Response to review by anonymous referee #2

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

> Work is performed carefully; however, I think some more information on the underlying physical sensitivities and (different) information content of the data would make the paper even more valuable.
>
> General comments:
> [1.] Physical meaning of retrieved products. The abstract states: “The accuracy in FWP should be unprecedented among estimates based on passive satellite data”. This is a strong statement that requires more argumentation. The manuscript needs to stress the complexity of ice particle size distributions (see Bartolomé García et al. 2024), transitioning from tiny particles (sub-visual cirrus) to large snow aggregates, and the information content of different retrieval products. For example, the selective sensitivity of the used data records (MODIS, CICM) that are affected by saturation in the VIS/IR signals.

- Regarding the complexity of particle size distributions (PSDs), we agree, there is considerable complexity and uncertainty in assumptions for PSD in relation to ice particle sizes. In our view, the method in this work presents a significant contribution toward tackling this and making these uncertainties explicit: by using an ensemble of particle models (habits + PSDs), and representing this uncertainty through distributions in the retrievals.
- The abstract and the introduction have been adjusted to make this contribution more clear.
- Introduction also includes a discussion on sub-mm and the gap it fills between MW and IR/Vis. There is a reference to Wu et. al 2024 that directly tackles the comparison between different data records and the role for sub-mm.
- The abstract has been updated to acknowledge that a proper validation can't be made (no co-located data). However, the analysis performed shows better accuracy over previous passive retrievals.

> As FWP is not a widely used variable, the difference to other product names (CIWP, see Stubenrauch et al., 2024) needs to be made clear.

- At first occurrence frozen water path is introduced, and it is stated that the new nomenclature is used to qualify that all frozen hydrometeors are taken into account. Product names like Cloud Ice Water Path, leave the ambiguity of what is meant by cloud.
- For a further motivation for FWP, refer to Eriksson et al. 2026. This is refereed in the text.

> Anticipating certain user questions: Could you make a rough estimate of how much of the global average FWP is missed by conventional products and how much more your product (CHIP-AWS) can sense?

- Thank you for the heads up on this anticipated question. In the introduction, we now added a paragraph that states conventional passive datasets have been shown to be roughly 50% compared to CloudSat CPR datasets in zonal mean comparisons. Two references are included. There it is also stated that our product gives values comparable to the CPR based datasets.

> and give some recommendations (I don’t expect detailed answers) on...
> How can one identify a “clear-sky” pixel and with which uncertainty?

- We thank the reviewer for this clarifying question. This is a point we would like to make more clear in our manuscript. The answer on what is clear-sky depends on what is meant by clear-sky. Clear-sky can mean different depending on the sensor's sensitivity. From an observational perspective taking sensor sensitivities into account, it makes sense to define clear-sky below the threshold where such an instrument would sense ice. Since quantiles are provided, this can be answered using any threshold one likes to produce a probability that a case is below the threshold.
- In order to demonstrate this, and make this application more clear, we have extended the sub section "Example Scene", with an example providing classification probabilities for clear sky and other classes.
- In addition, a new confusion matrix figure also provides recall and precision for FWP < 1g/m2.

> How many cirrus clouds will be missed?

- Again, the answer here depends on the definition for cirrus clouds in terms of FWP. But

taking (1g/m2 < FWP < = 10g/m2) to mean thin cirrus, our method struggles. There is not enough mass to get a strong signals for these channels.
- We hope that the added classification probability example and the confusion matrix help answer this more directly in the answer. The lower sensitivity for these cases is now acknowledged in the text.

> Do I also need to apply a threshold of 0.1 kg/m2?

- The retrievals give a distribution of values that expresses the uncertainty. When computing probabilities from these estimates no threshold should be applied.
- However, the retrievals for Z_m and D_m are generally more certain when FWP is higher, so the point estimate (the mean or most-probable) will likely be off when FWP is low. So, a threshold on FWP of 0.1 kg/m2 has been found to work to filter mean point estimates for Zm and Dm with more certainty.
- In our presented results and analysis, we have opted to filter with FWP 0.1 kg/m2 for Zm and Dm for certain comparisons (occurrence fractions and conditional PDFs), in order to not have means from uncertain retrievals skew the result.
- Other comparisons, such as the zonal mean, local means, and Hovmöller diagram do not use this filtering, instead they are weighted by the FWP.
- The Section "Conditional densities", its conditional probability densities Figure, and related text has been updated to address this. The figure now includes conditional means for two filtering levels: FWP > = 0.1 kg/m2 and FWP > = 0.01 kg/m2, to demonstrate how the filtering affects precision for different values.

> How much is a retrieval of 40 g/m2 worth?

- Added Figure showing the CDF for reference FWP values in the simulation database and the cumulative contribution to the total FWP.
- From this, values below 10 g/m2 and 40 g/m2 stand for 46% and 64%, respectively of all reference FWP values in the simulation database. However, values below 40 g/m2 contribute very little to the total FWP sum/average.
- Note, however, that even though we have a stated sensitivity down to 10 g/m2 or 40 g/m2, depending on how we compare, we still get statistically reasonable means when accounting for values below this. Since we are following a Bayesian approach, our distribution does account for values all the way down from the prior (simulation database).

> Is Dm basically only a measure of snowfall size?

- Dm is a measure of all ice hydrometeors in the atmosphere, a mass weighted size estimate. Not specifically snow, and not snow reaching the ground.

> [2.] Clearness and completeness of the provided information: At several instances (see my specific comments), more accurate formulations are needed. One example is the term bias, which is frequently used in the paper, but its actual meaning is not always clear, as truth is not really available.

- Agreed. We specified conditional bias for the conditional mean results and discussions. Other uses of bias where datasets are compared and there is no clear "true"-value have been replaced with "difference".
- Called it across-track dependence instead of across-track bias.

> Information on intercomparison data, e.g. MODIS, CLARA, is missing. The Discussion section is a bit confusing – here suddenly SPARe-ICE is appearing.

More information on accuracy and limitation these data is needed. For the outlook the, authors should also consider intercomparisons with ground-based reference sites, e.g., ACTRIS/Cloudnet.

- MODIS, CLARA and CCIC are now introduced in the "Data sources" section under the new subsection "Comparison datasets".
- SPARe-ICE is now also explicitly introduced. However, we opt to keep the introduction of this dataset in the discussion, since it is only there to discuss and place our MFE results into context.
- Mentioned EarthCARE and ACTRIS Cloudnet in the outlook for validation work.

3. Length of the record: It is never really stated how long the data record for the presented analysis of CHIP-AWS is. I find it really important to have exactly a (full) one-year period (March to March) to avoid seasonal imbalances. MODIS or ERA5 could be used to assess the interannual variability.

- Text has been updated to explicitly state that we provide data for the full year starting from January 2025.
- In the zonal mean, MODIS and DARDAR show the interannual variability.

Minor point, but in many instances: One sentence does not make a paragraph.

- Thanks. Manuscript has been revised to avoid this.

Specific Comments:

Abstract: (you should mentrion the product name (CHIP-AWS for branding)

- Thanks. Added.

L4: Detailed simulations- here information on the blend of Cloudsat retrievals and ERA5 should be mentioned. For later: Any retrieval can only be as good as the underlying database. The authors are a bit short on information (and motivation) how they construct the database. Explain why not only ERA5 information is taken.

- Clarified that simulations use hydrometeor information from CloudSat radar reflectivities and ERA5 for background-information.
- The "Simulation database" section now further explains the CloudSat CPR inversion approach and justifies its use.

L11:" The accuracy in FWP should be unprecedented among estimates based on passive satellite data". This is a strong message see major comment #1.

- See response to major comment 1.

L12: "significantly" could be nearly anything – quantify

- Thanks. Quantified swath size.

Introduction:

- L21 – first dataset, duration missing -> first full year data set

- Adopted as suggested.

- L34 225-1064 nanometer not micrometer; add (about 3mm) in parentheses after 94 GHz

- Adopted as suggested.

    - L33- maybe refer to the figure in the Buehler paper for the different wavelength respnse

- Added reference to Fig. 3a of Buehler 2007.

    - L35 – “spatial” is irritating – either vertical or radial

- Should have been vertical. Now fixed.

    - L39 – I would add climate and the changing water cycle as further motivatin

- We've added a short first paragraph to the introduction to motivate why FWP retrievals are important.

    - L43 – “applying machine learning” somehow sounds too fancy. ML is basically uses as regession

- Specified that Machine-Learning is used to train a non-linear predictor.

    - L47 – submillimetre vs far infrared?
    - L47 – for clarity mention the two conventional ones by name

- Specified that sub-mm fills the sensitivity gap to ice mass between MW and infrared.

    - L73 – “all hydrometeor classes” – hydrometeors haven’t been introduces yet. I think it is important to mention the arbitrary (model-dependent) distinction between ice and snow water content and the fact that kilometer-scale models introduce denser hydrometeor classes (graupel, sometimes hail) to enable the representation of strong convective events, i.e., thunderstorms.

- The text is rewritten to avoid using "hydrometeor classes". We agree that atmospheric models do not make a clear distinction between cloud ice and snow, but as we are here not comparing to any such data, a discussion of that topic would be a distraction.

    - L75 – “report” is not the right name – mostly they are even prognostic. Suggest to simply say “consider”

- Changed to GSRMs provide values for all hydrometeor classes.

    - L82 – retrieval PRODUCT

- Thanks. Changed to retrieval variable, since we've referred to the L2 files with all variables as the retrieval product.

    - L89 – I find the sentence confusing: “operational machine learning models for cloud retrieval have exclusively been trained on existing retrievals” . Maybe I misunderstood what is meant with “operational ML” (then it should be better explained) but tons of satellite retrievals are run on simulated measurements (mainly from atmospheric model

data).

- We have removed operational to avoid confusion. We meant ML methods that have been trained through co-location with existing retrievals.
- We ask the reviewer to point out if we've misunderstood the comment.

    - Similarly L90 “For reasons discussed below, this more complex approach could not be avoided” sounds weird.

- The sentence is rewritten.

Sources and reference data:

- It would be helpful for the reader to get a rough idea of the retrieval and validation strategy before presenting the subsections to understand for which purpose they are needed for.
- For example, I was first expecting that ERA5 output (including hydrometeors) would be used for the simulations.

- The data source subsection "Chalmers Cloud Ice Climatology (CCIC)" has been renamed to "Comparison datasets", which introduces CCIC, MODIS, and CLARA and the way they are used for comparisons: Zonal means, occurrence fractions, co-located samples.
- This also makes it more clear that ERA5 is not used for any comparisons. It is used for background simulation input (not hydrometeors).

    - L101: 145 over a +/- ??degree range

- Added. +/- 55 deg

    - L128: Give number of pressure levels. There is significantly more detail in the model layer output though I don’t believe it is important. More importantly, I am missing a discussing how physical consistency between thermodynamic state and ice hydrometeors is guaranteed? As an example could you have 20% relative humidity and high FWP?

Specified that all 37 pressure levels are used, but are interpolated into our simulation grid.

A good remark, indeed 20% relative humidity in a region with high ice mass could in principle be found in the input to our radiative transfer simulations. We spent considerable time trying to devise a correction procedure, starting with Rydberg et al (2007), but never found a satisfactory solution. We likely improved the realism for parts of the data, but with the cost of adding artefacts. For example, we applied corrections in Ekelund et al. (2020), but those resulted in decreased agreement with observations for "clear-sky" conditions (a disagreement around the "peaks" in their Fig 10), but we kept the corrections as they should hopefully increase the fidelity for the data of concern for that study (the lower end of Tb-s). However, for actual retrievals we cannot neglect the artefacts for the clear-sky part, which in fact constitutes a dominating part of the data. Therefore, we have for the time being put the corrections to the side. This is not ideal, but not critical. At the time of Rydberg et al (2007), output from ECMWF was known to have a low bias in upper tropospheric humidity, the quality of ERA5 is much improved. We have manually inspected a significant number of scenes, and found the agreement between CloudSat and ERA5 satisfactory. That said, we should be clearer about physical inconsistencies, and changes to the text have been done in Sec 4.1 and Sec 6 "Discussions".

Rydberg, B., Eriksson, P., & Buehler, S. A. (2007). Prediction of cloud ice signatures in

submillimetre emission spectra by means of ground-based radar and in situ microphysical data. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 133(S2), 151-162.

Ekelund, R., Eriksson, P., & Pfreundschuh, S. (2020). Using passive and active observations at microwave and sub-millimetre wavelengths to constrain ice particle models. Atmospheric Measurement Techniques, 13(2), 501-520.

- L131: why is only cloud water content used? I would also recommend the inclusion of rain as light rain amounts are frequent and even present in cases of no surface precipitation.

- Text updated to state that rain is populated in our simulation scenes through inversion from the CloudSat CPR reflectivities, with a reference for more details.

- L141: Why 2015? Maybe one could see from the longer-term MODIS, CCICs data set that this is an “average” year?

- The choice for 2015 was arbitrary, the data was available and ready to use at the start of this project.
- We have checked that this year was not an special with respect to ENSO. This is now stated in the text.

- Subsections 2.3 and 2.4 don’t provide information on the uncertainty of the described products. Specifically what is the limitation of CCICs wrt high FWP?
- CLARA and MODIS Data (and their uncertainties) are not described here and not mentioned in the data availability section.

- Sections 2.3 and 2.4 have been combined into one section "Comparison datasets".
- Here, brief comments on underlying methods, limitations, and if datasets provide uncertainty estimates are now added. References for more details are also included.

Data product:

- I would be curious how strongly the three main retrieval products are correlated?

- A good suggestion. We have now added a new table (Table 4) to show the correlations between the retrieval variables in the simulated training data, in the test retrievals, and in the CHIP-AWS retrievals.

- L159 – see my main comment on duration of the data set for analysis – full year

- Specified that we are providing data a full year of data from January 2025.

- L167 – provide also angle and swath width – then you don’t need it later

- Adopted.

- L188 – also Eriksson et al., 2020 don’t use the term spatial sensitivity but refer to vertical one..

- Thanks. Changed to radial. Those equations consider the radial footprint of the observations.

- L196 – “These restrictions are put in place to reduced the scope” -> These restrictions are put in place to reduce the complexity..

- Adopted.

- L197 - Variability can mean many different things, please explain your definition used here

- Thanks. Removed variability and state that the 75th percentile increases closer to the equator.

- L206 – at least mention that ERA5 is coarser

- Yes, it is a good point that ERA5 is much coarser than our discussion on footprint sizes of ~10km.
- However, ERA5 was used to populate background information in our simulations that are then used to train the ML model. CloudSat is still the source used for populating the simulations with hydrometeors. No ERA5 data is used to perform the actual retrievals once trained.
- We argue that ERA5 coarseness should not impact our footprint-by-footprint resolution and leave this discussion on ERA5 out, to keep the section focused.

- L214 – what does unreasonable mean – how can this be reproduced? I don’t expect a full description but maybe a link to a reference

- Text now specifies how unreasonable satellite altitudes and incidence angles are filtered out.

- L225 – How strongly are the variables correlated in the training, test and application? Do you see any changes? It is also very interesting to investigate the correlation with temperature, we also suggest including the zero degree line in the figures. One could also do a 2D histogram with zm and the height of the zero degree isotherm for consistency checks.

The correlation between variables, and whether the correlations are maintained in the retrievals, is an interesting point. We have now added a new table (Table 4) to show the correlations in the simulated training data, in the test retrievals, and in the CHIP-AWS retrievals.

We do in fact see a small change in the correlations, where the correlation between FWP and the other two variables increases for the test set retrievals. Encouragingly, the increase is much smaller in the CHIP-AWS retrievals. An observed increase could be explained by the fact that we may not always have enough information available in the observations to infer each variable entirely independently. The correlations are not so high to suggest that we can never infer independently, and we suspect that this behaviour varies across atmospheric conditions and/or the ranges of the variables. This is supported by the low degrees of freedom at low FWP, observed in Fig. 1 of Duncan et al. (2026). This is certainty interesting, and we plan to do a deeper analysis into the correlation behaviour moving forward.

- A 0degree isotherm for transect along the example scene has now been added.
- A nice suggestion regarding the 2D histogram with Zm and 0deg isotherm as an extra result. However, due to practical reasons and time limitations, we wont act on this suggestion. We hope the database comparisons, Hovmöller diagrams, and zonal means results sufficiently demonstrate the retrieval performance regarding this variable.

- L230: There needs to be more discussion on the choice of the threshold and implications for analysis – when do we have a (precipitating) cloud?.

- At first mention the reason behind the threshold of FWP >= 0.1 kg/m2 for Zm and Dm is explained, with a reference to the conditional distribution section for more motivation of the threshold level.

Retrieval Method:

- L295 – for me it would have been helpful to mention early that a binning of 0.01 for the CDF would be used
- What is the choice for the chosen output quantiles – why not 0.05 and 0.95?

- The binning of 0.01 is now stated in Section "Inversion with QRNN".
- Choice for quantiles 0.02, 0.16, 0.5, 0.84, and 0.98, is now provided: chosen to line up with with the median, +/- 1sigma, and +/- 2sigma.

- L 313: I guess randomly chosen?

- Correct. This is now stated in the text.

- L320 This is a good opportunity to advise the user if a “trustworthy” FWP is retrieved..or what the likelihood is that this pixel has no FWP?

- This is now introduced and demonstrated in the Example scene, where classification probabilities for a scene are shown.

- L330 – specifiy substantial

- This is now specified.

- L333: THE mean estimate

- Fixed.

Results:

- L336 – agrees is a too strong word -> represents the observed atmospheric state

- Adopted.

- L340 – what is meant with accurate? Just the RT? Then I would say first a representative atmospheric state and then an accurate RT.

- Thanks. Adopted.

- L343: best-performing – best understood surface interaction

- Thanks. Adopted.

- L353: As a rule of thumb values below….

- Thanks. Adopted.

    - L358: There is some slight discrepancy at the coldest temperatures already seen here over ocean – so one could start here already start the discussion on intense convection

- Thanks for the suggestion.
- We opt to keep the discussion of intense convection focused on land. Over ocean, we see that we have some simulated scenes with lower Ta than in observations.
- For intense convection over ocean being missed by our simulations, we would expect to see lower observed temperatures than our simulations can provide. Currently the opposite is the case, which is not as much of a problem since it gives us a greater range of cases to base retrievals on.
- However, when discussing land, the complexities of modelling deep convection are now brought up.

    - L360: You could also make a correlation matrix between different channels

- Our interpretation is that the reviewer is suggesting to show ill-posed-ness of the observations. To our understanding, we can have ill-posed cases that happen with uncorrelated channels. We opt to leave such a correlation matrix out of this manuscript, since that analysis is not comprehensive.

    - L362 – specify where one can see this? Furthermore, intense convective cores – you didn't describe beforehand why these are tricky, e.g. graupel, hail..

- Text has been updated, to specify that we attribute low Ta values for AWS21 with tropical deep convection.
- Added a short motivation as to why deep convection can cause issues.

    - L374: I think the figures (going down to 1 gm-3) overemphasize the importance of the really low values. Looking at the helpful Fig 5.

- Thanks for pointing this out.
- The new figure with the cumulative distribution of reference FWP from the simulation database and the cumulative contribution to the total FWP is used to help put importance of very low FWP values into context.

    - L409: bias-free – how can you judge as truth is there; away -> higher?

- From addressing the main comment on use of the word bias, we've decided to remove the paragraph discussing CCIC's bias.

    - L417: I am missing a discussion on the physical interpretation of Fig.8 Basically what can be seen what not?

- Figure 8 is now more thoroughly described and interpreted in the text. Now describing that for given CHIP-AWS estimate values, the co-located CCIC estimates tend to overestimate FWP below 1kg/m2 and underestimate above.

    - Section 5.34. The MODIS data set was not introduced before. To my knowledge it was never constructed to include precipitating ice so there is no surprise about the difference

- MODIS is now introduced in the "Data sources" section. The caveat that MODIS

retrieves the ambiguous cloud ice water path instead of all frozen hydrometeors (FWP), is now pointed out in the text, both in the introduction and in the results "Spatial distribution" section now called "Local means".

- L439 “The CHIP-AWS estimates have no coastline artefacts that cannot be attributed to orographic features. “ That is confusing. Orography makes an artifact?

- Artefact was an incorrect word choice. This word is removed.

- L442: cases : do you mean ocean and land? Maybe regimes is better: there is no clear indication where one can see the effect of the dry cases ?

- Thanks. Adopted

- L446: Better surface emissivities..there should be also some reference to the literature on that

- Thanks. Adopted.
- We assume the reviewer refers to references to what current emissivities we are using for ice and snow. This has now been specified in Section 4.1 and Table 2.

- L449 and following: I don’t understand the meaning of bias here because no truth is there.

- Following from main comment on use of bias, comparison is now made in terms if "difference" instead.

- L460: what is the correlation between DARDAR and 2C-ICE?

- It would be interesting to see this correlation. However, we want to keep manuscript focused, and a comparison between other datasets is not our goal.
- There are previous works that perform such comparisons, see Eriksson et al. 2026, and Duncan and Eriksson 2018.

- L469..and knowledge of the physics.

- Added that the results are also consistent with the expected limited information content deeper into clouds with the sensors used for MODIS and CLARA.

- L470: In case of multi-layer clouds radar – lidar is not a good reference

- Thanks. This is now mentioned.

- L484 . Refer to where the retrieval accuracy is mentioned.

- Thanks. This was referring to the sensitivity limits in Sec. 5.2.1 and 5.2.2. Revised to make this more clear and added reference the sections.

Discussion:

- I got a bit confused with this chapter which seemed to be more like a summary than a discussion.

- Removed repetition of results that in the beginning that are not really discussed.

- L487 and following: Which are the three orders of magnitude – be specific when one can “trust” the retrieval. As there are specific uncertainty values for each pixel, what would be the recipe for using them? What is meant with “entire relevant range”?

- Thanks. Specified the magnitudes, removed "entire relevant range".
- An example for how to use the uncertainty to obtain probabilities is now presented in Sec. "Example scene"

- L489: “The comparison database values are radar retrievals” What is meant by that?..or is this the start of a new paragraph?

- This sentence on radar retrievals is misleading. Now only referring to the simulation database reference values. This was referring to that the simulations are driven by radar measurements: ARTS onion peeling, that is now properly introduced in the "Simulation Database" section.

- L491: Now the discussion goes to Spare-Ice which has not been introduced before - why is it not included with the MODIS and CLARA comparison?

- Rephrased the first mention of SPARE-ICE in the discussion to explicitly introduce it and state that it is only included to provide context to our MFE results through their reported MFE results.
- SPARE-ICE is not included with the other datasets, since SPARE-ICE is currently not currently maintained operationally, whereas CLARA and MODIS are established datasets with operational processing. To the best of our knowledge, only SPARE-ICE provides a MFE estimate for FWP in literature.

- L501: Why not ground-based estimates from Cloudnet? This could also provide information on zm.

- Mentioned expanded comparisons the outlook.

- L504: “Recent machine-learning approaches trained on radar retrievals have achieved similar results” Sounds very generic – where does the information come from – new architecture, new training??

- Removed recent, and specified that we are talking in terms of zonal means. I.e. the ML-based approaches, thanks to their training loss-function, tend to minimize the average error.

Conclusions:

- L526: Simulations can mean anything – also atmospheric model simulations. In this sense the sentence in L534 should be reformulated” This is, to the best of our knowledge, the first simulation-based passive retrieval method”

- Reformulated to: "This is, to the best of our knowledge, the first passive retrieval method based on atmospheric radiative transfer simulations to do so (Duncan and Eriksson, 2018) (Eriksson et al. 2026)."

- L529: rather unspecific. Simulated Ta distributions agree with their climatological counterpart observed over one year globally.

- Thanks. Adopted.

    - L539: the effect of convective cores was never explained,

- This is now explained, see response to comment on L358.

    - L534: “all three variable??

- Thanks. Specified that we are referring to FWP, Zm and Dm.

    - There should also be some link to passive MW snowfall and SWP retrieval (Camplani et al., 2024) which use partly similar principles.

- Yes, this is very relevant related work to highlight. We've added it to the outlook section on interesting related paths.

Figures:

    - not always all lines are explained

- Thanks. All lines and areas are now either explained or have a clear legend entry.

    - Figure 1: Explain CDF on first occurrence. Give date and time. How was the transect line chosen? It would be very helpful for interpretation to plot the zero-degree isotherm (from ERA5) into the Zm transect.

- When discussing Figure 1 (example scene) and Figure 2 (classification probabilities) choice for transect is explained: arbitrary to intersect with regions of varying FWP.
- Date and time for start and end of pass is given in the first panel.
- A zero-degree isotherm is now added.

    - Figure 2: scaling of middle CDF/pdf is ns irritation – better go only to 1 or even lower FWP

- Adopted. Middle and bottom figure share axis range of 0 to 2 kg/m2.

    - Figure 6: Micrometer are missing, red and dashed lines need to be explained. The vertical scale of the MFE should be better visible for values of 100%

- Axis limits have been changed as suggested.

Tables

    - Table 1: main (top 3) variables. Why is the mirror angle not included?

- This is now included.

    - Table 3: Why not polarisation and bandwidth?

- The table lists inputs to our ML model. We refer to Table 1 in Eriksson et al. (2025) for complete channel specifications.

    - Table 4: Are these really the best metrics? From these I would say it can only retrieve snow? What about relative error and only calculated above the threshold of 0.1 kgm-2? In the abstract you mention down to 40

gm-2 – but how trustworthy is such an estimate?

We agree that, for a variable such as FWP which can vary over orders of magnitude, a relative error is more appropriate, and chose MFE for this reason. However, we admit that some confusion here may have arisen from how we have presented results in the table. MFE is in fact a unitless and relative measure. It is an error of 86%, not an error of 0.86 kg/m2, which perhaps is why the reviewer made the association with snow.

We thank the reviewer for bringing up this point, since it is now clear to us that Table 4 can be misinterpreted. Furthermore, we now realise that the metrics shown here simplify the situation too much, since RMSE, MFE, and bias can vary across the range of each variable. The variation of MFE and RMSE are already shown the lower panels of Fig. 8, and bias can be understood from the curves in the upper panels. We have therefore decided to remove MFE and RMSE from Table 4. We hope that the metrics, and their complexity, are now clearer to the reviewer, and help to answer the questions regarding FWP thresholds. Over 0.1 kgm-2, the MFE of FWP is between 80% and 50%. Around 40 gm-2, the MFE is around 100%, i.e. only 40 gm-2.