Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC2
https://doi.org/10.5194/amt-2021-60-RC2, 2021

Comment on amt-2021-60

Anonymous Referee #2

Referee comment on "Evaluating cloud liquid detection against Cloudnet using cloud radar Doppler spectra in a pre-trained artificial neural network" by Heike Kalesse-Los et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-60-RC2, 2021

The paper investigates the value of using Doppler radar to infer the presence of supercooled liquid layers in clouds. They test the performances of a ANN approach proposed by Luke et al. 2010 on a completely different dataset. As such the study is interesting because it tries to establish how ``portable'' these methodologies are when moving to different cloud regimes. However the paper lacks further analysis and therefore needs substantial improvements before being published.
See suggestions below.

Major comments:

1) The authors state that ``The objective of this study was to check the performance of the ANN trained with the MPACE observations in Luke et al. (2010)''. Isn't this objective a little bit too limited? For instance the conclusion that the algorithm does not work in convection is pretty obvious given the fact that the Luke's algorithm was trained in low turbulent conditions. But how if we train a ANN in convective regions? Is it more succesfull or still we have issues because of the intrinsic problem of convection (i.e. smearing of cloud peak)?
How do the statistical metrics improve overall?

2) I find the description of the four metrics at the end of page 8 a little bit confused. If I follow your guidelines: if precision < 1==> CD overestimation and if recall < 1==> CD underestimation then in all your cases precision and recall are lower than 1 which makes no sense. Please rephrase properly. Same for ``were classified correctly in an absolute and non-balanced way. (overall accuracy)'' not sure what you mean with ``non-balanced'' and what is the meaning of ''overall accuracy in the bracket?''

Maybe mention that recall is the same as ``probability of detection'' (which is a terminology used as well outside the ANN) (right?)
and precision =1-false alarm ratio (right?).
Finally it is not clear to me why the authors have not adopted
variables more commonly used in literature (like ETS) to assess the overall performances.
3) Conclusions 2 and 3 in the abstract are not really corroborated by proper statistical analysis. Which figures/tables prove these statements?
Of course we expect such results but we need to prove them.

4) Apart from the homogenization step to create more coherent liquid layer structures the overall concept underpinning the ANN
methodology is still based on point measurements.
My understanding is that there is more potential in these ANN techniques if we try to exploit local information (e.g. involving vertical and horizontal gradients, especially at cloud top e.g. Silber et al, IEEE 2019; Kalogeras et al., 2021 Remote Sensing) and not simply pixel variables (Doppler spectra). But this is not
explored at all here because we are still using the approach from 2010 (not much novelty). I see the merit of the current study but the authors
should discuss in more depth the way forward.

Minor comments:
Line 19-21 page 4: these instruments have not been used afterwards. I do not see the reason of including them here.
Fig4: Caption: ``Green and red dots near the bottom of the plots'' I cannot see them, where are they????