egusphere-2023-1085

Supercooled liquid water cloud classification using lidar backscatter peak properties

Authors' response to referee comments

We would like to thank the two anonymous referees for their valuable comments which we believe have improved the quality of the work and the manuscript. This work also received comments from an internal reviewer at the Australian Bureau of Meteorology, who we also thank for their valuable input.

The main changes are summarised here:

1) The most significant change in the revised manuscript is the new procedure and terminology for the reference cloud mask classification. We now use the terms supercooled liquid water cloud (SLWC), supercooled liquid water containing cloud (SLCC), mixed phase cloud (MPC), warm liquid water cloud (WLWC) and ice cloud (IC). The new reference mask classification is discussed in detail in Section 2.2.1 in the revised manuscript. Liquid cloud, MPC and IC are distinguished according to VDR thresholds of 0.1 and 0.4, which were determined following a sensitivity test on the values of VDR for all cloudy bins in our MPL dataset (shown in the new Figure 1). For liquid clouds, SLWC and WLWC are then distinguished using WRF air temperatures. The updated reference mask is described in detail in the revised manuscript in lines 120-136.
2) The XGBoost algorithm was retrained against the new reference mask, and now performs multiclass classification of backscatter peaks as SLWC, WLWC or neither (i.e. MPC/IC). Model setup was changed to include further hyperparameter testing to improve performance and avoid overfitting.
3) While model performance is discussed in Section 3.1, performance metrics relating to SLWC are the focus since the goal of this study is to use ceilometers to detect SLWC occurrence, not to provide a comprehensive cloud phase categorization mask. Given the implicit uncertainty in using a lidar-only reference mask to estimate MPC occurrence, we do not focus on the performance of MPC classification.

A tracked changes document (made with latexdiff) is attached along with the revised manuscript.

Reviewer comment 2

This review addresses the manuscript of Whitehead et al., 2023, entitled “Supercooled liquid water cloud classification using lidar backscatter peak properties”. The study deals with the identification of layers of liquid water based on machine learning techniques applied to observations of a polarization-capable micro pulse lidar (MPL). While still not being super-familiar to the emerging world of machine learning, I was able to get a bit of more impressions on how AI might assist in future data analyses. Central instrument of the study is, besides a toolset of existing Python-based machine learning APIs, a polarimetric MPL lidar system which was operated at Christchurch, CA, NZ. Three retrievals were cross-evaluated against each other. The default one, provided by the MPL software, one ML retrieval for the Antarctic site of Davis, and a newly retrieved one for the site of Christchurch.

As far as I got, the default MPL retrieval was set as the reference dataset. It was found that the ML dataset for Davis had worse scoring compared to the retrieval based directly on observations at Christchurch.

Being allover well structured and well written, the study in the end appears somewhat inconclusive. To my impression, the conclusions drawn are not suited to provide guidance for future studies. I’ve learned that machine learning classifications cannot be transferred between different sites. I wonder if this is a good basis for any future (comparative) studies. How should valid scientific conclusions be drawn when the underlying datasets are individually tuned to single sites? What do the authors think about this issue? Isn’t it more appropriate to use a physics-based retrieval which just requires a well-calibrated instrument? All further retrieval steps would then just be determined by the atmospheric state, without the inclusion of tuned ML-based decision trees.

Anyway – given the good structure and rather complete content, I consider the study as suited to be published in AMT. I nevertheless have a series of remarks and questions, which I would like the authors to reply on and to consider in the revised version of the manuscript. A second round of revision appears to me to be necessary in order to evaluate whether the identified issues/remarks could be accurately addressed.

Major comments:
I was missing an overview on the standard procedures of cloud detection. E.g., the widely used synergistic cloud retrievals such as ARSCL or Cloudnet use a combination of Att. BSC. threshold and gradient (https://doi.org/10.2172/1808567, https://joss.theoj.org/papers/10.21105/joss.02123 ... both just require the lidar for the liquid detection) and appear to be quite successful with this. In addition, both of these retrievals use different att. BSC thresholds which demonstrates that there's not the one and only solution. This information might be relevant to discuss, e.g., in line 101 of the manuscript.

R2C1: As per the reviewer’s suggestion, more detail was added to Section 1 at line 58 to describe the earlier methods:

“Operational networks of comprehensive observing systems, such as the Atmospheric Radiation Measurement (ARM) Climate Research Facility (Mather and Voyles, 2013) and Cloudnet (Illingworth et al., 2007), use synergistic radar-lidar algorithms to retrieve cloud properties including cloud phase. Within the Cloudnet retrieval, liquid water detection is based on empirically-derived thresholds of lidar attenuated backscatter (Hogan et al., 2003; Illingworth et al., 2007) and in recent versions, the attenuated backscatter profile shape (Tuononen et al., 2019; Tukiainen et al., 2020). “

Note also that Guyot et al. (2022) compared the new machine-learning algorithm to the Cloudnet retrieval (Tuononen et al., 2019) and found the new technique outperformed the previous method, as described in Section 1.

Equation 1: This expression is only true for an ideal lidar systems with known system constants, the absence of any cross-talk between the channels and a 100% perfect polarization state of the emitted light. A general treatment of the depolarization calibration procedure is described by https://amt.copernicus.org/articles/9/4181/2016/.

R2C2: We use the same definition of depolarization ratio for cloud phase determination as has been used in previous studies (e.g. Guyot et al., 2022; Lewis et al., 2020).

Lines 148-166: Multi layers. I don't quite get what the offset is between two subsequent layers. I'd calculate the actual offset between Q2prime and Q2second 1.1e-4-4.4e-5, which yields 6.6e-5. Can the authors please clarify? I also don’t understand the statement about the extinction in line 154. Extinction is not in units of m^-1 sr^-1.

R2C3: This section first reviews the preliminary analysis performed by Guyot et al. (2022) to compare the returned backscatter properties for primary peaks (first layer) and secondary peaks (all higher layers). In that study, their data showed that backscatter from primary peaks was higher than backscatter from secondary peaks, and that this was statistically significant (see Figure 3 from Guyot et al.). To ensure a fair comparison across primary and secondary peaks, the values of backscatter magnitude for secondary peaks were adjusted by adding an offset, which was calculated as the difference in the medians of the distributions. For our dataset, we compared the distributions of primary and secondary peaks magnitudes (shown in Figure 2) and found that they were not significantly different, so we chose not to adjust the secondary peaks.

The statement about extinction was rewritten at line 161:

“This was hypothesised to be the average reduction in lidar backscatter due to extinction from the lower layer(s)”

Minimum detection height: What is the minimum height for cloud layer detection? In Fig. 3e, it seems as if there is some certain gap between the ground and the first cloud bases. At least for the Arctic, low-level stratus clouds with a base below 100 m were recently highlighted to be a challenging but important puzzle piece in the Arctic cloud puzzle (Griesche et al., 2024; https://doi.org/10.5194/acp-24-597-2024). See the second case study in the manuscript, which indicates that there are issues near to the surface.

R2C4: The following line was added to Section 2.1.1 at line 86:

“The minimum range and detection height of the MPL is 100 m.”

Furthermore, the following was added to the conclusion at lines 518:

“Additionally, it should be noted that the minimum detection height of the MPL is 100 m. Therefore, the MPL analysis in this work potentially misses low-level cloud, which Griesche et al. (2024) identified as an important polar cloud regime for future observational studies.”

3f: Do the authors have an explanation for the gap in SLCC and WLCC at 0°C? Looking at Fig. 2f, this gap is not so pronounced in the overall LCC distribution. This discussion could be added to lines 182-183.

R2C5: We don't have an explanation for why this gap appears, although it could appear due to a difference in the normalization of the SLWC and WLWC KDE plots in Figure 4f. In the revised manuscript, a comment on the mode (i.e. peak in the KDE plot) of each SLWC and WLWC distribution was added to lines 281-283, as suggested.

Case studies 1&2, Figs 4 and 5: Based on the case studies I get really puzzled. The performance of the G22 retrievals is visually just really low, especially for the low-level clouds in the evening of case study 2 (Fig. 5). On the other hand, the reference VDR retrieval misclassifies high aerosol loads as liquid cloud in case study 1 (Fig. 4). Intuitively, I would presume that both issues could be mitigated just by using a simple combination of threshold value and gradient. Why the machine learning?

R2C6: Following reprocessing of the reference mask and retraining of the G22-Christchurch model, SLWC detection improved.

Lines 476-477: After applying the SHAP filtering, three parameters remain as relevant: temperature, peak prominence (BSC threshold) and peak width (gradient). It's funny to see that the cleaned version of the G22 retrieval just condenses to the same parameters as are used in the standard threshold/gradient methods.

R2C7: We consider this a good sign – the SHAP analysis shows that the most useful features are physically interpretable, which increases our confidence in the XGBoost model.

Figure 6: Would be nice to have a version of frequency vs. temperature. This could help to evaluate the impact of specular reflection on the retrievals, as specular reflection (false liquid) should be most prominent in the temperature range between -10 and -20°C.

R2C8: KDE plots showing the temperature distributions of peaks (and therefore cloud frequency as a function of temperature) are already presented in Figures 3f for liquid and non-liquid, and Fig. 4f for SLWC and WLWC.

Final statement in the conclusion could be added: Which retrieval will the authors use in future for their studies? Can the authors make a decision and motivate it?

R2C9: We added the following two lines to the conclusion, at line 501:

"For future work, such as the incorporation of this retrieval method in ALCF, the default XGBoost model will only use peak width, prominence and temperature as inputs, while a model using width, prominence and altitude as inputs will be an alternative for cases when temperature data is unavailable."

and at line 531:

“The G22-Christchurch model and algorithm will be incorporated in the next version of ALCF, so that future work can apply this retrieval technique to other lidar and ceilometer datasets.”

Minor comments:

Line 15: There's actually not only collision freezing. I suggest stating that INP needs to be involved in heterogeneous nucleation of ice at temperatures between 0°C and -40°C. Citation of Hoose and Möhler (2012, https://doi.org/10.5194/acp-12-9817-2012) should do the job.

R2C10: The revised manuscript now reads on line 14:

“Heterogeneous nucleation of ice in clouds occurs between -40 °C and 0 °C when SLW droplets interact with ice nucleating particles (INPs) such as dust and other aerosols, or other ice particles (Hoose and Möhler, 2012).”

Line 94: really just 21? I guess it's 61, isn't it?
https://www2.mmm.ucar.edu/rt/amps/information/configuration/configuration.html

R2C11: For the NZ grid used in this study, there are 21 levels.

Line 129: Existence of ice at T above 0°C. The statement is actually not quite true. Melting of ice depends on the dewpoint temperature. It can well exist at higher temperatures as long as dewpoint temperature (wet-bulb or ice-bulb temperature) is below 0°C. E.g., https://doi.org/10.1175/JAS-D-20-0353.1 Thus the WLCC liquid water statement should be treated somewhat relative.

R2C12: It is correct that ice can exist at temperatures above 0°C before the onset of melting. This section was rephrased and the line “ … since ice cannot exist above 0°C” was removed. See also R1C12.

Frequently, the term ‘cloud’ appears at positions where it should actually be placed in plural form (e.g., lines 16, 29, 48). Do variants of the English language exist, where ‘cloud’ is both, plural and singular? Or are these just typos?

R2C13: These examples sounded natural to the author, but were nonetheless changed to avoid ambiguity.

Lines 183 and 250: use citep

R2C14: Fixed.