

## Reviewer Response Document

### Reviewer 2

We thank the reviewer for their detailed analysis of our manuscript and the constructive feedback they provided. We have responded to each comment below along with the specific changes we will make to the manuscript.

**Precip phase: Sims and Liu (2015) show a simple scheme for estimating phase – it really depends on wet bulb temperature ( $T_w$ ), not  $T$ , and you've already got the ERA5 data in hand to make the calculation.**

We thank the reviewer for their suggestion and agree that this would be a beneficial technique to incorporate in further iterations of the model with more of a focus on precipitation phase classification. At this point, we feel that incorporating another full column of atmospheric data as an input would likely fall outside of the scope of this current project as it would require an entire new data alignment/preprocessing step, model training and hyperparameterization phase which would be a costly endeavor (in both time and computing resources). However, we have updated the manuscript in Section 2.5, paragraph 2 to include this reference along with a suggestion for further analysis in followup work.

**Undefined magic operations: These need a reference at a bare minimum, preferably with a 1-line “what they do” statement. I can cite: L.161: Adam optimization; L.181: squared errors elbow criterion method; L.214: using dropout; and I would encourage the authors to review the manuscript for other such mystery names.**

We agree that these topics should be described in more detail to make it clear to the reader what is actually being tuned during model development. We now include additional descriptions of each of these phrases along with regularization techniques which were used (i.e. Adam optimization, L2, dropout) where they appear in the manuscript.

**Connect text to the flow chart in Fig. 2b: In particular, it is not clear how Section 3.4 fits into the overall processing workflow.**

We have updated Figure 2.b to highlight where in the model pipeline the clustering occurs, and included a reference to this in the text. We also now include an additional reference to Section 4.2 of the manuscript where the clusters are later used for interpreting feature importances.

**Undercatch: The text keeps alluding to gauge errors, particularly for snow, but never really confronts the beast. Essentially all operational gauges bias low, most acutely for snow. This low bias affects the statistics, and there ought to be at least an organized, if short qualitative statement about this issue.**

We agree that this topic should be discussed in more detail and have now expanded our description of undercatch and the methods we have taken to minimize its impact through additional sentences and references added in Section 2.5, paragraph 1.

**References in the text: These are mostly in a form that is non-standard in my experience (no surrounding parentheses, such as L.16, but not always).**

All references in the text have now been corrected.

**L.50-52: Stating the references here is unnecessary; just point to Table 1.**

We agree and now simply point to Table 1 as suggested by the reviewer.

**L.57: “NRT” is almost always “near-real-time”. This is the only place where it’s used, so just say “post-real-time” and don’t give an acronym.**

We have removed this acronym for clarity.

**L.70-82: The phrasing is awkward; I’d describe the standard situation first and then summarize the deviations.**

We agree that including these results from Ny-Ålesund does not flow well at this point in the manuscript. We have now moved the second paragraph of Section 2.3 to Section 4.2 (to where these details are important for understanding the vertical sub-sectioning analysis we perform).

**Write out all acronyms: TMP, WV, Ze, S/R, P, RF, MLP, CNN, perhaps plus others.**

We thank the reviewer for this comment and have gone through the manuscript writing out all acronyms upon first use.

**L.84-87: The language can be simplified.**

We agree with the reviewer and this section has now been rephrased and simplified.

**L.122-123: If Cold Lake and Ny-Ålesund are excluded, what \*is\* done there?**

The data from these sites are still used as part of the training and testing process for DeepPrecip in Section 4.1 and in the uncertainty quantification component of the paper in Section 4.2 when we consider different subsections of the vertical profile and the influence of

these regions on retrieval skill. This additional data helps the model learn about precipitation events that may be unique to certain regional climates. We do not derive Z-P power laws at these sites because of the limited sample at Cold Lake and because of the limited vertical extent at Ny-Ålesund.

**L.183: “event-types” would be more descriptive as something like “intensity classes”. And, in subsequent discussions, the same terminology should be used whenever intensity classes are referenced.**

We agree and this terminology has now been adopted here and in the Fig. 3 caption.

**L.203-206: Kindly eliminate the redundancy.**

These lines have now been adjusted for clarity.

**L.214: This is the only mention of mixed phase. If it’s going to come up here, it needs to be introduced back when rain and snow events are introduced and related to Fig. 4.**

We agree with the reviewer and now introduce mixed-phase precipitation and the challenges it presents in the first paragraph of Section 4.1, when Figure 4 is introduced.

**L.219-221: This feature isn’t really discernable in Fig. 5.**

We agree, and now include additional commentary in the text describing the biases in the Ze-P relationships across multiple sites on these lines.

**Figs. 4a, 5a: I would suggest bolder separation between the different sites so that it’s easier to see them as separate time series. Also, the graphic needs to be larger. Finally, the horizontal time axis needs better labeling showing time increments.**

Both Figures 4.a and 5.a have been updated with bolder lines to delineate between different stations. We have also increased the line width and visibility of lines in both figures to make them more readable and added further time information to the x-axis of Fig. 4.a, as requested by the reviewer.

**L.276: I think it’s “In all other cases, ...”**

This has now been corrected.

**L.280: In addition to SPW, the abbreviations for reflectivity and Doppler velocity should be defined here.**

These definitions have now been included in the text when first defined.

**Fig. 8: In addition to labeling for bin, the Y axis also needs labeling for height, since this is referenced in the text.**

We agree with the reviewer and have added an additional y-axis for height to Fig. 8.

**L.301: “assumption” isn’t quite the right word; “speculation”?; “prior inference”?**

We have updated this language to "prior inference" as suggested by the reviewer.

**L.305-308: This last sentence needs to be more straightforward. I think you’re suggesting more aggressive use of the un-blanked, or at least, less generously blanked, satellite radar data.**

This sentence has now been updated for clarity as suggested by the reviewer.

**L.323-324: The name usually includes “mission” as “Global Precipitation Measurement (GPM) mission”.**

The word "mission " has now been added on this line.

We thank the reviewer once again for their constructive comments and suggested improvements for further enhancing the quality and clarity of the article.