Review of “Dual-frequency spectral radar retrieval of snowfall microphysics: a physically constrained deep learning approach” by Billault-Roux et al., AMT-2022-199

This manuscript presents a novel technique to invert snowfall microphysics from dual-frequency Doppler spectra by using a deep learning approach. The approach is based on an autoencoder-like framework, which links two neural networks together – one reproducing a well-established forward model, while the other is trying to invert the first one. Besides the implementation of the framework, the manuscript also presents its application to a recent field campaign where ground-based cloud radar measurements are compared with airborne in-situ data to discuss the performance of the inversion.

I really enjoyed reading this manuscript, since it does a great job to introduce and explain its neural network approach instead of leaving the reader with a *dark box* like some other manuscripts. To my knowledge, this study is one of the first to apply the autoencoder concept to the inversion of cloud microphysics. The paper is well structured, easy to follow and engages the curiosity of the reader. Due to this fact, this paper could help to pave the way for further neural network assisted inversions of cloud microphysics. Like RC1, I was impressed by the retrieval results and the comprehensive consideration that went in the necessary assumptions for the forward problem.

My comments are only minor in nature and mainly are concerned with some aspect how these necessary assumptions were introduced and the occasionally overconfidence in the proposed method. Albeit powerful, there are certain limitations in the inversion of cloud radar measurements which even neural networks will not be able to solve. Anyway, these comments only concern the presentation and not the study itself and should therefore be easy to address.

Overall, I am convinced that the presented manuscript is a very valuable contribution to the scientific debate and will advance the application of neural networks in atmospheric science. I clearly suggest the manuscript to being published in AMT after my comments and questions have been addressed.

General comments:

1.) Throughout the manuscript you write that your approach can mitigate the ill-posedness of the problem, “*i.e. when several values of x may yield similar outputs y: in such cases, the [direct] retrieval may yield arbitrary outputs” (L152).* However, I am not convinced that your approach is able to do just that. The ill-posedness is determined by the number of measurements and the degrees of freedom and I do not see a way how your approach can overcome this mathematical principle. Where a “direct”, e.g., lookup table-based retrieval may yield a set of possible ***x*** which can explain ***y,*** your approach only yields one unique ***y = f(x)*** relationship, although multiple combinations of ***x*** would still be physically viable. In the first case I obtain a measure of uncertainty while in the latter I am just getting “blind” for other latent variable combinations which also could explain my individual measurement. What you are actually doing during the construction of your decoder (by important sampling of ***x****)* and your encoder (by using specific snowfall cases) is the limitation of the ill-posed problem by well-chosen priors. That is an appropriate strategy to deal with the ill-posedness of the problem but it is not unique to your technique. Maybe your manuscript could be a little bit more modest around these sentences and acknowledge your use of priors for ***x*** (*the need … to assume prior values for x … has led us to explore a different direction, L143).*

This includes an earlier mentioning of the fact that your approach is directly trained on the data of interest and might thus not be directly suitable as a general retrieval for arbitrary X- and Ka-band spectrograms of ice clouds (first mentioned at L402ff). Moreover, while the introduction to the theory of inverse problems (Sec. 2.2) is thorough and almost too detailed (L133-L145), I am missing a short literature review about current studies using neutral networks (e.g., Piontek, MDPI, 2021), especially auto-encoders (e.g., Behrens et al, JAMES, 2022) in climate research.

2.) An aspect that is in general a bit unclear to me is your statement in e.g., L94: *"… an important peculiarity of the encoder's architecture is its ability to leverage the spatial consistency of the radar variables, which reduces the ill-posedness of the inversion problem*" which is iterated several times throughout the manuscript. Unless I have missed it you never mention what you mean with "*spatial consistency*" in the first place and only briefly describe its implementation around L340. This was also mentioned by the other referee, and we only can guess that you refer to the observation how spatial features (e.g., updrafts or fall streaks) are connected throughout adjacent range gates. Furthermore, it is not clear to me how this "*spatial consistency"* can overcome the ill-posedness of the inversion problem in each individual gate since we simply have too few measurements compared to the natural variability of realistic ice crystals (see my point 1). In my opinion, your spatial convolution kernels (which you only use for your encoder) lead to a smoother profile of latent variables (which is desired!) but cannot overcome the underdetermination of the problem (e.g., DFR ambiguity between aspect ratio vs. D_0).

3.) Like RC1, I am missing a discussion how attenuation by hydrometeors (especially at W-band) could introduce biases in your inversion. In your outlook, you could briefly discuss potential approaches to this problem in the framework of neural networks. Here, it might become necessary to train the decoder on full spectrograms instead of treating each range gate independently to capture the gate-to-gate interdependence cause by attenuation effects. Moreover, you promise (in your abstract) to relax constraints on beam matching with your approach, but only seem to consider a potential doppler shift (hopefully independently between X- and W-band) caused by a beam misalignment. The more fundamental limitation caused by a nonuniform beam filling in the context of different radar beam widths (0.53° vs. 1.8°, 50 m vs. 150 m beam diameter at 5 km altitude) are not handled but should shortly be mentioned.

4.) Throughout the manuscript the used definition on the particle diameter is not entirely clear to me and needs a more precise treatment. You introduce D_0 as size parameter of the PSD of Straka (2009) and you call it its "*mean diameter*" in L191. According to Straka (2009), however, D_0 is the median volume diameter of the exponential size distribution if $D = D_{max}$. Furthermore, your fixed relation $D_{eff} = 3D_0$ seems wrong to me for ice particles regardless your definition of D_0. The *effective diameter* regarding solar radiation depends rather on the area size relationship. As pointed out by RC1, your strategy how you relate the aspect ratio with particle mass and size is not clear to me – it appears to be based on the common soft spheroid approximation. Regarding this approach and the discrepancies found between your retrieval and in situ measurements you should consider the remarks made e.g., by Hogan et al (2012). They show how particle shape and orientation influence the mean diameter retrieved by the DWR technique compared to in situ measurements. Likewise, I am missing a more explicit description how the diameter, the area and the terminal velocity are connected. While I am convinced that the authors are aware of these points and have put sufficient care into their implementation, their manuscript lacks the necessary diligence regarding the particle diameter.

Specific comments:

L37: What do you mean when you write: "*the retrieval of snow microphysics from radar variables is not explicit"?*

L47: "[DFR] *can thus be used to identify populations of snow particles with a larger size or density."* How does density influence the DFR sensitivity? I always thought that the DFR technique is inherently insensitive to density of ice particles?

L76: "*especially when turbulent broadening is observed […], a direct computation of the dual-frequency spectral ratio is meaningless*." This statement is too harsh in my opinion. While turbulent broadening can severely hamper the exploitation of the dual-frequency spectral ratio, you should give an estimation at which magnitude its value becomes "*meaningless*".

Fig. 1: A little bit more descriptive caption would help the reader here.

L232: On what observations or studies are these numbers based? Could the choice to give aggregates more weight during the training not also bias your retrieval towards the property of aggregates? The (40/20/20/20%) distribution is thus part of the implicit prior of your approach, correct?

Sec 3.2: Although your accompanying paper Billault-Roux et al (2022) might provide these numbers, a reader might be interested in the size (hours or number of profiles) of your training dataset.

L314: I really appreciated your clear and comprehensible description of your neural network approach. I would be delighted if you could spend 1-2 more sentences on the role of these *residual blocks.* Do we know why or how they facilitate the training process?

L318ff: You mentioned the separation of your measurements into an 80% training, 10% validation and 10% testing data set. In the following, you no longer refer to how you used the 10% validation and 10% testing data set or did I miss something?

Tab 3. Could you elaborate where the choice of 30 for the number of channels comes from? I thought that the channel dimension relates to the radar bands used (X- and Ka-band).

L377-L381: The description of the loss function *O(S, ~S)* is quite complicated and hard to understand. Explain the problem with the difference between normalized spectra and why your loss function consists of two integrals. How can you spot discrepancies in the absolute reflectivity at all when you normalize the simulated and measured spectra with each other?

Fig 4.: While impressive, I would prefer to see the profile of some latent variables like *IWC* and D_0 in panel b) instead of showing the same spectrogram three times. The argument that PAMTRA can be imitated reasonably well (panel c) is already demonstrated in Fig. 3.

Sec 5.3.2 Size parameter: Several times you mention "*cofluctuation"* (e.g., L500) between ground-based retrieved and airborne measured properties. Have you averaged the airborne dataset after the selection of nearby overpasses? Otherwise, I would not expect a good correlation between a ground-based and airborne platform. This could also explain the much higher variability of the RASTA retrieval.

Sec 5.3.4 Aspect ratio: While reading this section I noticed that you never mentioned the average particle orientation in your PAMTRA simulations. As this is obviously fixed, this could be a further origin for the observed bias. Furthermore, I noticed here that you limited your AR to oblate particles (see Fig. B5, last panel). In the presence of rice shaped particles, e.g., needles, bigger biases in your size and IWC retrieval may be expected and could explain the observed discrepancies. Please include this in your discussion.

Typos and wording:

Throughout the text you are using the saxon genitive with objects, some examples:

L11: "*the problem's ill-posedness*" > "*the ill-posedness of the problem*"
L13: "*the retrieval's accuracy*" … "*the accuracy of the retrieval*" or "*the retrieval accuracy*"
L104: *"method's sensitivity" … "the sensitivity of the method"*
L196: *"the radar's properties" … "the radar properties"*

L52: "*comforted through" … "confirmed through"*

L67: "*The scattering regime transition in high frequencies is in principle visible*" … "*The transition of the scattering regime at higher frequencies is visible"*

L124: *"we use as a forward model the radiative transfer code PAMTRA" … "we use the radiative transfer code PAMTRA as a forward model"*

L207: *"leaving to future studies the possible improvements of the forward model" … "leaving possible improvements of the forward model to future studies"*

Tab 1: α_a and β_a should probably be the pre-factor and exponent of the area-size relationship?

L246: This sentence is awkward and hard to follow, please rephrase.

L458: *"Sect. 3.2"* is now *"Appendix A"*