Detection of supercooled liquid water clouds with ceilometers: Development and evaluation of deterministic and data-driven retrievals

5 Adrien Guyot[1], Alain Protat[1], Simon P. Alexander[2], Andrew R. Klekociuk[2], Peter Kuma[3], Adrian McDonald[4]

[1] Australian Bureau of Meteorology, Melbourne, Victoria, Australia

[2] Australian Antarctic Division, Kingston, Tasmania, Australia

[3] Department of Meteorology, Stockholm University, Stockholm, Sweden

10 [4] University of Canterbury, Christchurch, New Zealand

Correspondence to: Adrien Guyot (adrien.guyot@bom.gov.au)

Abstract. Cloud and aerosol lidars measuring backscatter and depolarization ratio are most suitable instruments to detect cloud phase (liquid, ice, or mixed phase). However, such instruments are not widely deployed as part of operational networks. In

15 this study, we propose a new algorithm to detect supercooled liquid water clouds based solely on ceilometers measuring only co-polarisation backscatter. We utilise observations collected at Davis, Antarctica, where low-level, mixed phase clouds, including supercooled liquid water (SLW) droplets and ice crystals remain poorly understood, due to the paucity of ground-based observations. A 3-month set of observations were collected during the austral summer of November 2018 – February 2019, with a variety of instruments including a depolarization lidar and a W-Band cloud radar which were used to build a 2-

20 dimensional cloud phase mask distinguishing SLW and mixed phase clouds. This cloud phase mask is used as the reference to develop a new algorithm based on the observations of a single polarisation ceilometer operating in the vicinity for the same period. Deterministic and data-driven retrieval approaches were evaluated: an extreme gradient boosting (XGBoost) framework ingesting backscatter average characteristics was the most effective method at reproducing the classification obtained with the combined radar-lidar approach with an accuracy as high as 0.91. This study provides a new SLW retrieval

25 approach based solely on ceilometer data and highlights the considerable benefits of these instruments to provide intelligence on cloud phase in polar regions that usually suffer from a paucity of observations.

1 Introduction

Mixed-phase clouds play a critical role in the earth radiation budget, through their complex interactions with incoming and outgoing shortwave and longwave radiation. This effect is particularly important at higher latitudes with variation in radiation affecting the snow or ice mass balance in the polar regions (Lawson and Gettelman, 2014). Despite their importance in the global climate system, the occurrence, amount, and nature of mixed-phase clouds remain poorly simulated in global climate models due to the paucity of reliable mixed-phase clouds observations, especially in remote regions of the globe such as Antarctica (Bodas-Salcedo et al., 2016; Hyder et al., 2018). Until recently, global climate models assumed that low-level clouds over the Antarctic Ice sheet essentially contained ice crystals, but Lawson and Gettelman (2014), and later Ricaud et al. (2020) both showed from their observations that around 50% of clouds contained supercooled liquid water (SLW) during the austral summer. Satellite-based lidar observations of mixed-phase clouds suffer from severe biases (Hu et al. 2009; Mace et al., 2020; Mace et al., 2021; McErlich et al, 2021) but points towards a high-frequency occurrence of mixed-phase clouds in Southern latitudes. Lawson and Gettelman (2014) and Ricaud et al. (2020) highlighted the significant impact the increased proportion of SLW clouds had on climate model simulations. The parameterization of cloud microphysical processes and precipitation remains challenging in Antarctica, given the limited observations; recent work (Sotiropoulou et al., 2021; Vignon et al., 2021) focused on improving the parameterisation of SLW, showing how simulations and observations can be combined to improve our understanding of underlying processes leading to its formation. Kay et al. (2016) and Frey et al. (2018) also highlighted the importance of Southern Ocean mixed-phase clouds in global coupled climate models, under the predicted increase of greenhouse gases concentrations.

Depolarisation lidar is the most reliable means of observing non-spherical shape for randomly oriented cloud particles (Mishchenko et al., 2000; Hu et al. 2009; Mace et al. 2020). Typically, a depolarization ratio below 10% is characteristic of SLW clouds, while higher values are produced by ice particles. However, the remoteness and year-round harsh conditions for operating ground or aircraft operations have limited the frequency of cloud observation campaigns in Antarctica. Only over the last decade have coordinated ground-based cloud and precipitation studies been conducted in various regions of Antarctica, including in Adelie Land (Grazioli et al., 2017; Genthon et al., 2018), Dronning Maud Land (Gorodetskaya et al., 2015), Ross Island (Scott and Lubin, 2014; Zhang et al., 2019), the South Pole (Lawson and Gettelman, 2014), the Antarctic Peninsula and Larsen Ice Shelf (Grosvenor et al., 2012; Lachlan-Cope et al., 2016) and East Antarctica (Alexander et al., 2021, Gehring et al., 2022). Complementary to ground-based observations, satellite-borne remote sensing capabilities including depolarisation lidar and cloud radar can be combined to generate cloud phase products in remote regions such as the Arctic and the Antarctic, capitalising on polar orbit satellites revolutions, with high frequency flights over the poles (Litowski et al., 2019, 2020). Although these new active remote sensing satellites enable generation of cloud phase products covering large areas, the

Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

60 drawbacks are reduced temporal and spatial resolutions as compared to ground-based remote sensing capabilities such as lidars and radars, and strong extinction of the lidar backscatter, making lower layers closer to the ground not observable from space if highly attenuating layers are present above. Satellite observations also suffer from the effect of ground clutter for observations closer to the ground (Bennartz et al., 2019). These recent cloud observations campaigns all pointed towards a higher than anticipated occurrence of SLW and mixed phase clouds (Scott and Lubin (2016); Ricaud et al., 2020; Zhang et al., 2019; Alexander et al., 2021; McErlich et al., 2021; Cossich et al., 2021).

65

Extensive observation campaigns require the deployment of dedicated instruments to determine the cloud phase, including usually, at least, a depolarization lidar. Hogan and Illingworth (1999) proposed to detect supercooled liquid water with ceilometers despite the absence of depolarisation data. Ceilometers are widely used and deployed by national weather services, typically at airports, to provide information on cloud cover and cloud base height. The manufacturers of ceilometers directly 70 provide the cloud base height and cloud cover, using proprietary algorithms. These variables have been derived from the attenuated backscatter profile measured by the ceilometer. In an operational context, the attenuated backscatter profile is generally not used, even though it contains valuable information on the structure of the atmospheric boundary layer and the thermodynamics of the cloud phase (Hogan et al., 2003, 2004; Morille et al., 2007; Munkel et al., 2007; Van Tricht et al., 2014), as well as the presence of aerosols. Following the initial work from Hogan and Illingworth (1999), further studies led 75 to the development and deployment of new detection algorithms for liquid cloud base layers and SLW (O'Connor et al., 2004), as part of the Cloudnet initiative (Illingworth et al., 2007). Recently, Tuononen et al. (2019) (thereafter referred to as T19) proposed an improved approach from the Cloudnet retrieval (Illingworth et al., 2007), utilising the shape of the attenuated backscattered profile instead of relying on finding the first value of backscatter above a given threshold value (Illingworth et al., 2007). Hamalainen et al. (2020) further applied the T19 approach combined with vertical meteorological profiles of 80 temperature to build a hydrometeor classification scheme, to detect supercooled liquid water in the clouds. These recent studies were the starting point and the motivation for the present paper: to evaluate if the T19 approach could be successfully applied to ceilometer observations from Antarctica.

The aims of this study were to: (i) utilise high-resolution observations of cloud phase combining a set of ground-based 85 instruments including a depolarization lidar and a cloud radar, to better understand cloud processes and microphysics; (ii) to evaluate the T19 supercooled liquid water retrieval for a ceilometer dataset collected at Davis, Antarctica; (iii) to develop, train, and test a new enhanced algorithm to retrieve supercooled liquid water, using only the attenuated backscattered signal measured by a ceilometer.

Atmospheric
Measurement
Techniques
Discussions

2 Methods

2.1. Data

2.1.1 The PLATO data

As part of the Australian Antarctic Division's Precipitation over Land and The Southern Ocean (PLATO) field campaign, which operated during the Year of Polar Prediction (YOPP, Bromwich et al., 2020), a suite of ground-based remote sensing instruments was deployed at Davis (68.5762 ºS, 77.9696 ºE), one of the three permanent Australian Antarctic bases on the continent (Gehring et al., 2022), during the southern hemisphere summer 2018/2019. These included a W-Band radar (Delanoë et al. 2016), a 355 nm depolarisation lidar operating for 3 months, and a ceilometer operating during a full year from November 2018 to October 2019. This set of instrumentation provides a unique opportunity to understand the physics of clouds over that region, improve existing or develop new cloud phase retrieval algorithms, and test a variety of models. All these instruments are non-scanning devices, therefore only able to perform acquisitions along a single path.

During the early 2010s, the World Meteorological Organisation Weather and Research Program initiated a 10-year collaborative research project (2013-2022), the Polar Prediction Project (PPP), aiming at improving the prediction of weather for polar regions, at short (hourly) to longer (seasonal) timescales (Jung et al., 2016). The pinnacle of the PPP are intense observational, modelling, and other related activities conducted under the umbrella of the Year of Polar Prediction (YOPP). Three Special Observing Periods (SOPs) occurred between mid-2017 and mid-2019, and included deployment of dedicated instrumentation on the ground, together with higher frequency routine observations (Bromwich et al., 2020).

The cloud radar, namely the Bistatic Radar System for Atmospheric Studies (BASTA) was initially developed within a research laboratory and further became a semi-operational instrument (Delanoë et al., 2016). The radar has a sensitivity of around – 50 dBZ at 1 km and a vertical resolution of 25 m. The BASTA was deployed on a dedicated concrete slab, and oriented to point vertically with an accuracy better than 0.1°, as shown in Figure 1b.

< Figure 1 here >

The Leosphere R-MAN510 cloud and aerosol depolarisation lidar (Royer et al., 2014) consists of elastic transmission and reception (parallel and perpendicular) at 355 nm and inelastic (Raman) reception at 387 nm. The lidar was operated with a 4-degree off zenith angle to avoid ambiguity between SLW clouds and oriented ice plates (Hogan and Illingworth, 2003). We did not consider the Raman channel further in this analysis due to persistent daylight during its Davis deployment rendering the weak Raman return signal unusable (Alexander and Protat, 2019). Raw lidar data were processed to provide vertically

120 resolved profiles of cloud phase, which broadly followed the algorithms developed by Alexander and Protat (2018) and Noh et al. (2019). The raw 355nm lidar backscatter profiles were first processed to remove background noise and correct for beam overlap. We calibrated the lidar following the method of O'Connor et al. (2004), who demonstrated that the lidar ratio is constant within optically thick liquid non-precipitating stratocumulus clouds. We scaled the raw signal until the observed lidar ratio matched the theoretical lidar ratio within these clouds. Then, the calibration values obtained during stratocumulus for the

125 limited number of optically thick clouds present above Davis were used to calibrate the three months of data collected during the summer.

Following calibration, we used a speckle removal technique to flag spurious noise which is ubiquitous at high altitudes in both the parallel and perpendicular channels. We followed the method of Alexander et al. (2021) who used a first pass of the algorithm to extract bright clouds (with large vertical gradient in backscatter) in the co-polarised channel, and then assigned

130 cloud phase based upon the layer-averaged backscatter and depolarisation (Hu et al., 2010). We isolated additional hydrometeors and aerosols based on pixels which had depolarisation ratios exceeding molecular backscatter and variances within empirically determined thresholds. A region-of-interest analysis to extract conjoined regions removed any spurious pixels initially flagged as hydrometeors or aerosols. The result of these steps was a much greater detection of ice virga than only using the parallel backscatter because thin ice virga has large depolarisation ratios, making them readily detectable in the

135 perpendicular channel. It also allowed attribution of liquid precipitation reaching the surface, because this second stage of the algorithm didn't require vertical gradients of backscatter to determine the presence of hydrometeors.

2.1.2 Vaisala CL51 ceilometer observations

In operational settings, ceilometers usually report cloud base heights and oktas (percentage of cloud cover over a given area) without providing information on cloud phase. However, the instruments operating in the near infrared spectrum record the

140 full backscattered profile from which the cloud base and okta have been derived. In this study, raw data was collected with the University of Canterbury Vaisala CL51, e.g., the full backscattered profile with a range of 15 km, a vertical resolution of 10 m and a time resolution of 15 s. These observations covered the PLATO period (November 2018 to February 2019) for which the depolarization lidar and the W-Band radar operated and extended till October 2019. The data was pre-processed using a dedicated software developed by Kuma et al. (2021), namely the Automatic Lidar and Ceilometer Framework (ALCF). This

145 software allows the processing of raw data from a variety of lidars and serves as a platform for comparing observations and models. Here, we used the version 1.1 of the software for processing the raw data generated by the Vaisala CL51 to: (1) produce daily netCDF files from the hourly Vaisala file format; (2) remove noise by applying a noise removal algorithm and subsampling the data to 5 min, 50 bins; (3) calibrate the attenuated backscatter using the approach of Hopkin et al. (2019). The

final pre-processed products were daily netCDF files including the total attenuated volume backscattering coefficient (ß, m^{-1} sr^{-1}) at a resolution of 5 min and bin vertical resolution of 50 m.

2.1.3 ECMWF ERA5

The latest-generation reanalysis product ERA5 from the European Centre for Medium-Range Weather Forecasts (ECMWF) was used in this study (Hersbach et al., 2020). The ERA5 hourly data on pressure levels were extracted via the Copernicus portal (https://cds.climate.copernicus.eu) as monthly netCDF files containing the geopotential, potential vorticity (pv, K m^2 kg^{-1} s^{-1}), relative humidity (r, %), air temperature (t, K), the specific cloud ice water content ($ciwc$, kg kg^{-1}), the specific cloud liquid water content ($clwc$, kg kg^{-1}), the specific rain water content ($crwc$, kg kg^{-1}), the specific snow water content ($cswc$, kg kg^{-1}), the horizontal components of the wind speed (u and v, m s^{-1}) and the vertical velocity (w, Pa s^{-1}). During the YOPP, enhanced observations were conducted including four radio soundings per day at Davis, instead of two during normal periods. The YOPP covered approximately the period with a concomitant operation of the W-Band radar and depolarisation lidar.

The nearest ERA5 grid point (located at 68.5 ºS, 78.0 ºE) to the location of the ceilometer, W-Band radar and depolarisation lidar was used as the centre of 9 neighbouring grid points forming a square. All the extracted variable fields were averaged to reduce potential spatial variability effects and reduce noise. A sensitivity to this averaging approach was also performed using only the central grid point and the averaging effect on the temperature and humidity fields was considered negligible. The vertical pressure level fields were linearly interpolated to the 50 m vertical resolution grid of the ceilometer, and the hourly variables were linearly interpolated to 5 min to match the ceilometer time resolution.

2.2 Cloud phase masks

2.2.1. Radar-lidar merged cloud phase mask

This cloud phase product is obtained by merging information obtained from the W-Band radar and the depolarisation lidar. The principle is the same as the approach from Delanoë and Hogan (2010) with satellite-based sensors, which combined observations from CloudSat and CALIPSO, taking advantages of the different sensitivities of the radar and the lidar. The underlying principle for SLW versus mixed-phase classification of a grid point is that the W-Band radar is not sensitive enough to detect very small, supercooled liquid water droplets. As a result, when a value of reflectivity is measured for the grid point labelled as SLW by the lidar, it implies that there must be ice particles in the volume generating a backscattered radar signal, mixed with SLW droplets as detected by the lidar. In this paper, the pipeline to produce the cloud phase mask was based on the procedure described in Noh et al. (2019), with some modifications.

Atmospheric
Measurement
Techniques
Discussions

In the first stage, the lidar and radar are re-gridded to the same temporal and vertical resolution, to create a new merged grid at 15-m vertical resolution, and 1-min temporal resolution. ERA5 reanalysis data (hourly on pressure levels) were extracted
180   for the closest grid point to Davis and linearly interpolated in time and space onto the merged radar-lidar grid.

The second stage is to incorporate the cloud phase product from the RMAN lidar, as described in detail in Section 2.1.1 above. In the third stage, the original lidar-only cloud phase labelling is refined utilising the cloud radar reflectivity field. If there is no measurement of reflectivity above the noise level for the grid point, we assume that there is no ice. In that case, points labelled "SLW" in the lidar-only cloud phase classification remain labelled as "SLW". Conversely, the presence of an observed
185   radar reflectivity implies that there is ice in the volume as well, which triggers a new classification of the grid point as "Mixed-phase". Finally, the last stage consists in utilising the radar reflectivity to identify signals at subfreezing temperatures, while the lidar backscatter is fully attenuated by lower clouds and doesn't provide any information on the cloud phase. This case triggers the grid point to be labelled "Unknown" as there is no possibility to distinguish Ice particles-only from a mixed-phase, although there is certainty that these grid points are not containing only SLW (Noh et al., 2019). "Unknown" could therefore
190   be interpreted as "Ice or mixed-phase" if needed.

2.2.2. Ceilometer cloud phase mask based on T19

The first cloud phase mask presented herein is based solely on ceilometer observations, following the work from T19. Liquid
195   water droplets generate very high values of the ceilometer backscatter signal, and subsequent strong attenuation in the vertical profile above the altitude of liquid water. T19 proposed a modification from the Cloudnet approach (Illingworth et al., 2007), utilising the shape of the attenuated backscatter profile, instead of using a single threshold value. The input to the technique is the pre-processed ceilometer dataset, e.g., the 50 m gate resolution, 5 min calibrated attenuated backscatter processed with ALCF as explained previously.

200

The exact approach proposed by T19 was implemented herein: the maximum of a localised peak value in the vertical profile of the backscatter is found, instead of the first value above a given threshold as in Cloudnet. However, the maximum of the peak value needs to exceed the same threshold value as in Cloudnet, namely the pivot ß value of $\beta = 2 \times 10^{-5}$ m^{-1} sr^{-1}, together with a maximum peak width at half height set at 150 m. The combination of these two criteria allows the identification of a
205   rapidly attenuating signal, which is typical of liquid water layers. In the case of T19, this method of identification enabled capture of the base of precipitating clouds, but the authors also noted its potential application for in-cloud icing detection. The authors also showed the possibility of identifying multiple peaks within the same profile with this method. However, they did not specify how to label the 50 m bins. We decided that based on the above, the altitude bin corresponding to the location of the peak (if found) was labelled as liquid water. Further to this, a reclassification was done to distinguish supercooled liquid

Atmospheric
Measurement
Techniques
Discussions
Open Access
EGU

210   water from other liquid water based on the interpolated re-analyses temperature fields: if the temperature T was between 0 ℃

and above – 38 ℃, the grid points were classified as supercooled liquid water, otherwise they were classified as liquid water.

In addition to liquid water, precipitation and ice clouds were identified following the same approach as T19, by selecting grid

points with values of backscatter above $\beta = 3 \times 10^{-6}$ m^{-1} sr^{-1} with a thickness of at least 350 m, e.g. 7 consecutive grid points

215   satisfying this criteria, therefore showing no attenuation within at least 350 m. The base of these clouds was accordingly the

lowest grid point of the points within the profile satisfying these criteria. As noted by T19, liquid layers can be identified within

precipitation and ice clouds as defined utilising our algorithm.

Fog is a phenomenon that probably occurs relatively frequently in the Southern Ocean and some regions of Antarctica (Lazzara,

220   2008), although few studies are available in the literature to accurately quantify its occurrence (Kuma et al., 2020). The same

method as T19 was again used here, detecting fog layers by identifying values of backscatter above $\beta = 10^{-5}$ m^{-1} sr^{-1} for the

lowest grid point (corresponding to 0-50 m above the surface) and a ß value 250 m above the instrument of $\beta < 3 \times 10^{-7}$ m^{-1} sr^{-1} (to restrict the identification to fog, and exclude low-level thicker clouds).

2.3 Enhanced data-driven ceilometer cloud phase mask

225

Based on detailed observations of the cloud phase mask from T19 for days with substantial amounts of clouds, a large amount

of speckle in the retrieved SLW phase was observed, corresponding to timesteps for which the radar-lidar cloud mask did not

observe any SLW. This led us to investigate if an alternative algorithm could perform better for these conditions. Importantly,

the concomitant high-resolution and robust observations of cloud phase using the combination of radar and lidar provided us

230   with a reference that could be used to develop and validate our new algorithm.

This new algorithm relies on an initial signal analysis of each attenuated backscattered profile, as in T19, but also makes use

of the statistical properties of the full dataset. It is based on a data-driven model including a learning and testing phase using

the reference radar-lidar cloud mask.

235

< Figure 2 here >

The first step is to build a dataset and design, train and test a supervised model. This first step is summarised in the flowchart

in Figure 2. First, we detected all peaks in the dataset that had at least a width of 50 m, and a peak value of $\beta = 2 \times 10^{-5}$ m^{-1} sr^{-1}

240   [1] (similar threshold value as in Tuononen et al. 2019). Several peaks can be detected for the same profile. For each peak, seven

features were attributed to characterise peak properties: the value of the backscatter at peak location, the peak width, the value of the backscatter at peak width, the peak prominence (e.g. the difference between the peak value and the surrounding baseline), the peak altitude above ground level, the total number of detected peaks for a given profile and if several peaks, the order of the peak within that total number, with the lowest peak taking the number '0'. In addition to this, the peak temperature was

245 also extracted using interpolated ERA5 fields. This 8-feature dataset of peak properties was then labelled for each row with a Boolean attribute based on the detection of either supercooled liquid water or mixed phase by the radar-lidar mask for that timestep (True: detection of SLW/Mixed Phase, False: no detection of SLW/Mixed Phase). We also accounted for the problem caused by signal extinction in multi-layer SLW situations: peak properties of a single peak corresponding to SLW with no extinction other than molecular in the lower levels, cannot be directly compared in terms of backscatter values to peak

250 properties of a presumed SLW peak at higher altitude for which the signal would have undergone substantial extinction by clouds or the presence of SLW of mixed phase below. The properties of a peak that would have undergone extinction will therefore see lower values of the value of the attenuated backscatter at peak location. To account for this extinction effect, we compared the value of the backscatter at peak location for single peaks, and for multiple peaks that would have experienced extinction (peaks with a peak order > 0). For single peaks, SLW data-only were selected based on the Boolean condition

255 defined using the radar-lidar cloud mask. For multiple peaks, an arbitrary set of conditions must be defined to extract only potential SLW peaks from the multiple peaks. These conditions were based on the observed statistical distribution of peak properties and were empirically set as: the width of the peak must be < 4, the peak width height must be > 40 x 10^{-6} m^{-1} sr^{-1}, and the peak prominence must be > 60 x 10^{-6} m^{-1} sr^{-1}. The distributions of the two sets of data (single peaks and multiple peaks) and their Kernel Density Estimates (KDEs) are shown in Figure 3.

260

< Figure 3 here >

For single peaks, one would expect the distribution of values of the peak to vary based on the concentration of liquid water. One would expect the same effect for multiple peaks, but the values at peak would be smaller due to varying degrees of

265 extinction of the backscatter signal at lower levels. The difference between the median value of the single peak distribution and the multiple peak distribution can be calculated and is equal to 4.20 x 10^{-5} m^{-1} sr^{-1}. Our hypothesis was that for such a large amount of data in both cases (single and multiple peaks), this difference, or offset, was the value of the average extinction due to the presence of various phases in the cloud on the lower altitude that affects the potential SLW peaks. Adding this offset to the value of the backscatter at peak location for all datapoints of the multiple peaks' distribution would therefore "adjust" their

270 peak values. This is shown as well in Figure 3; with the added offset, the adjusted distribution of multiple peaks covers roughly the same area of the single peak distribution. While not perfect, this adjustment enabled us to modify our 8-feature dataset to allow a fairer comparison across datapoints for the peak value, a critical feature for further analysis. Additionally, to this

Atmospheric
Measurement
Techniques
Discussions

adjustment of the peak value for some of the peaks, the multiple peaks for which the timesteps were labelled "true" which did not meet the arbitrary criteria also had their labelling changed from "True" to "False". Following the above, a new "adjusted" peak properties dataset can be used for further analysis.

The next step of our algorithm development was to design, train and test a data-driven model that could predict the label of each of the peaks. A relatively novel tree-based ensemble method was proposed by Chen and Guestrin (2016), e.g. extreme gradient boosting or XGBoost, which is an improved version of gradient boosting with the advantages of reducing overfitting and computational costs. The excellent performances of this method for a wide range of applications, consistently outperforming other methods such as Support Vector Machines or Random Forest led us to select this approach here. The principle of this algorithm relies on a "boosting" strategy, where predictions of "weak" learners are combined to produce a "strong" learner by utilising additive training strategies. The computational cost is reduced by allowing parallel computations during the training phase (Chen et al., 2015). Here, we only cover the fundamental principles of the additive learning, and the reader should refer to Chen and Guestrin (2015) for more details. The first learner was initially fitted to all input data; a second model was then fitted to the residuals to reduce the disadvantage of the "weak" learner. This process of fitting was repeated several times until the model satisfied a predefined criterion. The prediction of the model for a given set of hyperparameters was obtained by combining the predictions of each learner. The function that describes the prediction at each step t can be written as eq. (1):

$$f_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = f_i^{(t-1)} + f_t(x_i) \qquad (1)$$

where $f_t(x_i)$ is the learner at step t, $f_i^{(t-1)}$ and $f_i^{(t)}$ are the predictions at steps t-1 and t, and x_i is the input variable.

The extreme gradient boosting model uses the below expression to evaluate the model performance, eq. (2):

$$Obj^{(t)} = \sum_{k=1}^{n} l(\overline{y_i}, y_i) + \sum_{k=1}^{t} \Omega(f_i) \qquad (2)$$

where l is the loss function, n is the number of observations and Ω is the regularisation term defined as eq. (3):

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda||\omega||^2 \qquad (3)$$

300 where λ is the regularisation parameter, γ is the minimum loss needed to partition the leaf node and ω is the vector of the scores in the model leaves.

Features were passed to the model for training using a 3 Stratified k-folds cross validation. Stratified K-fold is a variation of k-fold (Ojala and Garriga, 2010) where each training/testing set (fold) contains approximately the same percentage of each

305 target class as the complete set. The cross-validation k-fold approach allows us to train and test the model three times. In this approach, testing folds never share the same data with other folds. Given the low number of features and small size of the dataset, computational cost was not a limitation, therefore no principal component analysis was applied to the data prior to model training and testing, and an extended grid search over a wide range of hyperparameters was implemented utilising both accuracy and balanced accuracies as the scoring methods. The hyperparameters maximum depth, minimum child weight and

310 learning rate (eta) were explored with respective chosen values of [0.3, 0.2, 0.1, 0.05, 0.01, 0.005] for the learning rate, [9,12] for the maximum depth and [5,8] for the minimum child weight. Other parameters were set at default values. All simulations were performed using a 1.3 GHz dual core Intel Core M and 8 GB of RAM memory.

Finally, once our model has been trained and tested on the data, the third step was to apply the algorithm including the trained

315 model subsequently to each profile. This approach is summarised in the flowchart in Figure 4.

< Figure 4 here >

The algorithm treated each vertical profile of attenuated backscatter sequentially: in the first step, peaks were detected, and

320 their associated properties computed. If no peaks were detected, the timestep corresponding to that backscatter profile was labelled as "not SLW". If several peaks were detected, a sequential pipeline as seen in Fig. 4 was implemented to correct potential SLW peaks for extinction using the statistical properties obtained in the pre-processing stage. For that given timestep, one or more peaks could be identified together with their properties. These peak features were passed to the previously trained XGBoost model for labelling, either "SLW" or "not SLW". If a given peak was labelled as "SLW", the corresponding bin at

325 peak altitude was labelled 'SLW", as well as the surrounding bins, with the SLW lower and upper boundaries defined as twice the peak width (at mid height) value. With this algorithm, several layers of SLW can be identified within a single profile, and SLW layers can be identified within or outside a cloud.

2.4 Strategies for intercomparison of cloud masks

330 As mentioned previously, the resolution of the merged radar-lidar mask was 1-min and 15-m, while the resolution of the ceilometer cloud mask was coarser (5-min, 50-m). For both masks, linearly interpolated hourly ERA5 variables fields have been used. In order to compare both masks various strategies regarding resolutions can be considered.

For the masks intercomparison, the coarser resolution of the ceilometer mask was used, and the radar-lidar mask was subsampled to 5-min timestamps. Since spatial variabilities could occur at the finer vertical resolution of the W-Band,
335 depolarisation lidar and ceilometer, grid to grid comparison of the masks was not considered suitable and relevant given the objectives of this study. Instead, a comparison timestep to timestep was performed, integrating the information available over each vertical column. Then, two different strategies were used to subsample the merged radar-lidar mask: (i) the matching timestamps of both masks were found, and if SLW or mixed-phase was identified in one bin of the merged radar-lidar mask, that timestep was labelled as positive, otherwise it was labelled as negative. Similarly, if SLW was identified in one of the bins
340 of the vertical column for the ceilometer mask, that timestamp was labelled as positive; (ii) a condition on the spatio-temporal structure of SLW and mixed-phase bins was considered over 5-min periods: for the timestamp to be labelled as positive, a given number of consecutive bins labelled SLW or mixed-phase need to be found at the same height. Threshold values for consecutive bins were set at 3, 4 and 5 and this criterion was applied as a 5-min moving window on the merged radar-lidar cloud mask to produce three subsets of data corresponding to this second strategy. Labelling of SLW for the ceilometer cloud
345 mask was performed similarly to the first strategy.

2.5 Metrics to evaluate mask intercomparison and model performances

350 Mask intercomparison, physical model evaluation and data-driven training and testing involve evaluating performance of predictions by comparing two one dimensional Boolean vectors (positive (true) for the presence of SLW, negative (false) for the absence of SLW). Several metrics were used for the evaluation: first, a confusion matrix was calculated from which the precision, the recall and the f1 scores were derived. The ratios of positive to negative in the dataset were closely monitored for each of the evaluations, and both the accuracy (the harmonic mean of the f1 scores for positive and negative) and the balanced accuracy (the arithmetic mean of the recall) were calculated. When the dataset was imbalanced, a focus was put on the balanced
355 accuracy to closely monitor the performance of the prediction of positive cases.

A true positive (TP) is defined as a test result indicating a correct prediction, a true negative (TN) is defined as a test result correctly indicating a wrong prediction, while a false positive (FP) is defined as a test result wrongly indicating that a prediction is correct, and a false negative (FN) is a test result wrongly indicating that a prediction is not correct.

Atmospheric
Measurement
Techniques
Discussions

360

The precision is the ratio of the number of true positives over the number of true positives and false positives, i.e. the ability of the classification not to label as positive a negative sample and is defined as eq. (4):

$$precision = TP \ / \ (TP + FP)$$
(4)

365   The recall is the ratio of true positives over the number of true positives and false negatives, e.g. the ability of the classification to find all the positive samples and is defined as eq. (5):

$$recall = TP \ / \ (TP + FN)$$
(5)

The f1 score (or hereafter also described as "accuracy") is defined as the arithmetic average of the precision and the recall,
370   with its best value at 1 and its worst score at 0, with equal contribution of recall and precision (eq. 6):

$$f1 = 2 \times (precision \times recall/(precision + recall)$$
(6)

Note that in the current case of binary classification, positive and negative labelling can be inverted so that f1 scores can be calculated both for positive and negative cases. The accuracy is then calculated as the harmonic mean of the positive and
375   negative f1 scores. The balanced accuracy on the other hand is defined as the arithmetic mean of the positive and negative recall.

3 Results

3.1 Ceilometer backscatter profile analysis: the 6th of January 2019 case study

380
<div align="center">< Figure 5 here ></div>

To illustrate the disparity between the various cloud phase retrievals, observations from the 6th of January 2019 were selected
385   as an example, as these included both low and higher clouds, with SLW present both within deep clouds or isolated from deep

Atmospheric
Measurement
Techniques
Discussions

clouds. In figure 5, one can see the calibrated attenuated backscatter from the ceilometer (Fig. 5a), the reference cloud phase mask combining radar and lidar (Fig. 5b), and the cloud phase attribution from two retrievals following T19 (Fig. 5c), and the trained XGBoost model described in this work (Fig. 5d). The XGBoost algorithm was trained on a different set of the data, which excluded the 6th of January 2019. During the first part of the day, until around 11:30 UTC, the radar-lidar cloud phase

390   mask shows little occurrence of SLW or mixed phase, except at the very beginning of the day (the first 10 min) and around 2:00 UTC. In the second part of the day, clear horizontal bands of SLW can be observed, including times with clouds or precipitation below the SLW bands. The ceilometer backscatter (Fig. 5a) showed distinct signal patterns for the first and second part of the day, but visually it remains difficult to clearly distinguish strong backscatter during the first part of the day, that could indicate the presence of SLW.

395

< Figure 6 here >

400   Figure 6 shows five selected vertical profiles of attenuated backscatter (Fig. 5a) for that day, chosen to illustrate the diversity in attenuated backscattering signal profiles. It also shows the theoretical molecular backscatter at the wavelength of the ceilometer, the identified peaks, as well as their average properties. For the profiles A, D and E, SLW or mixed phase were identified in the reference radar-lidar mask. These three backscatter profiles presented common characteristics, such as a narrow peak (low value of the peak width at mid-height), relatively high values of the attenuated backscatter (above 10^{-4} m^{-1}

405   sr^{-1}), and high prominences (high values of the difference between peak value and the surrounding baseline). Conversely, peaks B and C present much wider peaks (higher values of peak width at mid-height), and smaller values of the prominence and were classified as ice.

< Figure 7 here >

410

In Figure 7, the average peak properties such as peak value, peak width, peak width height, the number of peaks per profile, peak altitude for each of the attenuated backscatter profiles for which peak had been identified for the 6th of January 2019 are presented as scatter plots with peak β value as the y-axis. We chose to use peak value as the y-axis as this is the most

415   discriminant peak characteristic, and by analysing scatter plots, we can visually observe clustering patterns. In Figure 7, a single ß profile can generate multiple datapoints if several peaks were observed for that profile. The points A, D and E were well clustered within the same region in Fig. 7a and Fig. 7b, corresponding to high values of ß at peak, low values of the peak width and peak width height. Peaks A and E also appeared in the same region in Fig. 7c. Generally, SLW was observed when

420 a single peak was observed for the ß profile. This was not always the case, as sometimes, SLW can be observed when several peaks are present: for instance, on the 6th of January, the isolines showed that several instances of SLW were observed when two peaks were present, but this dropped dramatically for three or four peaks. For that specific day, the scatter plot with peak altitude (Fig. 7d) shows that SLW is more frequent at higher altitudes with three clusters around 1000 m, 2000 m and 3500 m AGL, corresponding to the spatial organisation of SLW (Fig. 5b), while non-SLW peaks are far more frequent at lower altitudes (cluster located between 0 and 1000 m AGL). These non-SLW peaks are associated to low-level clouds as no fog was identified

425 over that three-month period.

The data shown in Figures 5, 6, 7 demonstrate that peak average characteristics exhibit very specific features that could be used to detect the occurrence of SLW. The original approach from T19 was already skilful in identifying SLW. For the second half of the day on January 6th, 2019, there was a very good match between the cloud mask based on T19 (Fig. 5c) and the

430 reference cloud mask (Fig. 5b). Conversely, for the first half of the day, the T19 approach identified multiple SLW regions within the cloud producing a speckle pattern of SLW. This was not observed by the radar-lidar observations and algorithm and was thus probably wrongly labelled by the T19 approach. Our new approach based on the use of average peak characteristics and a dedicated trained algorithm using the radar-lidar reference showed a great improvement in the retrievals: the second half of the day remained like the retrieval of T19, although labelling thicker bands of SLW utilising the peak width property. These

435 thicker horizontal bands of SLW were more in line with the radar-lidar reference, which showed occurrence of SLW or mixed phase of about the same thickness. For the first half of the day, our approach outperforms the T19 approach, by removing the spurious speckle patterns while keeping the correct detection of SLW at the very beginning of the day, also observed in the radar-lidar cloud mask (profile A in Fig. 6a). The SLW around 2:00 UTC found in the radar-lidar cloud phase mask was also retained by our new technique. An occurrence of SLW at around 5:00 UTC was present in our retrieval but not in the reference

440 radar-lidar cloud mask. This may be due to an error in our retrieval, an error in the phase assignment in the lidar product, or different observations made by the lidar, radar and ceilometer at that timestep. Nonetheless, our retrieval performed very well: computed accuracies for that day (6th of January 2019) using the radar-lidar mask as the reference were equals to 0.84 for our new data-driven retrieval, compared to an accuracy score of 0.65 with the T19 approach. In the next section, we evaluate the performances of T19, a data-driven threshold approach and our XGBoost retrieval for our full dataset covering almost three

445 months of data during the Southern Hemisphere summer.

3.2 Evaluation of retrievals for the PLATO period

In Figure 8, similarly to what was presented in Figure 7, the average peak properties such as peak value, peak width, peak width height, the number of peaks per profile, peak altitude for each of the attenuated backscatter profiles for which peak had

15

450   been identified are presented as scatter plots with peak β value as the y-axis, but this time for the full dataset, equivalent to 11,327 datapoints. The two-dimensional distributions are like that of Figure 7, showing that the results of the case of January 6th can be extrapolated to the full dataset. The value of β at peak is directly correlated to the peak width height, making that feature redundant. The value of β at peak is clearly the best feature to separate the SLW-labelled datapoints (True) to the non-SLW datapoints (False). The other features, e.g., peak width, number of peaks and peak altitude are weaker discriminants, but

455   they reveal that non-SLW datapoints (False) are much more widespread than SLW datapoints (True), showing that extreme values of the features are usually associated with the non-occurrence of SLW. Typically, many peaks (>2), or very wide peak widths (> 3) for each profile are associated to non-SLW. The presence of a dense concentration of non-SLW datapoints at lower altitudes (< 1000 m) shows that SLW is usually not observed at these low altitudes above ground level.

460

< Figure 8 here >

As discussed previously, various sub-sampling strategies were considered to compare the ceilometer mask and the reference radar-lidar cloud mask. The comparison between the T19 algorithm and the reference dataset using the various subsampling

465   strategies showed minimal variability with accuracies varying between 0.84 and 0.85 for threshold consecutive values of 3, 4 and 5 and instantaneous comparisons. For the full dataset, the accuracy score was 0.84 (or 0.85 depending on the subsampling strategy used for the comparison), while for the dataset with peaks only, the accuracy was 0.72.

470   Based on the clustering as observed in Figure 8 for the full dataset, we decided to implement a second classification using arbitrary thresholds for each of the peak features. This has the advantage of not having to train and test a model and circumnavigates the need for a high-resolution reference dataset (although the arbitrary thresholds are based here on the cluster plots where the labelling has been done using the reference dataset). The arbitrary thresholds that we used to label a timestep as "SLW" were for peak features such as: value of β at peak > 5 x 10^{-5} m^{-1} sr^{-1}, peak width < 4, peak width height > 40 x 10^{-6}

475   m^{-1} sr^{-1}, peak prominence value > 60 x 10^{-6} m^{-1} sr^{-1} and the total number of peaks < 3. We evaluated the arbitrary threshold approach on both the full dataset and the dataset with only identified peaks. For the full dataset, the accuracy score was 0.89, while for the dataset including peaks only, the accuracy was 0.76.

Our extreme gradient boosting model was trained and tested using a 3-fold stratified cross validation approach as described

480   previously. The model was trained and tested for both the full dataset, and the dataset containing only timesteps for which peaks had been identified. For the full dataset, the best testing accuracy score was 0.91 (with learning rate = 0.005, max depth = 12, child weight = 8) for an average training accuracy score of 0.95. For the dataset for which peaks were identified, the total

Atmospheric
Measurement
Techniques
Discussions

dataset was made of 11,327 datapoints. The best testing score was of 0.81 (with learning rate = 0.01, max depth = 12, child weight = 8), for training accuracy scores (or f1) of 0.94.

485

3.3 Evaluating the importance of predictors for the XGBoost algorithm

While our designed, trained and tested extreme gradient boosting model performs remarkably well, we want to understand which of the average profile characteristics are the most important for skilful model predictions. Lundberg et al. (2020) recently proposed an explanation method for tree-based models, building on work based on classic game theoretic Shapley values (Lundberg and Lee, 2017). We implemented this explanation method, also known as "TreeExplainer" (Lundberg et al., 2020)
490 in its original python version. With TreeExplainer, we were able to provide local explanations for each prediction by calculating their Shapley (SHAP) values. The input features were the same as described for the implementation of our XGBoost model, e.g.*,* the average adjusted peak properties. Figure 9 shows the distribution of SHAP values for these 8 input features, together with their normalised value represented as a colour bar. The features were ranked from the most important at the top of the graph (value of backscatter at the peak location) to the least important at the bottom (the peak number, or rank).
495

< Figure 9 here >

500

As expected, the value of ß at the peak location was the most important feature to produce accurate predictions: high values of ß were important to detect the presence of SLW, while low values of ß were skilful in predicting the absence of SLW. The total peak number was the second feature of importance, with low values of the total peak number contributing to better predictions of the presence of SLW. In fact, very well-defined horizontal bands of SLW showed these typical characteristics
505 with a single narrow peak of high values. Peak prominence had a similar SHAP values distribution pattern as peak value, with less defined clusters. This result is consistent with the scatter plot of peak value versus peak prominence in Figure 8b, showing a strong correlation between these two features. Low values of peak width were important in the prediction of the presence of SLW, showing the importance of the peak shape (narrow) in indicating the presence of SLW. Low values of the peak width height tend to also help predict the presence of SLW. Conversely, peak altitude and peak temperature are associated with
510 average SHAP values close to 0, indicating that these two features were not important in the production of accurate predictions. While we saw a tendency of SLW to be located within a preferred range of altitudes (between approximately 1000 and 4000 m ASL), this range was too wide to make altitude of the peak a useful criterion to help identify SLW. Low values of peak altitude however seemed to help identify conditions with the absence of SLW (as indicated by the negative and blue SHAP values for the peak altitude).

515

4 Discussion and conclusions

We have demonstrated that the presence of supercooled liquid water can be detected using solely the measured attenuated backscatter signal from a ceilometer, even in the absence of information on the signal depolarisation. We utilised coincident

520 observations from a W-Band radar and a depolarisation lidar to build a reference cloud phase mask that was used as a benchmark to compare to our ceilometer-retrieved supercooled liquid water detections. Utilising the ceilometer data and an existing approach proposed by T19, we obtained an overall accuracy of 0.84 and an accuracy for days with detection of strong backscatter signals of 0.72. We then developed an enhanced method, utilising the benchmark dataset of cloud phase observations to develop, train and test an extreme gradient boosting model.

525

Utilising the ceilometer observations and this model, we increased the overall accuracy of correctly identifying SLW layers to 0.91 and the accuracy for days with a detection of strong backscatter signals to 0.81. This enhanced model greatly improved detection accuracy, for the cases where multiple peaks in the backscatter were observed and were erroneously classified as SLW by the method from T19. The most important input features were the value of the backscattered signal at the peak,

530 followed by the total number of peaks within that profile. In the current approach, we considered each profile (or timestep) independently from one another (although we did consider all points together in the statistical analysis and in the model development). Alternative or complementary to our existing algorithm, data-driven approaches such as Recurrent Neural Networks (LSTM or GRU) could consider the spatio-temporal patterns of peak properties to predict the occurrence and location of SLW at the following timestep.

535

Ceilometers are relatively low-cost ground-based active atmospheric remote sensing tools as compared to Weather radars or depolarization lidars. They are commonly deployed at aerodromes but also at other operational or research atmospheric monitoring facilities. Here, we showed that the raw backscatter signal can be utilised to detect supercooled liquid water, thus broadening the observational capabilities of such instruments, for regions where observations are scarce, like Antarctica. The

540 present work is the first of its kind utilising a benchmark radar-lidar cloud phase mask to train a dedicated model to detect supercooled liquid water from the ceilometer backscatter only. It will be important to test the application of the same approach elsewhere, especially for current monitoring sites or for historical data including the same set of instruments presented here, that is weather radar, depolarisation lidar and ceilometers. Our approach was developed for a polar environment, and it would be important to see how the developed technique transfers to regions at mid or low latitudes.

545

Ground-based observations of supercooled liquid water are complementary to spaceborne observations. Satellite detection of supercooled liquid water suffers from the attenuation of the signal in the lower layers, and from a lower spatial and temporal

resolution. The combination of satellite and ground observations has the potential to improve cloud phase products. Knowledge of the cloud phase including supercooled liquid water at high-resolution can help develop and validate icing

550 algorithms, with the objective of predicting aircraft airframe icing potential (Morcrette et al., 2019) or predicting the potential icing of wind turbines for wind production (Hamalainen et al., 2020). Detection capabilities developed in this paper will enable important studies to examine the seasonal variability of the occurrence of SLW and to develop aircraft icing potential nowcasting capabilities.

Code and data availability

555 The ALCF is open-source and available at https://alcf-lidar.github.io (last access: 16 November 2021) as well as permanent archive of code and technical documentation on Zenodo at https://doi.org/10.5281/zenodo.4411633 (Kuma et al., 2021). A tool for converting Vaisala CL31 and CL51 data files to NetCDF cl2nc is open-source and available at https://doi.org/10.5281/zenodo.4409716 (Kuma, 2020a). The observational data (ALCF-processed netCDF ceilometer files and the radar-lidar mask netCDF files) are available on Zenodo at https://doi.org/10.5281/zenodo.5832199 and will also be

560 available on the AAD datacentre. The ERA5 data are available through the Copernicus data portal at https://cds.climate.copernicus.eu (last access: 7 January 2022). The new ceilometer algorithm described herein has been developed at the Bureau of Meteorology and is not publicly available.

Author contribution

AG performed the ceilometer code development and overall data analysis. AP and SA analysed the radar and depolarisation
565 lidar data and produced the radar-lidar cloud mask. AP, SA, and AK provided regular scientific inputs on the analysis. AP, SA, AK, and AMD designed the experimental setup at Davis to gather ground-based observations. PK developed the code to pre-process the ceilometer data and provided scientific inputs on the ceilometer data analysis. AG prepared the manuscript with contributions from all co-authors.

570 **Competing interests**

The authors declare that they have no conflict of interest.

575 **Acknowledgments**

Atmospheric
Measurement
Techniques
Discussions

585   **References**

Alexander, S. P. and Protat, A.: Cloud properties observed from the surface and by satellite at the northern edge of the Southern Ocean. *J. Geo. Res. Atm.*, 123, 443–456, doi:10.1002/2017JD026552, 2018.

Alexander, S. P. and Protat, A.: Vertical profiling of aerosols with a combined Raman-elastic backscatter lidar in the remote

590   Southern Ocean marine boundary layer (43–66∘S, 132–150∘E). Journal of Geophysical Research: Atmospheres, 124, 12,107–12,125. doi:10.1029/2019JD030628, 2019.

Alexander, S. P., McFarquhar, G. M., Marchand, R., Protat, A., Vignon, É., Mace, G. G. and Klekociuk, A. R.: Mixed-phase clouds and precipitation in Southern Ocean cyclones and cloud systems observed poleward of 64°S by ship-based cloud radar

595   and lidar. *J. Geo. Res. Atm*, 126, e2020JD033626, doi:10.1029/2020JD033626, 2021.

Bennartz, R., Fell, F., Pettersen, C., Shupe, M. D., and Schuettemeyer, D.: Spatial and temporal variability of snowfall over Greenland from CloudSat observations, Atmos. Chem. Phys., 19, 8101–8121, doi:10.5194/acp-19-8101-2019, 2019.

600   Bodas-Salcedo, A., Hill, P. G., Furtado, K., Williams, K. D., Field, P. R., Manners, J. C., Hyder, P., and Kato, S.: Large Contribution of Supercooled Liquid Clouds to the Solar Radiation Budget of the Southern Ocean, J. Climate, 29, 4213–4228, doi:10.1175/jcli-d-15-0564.1, https://doi.org/10.1175/jcli-d-15-0564.1, 2016.

Bromwich, D. H., Werner, K., Casati, B., Powers, J. G., Gorodetskaya, I. V., Massonnet, F., Vitale, V., Heinrich, V. J., Liggett,

605   D., Arndt, S., Barja, B., Bazile, E., Carpentier, S., Carrasco, J. F., Choi, T., Choi, Y., Colwell, S. R., Cordero, R. R., Gervasi, M., Haiden, T., Hirasawa, N., Inoue, J., Jung, T., Kalesse, H., Kim, S., Lazzara, M. A., Manning, K. W., Norris, K., Park, S.,

Reid, P., Rigor, I., Rowe, P. M., Schmithüsen, H., Seifert, P., Sun, Q., Uttal, T., Zannoni, M. and Zou, X.: The Year of Polar Prediction in the Southern Hemisphere (YOPP-SH), *Bul. Am. Met. Soc.*, *101*(10), E1653-E1676, doi.org/10.1175/BAMS-D-19-0255.1, 2020.

610

Chen, T., and Guestrin, C.: Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794), 2016.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., and Cho, H.: Xgboost: extreme gradient boosting. *R package version 0.4-2, 1*(4), 1-4, 2015.

615

Cossich, W., Maestri, T., Magurno, D., Martinazzo, M., Di Natale, G., Palchetti, L., Bianchini, G., and Del Guasta, M.: Ice and mixed-phase cloud statistics on the Antarctic Plateau, Atmos. Chem. Phys., 21, 13811–13833, doi: 10.5194/acp-21-13811-2021, 2021.

620

Delanoë, J., A. Protat, J.-P. Vinson, W. Brett, C. Caudoux, F. Bertrand, J. Parent du Chatelet, R. Hallali, L. Barthes, M. Haeffelin, and Dupont, J.-C.: BASTA, a 95 GHz FMCW Doppler radar for cloud and fog studies. J. Atmos. Oceanic Tech., 33, 1023-1038, doi:10.1175/JTECH-D-15-0104.1, 2016.

625 Frey, W. R., and Kay, J. E.: The influence of extratropical cloud phase and amount feedbacks on climate sensitivity. *Climate Dynamics*, *50*, 3097–3116. https://doi.org/10.1007/s00382-017-3796-5, 2018.

Genthon, C., Berne, A., Grazioli, J., Durán Alarcón, C., Praz, C., and Boudevillain, B.: Precipitation at Dumont d'Urville, Adélie Land, East Antarctica: the APRES3 field campaigns dataset, Earth Syst. Sci. Data, 10, 1605–1612, doi:10.5194/essd-10-1605-2018, 2018.

630

Gehring, J., Vignon, E., Billault-Roux, A.-C., Ferrone, A., Protat, A., Alexander, S.P. and Berne: The influence of orographic gravity waves on precipitation during an atmospheric river event at Davis, Antarctica, accepted in *Journal of Geophysical Research: Atmospheres,* 10.1029/2021JD035210, 2022.

635

Gorodetskaya, I. V., Kneifel, S., Maahn, M., Van Tricht, K., Thiery, W., Schween, J. H., Mangold, A., Crewell, S., and Van Lipzig, N. P. M.: Cloud and precipitation properties from ground-based remote-sensing instruments in East Antarctica, The Cryosphere, 9, 285–304, doi:10.5194/tc-9-285-2015, 2015.

Atmospheric
Measurement
Techniques
Discussions

640   Grazioli, J., Genthon, C., Boudevillain, B., Duran-Alarcon, C., Del Guasta, M., Madeleine, J.-B., and Berne, A.:
Measurements of precipitation in Dumont d'Urville, Adélie Land, East Antarctica, The Cryosphere, 11, 1797–
1811, doi:10.5194/tc-11-1797-2017, 2017.

Grosvenor, D. P., Choularton, T. W., Lachlan-Cope, T., Gallagher, M. W., Crosier, J., Bower, K. N., Ladkin, R. S., and
645   Dorsey, J. R.: In-situ aircraft observations of ice concentrations within clouds over the Antarctic Peninsula and Larsen Ice
Shelf, Atmos. Chem. Phys., 12, 11275–11294, doi:10.5194/acp-12-11275-2012, 2012.

Hämäläinen, K., Hirsikko, A., Leskinen, A., Komppula, M., O'Connor, E.J. and Niemelä, S.: Evaluating atmospheric icing
forecasts with ground-based ceilometer profiles, *Met. Appl.*, 27:e1964, doi:10.1002/met.1964, 2020.
650

Hersbach,   H., Bell,   B., Berrisford,   P.,   et   al.: The   ERA5   global   reanalysis. *Q.   J.   R.   Meteorol.
Soc.*, 146: 1999– 2049, doi:10.1002/qj.3803, 2020.

Hogan, R. J. and Illingworth, A. J.: A climatology of supercooled layer clouds from lidar ceilometer data. In *CLARE'98*
655   *Final workshop, 13–14 September 1999* (pp. 161-165), 1999.

Hogan, R. J., Illingworth, A. J., O'connor, E. J., and Baptista, J. P. V. P.: Characteristics of mixed-phase clouds. II: A
climatology from ground-based lidar, Q. J. Roy. Meteorol. Soc., 129, 2117–2134, doi:10.1256/qj.01.209, 2003.

660   Hogan, R. J. and Illingworth, A. J.: The effect of specular reflection on spaceborne lidar measurements of ice clouds, Report
of the ESA Retrieval algorithm for EarthCARE project, 5 pp., 2003.

Hogan,   R.   J.,   Behera,   M.   D.,   O'Connor,   E.   J.,   and   Illingworth,   A.   J.:   Estimate   of   the   global   distribution   of   stratiform
supercooled liquid water clouds using the LITE lidar, *Geophys. Res. Lett.*, 31, L05106, doi:10.1029/2003GL018977, 2004.
665

Hopkin, E., Illingworth, A. J., Charlton-Perez, C., Westbrook, C. D., and Ballard, S.: A robust automated technique for
operational calibration of ceilometers using the integrated backscatter from totally attenuating liquid clouds, Atmos. Meas.
Tech., 12, 4131–4147, https://doi.org/10.5194/amt-12-4131-2019, 2019.

670   Hunter, J. D.: Matplotlib: A 2D graphics environment, Comput. Sci. Eng., 9, 90, doi :10.1109/MCSE.2007.55, 2007.

Hu, Y., Winker, D., Vaughan, M., Lin, B., Omar, A., Trepte, C., … Holz, R.: CALIPSO/CALIOP cloud phase discrimination algorithm. *Journal of Atmospheric and Oceanic Technology*, 26, 2293–2309. https://doi.org/10.1175/2009JTECHA1280.1, 2009.

675

Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., Wood, R. A., Meijers, A. J. S., Mulcahy, J., Field, P., Furtado, K., Bodas-Salcedo, A., Williams, K. D., Copsey, D., Josey, S. A., Liu, C., Roberts, C. D., Sanchez, C., Ridley, J., Thorpe, L., Hardiman, S. C., Mayer, M., Berry, D. I., and Belcher, S. E.: Critical Southern Ocean climate model biases traced to atmospheric model cloud errors, Nat. Commun., 9, 3625, doi:10.1038/s41467-018-05634-2,

680 2018.

Illingworth, A. J., Hogan, R. J., O'Connor, E., Bouniol, D., Brooks, M. E., Delanoé, J., Donovan, D. P., Eastment, J. D., Gaussiat, N., Goddard, J. W. F., Haeffelin, M., Baltink, H. K., Krasnov, O. A., Pelon, J., Piriou, J.-M., Protat, A., Russchenberg, H. W. J., Seifert, A., Tompkins, A. M., van Zadelhoff, G.-J., Vinit, F., Willén, U., Wilson, D. R., and Wrench,

685 C. L.: Cloudnet, B. Am. Meteor. Soc., 88, 883–898, doi:10.1175/BAMS-88-6-883, 2007.

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H. F., Holland, M., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P., Perovich, D., Reid, P., Renfrew, I. A., Smith, G., Svensson, G., Tolstykh, M. and Yang, Q.: Advancing Polar Prediction

690 Capabilities on Daily to Seasonal Time Scales, *Bulletin of the American Meteorological Society*, *97*(9), 1631-1647, doi:10.1175/BAMS-D-14-00246.1, 2016.

Kay, J. E., Wall, C., Yettella, V., Medeiros, B., Hannay, C., Caldwell, P., and Bitz, C.: Global climate impacts of fixing the Southern Ocean shortwave radiation bias in the Community Earth System Model (CESM). *Journal of Climate*, *29*, 4617–

695 4636, doi:10.1175/jcli-d-15-0358.1, 2016.

Kuma, P., McDonald, A. J., Morgenstern, O., Alexander, S. P., Cassano, J. J., Garrett, S., Halla, J., Hartery, S., Harvey, M. J., Parsons, S., Plank, G., Varma, V., and Williams, J.: Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations, Atmos. Chem. Phys., 20, 6607–6630, doi:10.5194/acp-20-

700 6607-2020, 2020.

Atmospheric
Measurement
Techniques
Discussions

Kuma, P., McDonald, A. J., Morgenstern, O., Querel, R., Silber, I., and Flynn, C. J.: Ground-based lidar processing and simulator framework for comparing models and observations (ALCF 1.0), Geosci. Model Dev., 14, 43–72, doi: 10.5194/gmd-14-43-2021, 2021.

705

Lachlan-Cope, T., Listowski, C., and O'Shea, S.: The microphysics of clouds over the Antarctic Peninsula – Part 1: Observations, Atmos. Chem. Phys., 16, 15605–15617, doi:10.5194/acp-16-15605-2016, 2016.

Lawson, R. Paul and Gettelman, Andrew: Impact of Antarctic clouds on climate Proceedings of the National Academy of
710 Sciences Dec 2014, 111 (51) 18156-18161; DOI:10.1073/pnas.1418197111, 2014.

Lazzara, M. A.: *A diagnostic study of Antarctic fog*, PhD thesis, The University of Wisconsin, Madison, 2008.

Listowski, C., Delanoë, J., Kirchgaessner, A., Lachlan-Cope, T., and King, J.: Antarctic clouds, supercooled liquid water and
715 mixed phase, investigated with DARDAR: geographical and seasonal variations, Atmos. Chem. Phys., 19, 6771–6808, https://doi.org/10.5194/acp-19-6771-2019, 2019.

Listowski, C., Rojo, M., Claud, C., Delanoë, J., Rysman, J.-F., Cazenave, Q., et al.: New insights into the vertical structure of clouds in polar lows, using radar-lidar satellite observations, *Geo. Res. Let.*, 47,
720 e2020GL088785, doi:10.1029/2020GL088785, 2020.

Loeb, N. A. and Kennedy, A.: Blowing snow at McMurdo station, Antarctica during the AWARE Field Campaign: Surface and ceilometer observations. *J. Geo. Res. Atm*, 126, e2020JD033935, https://doi.org/10.1029/2020JD033935, 2021.

725 Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4768–4777, 2017.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.-I.: From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* **2**, 56–67,
730 doi:10.1038/s42256-019-0138-9, 2020.

Mace, G.G. and Protat, A.: Clouds over the Southern Ocean as Observed from the R/V Investigator during CAPRICORN. Part I: Cloud Occurrence and Phase Partitioning, *J. App. Met. Clim.*, 57, 8, 1783-1803, doi:10.1175/JAMC-D-17-0194.1, 2018.

735   Mace, G.G. and Protat, A.: Clouds over the Southern Ocean as Observed from the R/V Investigator during CAPRICORN. Part II: The Properties of Nonprecipitating Stratocumulus, *J. App. Met. Clim.,* 57, 8, 1805-1823, doi:10.1175/JAMC-D-17-0195.1, 2018.

Mace, G. G., Benson, S., and Hu, Y.: On the frequency of occurrence of the ice phase in supercooled Southern Ocean low
740   clouds derived from CALIPSO and CloudSat. *Geophysical Research Letters*, 47, e2020GL087554. doi:10.1029/2020GL087554, 2020.

Mace, G. G., Protat, A. and Benson, S.: Mixed-phase clouds over the Southern Ocean as observed from satellite and surface based lidar and radar. *Journal of Geophysical Research: Atmospheres*, *126*, e2021JD034569, doi:10.1029/2021JD034569,
745   2021.

McErlich, C., McDonald, A., Schuddeboom, A. and Silber, I.: Comparing satellite- and ground-based observations of cloud occurrence over high southern latitude. *Journal of Geophysical Research: Atmospheres*, 126, e2020JD033607, doi:10.1029/2020JD033607, 2021.
750

Mellor, M.: *Notes on Antarctic aviation*. Cold regions research and engineering, Lab Hanover NH, 1993.

Morille, Y., Haeffelin, M., Drobinski, P., and Pelon, J.: STRAT: An automated algorithm to retrieve the vertical structure of
755   the atmosphere from single-channel lidar data, J. Atmos. Ocean. Tech., 24, 761–775, doi:10.1175/JTECH2008.1, 2007.

McDonough F., Wolff C.A., Politovich M.K.: Forecasting supercooled large drop icing conditions. In *13th Conference on Aviation, Range and Aerospace Meteorology*, New Orleans, LA, 11–15 January, American Meteorological Society, 2008.

760   McErlich, C., McDonald, A., Schuddeboom, A. and Silber, I.: Comparing satellite- and ground-based observations of cloud occurrence over high southern latitude, *J. Geo. Res. Atm*, 126, e2020JD033607, doi:10.1029/2020JD033607, 2021.

Münkel, C., Eresmaa, N., Räsänen, J., and Karppinen, A.: Retrieval of mixing height and dust concentration with lidar ceilometer, Bound.-Lay. Meteorol., 124, 117–128, doi:10.1007/s10546-006-9103-3, 2007.
765

Noh, Y.-J., Miller, S. D., Heidinger, A. K., Mace, G. G., Protat, A. and Alexander, S. P.: Satellite-based detection of daytime supercooled liquid-topped mixed-phase clouds over the Southern Ocean using the Advanced Himawari Imager. *J. Geo. Res. Atm*, 124, 2677– 2701, https://doi.org/10.1029/2018JD029524, 2019.

770   Mishchenko, M. I., Hovenier, J. W., and Travis, L. D. (Eds.): Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications, Academic Press, chap. 14, 393–416, 2000.

Morcrette, C., Brown, K., Bowyer, R., Gill, P., and Suri D.: Development and Evaluation of In-Flight Icing Index Forecast for Aviation, *Weather and Forecasting* 34, 3, 731-750, doi:10.1175/WAF-D-18-0177.1, 2019.

775

O'Connor, E. J., Illingworth, A. J., & Hogan, R. J.: A technique for autocalibration of cloud lidar. *Journal of Atmospheric and Oceanic Technology*, 21, 777–786, 2004.

Ojala, M. and Garriga, G.C.: Permutation Tests for Studying Classifier Performance, *Journal of Machine Learning Research*,
780   11, 1833-1863, 2010.

O'Shea, S. J., Choularton, T. W., Flynn, M., Bower, K. N., Gallagher, M., Crosier, J., Williams, P., Crawford, I., Fleming, Z. L., Listowski, C., Kirchgaessner, A., Ladkin, R. S., and Lachlan-Cope, T.: In situ measurements of cloud microphysics and aerosol over coastal Antarctica during the MAC campaign, Atmos. Chem. Phys., 17, 13049–13070,
785   https://doi.org/10.5194/acp-17-13049-2017, 2017.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J.: Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, *12*, pp.2825-2830, 2011.
790

Rew, R. and Davis, G.: NetCDF: an interface for scientific data access, IEEE Comput. Graph. Appl., 10, 76–82, https://doi.org/10.1109/38.56302, 1990.

Ricaud, P., Del Guasta, M., Bazile, E., Azouz, N., Lupi, A., Durand, P., Attié, J.-L., Veron, D., Guidard, V., and Grigioni, P.:
795   Supercooled liquid water cloud observed, analysed, and modelled at the top of the planetary boundary layer above Dome C, Antarctica, Atmos. Chem. Phys., 20, 4167–4191, https://doi.org/10.5194/acp-20-4167-2020, 2020.

Atmospheric
Measurement
Techniques
Discussions

EGU

Royer, P., A. Bizard, L. Sauvage, and L. Thobois: Validation protocol and intercomparison campaigns with the R-MAN510 aerosol lidar. Proc. 17th Int. Symp. for the Advancement of Boundary-Layer Remote Sensing, Auckland, New Zealand, 2014.

800

Rossum, G.: Python reference manual, Centre for Mathematics and Computer Science, Amsterdam, Netherlands, 1995.

Scott, R. C., and Lubin, D.: Unique manifestations of mixed-phase cloud microphysics over Ross Island and the Ross Ice Shelf, Antarctica, *Geo. Res. Lett.*, 43, 2936– 2945, doi:10.1002/2015GL067246, 2016.

805

Sotiropoulou, G., Vignon, É., Young, G., Morrison, H., O'Shea, S. J., Lachlan-Cope, T., Berne, A., and Nenes, A.: Secondary ice production in summer clouds over the Antarctic coast: an underappreciated process in atmospheric models, Atmos. Chem. Phys., 21, 755–771, doi:10.5194/acp-21-755-2021, 2021.

810   The pandas development team: Data structures for statistical computing in python, McKinney, Proceedings of the 9th Python in Science Conference, Volume 445, 2010.

Tuononen, M., O'Connor, E. J., and Sinclair, V. A.: Evaluating solar radiation forecast uncertainty, Atmos. Chem. Phys., 19, 1985–2000, doi:10.5194/acp-19-1985-2019, 2019.

815

Van Tricht, K., Gorodetskaya, I. V., Lhermitte, S., Turner, D. D., Schween, J. H., and Van Lipzig, N. P. M.: An improved algorithm for polar cloud-base detection by ceilometer over the ice sheets, Atmos. Meas. Tech., 7, 1153– 1167, doi:10.5194/amt-7-1153-2014, 2014.

820   Van Der Walt, S., Colbert, S. C., and Varoquaux, G.: The NumPy array: a structure for efficient numerical computation, Comput. Sci. Eng., 13, 22–30, https://doi.org/10.1109/MCSE.2011.37, 2011.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, 825   R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors: SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python, arXiv [preprint], arXiv:1907.10121, 11 December 2019.

Atmospheric
Measurement
Techniques
Discussions

Vignon, É., Traullé, O., and Berne, A.: On the fine vertical structure of the low troposphere over the coastal margins of East
830 Antarctica, Atmos. Chem. Phys., 19, 4659–4683, doi:10.5194/acp-19-4659-2019, 2019.

Vignon, É., Alexander, S. P., DeMott, P. J., Sotiropoulou, G., Gerber, F., Hill, T. C. J., et al.: Challenging and improving the
simulation of mid-level mixed-phase clouds over the high-latitude Southern Ocean. *Journal of Geophysical Research:
Atmospheres*, 126, e2020JD033490, doi:10.1029/2020JD033490, 2021.

835

Zhang, D., Vogelmann, A., Kollias, P., Luke, E., Yang, F., Lubin, D. and Wang, Z.: Comparison of Antarctic and Arctic
single-layer stratiform mixed-phase cloud properties using ground-based remote sensing measurements, *J. Geo. Res.
Atm*, 124, 10186– 10204, doi:10.1029/2019JD030673, 2019.

840

Figures

845

28
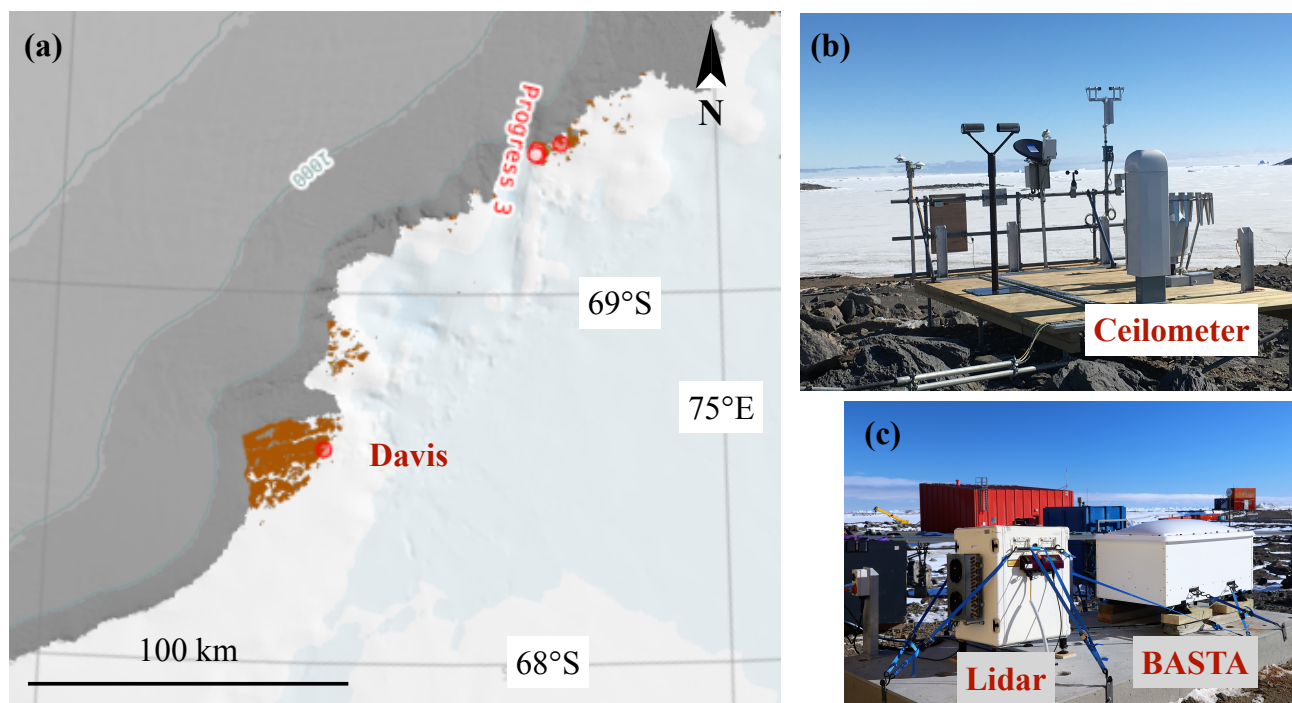
Figure 1: (a) Location of Davis in East Antarctica; the basemap was extracted from the Antarctic Digital Database from the Scientific Committee on Antarctic Research (SCAR) from the British Antarctic Survey; rocky outcrops are shown in brown and elevation as contour lines every 500 m; (b) Photograph of the Vaisala CL51 ceilometer (foreground) installed on the meteorological platform together with other instruments not used in this study; credit: Andrew Klekociuk, Australian Antarctic Division; (c) Photograph of the W-Band radar 850 (BASTA) and the Raman depolarisation lidar mounted on a dedicated concrete slab; credit: Simon Alexander, Australian Antarctic Division.
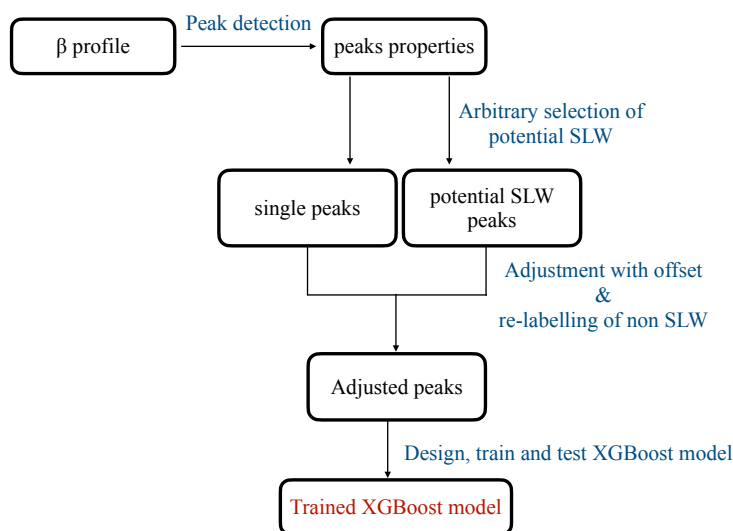
855   Figure 2: Flowchart describing the pre-processing stage in our data-driven algorithm: data preparation, model development, training, and testing.
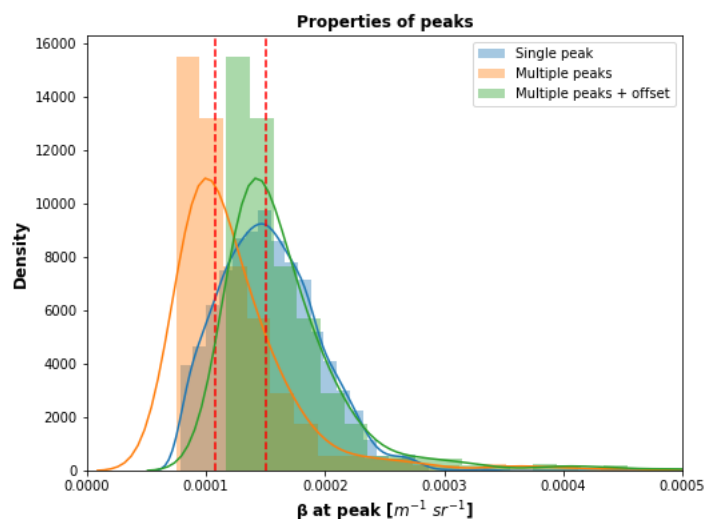
860   Figure 3: Distributions and Kernel Density Estimates of values of attenuated backscatter for identified peaks. For profiles with a single peak identified, these are labelled in blue (4,346 datapoints), while profiles including multiple peaks are shown in orange (2746 datapoints). Vertical dashed red lines indicate the median values of the single and multiple peak distributions. Adjusted multiple peaks (multiple peak attenuated backscatter values + offset) are shown in green.
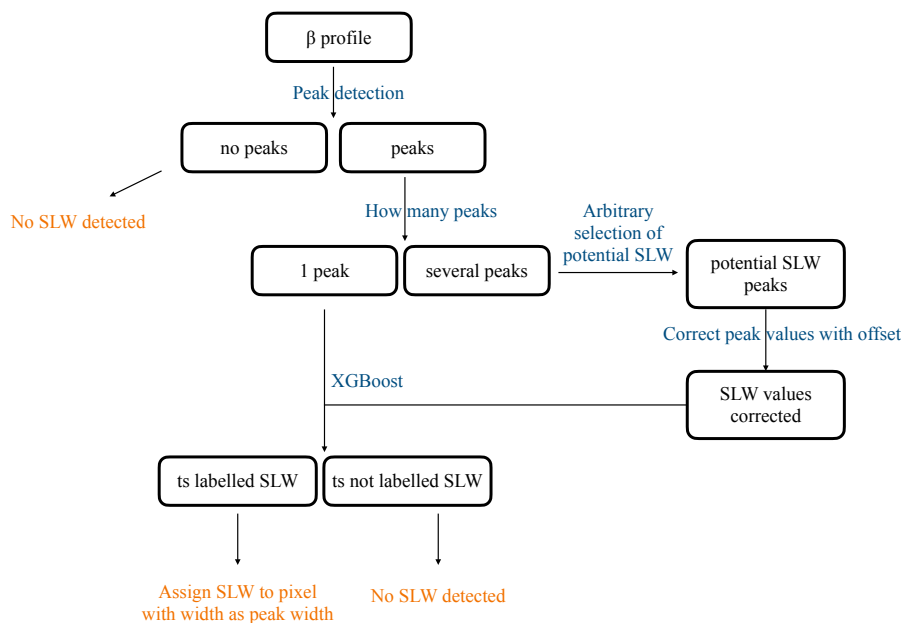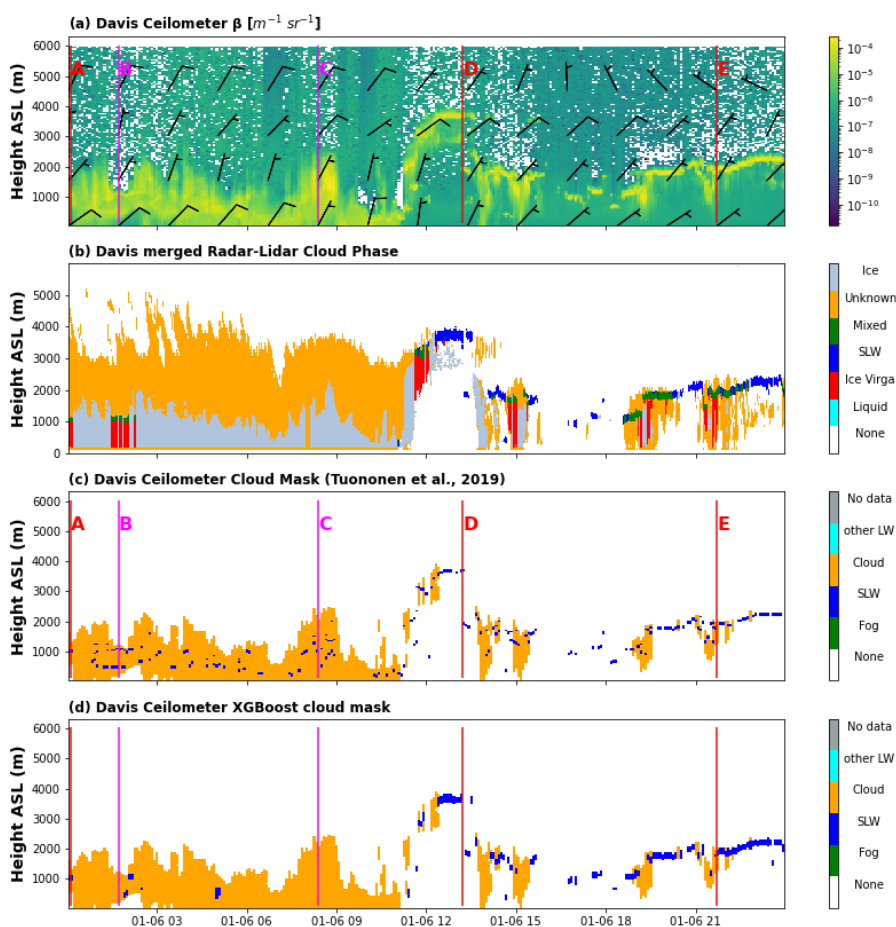
865

Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

Figure 4: Flowchart describing the second phase of the data-driven algorithm: SLW detection as part of the cloud phase mask algorithm.

870

Figure 5. For the selected day of the 6th of January 2019: (a) Calibrated attenuated total volume backscatter and wind barbs (wind data extracted from ERA5); (b) Cloud phase mask following the algorithm adapted from Alexander and Protat (2018); (c) Cloud mask based on ceilometer data using arbitrary classification following the approach from T19; (d) Our newly proposed data-driven cloud mask based on ceilometer observations and using a trained extreme gradient boosting method. Selected vertical profiles have been selected (A, D, and E in red correspond to identified SLW or mixed phase, while B and C in fuchsia correspond to non-SLW occurrences).

880
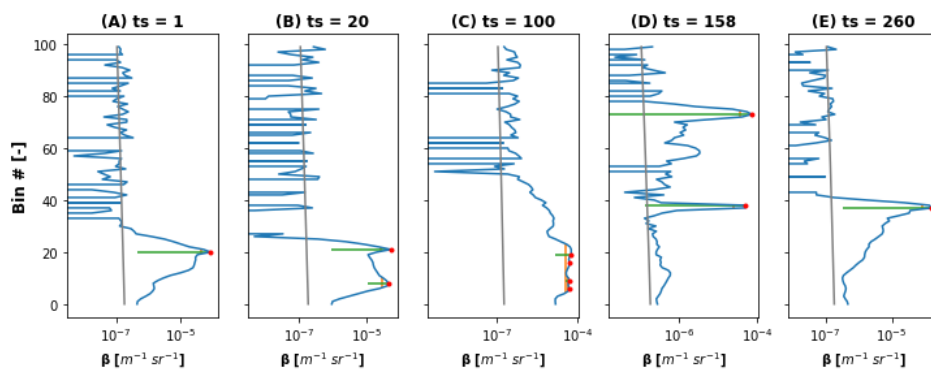
Figure 6. For the 6th of January 2019: selected vertical profiles of backscatter (A, B, C, D, E). Timesteps are indicated as "ts" and can be converted to time by multiplying by 5 min. Peak are identified with a red dot, peak widths with a vertical orange line and peak prominences with a horizontal green line. The theoretical molecular backscattering was computed following the equation as found in Kuma et al. (2021) and is shown on each subpanel as a solid grey line.
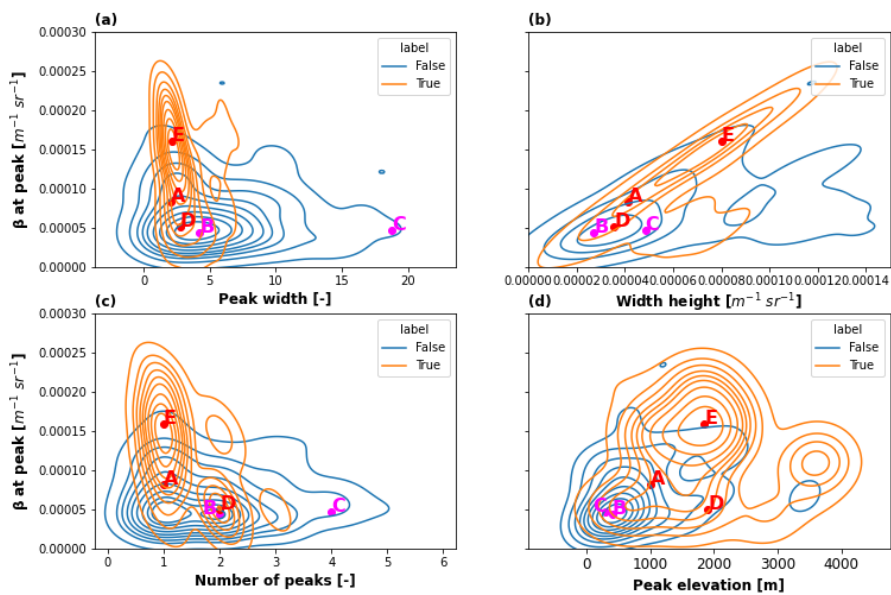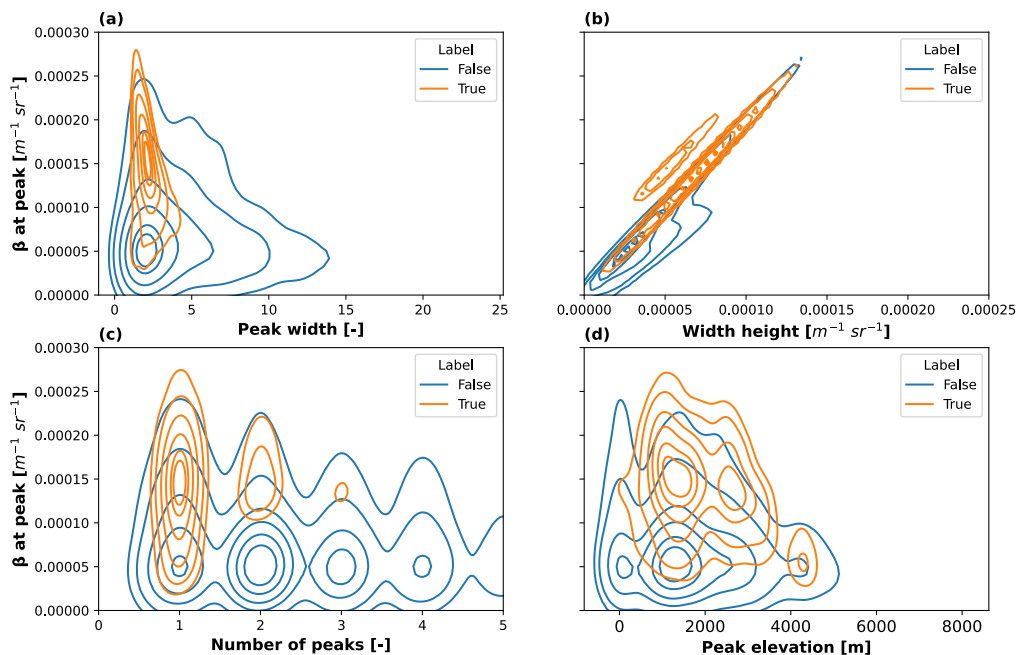

Figure 7. Scatter plots showing the distribution of peak average properties for the 6th of January 2019: (a) attenuated backscatter versus peak width; (b) attenuated backscatter versus peak width height; (c) attenuated backscatter versus the number of peaks within the profile; (d) attenuated backscatter versus peak altitude above ground level. Isolines are shown with a spacing of 0.1 and the label "True" corresponds to SLW or mixed phase observed by the reference radar-lidar mask, while "False" indicate no detection by the reference

mask. Selected vertical profiles from Figures 5 and 6 are also shown with red dots (A, D and E: SLW occurrences, e.g., "True") and fuchsia dots (B and C: non-SLW occurrences, e.g., "False").

900

Figure 8. Scatter plots showing the distribution of peak average properties for full period of the PLATO observations (November 2018 to February 2019: equivalent to a total of 11,327 datapoints): (a) attenuated backscatter versus peak width; (b) attenuated backscatter versus peak width height; (c) attenuated backscatter versus the number of peaks within the profile; (d) attenuated backscatter versus peak altitude above ground level. Isolines are shown with a spacing of 0.1 and the label "True" corresponds to SLW or mixed phase observed by the reference radar-lidar mask, while "False" indicate no detection by the reference mask.

910

Atmospheric
Measurement
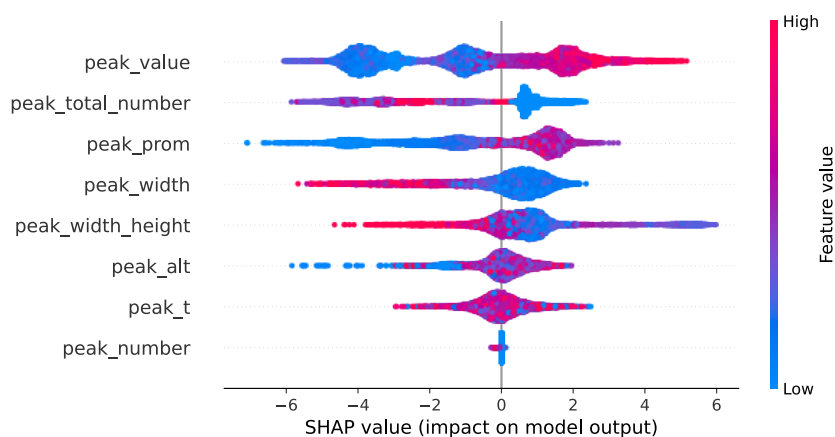Techniques
Discussions

Figure 9: Distribution of Shapley values (SHAP values) calculated by TreeExplainer applied to our XGBoost model. Features are ranked from the most important (top of the list, e.g. peak_value) to the le important (bottom of the list, e.g. peak_number). The normalised feature values are shown with a blue to pink gradient as indicated by the right-hand side colour bar.

915