Dear Dr. Hendricks,

We thank you for your second report and for your very useful comments. Following your new suggestions, we have revised our manuscript and we provide a point by point response to your comments. As for the first review-round, we follow the shape of your review and we provide our answer in blue :

1: On the use of penetration factors to describe freeboards errors

My concern was that the forward model used penetration factors to incorporate systematic freeboard into the inversion, while additional freeboard errors exist that are not dependent on snow depth. The authors have responded to this by replacing the term “penetration factor” with bias factor and mentioned additional freeboard error contribution in the manuscript. This addresses my comment adequately. I would still recommend to the authors to further discuss the range of bias factors in more detail. As an example, bias factors greater than 1 are plausible for radar altimetry in cases of off-nadir ranging biases. But negative bias factors for ICESat-2 are more difficult to explain, giving the low uncertainty bounds of laser freeboard.

We appreciate the reviewer’s positive assessment of our terminology update. To address your recommendation regarding the range of the bias factors, we have expanded our discussion in Sections 4.3 and 5.2. Specifically, we now explicitly discuss the physical implications of a bias factor greater than 1 for radar altimetry, noting its potential link to off-nadir backscatter (Section 5.1). Furthermore, we acknowledge that while our weakly constrained priors allow for a wide range of values, physically inconsistent results—such as negative bias factors for ICESat-2—indicate that future iterations may require adjusted prior distributions to enforce stricter physical consistency (Section 5.2) but can also be a reflection of partial incorrect forward model. We hope these additions provide the necessary nuance to this discussion.

2: Model setups with dimensions 3 and 4 may create false minima

My concern was that the forward model is ambiguous for dual-altimeter observations: (*Quote Stefan Hendricks from review): E.g. the observed freeboard differences between the two altimeters can be easily explained by a range of suitable combinations of snow depth and penetrations factor(s). The ice thickness than just can be chosen within the substantial valid range to account for the varying snow mass and to place the freeboards at the correct magnitude. With just two freeboard measurements there is no possibility to resolve this ambiguity.)*

The author response is that *“(Quote Authors from response): The inverse method (RJ-McMC) introduced in the paper can provide probabilistic inversions even for seemingly unconstrained problems: this comes from the existence of prior information, the inherent correlation scales that are introduced in the Bayesian inversion approach.”*

It is understood that the inversion will deliver solutions within the prior distributions, but I cannot follow the argument of how “inherent correlation scales” are able to resolve the ambiguity in the forward model. If this were the case, then I expect that the geophysical parameters (snow & sea ice thickness) will be roughly consistent and in a plausible range for CS2_IS2_2p/3p/4p. But they aren’t. Especially the cases of CS2_IS2_4p results in a forbiddingly large sea ice thickness bias (-1.45 m).
I am stressing this point because the authors state in the conclusions: “*We showed that this algorithm is also able to compute maps for the bias factors of the satellites used. This could help future research into the behaviour of α and its link with SIT and SD (using the covariances for example or with more sensitivity studies)”.* Again, I understand that a significant part of this work is to demonstrate the ability to invert for bias factors along geophysical variables. But my concern remains that a systemic flaw in the combination of forward model and observations prevents a basic level of level of confidence in the bias factors in the 3p and 4p cases. And the source of my mistrust is that resulting ice thicknesses are so far of the mark. This is vital information for assessing the method and I don’t find it helpful that this information has been moved to the supplement.
Regarding sea ice thickness metrics, the authors have also responded *“that our results are not meant to compete with a mature product such as AWI”.* I think that they should. The probabilistic inversion has access to the same data as the AWI product and more (ICESat-2). The AWI product is based only single-altimeter data, while the inversion technique can use dual-altimeter data. In the synthetic tests the probabilistic inversion can at least visually reconstruct a known truth from a subset. The CS2_IS2_2p sea ice thickness comparison to the AWI product shows a level of differences that one would expect from better snow information. But with the addition of the bias factors the sea ice thickness quality metrics turn substantially worse. The authors mention that “further tuning” is needed but at this stage the chances of success are not clear to me. I therefore recommend the authors to reconsider certain statements in their manuscript, and I have highlighted them in the annotated manuscript.

We acknowledge the reviewer’s concern regarding the potential for false minima and model ambiguity in the 3p and 4p configurations. We agree that while the Bayesian framework provides a probabilistic solution, the large SIT biases in the 4p case indicate that these specific setups currently lack the observational constraints needed for a robust retrieval.

We decided to keep most of the 3p/4p inversions results in the supplement, in order to focus on the 2p inversions in this paper. However, we still show one example of such 3p/4p inversion in Figure 12.

We have also revised Sections 4.3, 5.2, and the Conclusions (pp. 23-24) to emphasize that these higher-dimensional inversions are experimental and intended to showcase the algorithm's flexibility rather than to serve as a mature product.

Furthermore, we have updated the text to state that our method should indeed be evaluated against existing mature products like AWI, while noting where differences in snow information and dual-altimeter data lead to diverging results.

3: Operation IceBridge is very limited use for sea ice thickness validation

My concern was that the sea ice thickness computation in the Operation IceBridge Quicklook Product and the forward model was inconsistent in the initial submission. This has been changed in the revised version and additional validation data sources have been added. This comment has been addressed adequately.

We thank the reviewer for their positive assessment of our revisions regarding this comment. We have updated the OIB data and added BGEP validations (Figure 8).

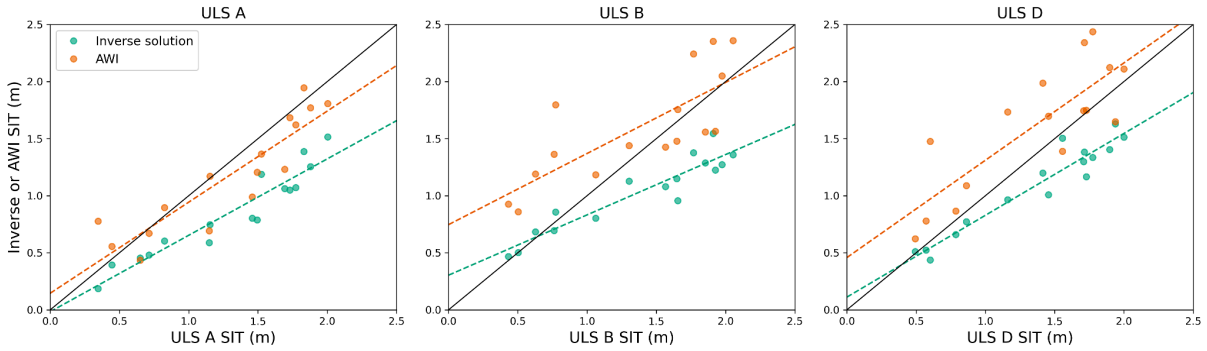

Figure 8. Validation of inverse SIT against BGEP moorings for the winter periods from 2018 to 2021. We also plot the results from AWI.

4: Delaunay Triangulation - More information needed

My comment was that the method description lacked substantial detail. The authors have amended the manuscript and added a section with synthetic inversion tests. The new section is a useful addition to the manuscript but lacks a quantitative comparison between truth and inversion result. The inversion technique is also compared to Gaussian Process Regression interpolation in the form of the software GPSat. GPSat has been designed also for gap-filling of radar altimeter of sea ice, and it is a good choice for comparison to the inversion results. But in the manuscripts, it used as an example for a “classical” or “traditional” interpolation, which I find an odd choice of words. GPSat does indeed only interpolate values and does not convert them with a built-in forward model and maybe this is what the authors mean here. But the comparison between the two does not seem on equal ground to me as the GPSat interpolation is limited only to 4 grid cells, while the inversion does not have this constraint for the same input data. My recommendation here is to update this section to mainly provide numbers that can service as

interpolation accuracy metrics. I have placed recommendations for improving this section in the annotated manuscript.

We thank the reviewer for their positive assessment of our revisions regarding this comment. To answer the new suggestions, we have re-runned the GPSat interpolation approach without the 4 grid cells constraint and updated the manuscript with the new results, in order to unbiased the comparison against the inversion. We agree with the reviewer that “classical” or “traditional” approaches were not an accurate choice of words and we thus have changed them by “others” or “alternatives”.

Furthermore, we have added a new figure to display a quantitative analysis of the Inversion/GPSat solutions (difference maps and histograms, computation of the RMSE and standard deviation for each case).

Summary :

The authors have substantially revised and improved the method description. I see the value of publishing a description of this method as it is a novel application. But I strongly recommend adding quantitative results for the new synthetic tests and reconsider a few conclusions that I find are not supported by the results. The paper should be suitable for publication after these and the more specific comments in the attached manuscripts are addressed.

In my editorial opinion, I noticed that many figure captions, especially in the supplement, are not descriptive enough. Additionally, other products or methods are frequently called “conventional,” “traditional,” or “classical,” which can come across as unnecessarily judgmental. It would be better to refer to them by name or as “alternatives.” Each method has its own advantages and disadvantages, and some may represent the state-of-the-art depending on their specific purpose.

We thank you for your second report and your encouraging remarks regarding the novelty and value of our Bayesian inversion approach. We have carefully considered your overarching recommendations and substantially revised the manuscript to ensure all conclusions are strictly supported by our findings.

First, we significantly strengthened the synthetic tests by adding a rigorous quantitative comparison (Figure 4) between our inversion and the GPSat interpolation. To ensure an unbiased evaluation on equal ground, we re-ran GPSat without its previous spatial constraints and now explicitly report the RMSE and standard deviation for both methods. Second, we critically reconsidered our conclusions regarding the higher-dimensional 3p and 4p multi-parameter setups. Acknowledging the ambiguity and the resulting sea ice thickness biases in these

configurations, we have moved the bulk of these results to the Supplement. We have reframed these specific setups in the main text as experimental demonstrations of the algorithm's flexibility, thereby keeping the manuscript focused on the robust 2p retrievals.

Additionally, we have implemented your editorial suggestions. We removed subjective terms like "conventional" or "classical," replacing them with objective identifiers to ensure a fair comparison between methodologies. Finally, we have thoroughly expanded the figure captions, particularly within the Supplement, to ensure they are fully descriptive and standalone.

We appreciate your guidance in refining this manuscript.