Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC2
https://doi.org/10.5194/amt-2022-10-RC2, 2022

Comment on amt-2022-10

Anonymous Referee #2

Referee comment on "Detection of supercooled liquid water containing clouds with ceilometers: development and evaluation of deterministic and data-driven retrievals" by Adrien Guyot et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-10-RC2, 2022

The manuscript presents a new machine learning approach for the classification of supercooled liquid water from ceilometer observations. Based on three months observation ins the Arctic, the approach shows improved performance compared to a previous method which also analysis ceilometer profile data. As reference, the authors use a mask derived from a combination of radar and depolarisation lidar observations. The study is nicely presented with convincing scientific quality. To highlight applicability of the novel tool at the large number of ceilometer data being collected globally, the authors could give perspective on the expected performance in other geographical settings and maybe measurements from other ceilometer types. The manuscript can be published after a series of minor comments are addressed.

Line 180: What is the vertical and temporal resolution of the RMAN cloud classification? How is this alight to the ‘re-gridded’ data of Radar and ERA5? How can a classification be ‘interpolated’ or ‘averaged’?

Line 208: Not clear what is meant by ‘how to label the 50 m bins.’ Please rephrase.

Line 209: So only one bin is classified as liquid water?

Line 211: In line 195 you state that the cloud phase mask utilises Ceilometer data only. But now you state that SLW and liquid water is differentiated according to the reanalysis temperature profile. This seems contradictory.

Line 239: if you say minimum peak width is 50m , this means only one range bin as you are operating an a grid of 50m?

Line 241: Where do you define the peak width? At half maximum or base?

Line 244: 'Lowest' peak defined by peak magnitude or altitude?

Line 245: Again, you are using ERA5 temperatures. I think you need to be careful calling the algorithm to utilise "ceilometer data only"?

Line 254: rephrase "For single peaks, SLW data-only were selected based on the Boolean condition defined using the radar-lidar cloud mask". What is meant by "data-only"?

Line 255: remove "arbitrary"? Your conditions have an empirical basis.

Line 257: what is meant by "width < 4"? Bins? Maybe better to use width in units of meters?

Line 265: Why is the multiple-peak distribution so narrow? Should there not be a dependence on the order of peak in the profile? i.e. could be not expect the peak at the lowest altitude in a multiple-peak profile to resemble the signature of a single-peak? Do you account for the order or altitude of the peaks?

Line 274: how often do you find this mismatch between peak-criteria and cloud classification mask that leads to an adjustment of the "true" indicator? What does this mean physically?

Caption Figure 4: Introduce meaning of 'ts'.

Line 330: This seems like an artificial problem. The masks are created based on higher-resolution data. Why would you create a 50m vertical resolution grid for the ceilometer-based mask if the observations have a resolution of 10m? would it not be more appropriate to map all data to the same vertical resolution in the beginning so they could now be compared more easily?

Line 334: same question for the temporal resolution.

Line 352: introduce meaning of confusion mask indicators

Line 357: state clearly what you are referring to with the term "prediction". Is true negative the case when the mask correctly indicates the absence of SLW? Then why call this "wrong prediction"? Also, if "false positive" refers to the mask wrongly assigning SLW, then why would you call this "wrongly indicating a correct prediction". Please clarify this paragraph.

Line 388: what is the number of samples in the training data? Is this stated in the methods section?

Line 411: Of course figures should be explained when they are being discussed, but please avoid repeating content of figure captions in the text.

Line 424: How did you evaluate presence of fog?

Line 444: Given you are using various different products, please use consistent labels throughout. E.g. in a similar way you are using T19, please use one label (such as XGBoost) for the "new algorithm". Also, please use one consistent label for the reference data. Right now, the reader can get easily confused. E.g. here I am wondering if "a data-driven threshold approach" has already been introduced or if this is yet another method.

Line 473: So the thresholds are not actually "arbitrary", but rather empirical values determined based on the previous analysis. The fact that they work well for your dataset is hence not surprising. It would now be the next step to assess whether these thresholds are more widely applicable, e.g. to perform SLW detection for a different time period or different location.

Line 484: After presenting these values, please put results into context e.g. to performance of the other approaches.

Line 507: This seems to contradict your statement from line 453: "The value of β at peak is directly correlated to the peak width height, making that feature redundant." Please explain.

Line 509: If peak temperature is not an important predictor, would it be possible to omit the ERA5 data and work solely on ceilometer observations as input?

Line 562: You are using software and algorithms developed elsewhere, yet you are not intending to share the code? Especially as you are claiming your algorithm has better performance than an existing approach of T19, it would be important to the community to be able to test your algorithm and verify your findings.

Figure 6: how is the "baseline" determined based on which you quantify the "peak prominence"?

Figure 7 and Figure 8: these are not a "scatterplots" because the individual sample pairs are not shown. Rather you are comparing isolines for the two cases.