

Review of amt-2022-10

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

March 2022

Overall comment

New algorithm has been developed to detect supercooled liquid water (SLW) by just using ceilometers. New method is compared with the existing method to observe SLW with more extensive observational data set and existing algorithms to identify SLW with only ceilometer. Authors have developed new method, utilising machine learning, that perform better compared to the existing algorithm for simple instrumentation. Manuscript is well written and mainly clear - some specific clarifications are requested in the followed section in detailed comments. I am suggesting this manuscript will be accepted after minor revision (mainly clarifications).

Detailed commets

How does authors see the potential of using this method elsewhere i.e would it require location-specific training set to detect SLW? Would be interesting to see test results for other location with / without specific training set. In addition, two models - trained with location specific data - could give different results on the same attenuated backscatter profile, right? Could authors elaborate on these aspect a bit in the manuscript. Would this cause some problem in some applications? Is this something to be accounted for?

line 49: Please add reference for: “Typically, a depolarization ratio below 10% is characteristic of SLW clouds, while higher values are produced by ice particles”

line 147: it is stated: “(2) remove noise by applying a noise removal algorithm and subsampling the data to 5 min, 50 bins;”

Could you please describe what kind of noise removal is applied – how it is done

– and same for bin sampling. More information is needed.

line 148: “(3) calibrate the attenuated backscatter using the approach of Hopkin et al. (2019)”

Please clarify if this is done after subsampling and if subsampling is generating some effect to the method?

line185: Please clarify how ice virga is defined?

line 257: It is stated that “the width of the peak must be < 4 ,”

It remains unclear, what units are in case of 4? Range gates (4*50m)? Please clarify, for example stating “, corresponding to X meters”.

line 249: “extinction other than molecular in the lower levels, cannot be directly compared in terms of backscatter values to peak”

Can you please clarify the terminology usage, why “extinction” and not “attenuation”? It is also stated that there is no other extinction than molecular – how about effect of aerosols?

Line 258: Does this “multiple peaks” -group consist all cases where number of peaks > 1 ? Did you check if there is any difference between beta value in cases of 2nd peak and 3rd peak? Is “multiple peaks” usually only 2 peaks. Will there rise some implications in case of 3 peaks?

Lines 265-270 “The difference between the median value of the single peak distribution and the multiple peak distribution can be calculated and is equal to $4.20 \times 10^{-5} \text{ m}^{-1} \text{ sr}^{-1}$.”

It remains unclear how this was calculated, please clarify. Is there difference between 2nd and 3rd peaks (see previous comment)?

Section 2.3 Enhanced data-driven ceilometer cloud phase mask

More information is needed in this section. Please clarify and describe:

- line 282: what are these “learners”? Please clarify. - How are the training and validation data sets selected? How long training set? How long validation data?

- Can you please tie these ML world terms into observations / data used. Many aspects remains unclear for the people unfamiliar with machine learning and it would be impossible for people to reproduce this algorithm. Any comment on this aspect?

figure 5: would it be possible to show the depolarization values in this plot? Would be interesting to see the data especially in the first half of the data period when T19 method gives lots of SLW detection and reference not.

line 435: How good is the reference in the first place? This reference method could be described in more detail as it remain unclear how the reference combi-

nation algorithm is working. Is this method seen as reliable reference? Please offer some references.

line 538: "raw" means different things for different people. I suggest to clarify, do you mean the data in output files or "attenuated backscatter" - meaning the data after calibration. Is it necessary to calibrate the signal before use?

line 539: Just out of curiosity: how many ceilometers there are in Antarctica where no other potential devices are installed?

line 561: "The new ceilometer algorithm described herein has been developed at the Bureau of Meteorology and is not publicly available."
For what use this algorithm is developed if it is not publicly available and cannot be reproduced based on this manuscript?