the presence of ice as target category (Griesche et al., 2020). Such a situation is depicted in the synergistic radar-lidar observables and the resulting Cloudnet target classification in Fig. 1. The signals of the PollyXT lidar /ceilometer were fully/partially attenuated by the near-surface fog occurring after Nov 18, 2014 07:30 UTC so that the cloud in 1.5-2.5 km around the 0 °C-isotherm was classified as ice cloud.

2.2 Description of the used Artificial Neural Network

Luke et al. (2010) use collocated measurements with profiling cloud Doppler radar and polarization lidar in thin mixed-phase clouds or lower layers of thick mixed-phase clouds to provide information about the existence of liquid water in higher cloud layers by predicting the lidar backscatter and depolarization signal from morphological features in the cloud radar Doppler spectrum. The procedure to determine the existence of supercooled-liquid droplets from cloud radar Doppler spectra is a two-step technique. In the first step, morphological feature extraction from cloud radar Doppler spectra is done by applying a second order Gaussian continuous wavelet transform (CWT) to each measured radar Doppler spectrum. In that way, the spectral power is decomposed into a 2D-array with feature localization in Doppler velocity and spectrum width; each Doppler spectrum can thus be regarded as a sum of different Gaussians. In the second step, a selected subset of bins from six different scales of the CWT as well as the first three radar moments (reflectivity factor (Z_e [dBZ]), mean Doppler velocity (V_D [m s^{-1}]), and Doppler spectrum width (σ [m s^{-1}])) of each Doppler spectrum are the input to the ANN used in this work to predict the existence of liquid. The ANN is of the ~~multilayer~~ multi-layer perceptron (MLP) type consisting of 256 input nodes, five hidden layers, and two output nodes. Each of the five hidden layers consists of 32 nodes. Lidar particle backscatter coefficient (β [sr^{-1} m^{-1}]) and lidar depolarization ratio (δ) are the two output variables from which the existence of liquid is predicted using appropriate thresholds of β and δ later on. In the training phase (which was performed on data from the Mixed Phase

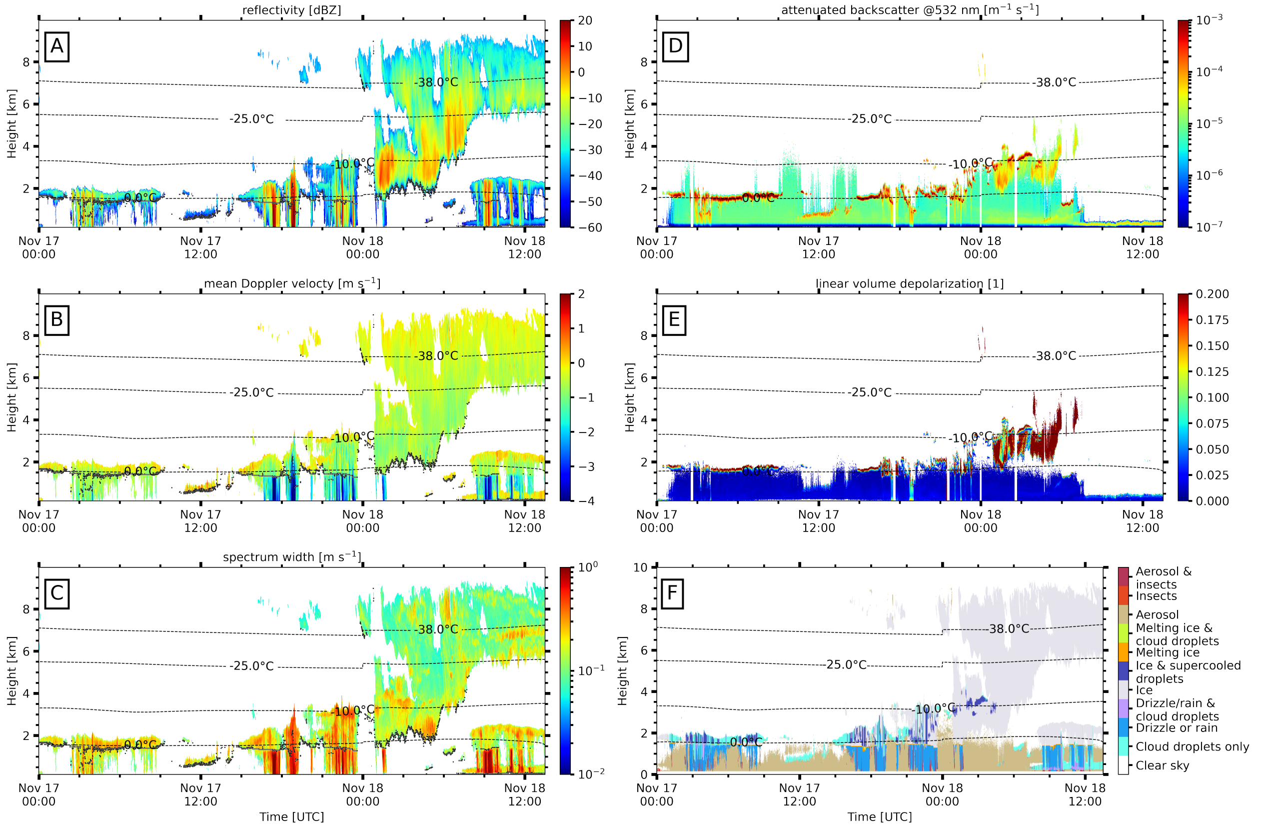

Figure 1. ~~left:~~ (A) MIRA-35 radar reflectivity factor, (~~top~~B) ~~,~~ radar mean Doppler velocity (middle), (C) radar spectrum width, (~~bottom~~D) ~~, right:~~ PollyXT lidar attenuated backscatter coefficient at 532 nm, (~~top~~E) ~~,~~ PollyXT lidar linear volume depolarization, (~~middle~~F) ~~and~~ Cloudnet target classification ~~(bottom)~~ of Nov 17, 2014 00:~~00~~ UTC to Nov 18, 2014 13:30 UTC observed during the ACCEPT experiment in Cabauw, Netherlands. Black ~~stars~~ dots (~~in the lidar variable panels~~A-C) indicate the first cloud base detected by the ceilometer.

Arctic Clouds Experiment (MPACE, Verlinde et al. (2007)) obtained in fall 2004 at the U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement (ARM) North Slope of Alaska (NSA) permanent site in Utqiagvik (formerly known as Barrow), Alaska, the backpropagation of errors algorithm was applied. In short, the β and δ output of the ANN for each time and height pixel were compared to values measured with a High Spectral Resolution Lidar (HSRL, Eloranta (2005)). The difference between ANN-predicted and lidar-observed (i.e., the error) was monitored and the internal weights of the nodes were adjusted until the error did not decrease any further during the successive cycling through the Doppler spectra training data set. Only a fraction of the MPACE data was considered in the training phase, most of the data was used for validation. Turbulent broadening of the cloud radar Doppler spectrum (e.g. in strong convection) decreases the imprint of cloud microphysics on the Doppler spectra. The MPACE dataset was characterized by largely stratiform conditions. As stated in Gardner and Dorling (1998), the ability of an ANN to predict cloud properties does not only dependent on an informed choice of predictors but also requires sufficient data that fully represent all cases that the ANN is required to generalize, as ANNs perform well for interpolation but poorly for extrapolation. We can thus only expect good predictions of liquid in low-turbulent clouds but not in strongly convective clouds. The objective of this study was to check the performance of the ANN trained with the MPACE observations in Luke et al. (2010) on a new data set, the ANN was thus not re-trained.

2.3 Classifying liquid containing sections from ANN-predictions

The ANN-predicts backscatter coefficient and particle depolarization ratio. Thresholds need to be applied to these predicted β and δ in order to identify regions which show optical properties similar to the ones produced by liquid water.

For visual illustration of the mapping from predicted lidar variables to hydrometeor class labels, a scatter plot of predicted β and δ was created (Fig. 2 (A)). As previously mentioned, lidar observed or ANN-predicted high values of β and near-zero δ are reliable indicators of liquid-dominated cloud regions; they clearly stand out as a feature in Fig. 2 (A). The scatter plot of predicted β and δ shows two more distinct features, one between the functions "linear-1" and "linear-2" with higher values of δ and lower values of β indicating ice and another feature of very high values of δ and very low values of β situated below the function "linear-2" that can be attributed to the optically thinner ice cloud with lower radar reflectivities above 7 km on Nov 18, 2014 (see Fig. 2 ~~, right)~~(B)). Similar to Luke et al. (2010), fixed thresholds of β and δ were used to derive a binary mask separating liquid predictions from other target types. For a sensitivity study of ANN-predicted liquid occurrence for the entire ACCEPT data set, different HSRL-based published thresholds (Shupe, 2007; de Boer et al., 2009; Luke et al., 2010) as well as a new linear function threshold (labeled "linear-1" in Fig. 2) were employed (see Table 1). Threshold values for β of all three published studies are similar. Shupe (2007) and Luke et al. (2010) use the same δ threshold of 0.1 for liquid classification while de Boer et al. (2009) with a value of 0.03 is much more stringent. The studies are subsequently referred to as "Shupe2007", "deBoer2009", and "Luke2010". The linear-1 threshold function was found by a sensitivity study and gave the most similar classification results to the three cited published threshold values. Figure 2 (B) shows the corresponding time-height representation color coded by linear separation of the predicted (backscatter vs. depolarization) dimension using linear functions. ~~Slightly paler color indicate regions of complete lidar attenuation by PollyXT and can also be seen as *extended lidar signal* predicted from cloud radar spectra, are excluded by the latter analysis.~~

Table 1. Published thresholds of β and δ for lidar-based liquid classification and linear-1 function threshold used for ACCEPT data set.

method	thresholds
Shupe2007	$\log(\beta) > -4.5$, $\delta < 0.1$
deBoer2009	$\log(\beta) > -4.3$, $\delta < 0.03$
Luke2010	$\log(\beta) > -4.3$, $\delta < 0.1$
linear function-1 ($m\delta + \beta$)	$m = 12$, $\beta = -5.0$

(A) (B)

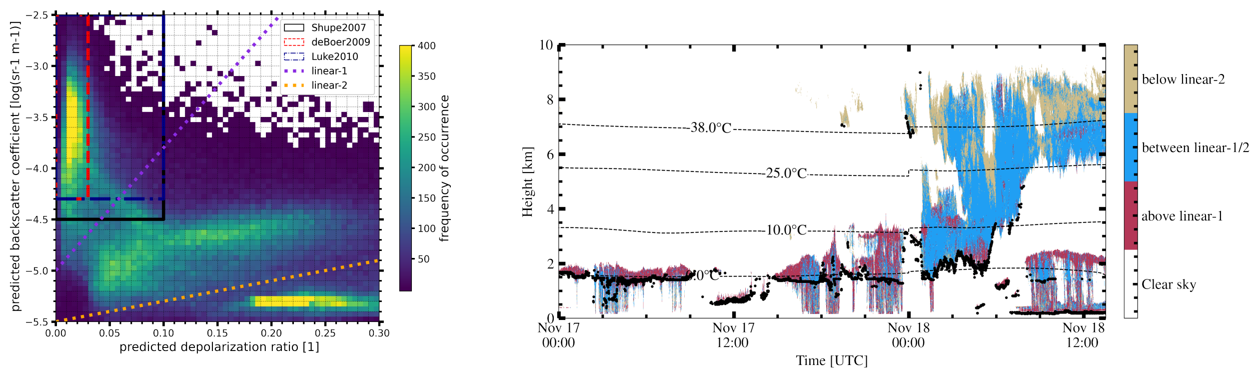

Figure 2. ~~Left:~~ (A) Frequency of occurrence of ANN-predicted lidar backscatter coefficient β vs. predicted lidar linear depolarization δ for the Nov 17 - 18, 2014 case study. ~~Right:~~ (B) Time-height mapping of predicted β and δ of the three corresponding areas in the ~~left~~ (A) panel, which are separated by the two linear thresholds. Black dots (B) indicate the ceilometer cloud base height.~~Light colors indicate regions beyond complete attenuation of the lidar beam.~~

The liquid classification methods were applied to the entire ACCEPT dataset. For doing so, the following pre- and post-processing steps were applied to the seven-week long data set. Firstly, to account for the effects of radar partial beam filling, cloud edges are excluded from the ANN input data by setting data in the first and last range gate of a detected cloud (i.e. cloud base and cloud top pixel) to "clear sky". ~~Secondly, only "cloud" pixels of the Cloudnet target classification mask (between first cloud base and last cloud top) are considered, so that pixels classified as rain/drizzle and~~ Secondly, pixels classified as aerosols/insects were explicitly excluded. ~~Thirdly, ANN liquid predictions for regions with good lidar echo and Cloudnet-classified as non-liquid class, are reclassified as non-liquid. Fourthly,~~ Thirdly, using model temperature data of the Global Data Assimilation System (GDAS1) employed by the Global Forecast System (GFS) model, unphysical liquid predictions below $-37\,°\mathrm{C}$ were re-classified as ice. The in-cloud pixels which were classified as liquid-containing by the ANN using the above-mentioned thresholds were sometimes quite patchy. Similar to Shupe (2007) a homogenization step to create more coherent liquid layer structures, by using a 5x5 pixel neighborhood smoothing was introduced. A pixel was kept as liquid-containing pixel, when at least 60 % of the pixels in the 5x5 box around the center one were also classified as liquid-containing.

3 Results and Discussion

To assess the performance of the Luke et al. (2010) ANN-based liquid prediction from cloud radar Doppler spectra using different published thresholds of lidar backscatter coefficient and depolarization ratio against the Cloudnet target classification and against independent observables, a two-step validation was performed. Firstly, a case study (Nov 17-18, 2014 consisting of

100.000 samples) was analyzed in depth, see Table 2. Secondly, statistical results for the ANN-based liquid-prediction for the entire ACCEPT data set (1070 hours of observations, i.e. 1.7 million samples) are given in Table 3 and discussed subsequently. In the following, the abbreviation CD is used for *cloud droplets bearing samples* and non-CD for *non-cloud droplets bearing samples*. It should be noted that no further distinction between other liquid-bearing samples such as drizzle/rain is made for the ANN-based liquid predictions.

Predictions that meet the criteria from Section 2.3 are compared to classifications from Cloudnet (treated as ground-truth)~~, resulting in an 2x2~~. The comparison yields an error matrix consisting of correctly classified ~~CD~~ predictions, i.e. *true positive* (TP) and ~~correctly classified non-CD predictions, i.e.~~ *true negatives* ~~. Additionally, CD predictions that are actually non-CD~~ (TN) as well as *false positives* (FP) ~~, and predicted non-CD which are CD~~ and *false negatives* (FN) which concern wrong predictions, respectively. Described below are four metrics used to evaluate the predictive performance against Cloudnets' liquid detection, three correlation coefficients $\rho_{\mathrm{a,b}}$, and the fraction of liquid predicted located within a relative humidity above 90%.

1. ~~precision = $\frac{\mathrm{TP}}{\mathrm{TP+FP}}$: The precision value~~ Error matrix: A 2 by 2 matrix consisting of the numbers for correctly identified CD (TP) and non-CD (TN) time-height grid cells, as well as falsely classified non-CD (FP) and CD (FN) cells respectively, i.e.

$$EM = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix}. \tag{1}$$

2. Precision: A real value between 0 and 1, where 1 is the perfect score. $prec = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$, i.e. the fraction of how many predictions where correctly classified as CD (i.e. TP) by the sum of TP and predictions falsely classified as CD (i.e. FP). In the context of this work, it measures the amount of CD overestimation. The closer precision gets to 1, the ~~fewer FP classification a method computes. (if precision $< 1 \Rightarrow$ CD overestimation)~~ more precisely actual CD cells are predicted as such. Precision can also be described as 1 minus the false alarm rate.

3. ~~recall = $\frac{\mathrm{TP}}{\mathrm{TP+FN}}$: The recall value~~ Recall or probability of detection: A real value between 0 and 1, where 1 is the perfect score. $recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$, i.e. the fraction of TP and the sum of TP and falsely classified non-CD (i.e. FN). In the context of this work, $recall$ measures the amount of CD underestimation. The closer recall gets to 1, the ~~fewer FN classification a method computes. (if recall $< 1 \Rightarrow$ CD underestimation)~~

4. ~~accuracy = $\frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$: The closer the accuracy value gets to~~ less likely it is missing actual CD cells. Note: The ceilometer lidar signal which is used as ground-truth indicator for CD availability, is much more sensitive to CD than Doppler cloud radar signals, thus recall values below 1 ~~the more pixel were classified correctly in an absolute and non-balanced way. (overall accuracy)~~ are expected.

5. ~~F1 − score = $\frac{2}{\frac{1}{\mathrm{recall}}+\frac{1}{\mathrm{precision}}}$: The balanced (harmonic) mean of CD over- and underestimation. (harmonic mean of precision~~ Accuracy: A real value between 0 and ~~recall)~~ 1, where 1 is the perfect score. $acc = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$, i.e. the

fraction of all correct predicted CD pixel and the sum of all samples. In the context of this work it measures the overall fraction of correct versus incorrect predictions, where $acc = 0.75$ if the retrieval correctly classifies 3 out of 4 inputs.

6. ~~$\rho_{\text{ceilo-CBH,LLH}}$: correlation of first ceilometer cloud base height (CBH) with first liquid layer height (LLH) of ANN and Cloudnet~~

7. ~~$\rho_{\text{MWR-LWP,LLT}}$: correlation of~~ Correlation between MWR-LWP ~~with ANN- and Cloudnet-derived~~ and retrieved liquid layer thickness ~~(LLT, product of sum of liquid containing pixel per profile times range gate resolution)~~

8. ~~$\rho_{\text{MWR-LWP,LWP}_{\text{ad}}}$: more physically meaningful correlation of~~ LLT: The MWR-LWP ~~with LWP calculated from LLT and profiles of temperature and pressure under adiabatic assumption (as in Karstens et al. (1994))~~

9. ~~Liq-Pxl at RH~~time series is correlated to the LLT time series computed as the sum of the vertical extend of CD containing volumes LLT= $N_{CD} \cdot \delta h$, with $\delta h = 40$ ~~> 90 %: fraction of liquid-classified pixels that overlap with pixels of radio-sounding based relative humidity (RH) with respect to water above 90 %~~m range resolution. Profiles in which rain was observed at ground were excluded from the correlation coefficient determination to avoid wrong MWR-LWP caused by a wet MWR radome.

10. Correlation between ceilometer first CBH and retrieved first liquid layer height LLH: The Ceilometer first CBH time-series is correlated to the first LLH time-series as retrieved from the CD mask.

3.1 Nov 17-18, 2014 Case Study Results

The 37.5 h long case study of Nov 17, 2014 ~~00:00~~0 UTC - Nov 18, 2014 13:30 UTC was characterized by a multitude of cloud types including pure liquid water clouds, stratiform mixed-phase clouds, high clouds, mid-level clouds and near-surface clouds (fog) as shown in Fig. 1. On Nov 17, 2014 between 3-9 UTC and ~~15-00~~15-24 UTC several rain showers from low mixed-phase clouds with cloud-top temperatures between -10 °C and -2 °C were observed. At around 12 UTC, a thin warm liquid cloud at 1 km altitude with a LWP below 30 g m^{-2} was present. On Nov 18, different multi-layer clouds with varying vertical extent were present, a high cloud in 6-9 km was firstly situated above a mid-level cloud in 2-5 km and later on over a precipitating stratiform cloud in about 2 km altitude with cloud-top temperature of -5 °C ~~which had~~. Below this cloud was a layer of near-surface fog~~below~~.

In Fig. 3 the comparison of the resulting liquid masks of the ANN of all presented thresholds and for Cloudnet for this case study are shown. There is mostly good agreement in liquid-detection for the stratiform mixed-phase ~~cloud~~ clouds on Nov 17 before 21 UTC and the liquid cloud ~~in 1 km~~at around 12 UTC on Nov 17. However, since the liquid-threshold boundaries of deBoer2009 are very strict, many potential liquid pixel candidates are not considered (e.g. around 3 UTC, and 18 UTC on Nov 17). For this particular case, the Cloudnet target classifcation algorithm was not able to fully identify the cloud-top layer at -10 °C during ~~0-6~~21-24 UTC ~~and at~~ on Nov 17 and at about 2 km during 9-12 UTC on Nov 18, as mixed-phase and/or supercooled liquid containing because of full lidar signal attenuation in the rain/fog below. The ANN-based liquid-detection clearly outperforms Cloudnet in these situations. ~~However, as stated in Luke2010, the ANN performance is expected~~

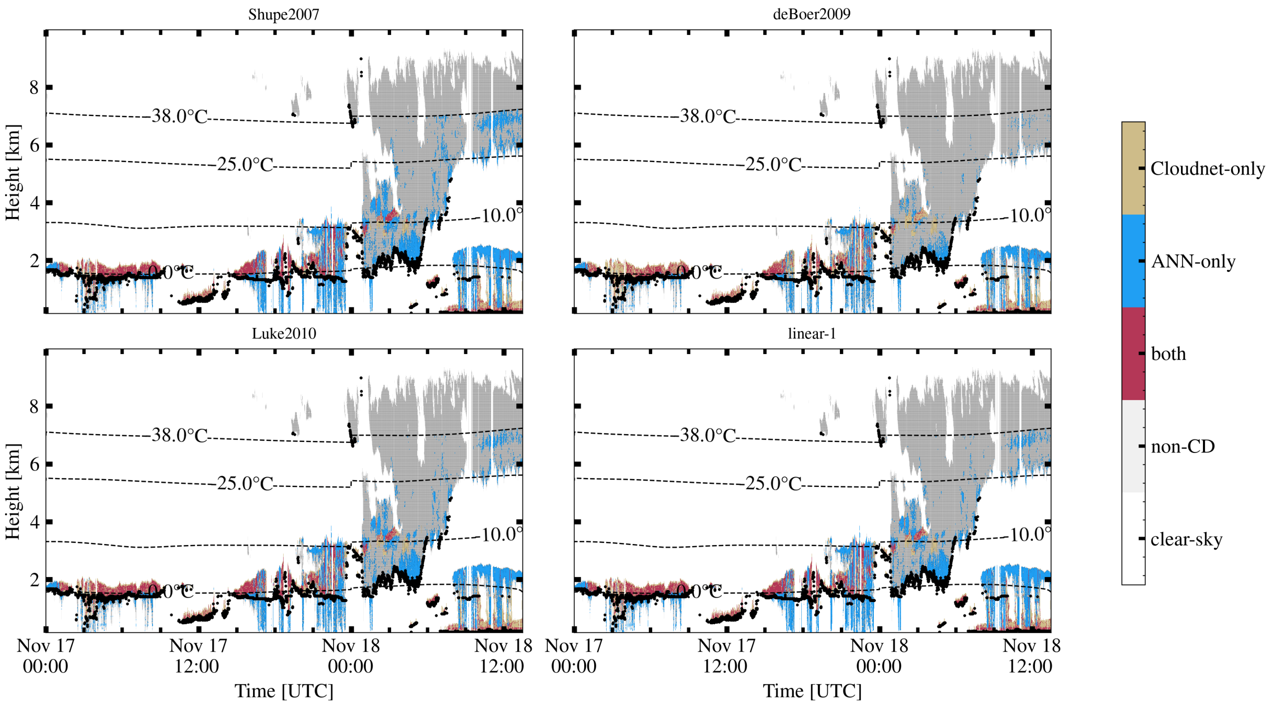

Figure 3. Sensitivity study of liquid pixel classifications of the Nov 17 - 18, 2014 case study using liquid mask thresholds of Shupe2007 (upper left), deBoer2009 (upper right), Luke2010 (lower left), linear-1 (lower right) on the ANN-predicted lidar variables. Light-brown: ~~Cloudnet only~~ Cloudnet-only liquid classifications, blue: ANN-predicted pixels using the given thresholds which were not classified as liquid droplets by Cloudnet, red: pixels classified as liquid by the ANN and Cloudnet, ~~black contour~~grey shading: pixel for which neither Cloudnet nor the ANN classified cloud ~~edges~~droplets, ~~green stars~~black dots: ceilometer first cloud base height, white: clear-sky~~or not classified as liquid by either method~~.

~~to decrease in more turbulent conditions leading to Doppler spectrum broadening. This is the case at 10-13 UTC on Nov 19 in a layer at about 7 km altitude around −37 °C (very low probability for liquid) for all but the deBoer2009 threshold.~~

For independent validation of the areas classified as liquid-containing, the summed up liquid layer thickness (LLT) of all pixels classified as liquid by the ANN or Cloudnet is compared to the MWR-LWP (Figure 4) as proposed by ~~Luke2010.~~ Luke et al. (2010). MWR-LWP uncertainty amounts to 25 $\mathrm{g\,m^{-2}}$. Profiles in which considerable amounts of rain/drizzle reached the ground were excluded in the LLT-determination to avoid ~~situation~~ situations with a wet MWR radome leading to an invalid MWR-LWP estimate (as indicated by the rain flag in Figure 4). In some situations the ANN and in others Cloudnet matches the ~~timeseries~~ time series of MWR-LWP better~~on Nov 17 while on Nov~~. A large discrepancy between ANN-LLT and MWR-LWP is obvious on Nov 18, 4-6 UTC: MWR-LWP are very low, while the ANN-LLT ~~mostly (except when using the deBoer threshold) overestimates LLT due to the explained misclassifications. More ACCEPT case studies have been analyzed in detail but are not presented here because they show very similar results.~~ is high. A misclassification of ice as liquid by the ANN in 2-3.5 km height can thus be concluded which is corroborated by the PollyXT lidar signal showing high depolarization

Table 2. Error matrix, performance metrics, and correlation coefficients for ceilometer-CBH vs. LLH, MWR-LWP vs. LLT, MWR-LWP vs. $LWP_{ad,cor}$, case study Nov 17-18. Statistic includes only valid pixel.

	Shupe2007	deBoer2009	Luke2010	linear-1	Cloudnet
TP	28684	22209	26803	28816	46215
TN	59342	60605	59615	59620	62424
FP	3082	1819	2809	2804	0
FN	17531	24006	19412	17399	0
precision	0.903	0.924	0.905	0.911	1
recall	0.621	0.481	0.580	0.624	1
accuracy	0.810	0.762	0.795	0.814	1
$\varrho_{\text{MWR-LWP,LLT}}$	0.436	0.533	0.490	0.489	0.471
$\varrho_{\text{MWR-LWP,LWP}_{\text{ad}}}$	0.275	0.471	0.335	0.345	0.399
$\varrho_{\text{ceilo-CBH,LLH}}$	0.775	0.725	0.738	0.755	0.913
Liq-Pxl at RH > 90 %	n/a	n/a	n/a	n/a	n/a

values indicating ice crystals. After 7 UTC on Nov 18, the lidar signals are fully attenuated by the fog near the ground and are thus not available for assessment of ANN classifications in higher layers. Analysis of radar Doppler spectra time- and height spectrograms in around 6-9 km altitude showed only monomodal spectra related to the falling ice crystals from above. In conclusion, most certainly, no formation of supercooled liquid in 7 km altitude at $-37\,°C$ occurred. The ANN thus most likely misclassified ice as liquid because the observed Doppler spectra at around 7 km were characterized by high spectrum width, small reflectivities and small mean Doppler velocities. High Doppler spectrum width might be related to more turbulent conditions which result in a decrease of the performance of the ANN because the microphysical imprint of the hydrometeors on the radar Doppler spectra is decreased.

The error matrix and evaluation metrics (first 8 rows in ~~table~~ Table 2) show the performance of ~~Luke et al. (2010) by comparing the~~ the ANN by comparing ANN-based liquid predictions to valid Cloudnet liquid detections for time-height cells with reliable radar and lidar signal status. Depending on the threshold given in Table 1, precision ranges between 0.9 (Shupe2007) and 0.92 (deBoer2009). Contrarily, recall values range between 0.53 (deBoer2009) and 0.67 (linear-1) indicating that more lose thresholds are better in detecting more TP, while keeping the number of FN comparably low. Overall accuracy ranges between 0.78 (deBoer2009) and 0.83 (linear-1)~~, with the more balanced F1-score showing the same threshold candidates for minimum 0.67 (deBoer2009) and maximum 0.77 (linear-1) values.~~.Regions with high ~~radar~~ Doppler spectrum width ~~at~~ near cloud base (see ~~figure 1 Nov~~Figure 1 Nov. 18~~03:00 - 06:00~~, 3-6 UTC between ~~2 and 3~~2-3 km altitude) contribute to a large portion of those FP for all thresholds. Lower recall values indicate a higher degree of underestimation of CD detections, which is caused by liquid layers with low LWP values below 50 g m^{-2}, e.g. the thin liquid cloud on Nov. 17 around 12~~:00~~ UTC in 0.5 to 1 km altitude. Profiles characterized by low precipitation rates of rain and drizzle have a negative Cloudnet rain flag and are thus not excluded from the analysis. For these drizzle/rain pixel the ANN often predicts liquid (see Figure 3 and A2 between 0-1.5 km). Since the ANN does not distinguish between different liquid classes such as drizzle/rain and cloud droplets (CD), the ANN classifies all these pixels as cloud droplets which are then counted as FP. FN often occur when Cloudnet classifies a certain hydrometeor class at low altitude and extends this target class for all pixel in the profile up to cloud top which e.g. either happens in low intensity precipitation (see Figure A1, misclassification of drizzle/rain as cloud droplets by Cloudnet, e.g. Nov 17, 2014 3–4 UTC, 0.5–2 km) and for the ice and supercooled droplets class on Nov 17, 17:30 UTC resulting in a 1 km deep mixed-phase layer in 1.5-2.5 km altitude). In such situations the ANN might be more accurate in determining the location of cloud droplets but since it is evaluated against Cloudnet as ground-truth, FN result.

In this work the ceilometer first cloud base height ~~variable~~ (CBH) is correlated to the predicted first liquid layer height (~~if liquid is present). However,~~ LLH). $\rho_{\text{ceilo-CBH,LLH}}$ of the four ANN methods are on the order of 0.86 (deBoer2009) to 0.92 (Shupe2007) for the entire ACCEPT dataset (see Table 3), i.e., there is a failure rate of 8–14 %. This failure rate can be explained by several conditions: Firstly, in some situations, like on Nov 18, 2014 between 1–4 ~~UTC~~ UTC, the ceilometer cloud base variable is not representing the base of the liquid layer but instead the base of precipitating ice crystals (Fig. 1). ~~Situations like these are the most important factor for leading to $\rho_{\text{ceilo-CBH,LLH}}$ of the four ANN threshold methods to be on the order of 0.86 (deBoer2009)to 0.92 (Shupe2007)~~ This is caused by specular reflection from the planar planes of horizontally aligned ice crystals as described in Westbrook et al. (2010). When the ANN is not misclassifying these ice crystals as liquid, the difference in ceilo-CBH and ANN-LLH is high. Secondly, there are situations where liquid layers with low LWP are only detected by the ceilometer but not by the cloud radar (Nov 17, ~~i. e., a failure rate of 8–14 %. Higher~~ 11 UTC, cloud at 1.7 km). Thirdly, there are cloud scenes where the ceilometer is fully attenuated by precipitation or low level fog (thus reporting the precipitation or fog base as first cloud base) which the radar can penetrate/is not sensitive to or which is below the first radar range gate. Fourthly, in situations where the ceilometer is still able to penetrate light precipitation to detect CBH (Nov 17, 3-9 UTC, 17-24 UTC) and the ANN misclassifies drizzle/rain as cloud droplets, further discrepancies arise. These conditions lead to a decrease of the $\rho_{\text{ceilo-CBH,LLH}}$. The $\rho_{\text{ceilo-CBH,LLH}}$ for ceilometer-CBH and Cloudnet ~~(which uses ceilometer data as input for the thermodynamic phase classification) are expected because the ceilometer is better able to detect thin liquid water layers with low LWP than the cloud radar is~~ for the entire ACCEPT data set is higher and amounts to 0.97. While the liquid layer base height variable in Cloudnet is based on the gradient of ceilometer attenuated backscatter coefficient, the internal ceilometer cloud base determination is not precisely documented in the ceilometer manual. Differences in cloud base height leading to a failure rate of 3 %may thus occur due to different backscatter coefficient thresholds. ~~This being said, $\rho_{\text{ceilo-CBH,LLH}}$ for Cloudnet is actually expected to be 100 % - it is slightly lower due to averaging of ceilometer data to the 30 s Cloudnet input file resolution.~~

The $\rho_{\text{MWR-LWP,LLT}}$ also shows positive correlations for all methods~~and thresholds~~. As shown in Table 2, it ranges between 0.44 (Shupe2007) to 0.53 (deBoer2009), for Cloudnet the $\rho_{\text{MWR-LWP,LLT}}$ amounted to 0.47. Converting the LLT to the physical more meaningful $\text{LWP}_{\text{ad,cor}}$ results in $\rho_{\text{MWR-LWP,LWP}_{\text{ad}}}$ that are very similar to $\rho_{\text{MWR-LWP,LLT}}$ with moderate correlation (0.47) for ~~Shupe2007 (0.28) , Luke2010 (0.34)~~deBoer2009, and weaker correlations for all other methods. Both $\rho_{\text{MWR-LWP,LLT}}$ and ~~"linear-1" (0.35) while they increase~~ $\rho_{\text{MWR-LWP,LWP}_{\text{ad}}}$ of deBoer2009 show the strongest relationship to the measured

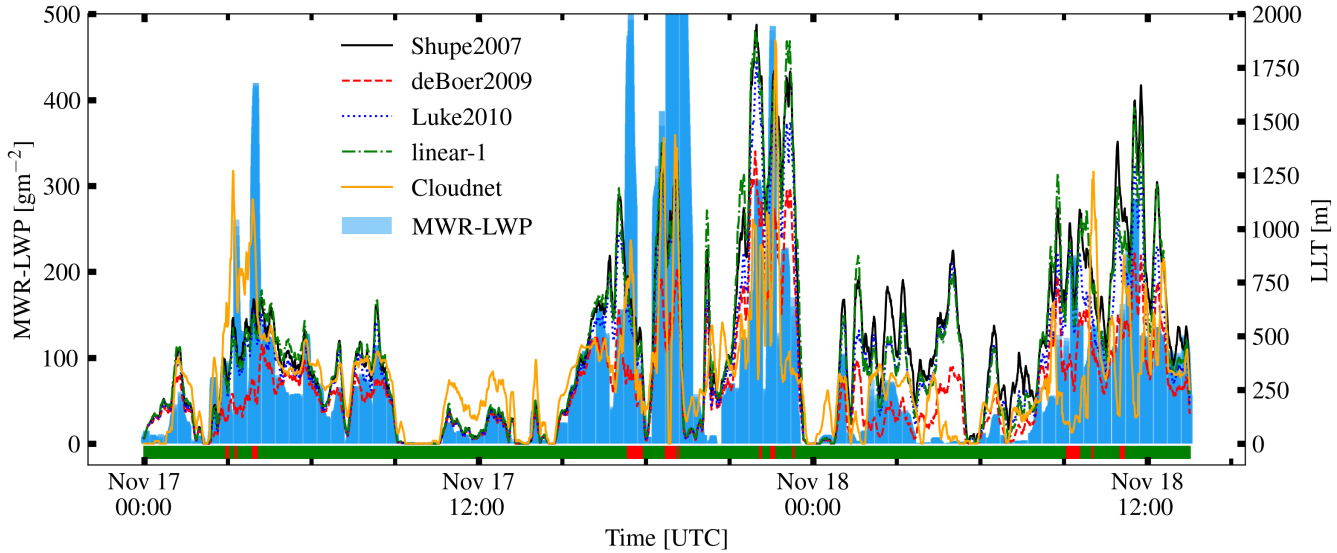

Figure 4. Comparison of MWR-LWP (left y-axis, blue bars) and liquid layer thickness (LLT, right axis) of the ANN-predicted liquid layer masks and Cloudnet-LLT (orange) for the Nov 17 - 18, 2014 case study for all used liquid-detection thresholds. The disdrometer-based Cloudnet rain flag is depicted by green and red markers near the bottom of the plot respectively indicating profiles with rain (red) and times where it was drizzle/rain free or precipitation rates were too low to be observed by the disdrometer.

MWR-LWP. The period Nov. 17, 21 UTC to Nov. 18, 12 UTC in Figure 4 shows the highest differences in LLT between the deBoer2009, Cloudnet and the other methods. The number of CD predictions in the precipitating system (Nov. 17, 20-23 UTC), the region with higher spectrum width (Nov. 18, 4-6 UTC at cloud base and Nov. 18, 10-13 UTC at 7 km altitude, see: Fig. 3) are lowest for deBoer2009~~(0.47) showing the~~ , therefore reflecting the MWR-LWP best. However, deBoer2009 also counts the least amount of ~~FP which cause lower correlations for other thresholdsand Cloudnet (0.40) which can only be explained by the fact that thin LLT (which are better detected by the latter one) result in rather low~~ ~~$LWP_{ad,cor}$~~TP, due to its tight thresholds, which seems to have minor effects on the correlation coefficient. Unfortunately, no radio sondes were launched during the presented case study~~.~~, so the relative humidity related measure could not be determined. Multiple other case studies had similar results.

3.2 Oct 5, 2014 Case Study Results

As previously mentioned, no validation of the ANN-liquid prediction can be made if the ground-based lidar signals are fully attenuated. We therefore use the unique opportunity to compare the Cloudnet and ANN liquid identifications in multi-layer cloud situations to a nearby (47 km distant) CALIPSO overpass on Oct 5, 2014 ~~case study for all used liquid-detection thresholds. Green and red dots near the bottom of the plots respectively indicate valid MWR-LWP data (no drizzle or rain detected~~ 01:05 UTC. On Oct 5, 2014 01-04 UTC multiple cloud layers were present. Besides warm stratiform liquid clouds

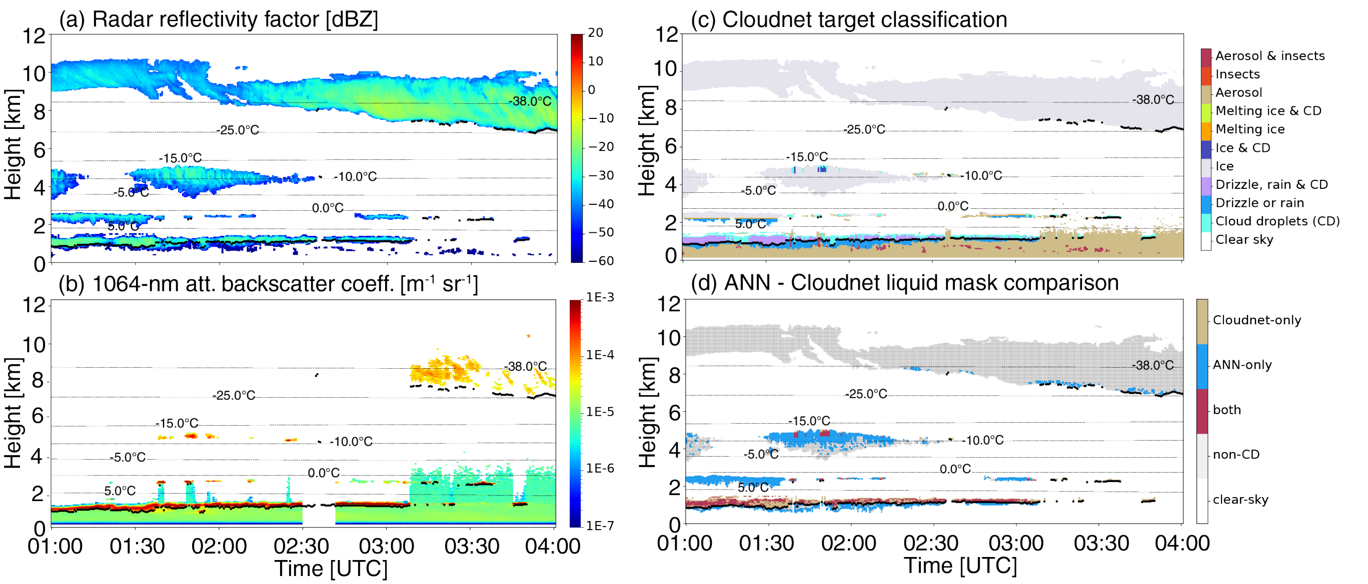

Figure 5. ~~Comparison of MWR-LWP (left y-axis~~Observations and retrievals on October 5, ~~blue bars~~2014 01 - 04 UTC: a) ~~and liquid layer thickness (LLT~~MIRA-35 radar reflectivity factor, ~~right axis~~b) ~~of the ANN-predicted liquid layer mask (red~~PollyXT 1064 nm attenuated backscatter coefficient, c) Cloudnet target classification, d) comparison of Cloudnet and ~~Cloudnet-LLT (~~ANN liquid masks. First ceilometer cloud base is indicated by black ~~) for the Nov 17 - 18~~dots.

below 3 km altitude, a midlevel cloud with cloud top temperature of -14 °C was observed in 5 km altitude. An extensive cirrus was present between 7-10.5 km altitude. From 01-03 UTC, the PollyXT lidar signal was mostly fully attenuated by the lowest liquid cloud in 1 km altitude leading to a misclassification of liquid as ice by Cloudnet for the warm cloud in 2.5 km altitude. Also, (except for a few pixels where the lidar had a valid signal) Cloudnet classified the midlevel cloud as ice-cloud. The ANN correctly predicted liquid for all warm clouds (note that below cloud base of the lowest cloud layer, ANN also predicts liquid which are counted as cloud droplets (CD) since it does not distinguish between different liquid classes such as cloud droplets and rain/drizzle). The ANN classifies the midlevel cloud as liquid-topped with ice precipitating from it below. The phase classification of the ANN in the cirrus is mostly ice except for some regions close to cloud base where high spectrum width and near-zero mean Doppler velocities result in a prediction of supercooled liquid.

The cloud fields were extensive so CALIPSO identified a very similar cloud situation with a cirrus of high vertical extent and a midlevel cloud in 3.5-5 km. The CALIOP signal was fully attenuated in this cloud layer so the low level warm clouds were missed by the satellite observation. The CALIPSO cloud phase index classified the high cloud as ice cloud and the midlevel cloud as liquid-topped cloud with liquid-only or liquid+~~15 min~~ice in the lower regions of this cloud. CALIPSO thus validates the ANN-based liquid prediction for the midlevel cloud. This hints to the usefulness of employing satellite-based hydrometeor target classifications as independent validation tool.

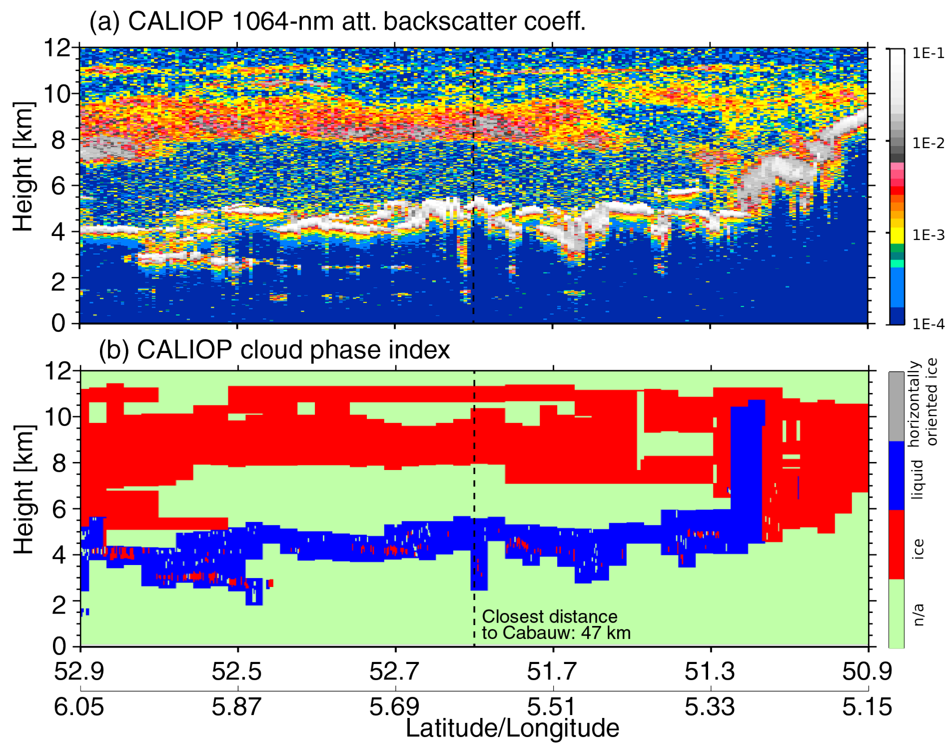

Figure 6. a) Curtain plots of CALIOP 1064 nm attenuated backscatter coefficient and b) CALIOP cloud phase index on Oct 5, 2014 where the closest distance of CALIPSO to Cabauw was 47 km at 01:05 UTC.

3.3 Statistical results for entire ACCEPT field campaign

A more general evaluation of all methods is done for the entire ACCEPT field campaign comprised of 1070 h ~~for the MWR radome to dry)~~of observations counting more than 1.7 M samples. The summary of this evaluation is presented in Table 3. All ANN-thresholds achieve high precision values > 0.9, indicating a low FP rate. Recall values are moderately lower compared to the Nov 17-18, 2014 case study, ranging from 0.4 (deBoer2009) to 0.54 (Shupe2007). Accuracy lies above 0.75 (three out of four predictions are correct) for all methods except slightly lower values for deBoer2009 (explained in Section 3.1). However, deBoer2009 achieves best correlation for $\rho_{\mathrm{MWR\text{-}LWP,LLT}}$ and ~~invalid MWR-LWP data masked for analysis due to precipitation. (Garrett and Peng, 2021)~~$\rho_{\mathrm{MWR\text{-}LWP,LWP_{ad}}}$, due to CD overestimation (larger numbers of FP) for Shupe2007, Luke2010 and linear-1. Overall, all methods achieve better correlation values for the entire data set compared to the case study of Nov 17-18, 2014, with high $\rho_{\mathrm{ceilo\text{-}CBH,LLH}}$, values ranging from (0.86-0.92) and (0.97) for Cloudnet respectively. The similar values of correlation coefficients and cloud droplet prediction error matrix elements in Table 2 and Table 3 indicate that the entire data set is well represented by the Nov 17-18, 2014 case study. As indicated in Section 3.1, mis-interpreted spectral signatures (small ice particles with low fall speed are misclassified as ice) and turbulence broadened radar Doppler spectra are the main driver for miss-classifications of the pre-trained Luke et al. (2010) approach.

Table 3. Error matrix, performance metrics, and correlation coefficients for ceilometer-CBH vs. LLH, MWR-LWP vs. LLT, MWR-LWP vs. $\mathrm{LWP}_{\mathrm{ad,cor}}$, for ~~case study Nov 17-18~~the entire ACCEPT data set. Statistic includes only valid pixel.

	Shupe2007	deBoer2009	Luke2010	linear-1	Cloudnet
TP	~~28684~~ 406235	~~22209~~ 302643	~~26803~~ 374880	~~28816~~ 401331	~~46215~~ 757342
TN	~~59342~~ 919571	~~60605~~ 938429	~~59615~~ 925740	~~59620~~ 925243	~~62424~~ 962586
FP	~~3082~~ 43015	~~1819~~ 24157	~~2809~~ 36846	~~2804~~ 37343	0
FN	~~17531~~ 351107	~~24006~~ 454699	~~19412~~ 382462	~~17399~~ 356011	0
precision	~~0.903~~ 0.904	~~0.924~~ 0.926	~~0.905~~ 0.911	0.915	1
recall	~~0.621~~ 0.536	~~0.481~~ 0.400	~~0.580~~ 0.495	~~0.624~~ 0.530	1
accuracy	~~0.810~~ 0.771	~~0.762~~ 0.722	~~0.795~~ 0.756	~~0.814~~ 0.771	1
~~F1-score 0.736 0.632 0.707 0.740 1~~ $\rho_{\text{MWR-LWP,LLT}}$	~~0.436~~ 0.490	~~0.533~~ 0.566	~~0.490~~ 0.515	~~0.489~~ 0.530	~~0.471~~ 0.473
$\rho_{\text{MWR-LWP,LWP}_{\text{ad}}}$	~~0.275~~ 0.348	~~0.471~~ 0.462	~~0.335~~ 0.370	~~0.345~~ 0.387	~~0.399~~ 0.432
$\rho_{\text{ceilo-CBH,LLH}}$	~~0.775~~ 0.915	~~0.725~~ 0.859	~~0.738~~ 0.897	~~0.755~~ 0.905	~~0.913~~ 0.974
Liq-Pxl at RH > 90 %	~~n/a~~ 0.602	~~n/a~~ 0.653	~~n/a~~ 0.626	~~n/a~~ 0.620	~~n/a~~ 0.816

3.4 ~~Statistical results for entire ACCEPT field campaign~~

~~A second, more general evaluation of the cloud droplet (CD) classification results of the ANN using the different mentioned thresholds against the Cloudnet liquid target classification via error matrix and several independent observations is presented in Table 3.~~

~~Overall, evaluation results are similar to the case study presented in Table 2) meaning the presented case study represents the data set well. Looking at the entire data set, all thresholds used in the ANN achieve high precision values > 0.9. Accuracy values are slightly lower than for the case study and range between 0.72 (deBoer2009) and 0.77 (Shupe2007, linear-1). Again, recall values are moderately lower ranging from 0.4 (deBoer2009) to 0.54 (Shupe2007). Multiple other case studies were investigated with the same conclusion for false predictions. Missing spectral signatures and turbulence broadened spectra are the main driver for miss-classifications of the pre-trained Luke et al. (2010) approach.~~

An additional independent validation is done using radiosonde launches from the campaign site as well as launches from DeBilt airport about 30 km away. Liquid-detected pixels are only evaluated in this way within ±30 min of a radiosonde launch, meaning only a small subset of data from the entire field experiment is considered. Radio sounding profiles with RH with respect to liquid water (w.r.t.l.) larger than 90 % overlapping with liquid detection layers occur only during 1.5 h out of 58 h of available liquid detection data, i.e. only during 2.5 % of the time is liquid classified. This validation method thus only has very limited utility for the quality of the thermodynamic phase classifications made, but is shown here for the sake of completeness as similar future evaluation studies might have larger ~~datasets available. However~~data sets available. As shown in the last row of Table 3, for all methods the majority of number of liquid-containing pixels occur when the radiosonde RH w.r.t.l. is larger than 90 % and liquid occurrence is thus likely. There are two explanations why the fraction of Cloudnet-classified liquid pixel overlapping with areas of radiosonde RH > 90 % is much higher (72 %) than for the ANN results (54-61 %). Firstly, with the radiosonde drifting away with height (and time), the assumption of having the same thermodynamic profile over the ACCEPT-site and the sounding location becomes less certain for liquid detections higher in the atmospheric profile (where the ANN is predicting more liquid than Cloudnet). Secondly, ~~the overlap fraction does include false positives(mostly caused by the ANN) but not true negatives.~~ not all elements of the error matrix are represented in the overlap fraction of pixel with liquid-detection and RH > 90 %. While liquid pixels unrecognized by Cloudnet (i.e. beyond lidar attenuation) are not included in the overlap fraction, wrongly detected ANN liquid pixels (i.e. false positives, FP) are included and thus reduce the fraction of overlap pixel for ANN-predicted liquid.

~~Error matrix, performance metrics, and correlation coefficients for ceilometer-CBH vs. LLH, MWR-LWP vs. LLT, MWR-LWP vs. $LWP_{ad,cor}$, for the entire ACCEPT data set. Statistic includes only valid pixel.~~

To understand the performance of the liquid prediction by the ANN more in depth, conditions under which enhanced spectrum width values lead to liquid-prediction error matrix elements TP, FP, and FN are described subsequently. The co-existence of multiple hydrometeor types with sufficiently different fall velocities in the same radar volume leads to multimodal Doppler spectra with a high total spectrum width. If the slow-falling hydrometeors have a low reflectivity and narrow peak width, the ANN likely predicts liquid. If there are indeed small cloud droplets and larger ice crystals in the volume, this results in TP. If however, there is a co-existence of multiple ice crystal types of which one is small and has a small fall velocity, this results in FP. Further, if the enhanced SW is not caused by multiple hydrometeor types but by turbulence, liquid peak signatures can be smeared thus leading to FN. In calm conditions (low turbulence) it is more likely that a bimodal Doppler spectrum with two ice classes is misclassified as one ice- and one liquid class leading to FP. This problem diminishes with increasing turbulence because of broadening of the peaks and smearing of the individual peaks. The latter (smearing) is the same mechanism for FN in high turbulent conditions.

However, considering only spectrum width is not sufficient as it is always a combination of radar reflectivity, mean Doppler velocity, spectrum width etc. that leads to correct or incorrect classification of liquid by the ANN. By discussing relative frequency of occurrence (FoO) plots of radar moments and environmental temperature of the liquid-prediction error matrix elements TP, TN, FP, FN as illustrated in Figure A3 in the Appendix, we assess which combinations of moments mostly lead to TP. As shown in Figure A3, the distribution of radar moments of TP is different from those of TN, FP, and FN while the FoO distribution of radar moments of the latter (TN, FP, FN) are mostly similar. Specifically, the radar reflectivities of TP of cloud droplets is monomodal with a maximum FoO at -25 dbZ to -30 dBZ, while it is bimodal for TN, FP, FN with the two maxima occurring at -25 dBZ and -10 dBZ. The second maximum at -10 dBZ can be attributed to situations in which the ANN predicted cloud droplets in drizzle/rain. With values between -2 $m\,s^{-1}$ and 0.5 $m\,s^{-1}$ the distribution of mean Doppler velocity of TP is narrower than of TN, FP, FN which have V_D values of about -4 $m\,s^{-1}$ and 1 $m\,s^{-1}$ and a maximum FoO at more negative values of around -0.5 $m\,s^{-1}$ than the TP (maximum FoO at -0.2 $m\,s^{-1}$). TP generally occur at larger spectrum width σ than TN, FP, FN with a maximum FoO of TP at 0.2–0.25 $m\,s^{-1}$ while the FoO of TN, FP, FN peaks at 0.05-0.1 $m\,s^{-1}$.

Spherical particles have a theoretical radar linear depolarization ratio (LDR) of minus infinity dB, however, due to technical limitations, the smallest detectable LDR of the MIRA cloud radar is -30 dB which corresponds to the peak of FoO of TP. While FN also peak at -30 dB, TN and FP are characterized by high FoO in the range of -30 dB to -25 dB which again can be attributed to drizzle/rain where perfect sphericity of the hydrometeors is not always given. The FoO distribution of error matrix elements in the environmental temperature space show that only a considerable fraction of TN are detected at very low temperatures which is plausible. Maximum FoO of all four error matrix elements occur at positive temperatures which is caused by the consecutive attenuation of ground-based lidar signal with height leaving more pixel at higher temperature in the Cloudnet-ANN comparison. Comparing the FoO of liquid detection error matrix with respect to the different backscatter- and depolarization thresholds (Shupe2007, deBoer2009, Luke2010, and linear-1~~Cloudnet TP406235 302643 374880 401331 757342 TN919571 938429 925740 925243 962586 FP43015 24157 36846 37343 0 FN351107 454699 382462 356011 0 precision 0.904 0.926 0.911 0.915 1 recall 0.536 0.400 0.495 0.530 1 accuracy 0.771 0.722 0.756 0.771 1 F1-score 0.673 0.558 0.641 0.671 1 $\rho_{\text{MWR-LWP,LLT}}$ 0.490 0.566 0.515 0.530 0.473 $\rho_{\text{MWR-LWP,LWP}_{\text{ad}}}$ 0.348 0.462 0.370 0.387 0.432 $\rho_{\text{ceilo-CBH,LLH}}$ 0.915 0.859 0.897 0.905 0.974 Liq-Pxl at RH > 90 % 0.602 0.653 0.626 0.620 0.816~~), the more stringent criteria of deBoer2009 generally lead to narrower FoO distributions of TP. Summarizing, as shown in the description of FoO of the radar moments of the error matrix components above, TP are mostly characterized by high spectrum width in combination with low absolute values of V_D and small radar reflectivities but due to the overlap of radar moments of all error matrix elements, the same combination of Z_e, V_D, and σ can be caused by TP, TN, FP, FN.

4 Summary and Outlook

The current study shows that synergistic observations of depolarization lidar and cloud Doppler radar in conjunction with machine learning techniques can be used to detect liquid beyond full lidar signal attenuation. ~~It was shown that this~~ This approach performs well in stratiform ~~(low-turbulent)~~ cloud situations but is not suited for ~~strongly convective~~ situations in which the imprint of different hydrometeor populations in the same cloud volume on the cloud radar Doppler spectrum is masked, e.g. by turbulent spectrum broadening. We demonstrated that the ANN of Luke et al. (2010) pre-trained with the MPACE data set in Alaska could successfully be applied to the ACCEPT data set obtained in Cabauw, the Netherlands and is able to improve the Cloudnet target classification for stratiform optically thick liquid-layers or situations in which multiple liquid layers exist. We applied different published lidar-based liquid-detection thresholds to the predicted lidar backscatter coefficients and depolarization lidars - all were found to perform better in some situations than others and could be seen as either to stringent (deBoer2009) missing thinner liquid layers or ~~to~~ too broad (Shupe2007, Luke2010, "linear-1") leading to ~~misclassification~~ misclassifications of ice as liquid. No suggestion on best thresholds can thus be made. To overcome limitations due to ambiguities caused by thresholding, focus should therefore be put on the development of techniques which do not rely on explicit lidar thresholds for liquid detection. This could be realizable by applying novel convolutional artificial neural networks which could be used to exploit the full information content of high-resolution cloud radar Doppler spectra. Additionally,

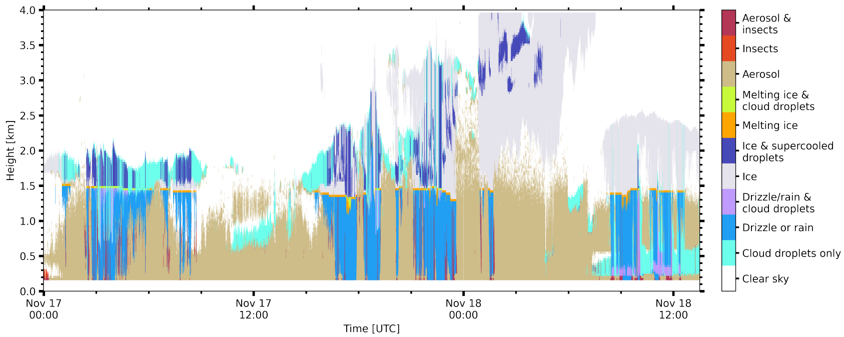

Figure A1. Zoom of Cloudnet target classification from 0–4 km altitude for Nov 17-18, 2014 case study in Cabauw, Netherlands.

radar Doppler spectra peak-separation techniques such as PEAKO (Kalesse et al., 2019) and peakTree (Radenz et al., 2019) are helpful assess the possibilities of liquid occurrence.

Furthermore, two recent studies also showed the benefit distinguishing between cloud-top liquid-bearing layers and embedded liquid layers when assessing the performance of liquid-detection retrievals (Silber et al., 2020) and (Kalogeras et al., 2021). Silber et al. (2020) retrieved cloud thermodynamic phase of Arctic clouds based on one year zenith-pointing Ka-band radar and HSRL observations. They found that cloud-top liquid-bearing samples can be more reliably detected than embedded liquid layers as the latter are more difficult to separate from falling ice signatures in the PDF of the first three radar moments as well as Doppler spectra left slope and right slope. Kalogeras et al. (2021) developed a Ka-band radar-only, moment-based technique for supercooled liquid water detection in Arctic mixed-phase clouds. The novelty of this method is that it is a neighborhood-dependent algorithm employing gradients of moments. They concluded that best skill levels for liquid detection are realized for combinations of spectral width and reflectivity vertical gradient and also found their algorithm to be most reliable when applied to cloud tops.

The identification of the presence of liquid layers in the entire vertical column of optically thick or ~~multilayered~~ multi-layered cloud situations is a first step to get a better understanding of which microphysical particle growth processes might occur in a mixed-phase cloud. The shown results will therefore be used in follow-up studies for characterization of microphysical hydrometeor growth processes.

Data availability. Cloudnet-processed data for the ACCEPT campaign are available via https://cloudnet.fmi.fi. The Mira-35 moment data as well as compressed (noise-removed) Doppler spectra are available upon request from Patric Seifert (seifert@tropos.de).

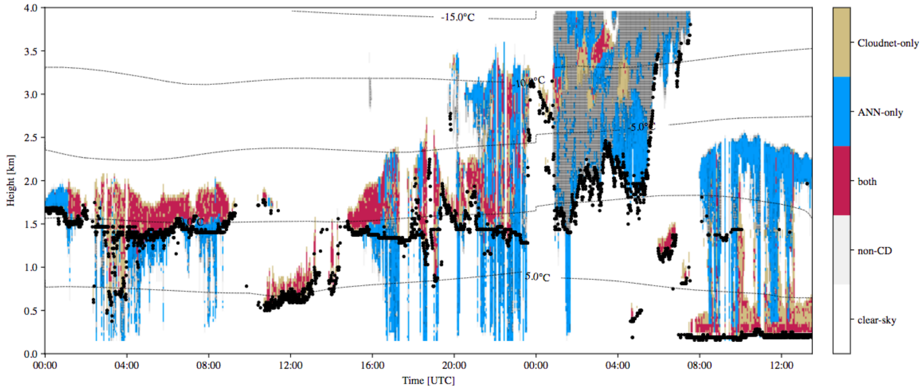

Figure A2. Zoom of comparison of cloud droplet detection of Cloudnet and ANN (using linear-1 thresholds) from 0–4 km altitude for Nov 17-18, 2014 case study in Cabauw, Netherlands. Black dots indicate ceilometer first cloud base height.

Appendix A

Author contributions. Heike Kalesse-Los and Willi Schimmel, did the data analysis and prepared the manuscript. Patric Seifert was PI of the ACCEPT field experiment and helped in preparing the manuscript. Edward Luke did the ANN simulations.

Competing interests. No competing interests are present.

Acknowledgements. Most of the work H. Kalesse-Los conducted for this study was within the framework of the DFG project COMPoSE, GZ: KA 4162/1-1. Willi Schimmel was funded through the ESF-Project ~~100339509. Das Vorhaben~~ 100339509 in which H. Kalesse-Los is a Principal Investigator. This ESF-Project "Bodengebundene Fernerkundung der Atmosphäre zur Verbesserung der Charakterisierung mikrophysikalischer Wolkeneigenschaften sowie der Leistungsprognose erneuerbarer Energien FKZ: 100339509" wird im Rahmen des Programms "Vorhaben in den Bereichen Hochschule und Forschung" vom Freistaat Sachsen und dem Europäischen Sozialfonds gefördert. This ~~project~~ study has also received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No ~~654109.~~ 654109 and from the European Union Seventh Framework Programme (FP7/2007–2013): People, ITN Marie Curie Actions Programme (2012–2016), in the frame of ITaRS under grant agreement 289923.

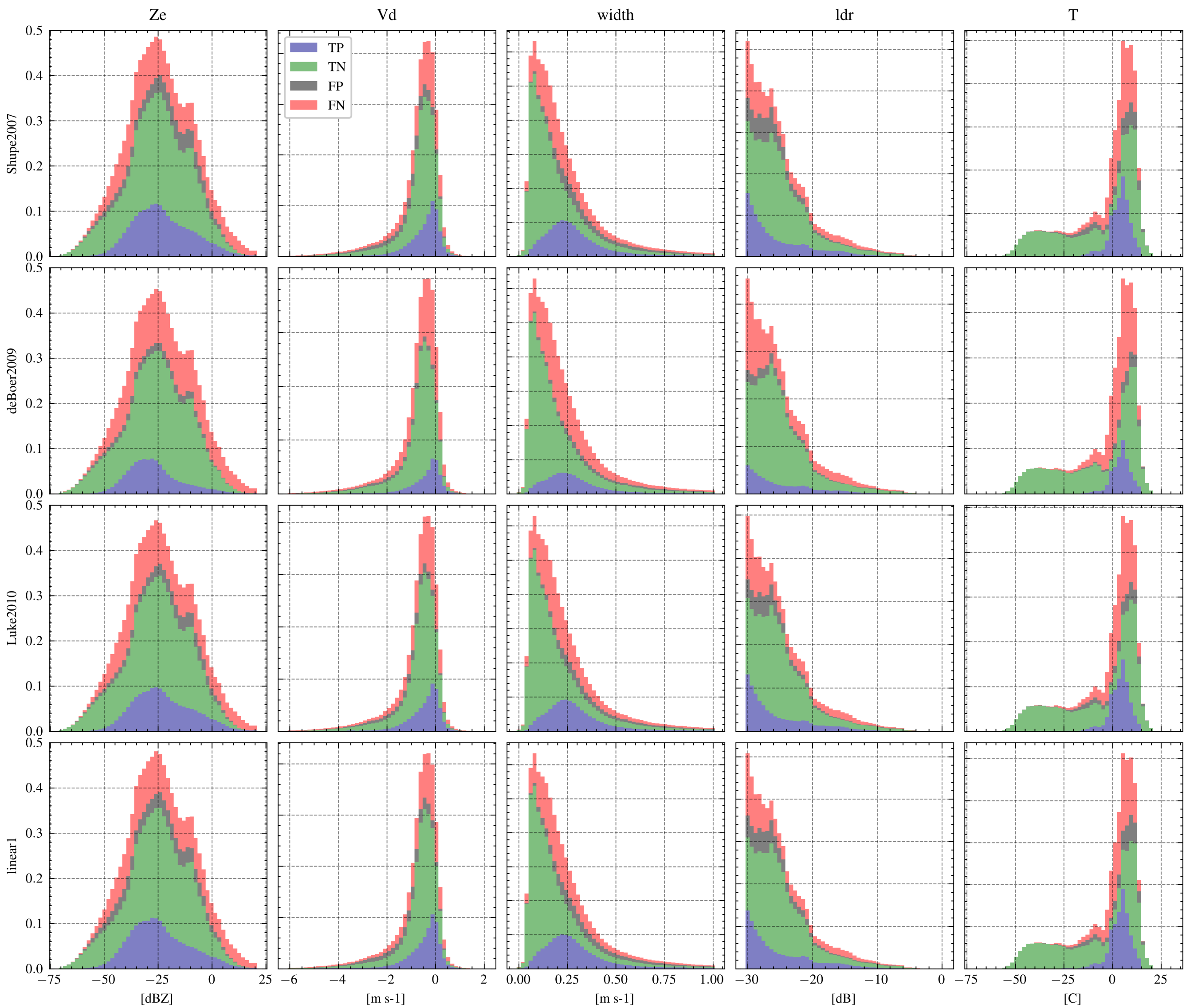

Figure A3. Relative frequency of occurrence plots of radar moments reflectivity (Ze, left column), mean Doppler velocity (Vd, second left column), spectrum width (middle column), linear depolarization ratio (LDR, second to right column), environmental temperature T (right column) of ANN-liquid-prediction error matrix elements TP (blue), TN (green), FP (grey), and FN (red) for the four utilized ANN-lidar variable thresholds of Shupe2007 (first row), deBoer2009 (second row), Luke2010 (third row), and linear-1 threshold (fourth row) for the entire ACCEPT field experiment.

References

Andersen, H., Cermak, J., Fuchs, J., Knutti, R., and Lohmann, U.: Understanding the drivers of marine liquid-water cloud occurrence and properties with global observations using neural networks, Atmospheric Chemistry and Physics, 17, 9535–9546, https://doi.org/10.5194/acp-17-9535-2017, https://www.atmos-chem-phys.net/17/9535/2017/, 2017.

Büehl, J., Seifert, P., Wandinger, U., Baars, H., Kanitz, T., Schmidt, J., Myagkov, A., Engelmann, R., Skupin, A., Heese, B., Klepel, A., Althausen, D., and Ansmann, A.: LACROS: the Leipzig Aerosol and Cloud Remote Observations System, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 8890, https://doi.org/doi:10.1117/12.2030911, 2013.

Cotton, W. R. and Anthes, R. A.: The mesoscale structure of extratopical cyclones and middle and high clouds. Storm and Cloud Dynamics,, International Geophysical Series, 44, 745—-787, 1989.

de Boer, G., Eloranta, E. W., and Shupe, M. D.: Arctic Mixed-Phase Stratiform Cloud Properties from Multiple Years of Surface-Based Measurements at Two High-Latitude Locations, Journal of the Atmospheric Sciences, 66, 2874–2887, https://doi.org/10.1175/2009JAS3029.1, https://doi.org/10.1175/2009JAS3029.1, 2009.

Eloranta, E. E.: High spectral resolution lidar, in: Lidar, pp. 143–163, Springer, 2005.

Engelmann, R., Kanitz, T., Baars, H., Heese, B., Althausen, D., Skupin, A., Wandinger, U., Komppula, M., Stachlewska, I. S., Amiridis, V., Marinou, E., Mattis, I., Linné, H., and Ansmann, A.: The automated multiwavelength Raman polarization and water-vapor lidar PollyXT: the neXT generation, Atmospheric Measurement Techniques, 9, 1767–1784, https://doi.org/10.5194/amt-9-1767-2016, https://doi.org/10.5194/amt-9-1767-2016, 2016.

Gardner, M. and Dorling, S.: Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, Atmospheric Environment, 32, 2627 – 2636, https://doi.org/https://doi.org/10.1016/S1352-2310(97)00447-0, http://www.sciencedirect.com/science/article/pii/S1352231097004470, 1998.

Garrett, J. D. and Peng, H.-H.: garrettj403/SciencePlots, https://doi.org/10.5281/zenodo.4106649, http://doi.org/10.5281/zenodo.4106649, 2021.

Görsdorf, U., Lehmann, V., Bauer-Pfundstein, M., Peters, G., Vavriv, D., Vinogradov, V., and Volkov, V.: A 35-GHz Polarimetric Doppler Radar for Long-Term Observations of Cloud Parameters—Description of System and Data Processing, Journal of Atmospheric and Oceanic Technology, 32, 675–690, https://doi.org/10.1175/JTECH-D-14-00066.1, https://doi.org/10.1175/JTECH-D-14-00066.1, 2015.

Griesche, H. J., Seifert, P., Ansmann, A., Baars, H., Velasco, C. B., Bühl, J., Engelmann, R., Radenz, M., Zhenping, Y., and Macke, A.: Application of the shipborne remote sensing supersite OCEANET for profiling of Arctic aerosols and clouds during Polarstern cruise PS106, Atmospheric Measurement Techniques, 13, 5335–5358, https://doi.org/10.5194/amt-13-5335-2020, https://doi.org/10.5194/amt-13-5335-2020, 2020.

Hobbs, P. V. and Rangno, A. L.: Microstructures of low and middle-level clouds over the Beaufort Sea, Quarterly Journal of the Royal Meteorological Society, 124, 2035–2071, https://doi.org/10.1002/qj.49712455012, https://doi.org/10.1002/qj.49712455012, 1998.

Hogan, R. and O'Connor, E.: Facilitating cloud radar and lidar algorithms: The Cloudnet Instrument Synergy/Target Categorization product, J. Atmos. Sci, Availableonlineatwww.cloud-net.org/data/products/categorize.html, 2006.

Hu, Y., Liu, Z., Winker, D., Vaughan, M., Noel, V., Bissonnette, L., Roy, G., and McGill, M.: Simple relation between lidar multiple scattering and depolarization for water clouds, Optics Letters, 31, 1809, https://doi.org/10.1364/ol.31.001809, https://doi.org/10.1364/ol.31.001809, 2006.

Illingworth, A. J., e. a.: CLOUDNET: Continuous Evaluation of Cloud Profiles in Seven Operational Models Using Ground-Based Observations, Bull. Amer. Meteor. Soc., 88, https://doi.org/DOI:10.1175/BAMS-88-6-883, 2007.

Kalesse, H., de Boer, G., Solomon, A., Oue, M., Ahlgrimm, M., Zhang, D., Shupe, M. D., Luke, E., and Protat, A.: Understanding Rapid Changes in Phase Partitioning between Cloud Liquid and Ice in Stratiform Mixed-Phase Clouds: An Arctic Case Study, Monthly Weather Review, 144, 4805–4826, https://doi.org/10.1175/MWR-D-16-0155.1, https://doi.org/10.1175/MWR-D-16-0155.1, 2016a.

Kalesse, H., Szyrmer, W., Kneifel, S., Kollias, P., and Luke, E.: Fingerprints of a riming event on cloud radar Doppler spectra: observations and modeling, Atmospheric Chemistry and Physics, 16, 2997–3012, https://doi.org/10.5194/acp-16-2997-2016, https://doi.org/10.5194/acp-16-2997-2016, 2016b.

Kalesse, H., Vogl, T., Paduraru, C., and Luke, E.: Development and validation of a supervised machine learning radar Doppler spectra peak-finding algorithm, 12, 4591–4617, https://doi.org/10.5194/amt-12-4591-2019, https://doi.org/10.5194/amt-12-4591-2019, 2019.

Kalogeras, P., Battaglia, A., and Kollias, P.: Supercooled Liquid Water Detection Capabilities from Ka-Band Doppler Profiling Radars: Moment-Based Algorithm Formulation and Assessment, 13, 2891, https://doi.org/10.3390/rs13152891, https://doi.org/10.3390/rs13152891, 2021.

Karstens, U., Simmer, C., and Ruprecht, E.: Remote sensing of cloud liquid water, Meteorology and Atmospheric Physics, 54, 157–171, https://doi.org/10.1007/BF01030057, http://dx.doi.org/10.1007/BF01030057, 1994.

Kollias, P., Miller, M. A., Luke, E. P., Johnson, K. L., Clothiaux, E. E., Moran, K. P., Widener, K. B., and Albrecht, B. A.: The Atmospheric Radiation Measurement Program cloud profiling radars: Second-generation sampling strategies, processing, and cloud data products, J. Atmos. Oceanic Technol., 24, 1199–1214, 2007.

Kollias, P., Clothiaux, E. E., Ackerman, T. P., Albrecht, B. A., Widener, K. B., Moran, K. P., Luke, E. P., Johnson, K. L., Bharadwaj, N., Mead, J. B., Miller, M. A., Verlinde, J., Marchand, R. T., and Mace, G. G.: Development and Applications of ARM Millimeter-Wavelength Cloud Radars, Meteorological Monographs, 57, 17.1–17.19, https://doi.org/10.1175/amsmonographs-d-15-0037.1, https://doi.org/10.1175/amsmonographs-d-15-0037.1, 2016.

Komurcu, M., Storelvmo, T., Tan, I., Lohmann, U., Yun, Y., Penner, J. E., Wang, Y., Liu, X., and Takemura, T.: Intercomparison of the cloud water phase among global climate models, Journal of Geophysical Research: Atmospheres, 119, 3372–3400, https://doi.org/10.1002/2013jd021119, https://doi.org/10.1002/2013jd021119, 2014.

Lamer, K., Kollias, P., and Nuijens, L.: Observations of the variability of shallow trade wind cumulus cloudiness and mass flux, Journal of Geophysical Research: Atmospheres, 120, 6161–6178, https://doi.org/10.1002/2014jd022950, https://doi.org/10.1002/2014jd022950, 2015.

Liljegren, M. P. C. . D. D. T. . J. C.: A Neural Network for Real-Time Retrievals of PWV and LWP From Arctic Millimeter-Wave Ground-Based Observations, IEEE Transactions on Geoscience and Remote Sensing, 47, 1887 – 1900, 2009.

Luke, E. P., Kollias, P., and Shupe, M. D.: Detection of supercooled liquid in mixed-phase clouds using radar Doppler spectra, Journal of Geophysical Research (Atmospheres), 115, D19201, https://doi.org/10.1029/2009JD012884, 2010.

Morrison, H., de Boer, G., Feingold, G., Harrington, J., Shupe, M., and Sulia, K.: Resillience of persistent Arctic mixed-phase clouds, Nature Geosci., 5, 11–17, 2012.

Myagkov, A., Seifert, P., Bauer-Pfundstein, M., and Wandinger, U.: Cloud radar with hybrid mode towards estimation of shape and orientation of ice crystals, Atmospheric Measurement Techniques, 9, 469–489, https://doi.org/10.5194/amt-9-469-2016, https://www.atmos-meas-tech.net/9/469/2016/, 2016a.

Myagkov, A., Seifert, P., Wandinger, U., Bühl, J., and Engelmann, R.: Relationship between temperature and apparent shape of pristine ice crystals derived from polarimetric cloud radar observations during the ACCEPT campaign, Atmospheric Measurement Techniques, 9, 3739–3754, https://doi.org/10.5194/amt-9-3739-2016, https://www.atmos-meas-tech.net/9/3739/2016/, 2016b.

O'Connor, E. J., Hogan, R. J., and Illingworth, A. J.: Retrieving Stratocumulus Drizzle Parameters Using Doppler Radar and Lidar, Journal of Applied Meteorology, 44, 14–27, https://doi.org/10.1175/jam-2181.1, https://doi.org/10.1175/jam-2181.1, 2005.

Pfitzenmaier, L., Dufournet, Y., Unal, C. M. H., and Russchenberg, H. W. J.: Retrieving Fall Streaks within Cloud Systems Using Doppler Radar, Journal of Atmospheric and Oceanic Technology, 34, 905–920, https://doi.org/10.1175/JTECH-D-16-0117.1, https://doi.org/10.1175/JTECH-D-16-0117.1, 2017.

Radenz, M., Bühl, J., Seifert, P., Griesche, H., and Engelmann, R.: peakTree: a framework for structure-preserving radar Doppler spectra analysis, 12, 4813–4828, https://doi.org/10.5194/amt-12-4813-2019, https://doi.org/10.5194/amt-12-4813-2019, 2019.

Ramelli, F., Henneberger, J., David, R. O., Bühl, J., Radenz, M., Seifert, P., Wieder, J., Lauber, A., Pasquier, J. T., Engelmann, R., Mignani, C., Hervo, M., and Lohmann, U.: Microphysical investigation of the seeder and feeder region of an Alpine mixed-phase cloud, 21, 6681–6706, https://doi.org/10.5194/acp-21-6681-2021, https://doi.org/10.5194/acp-21-6681-2021, 2021.

Rose, T., Crewell, S., Löhnert, U., and Simmer, C.: A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere, Atmospheric Research, 75, 183–200, https://doi.org/10.1016/j.atmosres.2004.12.005, https://doi.org/10.1016/j.atmosres.2004.12.005, 2005.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, Nature, 323, 533 EP –, http://dx.doi.org/10.1038/323533a0, 1986.

Sassen, K.: The Polarization Lidar Technique for Cloud Research: A Review and Current Assessment, Bulletin of the American Meteorological Society, 72, 1848–1866, https://doi.org/10.1175/1520-0477(1991)072<1848:tpltfc>2.0.co;2, https://doi.org/10.1175/1520-0477(1991)072<1848:tpltfc>2.0.co;2, 1991.

Seifert, P., Ansmann, A., Mattis, I., Wandinger, U., Tesche, M., Engelmann, R., Müller, D., Pérez, C., and Haustein, K.: Saharan dust and heterogeneous ice formation: Eleven years of cloud observations at a central European EARLINET site, Journal of Geophysical Research, 115, https://doi.org/10.1029/2009jd013222, https://doi.org/10.1029/2009jd013222, 2010.

Shupe, M.: A ground-based multiple remote-sensor cloud phase classifier, Geophys. Res. Lett., 34, 2007.

Shupe, M., Kollias, P., Matrosov, S., and Schneider, T.: Deriving mixed-phase cloud properties from Doppler radar spectra, J. Atmos. Ocean. Tech., 21, 660–670, 2004.

Shupe, M. D., Daniel, J. S., de Boer, G., Eloranta, E. W., Kollias, P., Long, C. N., Luke, E. P., Turner, D. D., and Verlinde, J.: A Focus On Mixed-Phase Clouds, Bulletin of the American Meteorological Society, 89, 1549–1562, https://doi.org/10.1175/2008BAMS2378.1, https://doi.org/10.1175/2008BAMS2378.1, 2008.

Silber, I., Verlinde, J., Wen, G., and Eloranta, E. W.: Can Embedded Liquid Cloud Layer Volumes Be Classified in Polar Clouds Using a Single- Frequency Zenith-Pointing Radar?, IEEE Geoscience and Remote Sensing Letters, 17, 222–226, https://doi.org/10.1109/lgrs.2019.2918727, https://doi.org/10.1109/lgrs.2019.2918727, 2020.

Strandgren, J., Bugliaro, L., Sehnke, F., and Schröder, L.: Cirrus cloud retrieval with MSG/SEVIRI using artificial neural networks, Atmospheric Measurement Techniques, 10, 3547–3573, https://doi.org/10.5194/amt-10-3547-2017, https://www.atmos-meas-tech.net/10/3547/2017/, 2017a.

Strandgren, J., Fricker, J., and Bugliaro, L.: Characterisation of the artificial neural network CiPS for cirrus cloud remote sensing with MSG/SEVIRI, Atmospheric Measurement Techniques, 10, 4317–4339, https://doi.org/10.5194/amt-10-4317-2017, https://www.atmos-meas-tech.net/10/4317/2017/, 2017b.

Sun, Z. and Shine, K. P.: Studies of the radiative properties of ice and mixed-phase clouds, Q. J. R. Meteorol. Soc., 120, 111–137, 1994.

Tan, I., Storelvmo, T., and Zelinka, M. D.: Observational constraints on mixed-phase clouds imply higher climate sensitivity, Science, 352, 224–227, https://doi.org/10.1126/science.aad5300, http://science.sciencemag.org/content/352/6282/224, 2016.

Vassel, M., Ickes, L., Maturilli, M., and Hoose, C.: Classification of Arctic multilayer clouds using radiosonde and radar data in Svalbard, Atmospheric Chemistry and Physics, 19, 5111–5126, https://doi.org/10.5194/acp-19-5111-2019, https://doi.org/10.5194/acp-19-5111-2019, 2019.

Verlinde, J., Harrington, J. Y., Yannuzzi, V. T., Avramov, A., Greenberg, S., Richardson, S. J., Bahrmann, C. P., McFarquhar, G. M., Zhang, G., Johnson, N., Poellot, M. R., Mather, J. H., Turner, D. D., Eloranta, E. W., Tobin, D. C., Holz, R., Zak, B. D., Ivey, M. D., Prenni, A. J., DeMott, P. J., Daniel, J. S., Kok, G. L., Sassen, K., Spangenberg, D., Minnis, P., Tooman, T. P., Shupe, M., Heymsfield, A. J., and Schofield, R.: The mixed-phase arctic cloud experiment, Bull. Amer. Meteor. Soc., 88, 205–221, https://doi.org/http://dx.doi.org/10.1175/BAMS-88-2-205, 2007.

Verlinde, J., Rambukkange, M. P., Clothiaux, E. E., McFarquhar, G. M., and Eloranta, E. W.: Arctic multilayered, mixed-phase cloud processes revealed in millimeter-wave cloud radar Doppler spectra, J. Geophys. Res., 118, 13 199–13 213, https://doi.org/10.1002/2013JD020183, 2013.

Westbrook, C. D., Illingworth, A. J., O'Connor, E. J., and Hogan, R. J.: Doppler lidar measurements of oriented planar ice crystals falling from supercooled and glaciated layer clouds, Quarterly Journal of the Royal Meteorological Society, 136, 260–276, https://doi.org/10.1002/qj.528, https://doi.org/10.1002/qj.528, 2010.

Williams, C. R., Maahn, M., Hardin, J. C., and de Boer, G.: Clutter mitigation, multiple peaks, and high-order spectral moments in 35GHz vertically pointing radar velocity spectra, Atmospheric Measurement Techniques, 11, 4963–4980, https://doi.org/10.5194/amt-11-4963-2018, https://doi.org/10.5194/amt-11-4963-2018, 2018.

Yong-Sang, C., Chang-Hoi, H., Chang-Eui, P., Trude, S., and Ivy, T.: Influence of cloud phase composition on climate feedbacks, Journal of Geophysical Research: Atmospheres, 119, 3687–3700, https://doi.org/10.1002/2013JD020582, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JD020582, 2014.