Response to review by anonymous referee #3

Opening remarks

Warm thanks to all four anonymous referees for taking the time to review our manuscript and for providing valuable and actionable feedback. Their comments and suggestions have resulted in revisions that present and discuss our results and limitations with greater precision, including important qualifications and specifications. We highlight some larger changes motivated by comments from multiple referees.

First, to answer how the provided uncertainty estimates can be applied. We previously illustrated the provided quantiles through uncertainty in an example scene. We now also demonstrate how to apply these quantiles to obtain probabilities for cloud FWP classifications.

More than one comment was concerning too strong claims on achieved horizontal resolution and effective resolution in the manuscript. We have reworded these claims with appropriate qualifications. To motivate that the method provides increased ability to resolve structure, we add a confusion matrix for FWP ranges. This matrix shows that low FWP estimates are never confused with high FWP estimates, and vice versa.

There were comments asking about the general value of achieving retrieval performance down to 10 or 40 g/m2. Now introduced a figure showing the cumulative distribution and cumulative contribution to the total FWP, derived from our simulation database. This helps clarify how sensitivity levels affect individual retrievals and the average global FWP.

Lastly, we have added clarifications, further details, and additional references when discussing other datasets, instrument information sensitivity, and the methods used to create our simulation database.

The remainder of this document directly addresses the referee's comments.

Replies on referee's comments

> The manuscript "The Arctic Weather Satellite, introducing a new wavelength range for ice hydrometeor retrievals" by McEvoy et al. introduces a new frozen water path (FWP) product based on observations from the Arctic Weather Satellite (AWS), which provides the first operational Earth-viewing microwave observations at submillimeter wavelengths. The product is based on a pixel-wise retrieval algorithm using a database of reference hydrometeor properties derived from CloudSat Cloud Profiling Radar (CPR) observations. The dataset provides retrievals for the central portion of the AWS swath and is restricted to latitudes equatorward of $\pm 60^{\circ}$. The retrieval also exhibits increased uncertainties over high-latitude land surfaces.
>
> The manuscript is generally well written and provides an adequate characterization of the dataset and retrieval methodology. However, several statements regarding the accuracy, spatial resolution, and physical information content of the retrieval go beyond what can be established from

the analyses presented in the manuscript. In particular, only limited comparisons against independent observations are provided, making it difficult to determine whether some of the additional small-scale variability seen in the AWS retrieval represents genuine FWP structure or retrieval noise. I therefore think that several claims regarding retrieval accuracy and resolved spatial structure should be removed, qualified, or reformulated.

Since these concerns primarily affect the interpretation and presentation of the results rather than the retrieval methodology itself, I recommend minor revisions.

Comments:

Title: The current title may overstate the novelty of the wavelength range. Submillimeter observations have previously been used for atmospheric and hydrometeor observations from airborne sensors and limb-sounding instruments. I suggest clarifying that the novelty concerns operational Earth-viewing observations, if that is what the authors intend.

- A good point. We have now specified that the sub-millimetre is the wavelength range and that it is novel in the context for operational ice hydrometeor retrievals.
- We opted not to qualify observations as earth-viewing, since limb-sounding can be argued to also belong in this category.

l. 11: Given the limited independent validation presented in the manuscript, I do not think this claim regarding retrieval accuracy is sufficiently supported. I suggest removing or qualifying it. The claim also seems difficult to reconcile with the stated limitations of the retrieval over land surfaces.

- We thank the reviewer for this point, which another referee also raised. To avoid overstating the claim, we have extended the abstract. We now directly concede that a proper validation was not made and why. We also state that the comparisons presented show that performance is better than other passive retrievals, but now qualify that there are issues over ice and intense convective cores over land.

l. 45: Please elaborate on the statement that passive retrievals do not provide new physical insights. It is not clear to me what distinction is being made here between the physical information provided by passive retrievals and that provided by the proposed approach.

- We thank the reviewer for pointing out the ambiguity in our wording. We intended to refer specifically to data-driven "black-box" machine-learning approaches, where assumptions are not necessarily transparent. This is in contrast to methods in which assumptions on, for example, microphysics are explicitly modelled, and therefore can provide physical insights. - We have now added a clarifying sentence to this paragraph, stating that data-driven "black-box" machine-learning retrieval approaches do not include independent and explicitly modelled assumptions about microphysics.
- We have also updated the introduction in a similar way, to more explicitly contrast with our approach.

l. 89: As I understand the methodology, the CHIPS-AWS retrieval is itself indirectly constrained by CloudSat-derived hydrometeor retrievals through the reference database used for its development. Please clarify this relationship when contrasting CHIPS-AWS with CloudSat-based products.

- Our simulations use CloudSat radar reflectivities as input. However, no CloudSat-derived retrieval product is used at any point. We are therefore not constrained by any of the CloudSat-based products to which we compare. We thank the reviewer for bringing this to our attention, since it is an important distinction, and have updated the text to clarify that our simulations are driven by the radar reflectivities.

> l. 91: Please explain more concretely how the proposed approach enables improved physical understanding. What additional physical information can be inferred from the retrieval beyond the presence and magnitude of the retrieved ice hydrometeor signal?

- We thank the reviewer for the interesting question. Since our synthetic database is produced from simulations, all assumptions (e.g. microphysical) are explicit, and this allows us to gain physical insight. For example, if poor retrieval performance is observed for a set of observations, these cases can be compared to similar cases in the synthetic database. This allows us to assess the impact of our assumptions and the sensitivity of the retrievals to them. Consequently, we can evaluate the quality of our assumptions, which in turn informs our understanding of the real world.
- The text has been updated to clarify that a synthetic database produced from simulations allows us to explain the mapping from a physical state to simulated instrument observation, with all microphysics and other assumptions being explicit.

> l. 194: Is the primary reason for these restrictions the difficulty of accurately simulating observations at high incidence angles over high-latitude surfaces? If so, I think this should be stated more explicitly.

- The restrictions were primarily chosen to limit scope and complexity in the initial version of CHIP-AWS. Our strategy was to first restrict to conditions which we are confident we can simulate well, validate our results, and then relax these conditions in future versions.
- With that said, the restrictions are chosen with certain concerns in mind. The limited latitude range was primarily chosen due to suspected difficulties in simulations over snow- and ice-covered surfaces. The scan angles are limited due to a concern that too high incidence angles would result in a too slanted beam through the atmosphere, affecting the vertical column-retrievals.
- The concerns behind these restrictions are now mentioned in the text.

> l. 205: I do not think an effective spatial resolution can be inferred directly from the nominal footprint size or sampling characteristics. Please quantify the effective resolution using appropriate test data, or reformulate this statement in terms of the nominal footprint or sampling resolution.

- We adopt the suggestion of reformulating in terms of nominal footprint,

> l. 214: Please consider reformulating this sentence for clarity.

- It is now more clearly specified what values are excluded from averaging.

> l. 216 and following: Please specify whether (Z_m) is calculated along the vertical radar column or along the slanted AWS viewing geometry.

- The text is updated to specify that all retrievals of FWP, Zm and Dm are estimates of the vertical column and not along the slanted AWS viewing geometry.

> Section 4.1: Please specify whether any spatial averaging or aggregation is

applied to the reference FWP values before they are used in the retrieval database.

- We have updated the text to clarify that radar inversions and simulation are performed in a grid that follows CloudSat's native sensor grid in its along-track trajectory. In the altitude dimension, radar reflectivities are interpolated to a grid defined by us.
- Regarding spatial averaging, simulated brightness temperatures are integrated and weighted with a Gaussian to simulate AWS's footprint, which we now address in section 4.1.

l. 321: Please specify the number of neurons in each layer of the network.

- The text has been updated to specify that each layer has 512 neurons.

l. 330: Please provide a quantitative range for the differences described here as "substantial."

- Provided a range of expected remap-distances.

Fig. 4: Please consider using frequency rather than channel names for the labels. This would make the figure easier to interpret without requiring familiarity with the AWS channel nomenclature.

- We thank the reviewer for pointing this out. We strived to do this in all figures, but must have missed this one. It has now been fixed.

Fig. 6: The micron symbol appears to be missing.

- We thank the reviewer for pointing this out. Unfortunately, this looks like an issue arising from the processing by Copernicus, since the original file does not have this issue. We will keep an eye out for this.

l. 404: The presence of greater spatial variability or finer-scale structure does not by itself demonstrate that the retrieval contains more information about actual cloud or FWP structure. Some of the additional variability in the CHIPS-AWS estimates also appears consistent with pixel-scale retrieval noise. I suggest avoiding an interpretation of increased spatial variability as increased physical information unless this can be demonstrated against an independent reference.

- We have extended the description in the text to more carefully talk about the suspected increase in ability to resolve smaller systems. We now differentiate between three levels of structure; the finest (pixel ~10km) based variation, slightly larger (at least 40km2) (comprises of multiple pixels/footprints 4-10) we think are genuine systems being resolved, and the larger scale, where CCIC and CHIP-AWS agree.
- That high FWP values are predicted in areas with low FWP through noise is unlikely. This is demonstrated by an added confusion matrix showing that high FWP values are never confused with low FWP values/clearsky.