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).
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).
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.
   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?
   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.
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.

5. Line 271-273. How many pencil beams are used in each 3D Gaussian footprint averaging?
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.
7. Line 298: μ should be calculated as the integral of x over the PDF, not the CDF.
8. Line 316-319. What about cases with values larger than 1 kg m^{-2}? There should be many cases in this range.
9. Figure 5. the '>=' sign is missing.
10. 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 different 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)?
11. 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.
12. 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?