

Atmos. Meas. Tech. Discuss., referee comment RC1
<https://doi.org/10.5194/amt-2021-137-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on amt-2021-137

Anonymous Referee #1

Referee comment on "Using artificial neural networks to predict riming from Doppler cloud radar observations" by Teresa Vogl et al., Atmos. Meas. Tech. Discuss.,
<https://doi.org/10.5194/amt-2021-137-RC1>, 2021

This work describes the development and evaluates the performance of artificial neural networks (ANNs) to diagnose riming using measurements from vertically-profiling Doppler radars.

This is a useful scientific objective. From a remote sensing perspective, knowing when and where riming is occurring can help improve estimates of precipitation rates and reduce uncertainties in retrievals based on radar observations. The work suggests that riming can be diagnosed without using mean Doppler velocity - this is useful because mean Doppler velocity will be perturbed in areas of ascent or descent such as convective cores or atmospheric waves. Finally, detecting and quantifying riming provides information about the microphysical processing experienced by ice-phase precipitation.

The method for development of the ANNs is sound, as is the testing of the ANNs, but I feel there are some shortcomings in terms of scientific quality. The analysis of the test results, particularly for the BAECC cases, seems unnecessarily limited and basic. While several sources of uncertainty are acknowledged, there is only a rudimentary discussion of quantified uncertainties, and there is no quantitative estimate of the propagation of these uncertainties into the results of the study. Additionally, one critical source of uncertainty is not addressed.

Figures are generally well-made and used well to convey information. The narrative overall is good, but there is some text that would benefit from better organization.

See my specific comments below for details. I think it would be worthwhile to provide more quantitative analysis of the BAECC results and make some minor reorganization to the section that is noted below. I'm recommending this be treated as a "major revision".

Specific comments

#####

The method for creating the FR values in the BAECC dataset depends on a mass-dimension relationship for unrimed snow, and neither the uncertainty of this relationship nor its impact on the FR predictions of the ANN are assessed. The mass-dimension relationships for snow particles are highly variable.

The description of the development of the ANN provided in section 2.3.2 would benefit from some reorganization. See my comments below about some explanations about the hyperparameters and loss function that seem to be out of order.

In the discussion of the testing with the TRIPEX-Pol data, there's very little in the way of quantitative comparisons between the results for ANN #1 vs. #2 and W-band vs. Ka-band. The description of the data (section 2.1.2) indicated that the data were regridded onto a common height-time grid. Wouldn't it be possible to make more quantitative comparisons, for example by looking at whether there is a 1:1 relationship among the four possible combinations? I think a quantitative comparison would be a more convincing demonstration rather than the statements that the "predicted riming index is very similar and the features are almost identical" (P13, L13-14).

P3, L2-3: It would be helpful to be clear how "terminal fall velocity" is being defined. Is this the fallspeed (speed at which a particle approaches the Earth's surface) or is the terminal speed (or terminal "velocity", which is relative to the motion of the fluid)? Also, in almost no case will the speed of a single particle be equal to MDV, since MDV is determined by an integration over the Doppler spectrum.

P4, L5: It is not fully clear to me what is meant by "developed models". Is this referring to the ensemble of ANNs?

P5, L9-10: Can you describe why these particular days were selected?

P6, L8: It is not clear here what is meant by "having the same D_{\max} as the observed particles". Since IWC is determined by integration over the size distribution, it is not necessary that the sizes of the observed and unrimed snow particle be the same. Do you mean "for unrimed snow having the same $N(D_{\max})$ as the observed particles"?

P6, L9-10: What was the total number of samples examined, and how many of those had $FR < 1$?

P6, L15-16: Since air below CBH is subsaturated, is the loss of ice mass by sublimation possibly significant?

P6, L18: Was it necessary to assume that MDV is constant between CBH and the surface? Doesn't the KAZR provide MDV profiles at bins below CBH?

P7, L10-15: Then, how are the impacts of these uncertainties quantified? Is this discussed later in the paper?

P7, L24: Not just moments, but also SEW, which is not a moment.

P8, L7: Is the use of "x" in " $\max(0,x)$ " intentional, and does it represent time as it did in P6, L16?

P8, L9-10: Could you provide a more complete description of the error? I'm not sure of the meaning of "This error is defined by the loss function". Is this simply a measure of the difference between the outputs of the ANN and the actual observed values? How is it calculated?

P9, L1-7: This paragraph begins by identifying the number of hidden layers and number of neurons as hyperparameters to be tuned, but the following description of k-fold cv says nothing about how these hyperparameters are tuned. Instead, the role of k-fold cv to produce multiple models for estimating average and variance is described.

P9, L16: Does "layer depth" = number of hidden layers?

P9, L16-19: This seems to be the content that should follow the first sentence of the paragraph starting at P9, L1 (see my comment above).

P9, L17: The RMSE of what?

P9, L20: So, is this the definition of the "loss function" (see my earlier question) and is the RMSE described above on L17?

P9, L22-24, P10, L1-2: There are a few points to ask regarding this paragraph. What is meant specifically by "less well represented"? How can high FR cases be "less well represented" but still be "clearly separated in the input feature space"? The RMSE *is* the error, or at least it is one way of measuring the error. It is probably not necessary to say that it "basically represents the error". Is the performance of ANN #0 actually "much worse"? The "test set RMSE" is only 0.02 larger than that for ANN #2.

P10, L2: Please clarify this. Why would you change the name for a quantity because there are many steps involved in its calculation?

P10, L12-13: I expect this 1.5 hour period on 21 Feb 2014 was not included in the training data. Is that correct? Is it part of the 10% of data that was retained for the testing phase (see P7, L28-29).

P13, L20-21: The differences in sensitivity between the W- and Ka-band radars have not been discussed previously (section 2.1.2). I'm not sure that I understand this comment about the differences in sensitivity. For either the W-band or Ka-band radar, if the backscattering from a radar bin is below the minimum detectable signal of the radar, I would guess that there is no data for that radar bin and there would be no ANN estimate of FR for that radar bin. How would *sensitivity* influence the *accuracy* of the predicted FR?

P13, L24-25: I think this needs some additional details. You seem to be describing the features of a plot, but it is not clear what is being plotted. Also, regarding "These signatures are attributed to changes in the particle density during the riming process", I assume that only applies to the riming signature. Is there an explanation also for the "hook-like" signature associated with aggregates?

P15, L1-2: I see a (weak) suggestion of a hook-like feature in the upper-left panel of Figure 6, but not so much in the other panels. I suggest maybe overplotting a line or an oval to indicate where the hook-like features are located in the panels.

P18, L13-14: I think it is probably necessary to acknowledge, though, that the presence of rimed particles at the surface indicates that riming is occurring *somewhere* in the column, but that it isn't strong evidence of the correctness of the vertical distribution of FR as predicted by the ANNs.

P18, L16: Are you saying that you are confident you can apply the ANN #2 to W-band radar observations and get accurate indications of riming?

P21, L20-25: It seems to me that application to airborne in situ observations (airborne radar plus particle imaging probes) would be a natural extension of this work and could help validate the ANN results for riming aloft.

Technical corrections

#####

P1, L1: SLW has not yet been defined.

P5, L5: "poiting" should be "pointing".

P5, L13: I believe "Snofall" should be "Snowfall".

P5, L25: I believe that "Lee" is not capitalized in this usage.

P7, L18: This is the first use of the term "fully-connected" (the term "fully connected" is used on P8) and the only use of the term "deep neural network". It would be helpful to define them here if they are important to this paragraph.

P10, L13: "17.00" should be "17:00".

P13, L22: "outruled" is more commonly expressed as "ruled out".

P15, Figure 5: Note that the righthand panels are mislabelled as "W-band".

P15, L7-8: Since the first sentence refers to findings in the previous section, it would be clearer to say "the 19 March 2021 Leipzig case".

P15, L9: I think the term "instability" would usually be used, rather than "lability"? Is that the intended meaning?

P16, Figure 6: Please check the arrangement of the panels of the plot. According to the caption, W-band should be on top and Ka-band on the bottom, but this doesn't seem to be the case.