Response to review by anonymous referee #1

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

> This paper presents CHIP-AWS, an ice cloud product retrieved from the AWS radiometer. The retrieval algorithm is based on a Quantile Regression Neural Network (QRNN) trained using a database simulated by the ARTS forward model. In addition to retrieving ice cloud properties, the algorithm provides uncertainty estimates for each retrieval. Its performance is evaluated using both the simulated database and other ice cloud products, such as CCIC. Overall, the methodology and results are clearly presented. Several aspects require further clarification or additional discussion, as outlined in the comments below.
>
> (1.) In the Incidence Angle row of Table 1, it is stated that the satellite zenith angle is reported in the Level-1B (L1B) file. However, lines 114–117 indicate that the L1B file provides only the Earth-surface incidence angle, rather than the satellite zenith angle. These descriptions appear to be inconsistent and should be clarified. In addition, it would be helpful to include the equation

> used to convert the Earth-surface incidence angle to the satellite zenith angle in the discussion around lines 114–117. This equation is also needed for the footprint size calculations in Eqs. (5) and (6).

- Clarified this by updating the phasing in what were originally lines 114-117 from "Instead, zenith angles from Earth-surface footprint locations towards the satellite are available" to "Instead, satellite zenith angles are available, i.e. the zenith angles from Earth-surface footprint locations towards the satellite.".
- Updated Table 1: Changed label from Incidence Angle to Satellite Zenith Angle.
- What is referred to the incidence angle theta, is the same satellite zenith angle. Updated text to now refer to theta as the satellite zenith angle when describing Eqs. (5) and (6) to avoid confusion. Thus, satellite zenith angles reported from L1b files, and the averaged satellite zenith angles in our L2 files can be directly used as theta in these Eqs.

> (2.) Please check Eq. (4) again. The equation here calculates the particle diameter for each layer rather than the column-averaged particle diameter. Dm should be the FWC- weighted average of the layer-resolved particle diameters, following a similar formulation as Eq. (3).

- In Delanoë et al. 2014 Eq. (3), Dm is the volume-weighted diameter (melted-equivalent) at a specific vertical level. Our Eq. (3) is the same, except integrated over the vertical dimension z. This should be the mass-weighted mean particle size for the entire column.
- In the text we've added a reference to a derivation that shows this is the column weight averaged particle diameter: Amell et al. 2022, appendix A (eq. A9)

> (3.) For the simulation database discussion in Lines 259-264, it would be helpful to discuss why the CloudSAT CPR reflectivity observations are used to first retrieve ice cloud properties instead of directly using the 2C-ICE or DARDAR products. One reason I can think of is that using CPR reflectivity can keep the particle habit and PSD selection in the radar retrieval consistent with the selections in the following TB simulations. If there are more considerations, please clarify them.

- Indeed, it is to ensure we model our uncertainty from particle model assumptions, both when populating FWC for our simulation scenes, and in our forward simulations of AWS observations.
- We have added a paragraph motivating this in Sec. 4.1.

> Also, it would be helpful to provide more details about the radar retrieval method. For example, is it based on the optimal estimation method (OEM) or another approach? Do you retrieve only ice water content, or both ice and liquid water content? Since W-band radar signals are strongly attenuated by liquid clouds, how is the liquid water treated in the retrieval? Do you also consider the melting layer, and how is it handled?

- In Section "Simulation database" the following changes have been made:
  - A reference to a description of the radar inversion method (ARTS onion peeling) is now given in the text.
  - We have also specified that Liquid water content (including super-cooled) is populated from ERA5, while rain water content (for liquid hydrometeors) is retrieved from the radar backscatter.
  - Both LWC and RWC is taken into account when retrieving from radar and for the

forward simulations producing the passive MW observations.
    - It is now also acknowledged that the melting layer is not included, as we don't have any scattering data for melting particles.

> Finally, have you compared the radar retrieval results with existing products such as DARDAR or 2C-ICE? The manuscript mentions that the mean values from DARDAR and 2C-ICE differ by about 20% (lines 459-460). It would be helpful to know whether your results are closer to DARDAR or 2C-ICE.

- As part of the introduction of the radar inversion method (Sect. Simulation database), it is now mentioned that FWP estimates are closer to DARDAR, in terms of zonal-means. References are given for further comparisons and details.

> (4.) For the surface emissivity model, apart from the TELSEM/TESSEM used for land/ocean surfaces listed in Table 2, lines 263-264 also indicate that the surface emissivity is changed based on the land-cover dataset. Do you use additional surface emissivity datasets for other surface type? If so, please describe this dataset around lines 263-264 and include it in Table 2 as well.

- Land and ocean are covered by TELSEM and TESSEM. For ice and snow, the emissivity is modelled by a flat surface informed by skin temperature and a scalar reflectivity. This is now specified in the paragraph and added to the table.

> (5.) Line 271-273. How many pencil beams are used in each 3D Gaussian footprint averaging?

- The number of pencil beam values and the along-track sampling distance is now specified.

> (6.) Figure 2: In the top panel of the right plot, the PDF shows an almost uniform distribution over approximately 5–20 kg m-2, which is quite broad. The uncertainty shown in the middle panel seems more reasonable. For cases with large FWP values, however, the uncertainty estimate would generally be expected to decrease, which is not consistent with the results shown here. Please provide further discussion on the possible reasons and the reliability of the uncertainty estimates.

- Possible causes for larger spread for this more intense case is now explicitly mentioned in the discussion of Figure 2. In short, the spread can be due to the particle models responding differently, contributing to a wider uncertainty. Another reason could be that some channels become saturated, preventing the FWP to be constrained better. Still, the estimate is certain that we have FWP greater than 5 kg/m2. It can be argued that this still is a fairly narrow estimate.

> (7.) Line 298: μ should be calculated as the integral of x over the PDF, not the CDF.

- Yes, the expected value can be calculated as suggested. However, in practice, we do in fact integrate quantile values x over the CDF (probability levels). This is formally expressed as a Lebesgue-Stieltjes integral.
- This is now better motivated in the text.

> (8.) Line 316-319. What about cases with values larger than 1 kg m-2? There should be many cases in this range.

- These are included, but in a linear scale instead. I.e values below 1kg m-2 are log scale, values above are linear. This is now said explicitly in the text.

> (9.) Figure 5. the '> =' sign is missing.

- Thank you. This appears to have happened during Copernicus processing, since submitted file does not have them missing. We will be double check for final submission.

> (9.) Figure 6 and 8. The figures here show the density distribution of the retrieved parameters versus the true values. The conditional PDF (p(Y|X)) in Bayesian theory, however, represents the probability of the simulations given the observations, which is a diˆerent concept. I do not think these two types of density distributions are equivalent. Can you give comments about why these plots are referred to as conditional density p(Y|X)?

- Figures 6 and 8 are not showing the joint density distribution of retrieved from database (Y) vs. true database value (X), but are showing the conditional PDF, i.e. the retrieved from database given true database value.
- In other words, these figures show the joint distribution P(X, Y) divided by the marginalised probability function for X (P(X)): $P(Y|X) = P(X,Y) / P(X)$.
- We don't think labels or the text is directly misleading in this regard. We have made sure the text surrounding these Figures are explicit that its the probability of an estimated value given the database value or in the case of Figure 8 P(CCIC|CHIP-AWS).

> (10.) Line 402-404. For the comparison of these two plots, does it indicate that the AWS- CHIP results have higher horizontal resolution, or are they noisier? The CCIC results appear to be smoother and more continuous.

- This is a valid concern that we have now tried to address in the text thanks to this comment and a similar comment by referee 3.
- 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.

> (11.) Line 416-417. In line 414, it is indicated that CCIC is trained with 2C-ICE. Intuitively, the CCIC results should best reflect the statistical characteristics of 2C-ICE. Since CHIP- AWS does not use 2C-ICE data, why do the CHIP-AWS results match 2C-ICE better than the CCIC results? Could you provide more discussion on this point?

- With CCIC trained on 2C-ICE, we can expect the retrievals to be calibrated against 2C-ICE. That the mean CCIC values should match the 2C-ICE mean. Other statistical properties depend on physical or practical limitations.
- We agree with the reviewer that "CHIP-AWS is closer to 2C-ICE than CCIC" is too strong of a statement. We do think though that for individual values that CHIP-AWS can be closer to 2C-ICE. We have now qualified and explained this in the text.