Review: Evaluating cloud liquid detection using cloud radar Doppler spectra in a pre-trained artificial neural network against Cloudnet liquid detection
By Authors: Heike Kalesse, Willi Schimmel, Edward Luke, and Patric Seifert

General comments:

This study compares the performance of different techniques (#1 lidar (Cloudnet product); #2 ANN applied to radar Doppler spectra) in detecting supercooled liquid water. The ANN method (Luke et al., 2010) trained with radar Doppler spectra data obtained in the Arctic was applied to radar data recorded during the ACCEPT in Cabauw, the Netherlands, 2014. The comparison was firstly conducted to an event with some analysis, and then statistical results for the whole campaign were presented. The results show that the ANN approach outperforms the Cloudnet algorithm in multi-layered stratiform clouds. This is actually expected since Cloudnet target classification is relied on lidar data which have been totally attenuated by the lowest liquid layer. I feel the most interesting part of the study will be assessing the performance of ANN method in presence of convection. I will list more specific comments below.

Overall, this manuscript is well-written and relevant studies are properly cited. I like the Introduction section, since it is quite comprehensive and reads very friendly for researchers who are not very familiar with this topic. Some parts in results section need further clarification. The scientific merit that just showing the (Luke et al., 2010) works well in a different climatology seems weak to me. I think the evaluation in presence of convection could be supplemented by more detailed analysis. Then, this could be a nice paper that I would recommend to be published on AMT.

Major comments:

1) It appears that the method works well as shown by rho_ceilo-CBH, LLH > 85% in Table 2. However, I think it is more important to know at what conditions the 15% fails. Although it is already well known that the liquid peak in Doppler spectrum can be blurred by turbulence, at what extent the turbulence can smear the liquid peak is still not very clear. This may be described by factors related to turbulence, such as spectrum width, velocity, Z, variance of V and so forth. Then, the scientific significance of this study will be improved.

   I think the current explanation is widely accepted knowledge. As the author wrote 'the objective of this study was to check the performance of the ANN trained with the MPACE observations in Luke et al. (2010) on **a new data set**'. The clouds over the Netherlands are definitely more convective than the Arctic, therefore the convective conditions should be well addressed.

   Also, one explanation for the FP of ANN is enhanced SW. To my understanding, the enhanced SW should smear the liquid signature, thus leads to FN. So, turbulence can lead to FP and FN. At what conditions those two 'bad' classes can be expected?

2) Figure 2a. I am curious how well the ANN can predict β and δ. This may also be a part of the

‘evaluation’. The accuracy of estimated β and δ may be as important as the selection of thresholds as presented in Table 1. Have you compared the predicted values with observations? At least with beta observed by the ceilometer.

3) P5 L6.’ Thirdly, ANN liquid predictions for regions with good lidar echo and Cloudnet-classified as non-liquid class, are reclassified as non-liquid.’ This step confuses me. I think it is of importance to know at what conditions the ANN misclassifies lidar-detected non-liquid to liquid. I would not simply ignore this scenario.

Minor commennts:

Figure 1. Numbers for subfigures are missing.

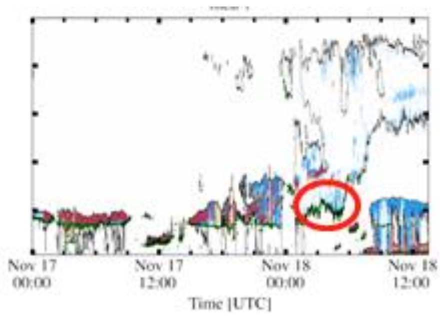
P2 L23 D^6 comes from Rayleigh approximation, which may not be valid for a large fraction of large ice crystals for a cloud radar

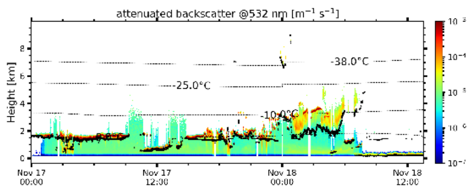
Figure 2 and 3. I suggest overlap the temperature isothermal lines which will greatly help the interpretation of the results.

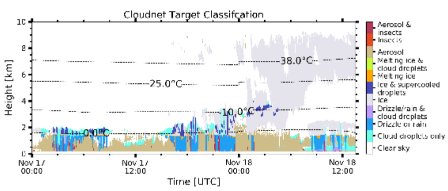
Figure 3. The green circles are hardly identifiable from black cloud edges. Please use the color which is more contrasting with others.

Figure 3. The overview of this precipitation event has already been presented on Figure 2 (b). I suggest the use of smaller yaxis range. Most interesting signatures are below 2 km. The current yaxis scale seems too large to me and the differences among these subfigures are difficult to recognize. The liquid layer above 4 km may deserve a separate figure.

Figure 3. (Although I doubt the reasonability of ‘Thirdly, ANN liquid predictions for regions with good lidar echo and Cloudnet-classified as non-liquid class, are reclassified as non-liquid.’) The region marked by red circle should correspond to ‘good lidar echo’ in Figure 1. Why ANN still identified liquid in this region?







P9 L22. ‘cloud-top layer at −10 °C during 0-6 UTC’. I am confused by this sentence. -10 °C is around 3.5 km. The cloud top during 0-6 UTC Nov 18 is definitely much higher. Do you mean 21-24 UTC Nov 17?

P9 L25. This is interesting. Turbulence favors liquid formation, but may lead to weakened liquid spectral signature if liquid is present. As shown in Figure 1, it is obvious that the SW is enhanced at this layer. However, given the weak signal in deBoer2009 and the rather low temperature, it is very unlikely that they are liquid layers. Could you please present examples of the radar Doppler spectrum in this layer as well as at 8 km 6 UTC Nov 18?

Figure 4. The rain flag is missing.

P10 L1. May not be the ‘Misclassification’. In some cases, e.g. after 7 UTC Nov 18, lidar signals are totally attenuated by the lowest liquid. So, ANN may be correct in the upper layer. Please carefully address this point.

P10 L4. ‘by comparing the predictions to valid Cloudnet liquid detections’. Do you mean the cloudnet product with ‘good lidar echo’? Or regardless of the lidar signal quality?

P11 L7. To my understanding, high ρ ceilo-CBH,LLH for ceilometer-CBH and Cloudnet is expected, because cloudnet uses ceilometer data as input. How is this linked to the sensitivity between lidar and radar? I am confused by the logic.

P11 L9. How the averaging affects the performance?

P13 L10. The first point may explain the difference between radiosonde and cloud/lidar method, but not the reason why the liquid pixel is higher in cloudnet than ANN.

P14 L15. It would be nice to refer the relevant machine learning techniques. For example, the work by the authors (Kalesse et al., AMT, 2019).

Typos:
P9 L25. Nov 18

P11 L7. The high value of rho_ ceilo-CBH,LLH … is expected, because…

P11 L21. Case; resulted