Dear Dr Hendricks,

We thank you for taking the time to review our paper and for your very constructive thoughts and comments. We have revised our manuscript following your suggestions and detail here these modifications. Your comments are in dark whereas we provide our answers in blue.

- **General feedback**

The paper Bayesian inversion of satellite altimetry for Arctic sea ice and snow thickness from René-Bazin et al. describes the application of a novel inversion scheme to simultaneously derive sea ice thickness and snow depth from multi-sensor (Ku/Ka-Band radar altimeter and Ku-Band radar/Laser altimeter) data. One of the novel aspects of the method is that is uses a Bayesian inversion and a data sampling strategy based on random triangulation of surrounding data and thus can work with gappy input data.

A lot of work has undoubtly went into this study and the method has been to my knowledge never applied to snow depth and sea ice thickness estimation. Nevertheless, from seeing the results I am not convinced that this method improves upon the results from much simpler methods.

You are totally right about this point and the point of this paper is not to improve on existing products but rather to showcase a new method. We hope that the updated version reflects this ambition and gives sufficient credit to existing products that are undoubtably more mature. We note nevertheless that while in some aspects our method is more complex (but could be optimised) in others it is simpler as it gets rid of unnecessary and unjustified assumptions (i.e. snow depth climatology).

There are several points leading to my conclusion: Firstly, the chosen forward models include penetration factors. These factors are dependent on snow depth and have a limited capability to describe the freeboard errors of laser and radar altimeters (General Comment 1). Equally important, the penetration factors also seem to introduce multiple solutions, which cannot be resolved with 2-dimensional data vector (General comment

Perhaps the most central and innovative point of our paper is that the rj-MCMC 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 transdimensional inversion approach. We deal with the comment on correlated snow and alpha parameters further down in our replies.

2). The validation is also very limited (General Comment 3) and may not be decisive enough to demonstrate that the CS-IS-2p SIT outperforms the AWI SIT. The CS-IS-3p and CS-IS-4p are definetely not outperforming the AWI SIT. On top of these concerns, I find the description of the methodology could be expanded substantially as key characteristics of the spatial sampling, such as the node and ensemble statistics (General Comment 4) are missing.

We have rephrased everywhere in the paper to make clear that our results are not meant to compete with a mature product such as AWI. This paper is a proof of concept and not a data paper. At this stage we are very happy to see that our results are comparable (albeit not as good) to existing products. We have also expanded following the reviewer's comments on the validation analysis (see further below).

The comment on method presentation is totally fair and we have rephrased and reorganised the paper accordingly.

I have provided more information in the General Comments below and added specific questions and recommendations in the pdf file attached to this review.
My recommendation for going forward with this would be remove the penetration factors from the forward model (I don't see how the ambiguity they introduce can be resolved) and to add another sensor sensititive to snow depth (e.g. a simple forward model used in passive microwave retrieval of snow on sea ice).

We hope that we have convinced the reviewer that the method proposed can tackle inversion problems that are seemingly underdetermined. Therefore, we want to retain this discussion of the 3d and 4d inversion problems but we now move it to the discussion section.

Then the ensemble statistics and residual inversion errors could be used as indication for snow depth uncertainties. Potentially sea ice density could also be inverted for and not prescribed by an external sea ice type mask.

This is an excellent suggestion especially in the context of CRISTAL which will have a radiometer onboard. The aim of this paper (already quite lengthy and rich) is to describe the application of the rj-MCMC method for the first time to a dual frequency altimetry retrieval case study. We want to present the method as clearly as possible and further extension for better products or new fusion approaches will be part of future research.

But as the manuscripts is now, I cannot recommend publication at this point.

- **General Comment 1: On the use of penetration factors to describe freeboards errors.**

Penetration factors are a central piece of this work. This factor is used to relate a freeboard estimate from any altimeter type to the sea ice freeboard (alpha=1) of snow freeboard (alpha=0), with the difference being the snow depth. Sea ice freeboard is the target variable for Ku-Band radar altimetry, while snow freeboard is the target variable for Ka-Band radar and laser (infrared and green) altimetry. Deviations from the 0 and 1 values are used here and in other literature to describe the freeboard error with respect to the corresponding target variable.

The issue with describing errors of radar-derived freeboard as a direct function of snow depth is that there are many error sources for which snow depth is not a suitable proxy. The authors write in L44+ as their motivation for using the penetration factor that the known backscatter from the snow at Ku-Band raises the backscatter elevation distribution and thus bias the radar freeboard value high. While snow backscatter is undoubtly relevant at Ku-Band, there are issues with the assumption that the freeboard error must be a direct function of snow depth.

The main issue is that sea ice surfaces are very rarely 1D layers at SAR altimeter footprint sizes and different sea ice surface have shown strong differences in backscatter values. Not only elevation distribution of backscatter matters, but also its azimuth distribution. Or more plainly, a patch of thin ice or open water with substantially larger backscatter coefficient than the surrounding snow-covered sea ice a bit off-nadir can dominate the main peak of the waveform and the corresponding retracker range is invariant of snow backscatter or the ice surface elevation distribution. The backscattered power arrives at later times and biases the retracker range high. And since the off-nadir angle is unknown (except for CryoSat-2 SARin with very limited coverage) the freeboard in these case will be be biased low, often by factors several times the snow depth thus providing a strong opposing bias to snow backscatter.

Additional issues are sea level anomaly errors directly affect the freeboard errors as well and that also for ice surface backscatter may have a relationship with elevation, resulting in a bias in waveform models akin to the sea state base in ocean altimetry. For gridded freeboard values there might be additional impacts based on the surface type classification and possibly the rejection rates of waveforms.

My main take away message that freeboard errors cannot be realistically modelled only from snow depth. Penetration factors may be used as an empirical bias correction, but they absolutely should not be used to infer the actual backscatter mechanism at local scale as the errors of gridded freeboards.

Most of the paper is focusing on the '2p' implementation (ice and snow depth inversion) with the commonly used assumption on penetration factors. Discussion of potential extension to generalising the inversion to 3p or 4p case studies to investigate potential biases in the freeboard retrievals are discussed in a secondary level. We feel that part is important to understand the potential of the method but agree with the reviewer that definitive conclusions cannot be drawn at this stage about actual penetration factors from this study alone. We nevertheless feel that this discussion and case studies are an important strength of our method, and we keep it in the discussion part of the paper with figures relegated now to the supplementary section.

While snow backscatter is undoubtly relevant at Ku-Band, there are issues with the assumption that the freeboard error must be a direct function of snow depth.

We want to clarify here that our use of alphas doesn't impose a covariance between snow depth and penetration factor as the alphas can evolve independently of snow and ice thicknesses (spatially and temporally). Alphas are therefore a

representation of where the main scattering horizon resides and are not tied by the snow thickness in any way. Our only constrain on the alphas are to impose a prior for α_{CS}= -0.5 to 0.5 and α = 0.5 to 1.5. Interestingly we have found that these priors are not critical on the results found.

As you can see below, the results are very similar between both inversions with different priors for the bias factor:

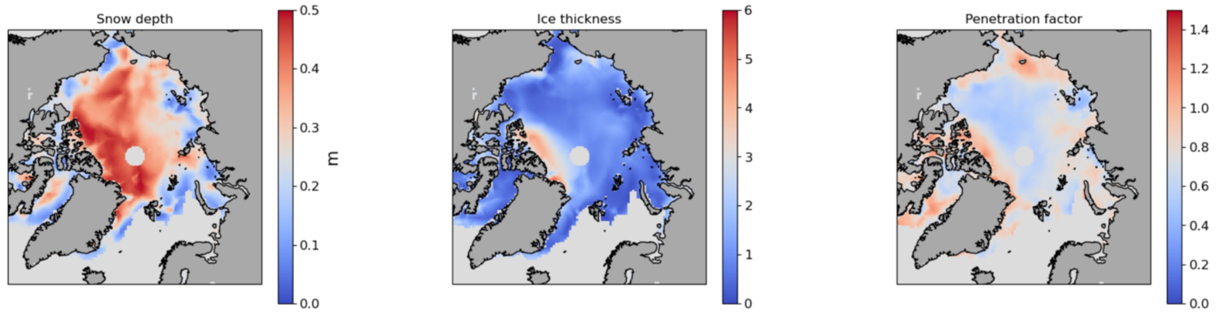


Figure 1: Inversion with CS-IS-3p for 2019/04. The prior used for the bias factor is [0, 1.5]

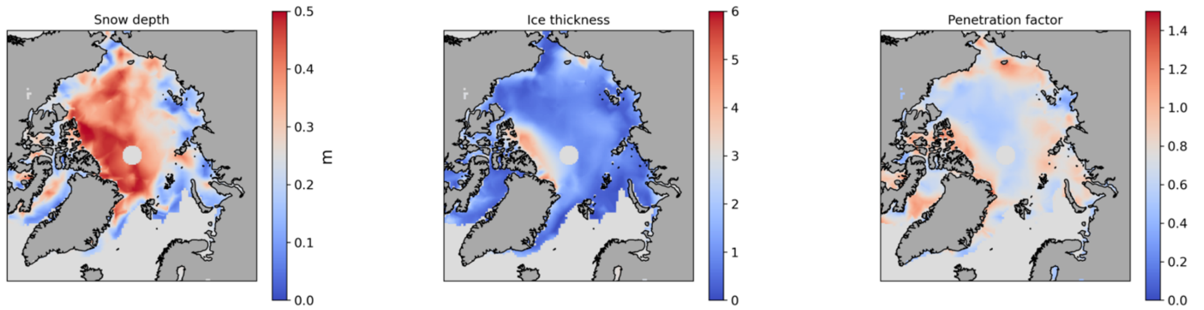


Figure 2: Inversion with CS-IS-3p for 2019/04. The prior used for the bias factor is [-0.5, 1.5]

The main issue is that sea ice surfaces are very rarely 1D layers at SAR altimeter footprint sizes and different sea ice surface have shown strong differences in backscatter values. Not only elevation distribution of backscatter matters, but also its azimuth distribution. Or more plainly, a patch of thin ice or open water with substantially larger backscatter coefficient than the surrounding snow-covered sea ice a bit off-nadir can dominate the main peak of the waveform and the corresponding retracker range is invariant of snow backscatter or the ice surface elevation distribution.

We will now include this comment in the paper (add references) but again our use of alphas doesn't preclude from this eventuality (see previous comment). In a way alphas can be incorporating other types of biases. We agree with the reviewer that this calls for nuancing their definition as 'penetration factors' and we have renamed these throughout as 'bias factors'. For example, the example given here of a thin ice patch dominating the return could be resulting in an artificially high 'penetration factor' or even larger than 1. In that sense we feel bias factor is more appropriate at this stage. Note that now all this discussion has been relegated to the discussion

and supplementary section. We also note the interesting similarities in the spatial/temporal patterns and magnitude of alphas to previous literature (Nab et all, 2025) and therefore we feel it is in an interesting addition to our study.

The freeboard in these case will be biased low, often by factors several times the snow depth thus providing a strong opposing bias to snow backscatter.

We had tried an alternative approach to add extra variables of bias (+B1 +B2 in the equations) but we felt that this was adding additional levels of degeneracy to the problem without enough additional physical justification.

Additional issues are sea level anomaly errors directly affect the freeboard errors as well and that also for ice surface backscatter may have a relationship with elevation, resulting in a bias in waveform models akin to the sea state base in ocean altimetry. For gridded freeboard values there might be additional impacts based on the surface type classification and possibly the rejection rates of waveforms.

We agree that sea level anomaly interpolation and roughness related biases remain that can pollute any altimetry approach (including ours). We note that we use as our main dataset physically retracted echoes (LARM, reference) to account for parts of the roughness effect but in line with Landy et al (2025) these roughness biases are not totally removed as evidenced by the roughness dependent differences between AltiKa and ICESat-2 freeboards.

Penetration factors may be used as an empirical bias correction, but they absolutely should not be used to infer the actual backscatter mechanism at local scale as the errors of gridded freeboards.

We totally agree with this statement. We have now rephrased these terms as bias factors that capture parts of these empirical biases. We feel that the patio-temporal maps of these bias are informative to understand potentially mechanisms that lead to deviations from the standard assumptions. We have relegated the corresponding figures to the supplementary and only keep discuss it in the main paper.

- **General Comment 2: Model setups with dimensions 3 and 4 may be create false minima.**

Inverting for 3 or 4-dimensional model space with only a 2-dimensional data space is an ill-posed problem. But the additional challenge I see is the promotion of false solutions by including the penetration factors. 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.

At least it is my guess that a wrong minimum is the reason why CS-IS-3p and CS-IS-4p do not generate competitive results compared to the other products and not against the CS-IS-2p setup without penetration factors.

One option to reduce the likelihood for false minima would be to define the prior distribution from the climatological values of snow depth and sea ice thickness and a realistic deviation from the mean value. The penetration factor range could also be narrowed to prevent the negative IS penetration factors seen in figures 12 & 14.

But the main take-away message from this analysis is, when taking the results from from CS-IS-2p/3p/4p as presented here is: The inclusion of penetration factors substantially degrades your retrieval. I am not sure that this is the message the authors intend to make

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. This is an important misconception from the reviewer. While this is obviously true for a local 2p equation with two unknowns it doesn't hold for non-local, prior informed Bayesian inversions as done here. This is the strength of the method! We have clarified this point in the revised version of the manuscript.

But the additional challenge I see is the promotion of false solutions by including the penetration factors.

We agree with this. We have now relegated this to the supplementary, but we find these sensitivity runs very useful to look at patterns. We also present now much more extensively the covariances of the solutions found and discuss these. For example alphas covaries snow depth in that way.

At least it is my guess that a wrong minimum is the reason why CS-IS-3p and CS-IS-4p do not generate competitive results compared to the other products and not against the CS-IS-2p setup without penetration factors.

This is a good insight and we now discuss that more extensively by showing the full solution probability maps. Another potential extension would be to inverse the anomalies to the prior.

But the main take-away message from this analysis is, when taking the results from CS-IS-2p/3p/4p as presented here is: The inclusion of penetration factors substantially degrades your retrieval. I am not sure that this is the message the authors intend to make.

We have revised the presentation of the results to emphasise more the novelty of the method and less on the specific applications to inversion problems that are challenging and will be further expanded in future research papers. These results have been moved to the supplementary.

- **General Comment 3: Operation IceBridge is very limited use for sea ice thickness validation.**

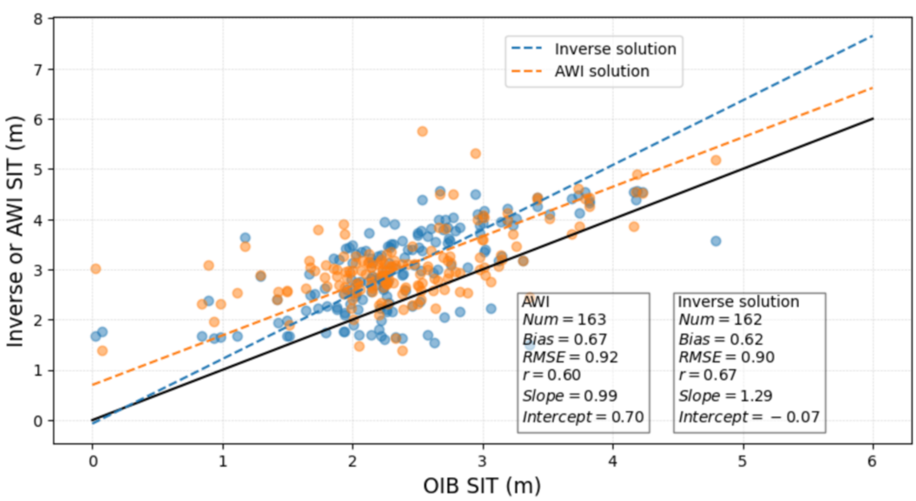
Airborne data from NASA Operation IceBridge is a good validation data source for snow depth on sea ice but not so much for sea ice thickness. The parameter is derived from the laser scanner and snow radar data, and the estimation of sea ice thickness uses one of the forward model functions. There are more suitable sensors for sea ice thickness validation such as upward looking sonars or EM-induction sounding sensors.

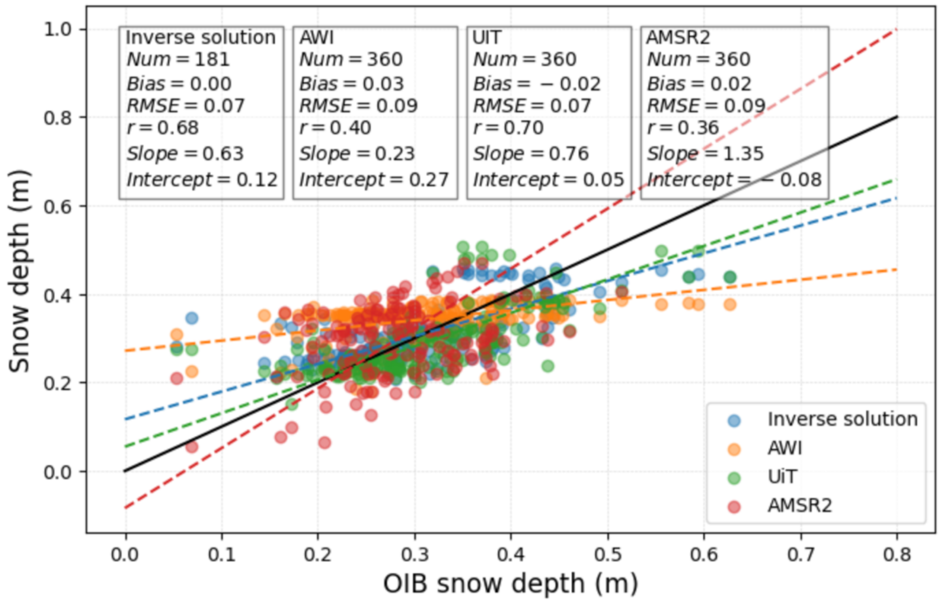
It is also my understanding from the manuscript that the authors use the OIB quick look product as is, and do not apply the sea ice density values for FYI and MYI as in the forward model functions. In this case the OIB sea ice thickness becomes inconsistent, since the OIB sea ice thickness is based on a single value for sea ice density (915kg/m3, Kurtz et al. 2015). An exact match between the inversion result and the OIB thicknesses would than rather point to an error in the freeboards/snow depth input.

I recommend here to use ULS observation, for example BGEP, for sea ice thickness validation. The required data should be available within the author team. That would provide a more robust validation, also in other month than Arctic spring.

- Kurtz, N., Studinger, M., Harbeck, J., Onana, V., & Yi, D. (2015). IceBridge L4 Sea Ice Freeboard,Snow Depth, and Thickness (IDCSI4, Version 1) [Data set]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center.

We thank the reviewer for this comment. We have re-computed the OIB QuickLook data using the same snow, ice and water densities as in the inversion. The results for SIT and SD are presented below:

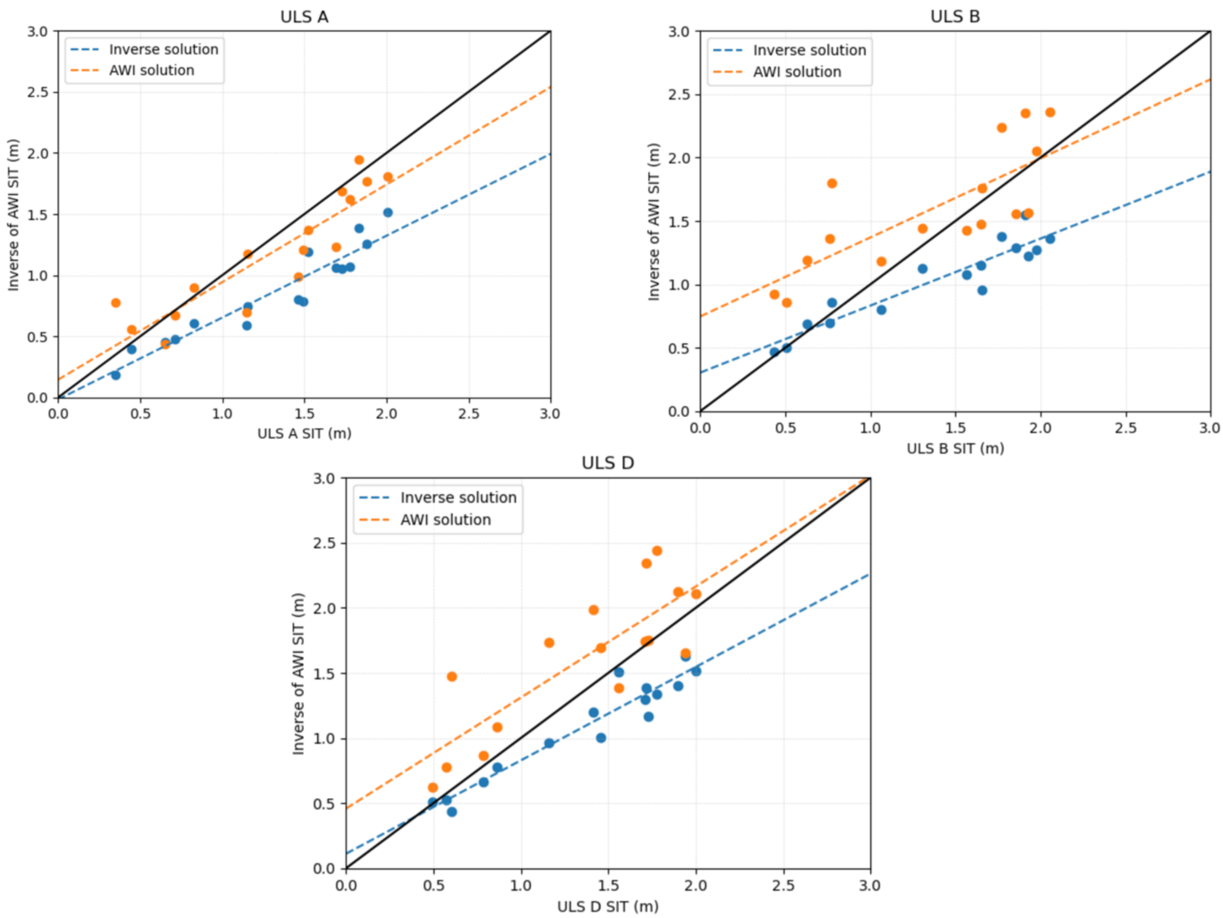




We note that, for both SIT and SD, the inversion produces results that are close to other products, and even better for some statistics (especially for the SD).

I recommend here to use ULS observation, for example BGEP, for sea ice thickness validation. The required data should be available within the author team. That would provide a more robust validation, also in other month than Arctic Spring.

We fully agree with the reviewer that we should extend our validations. We have added these validations against BGEP ULS data in the revised manuscript. The validations are presented below and will be included in the pa:



	ULS A		ULS B		ULS D	
	AWI	CS_IS_2p	AWI	CS_IS_2p	AWI	CS_IS_2p
Bias	-0.12	-0.44	0.23	-0.34	0.26	-0.27
RMSE	0.26	0.49	0.43	0.44	0.4	0.33
r	0.9	0.95	0.77	0.94	0.82	0.96
Slope	0.8	0.67	0.62	0.53	0.85	0.72
Intercept	0.15	-0.02	0.75	0.3	0.46	0.11

Table 1: Statistics of the validation

The upward-looking sonar (ULS) moorings were deployed by the Beaufort Gyre Exploration Project (BGEP). The BGEP data were collected at three locations in the Beaufort Sea (BGEP-A, BGEP-B, and BGEP-D) between 2010 and 2024. Sea ice thickness was derived from draft measurements by applying a ratio of 0.89, following the method of Rothrock et al (https://doi.org/10.1029/2001JC001208).
The inversion was then performed at mid-month intervals for the winter periods from 2018 to 2021, corresponding to the years for which comparison data are available from AWI.

- **General Comment 4: Delaunay Triangulation – More information needed**

I find it difficult to assess the maturity of sampling the observation space with the Delauny Triangulation. My understanding is that the number of triangles changes are variable variable, but what are the numbers of nodes/triangles in the ensemble? The only indication of triangle properties is in Figure 1, where the triangles look quite coarse and there seems to be extrapolation to the entire sea ice cover. Some triangles look substantially larger than correlation length scale of sea ice surface types and I am not sure that sea ice in triangles center can be accuratetely described by the node positions.

Another part of my confusion comes from the statement that “the dimension of the model space (the number of nodes n_Hi) will be treated as an unknown variable (L166)” and “At each iteration of the random walk, the algorithm makes a random choice between four perturbations of the model parameters (L196)”. If birth and death of nodes are chosen randomly than my assumption is that the number of nodes does not change substantially in the ensemble. Therefore, does the inversion for spatial resolution means to sacrifice spatial resolution in one area for another?

My recommendation is to describe the method in more detail and show the results more prominently. How are the nodes distributed and what are interpolation distances in the aggregation to the regular grid? What are the statistics of the ensemble spread?

We agree with the reviewer that we should provide a more detailed explanation of this method, particularly given that the purpose of this paper is to present an innovative approach. Therefore, we have added a section to the Results where we provide a synthetic simulation of the inverse approach (using as True SIT/SD fields results of previous inversions), emphasising the key aspects of the method and explaining in more detail how the inversion is performed. We have also included

a distribution of models for some of our main test cases to demonstrate the method's advantages with real data. Similarly, we have moved some standard deviation maps from the supplementary material to the main text.

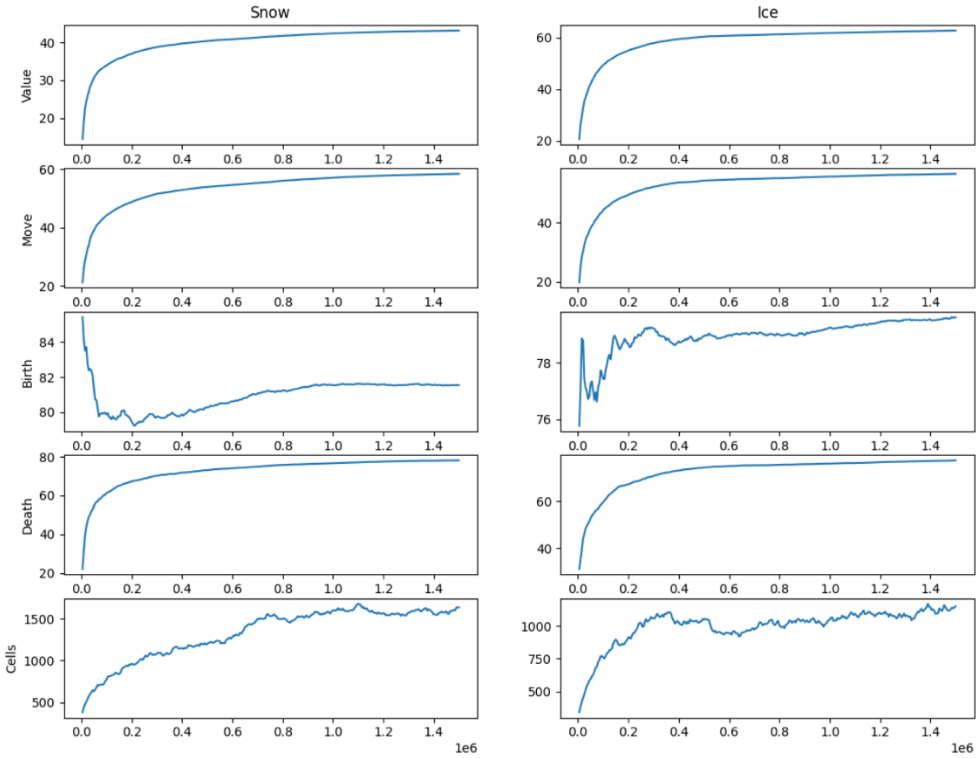


Figure 3: Acceptance rates of the inversion for 2019/04 and with the model CS-IS-2p.

My understanding is that the number of triangles changes are variable, but what are the numbers of nodes/triangles in the ensemble?

We have added the acceptance rates to the supplementary material. These are the rates at which each perturbation is accepted at each iteration, and they provide a better understanding of the sampling process for some inversions. An example is shown on Figure 3. As can be seen, convergence is reached for all the indicators, including the number of triangles, after around 500,000 iterations. Furthermore, the rates for Value, Move, Birth and Death are somewhat close to those found in previous studies using this method (Hawkins, 2018). These values are adjusted during calibration to ensure the rates accurately reflect our efficient exploration of the entire model space.

Some triangles look substantially larger than correlation length scale of sea ice surface types and I am not sure that sea ice in triangles center can be accurately described by the node positions.

The fact that some triangles are large is also part of the reason why apparently local underdetermined inversions can be solved within our framework. The exact position of the triangle is not important but rather the values of freeboards that fall within it.

If birth and death of nodes are chosen randomly than my assumption is that the number of nodes does not change substantially in the ensemble. Therefore, does the inversion for spatial resolution means to sacrifice spatial resolution in one area for another?"

Birth and rate are random but controlled by the cost function. So more cells are created where more data are present while less cells are available where there is less data or where the agreement between model and observations is poorer. This is true for each variable inverted for (snow, ice, alphas) independently of each other, thus there is no sacrifice of spatial resolution in one area for another. Again, that is the power of our method in that it is a purely statistical method including space.

We have added in the results section of the manuscript a description of a synthetic test we have performed in order to emphasize on the advantages of this approach, especially compared to other interpolation method.

Our inverse approach doesn't require an interpolation of the freeboards in order to compute the ice and snow depth. On the figures below, we can see that both methods produce similar results for the SIT and SD but with one additional step for the 'classical' approach.

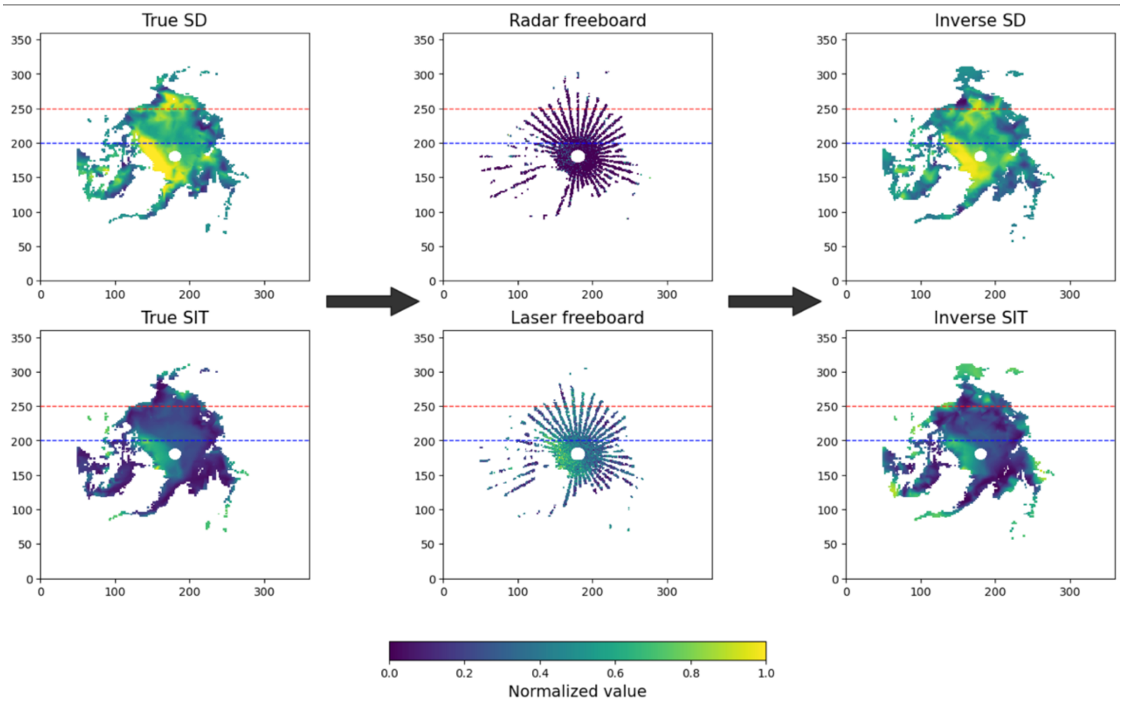


Figure 4: Schematic of the steps needed to retrieve SIT/SD when using the inverse approach. The True SIT/SD fields used for this synthetic test were computed using previous results of inversion. The blue and red lines correspond to the lines analysed in Figures 6 and 7. The values are normalized by the maximum of each parameter.

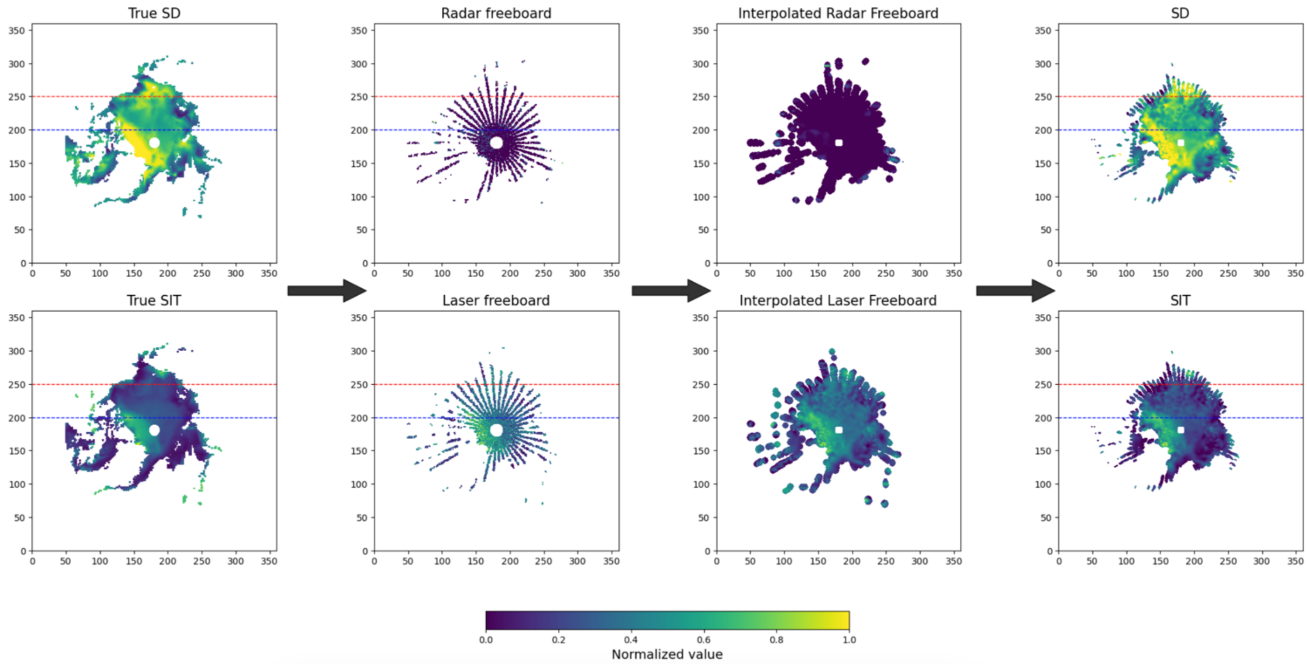


Figure 5: Schematic of the steps needed to retrieve SIT/SD when using a classical interpolation approach (in this case we use the GPSat interpolation method). The True SIT/SD fields used for this synthetic test were computed using previous results of inversion.

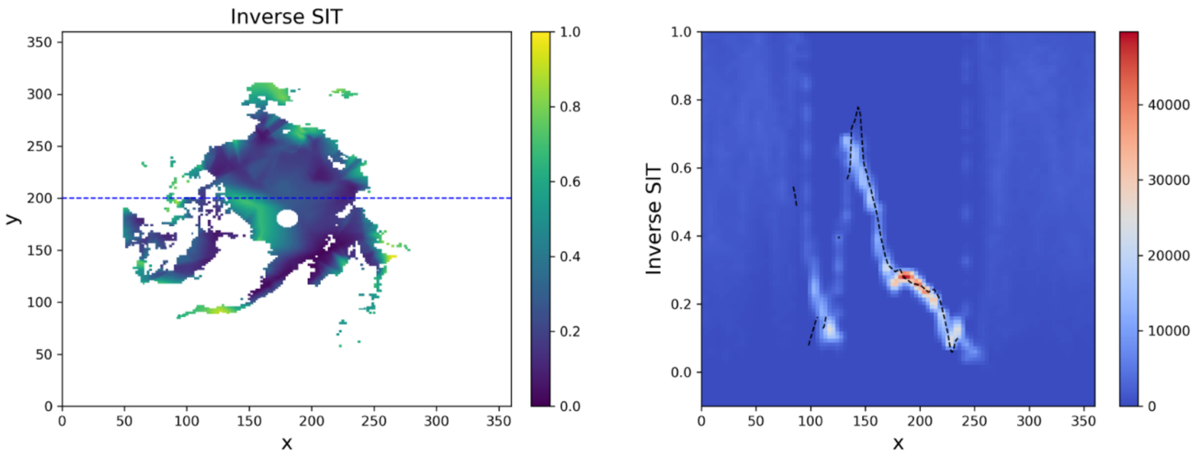


Figure 6: (right) Distribution of models on the horizontal line y=200. The black dashed line is the True SIT

The next two plots are presenting the distribution of models computed during the inversion. We chose these 2 examples with y=200 (Figure 6) and y=250 (Figure 7) to emphasize on the importance of the input data in order to obtain accurate results. For y=200, as we can see on Figure 4, the number of freeboards is important. Thus, the inversion is more constrained by the data and the distribution of models is tightened around the True SIT. On the other hand, for y=250, there is fewer input data (Figure 4) and the distribution of models is thus much more widespread.

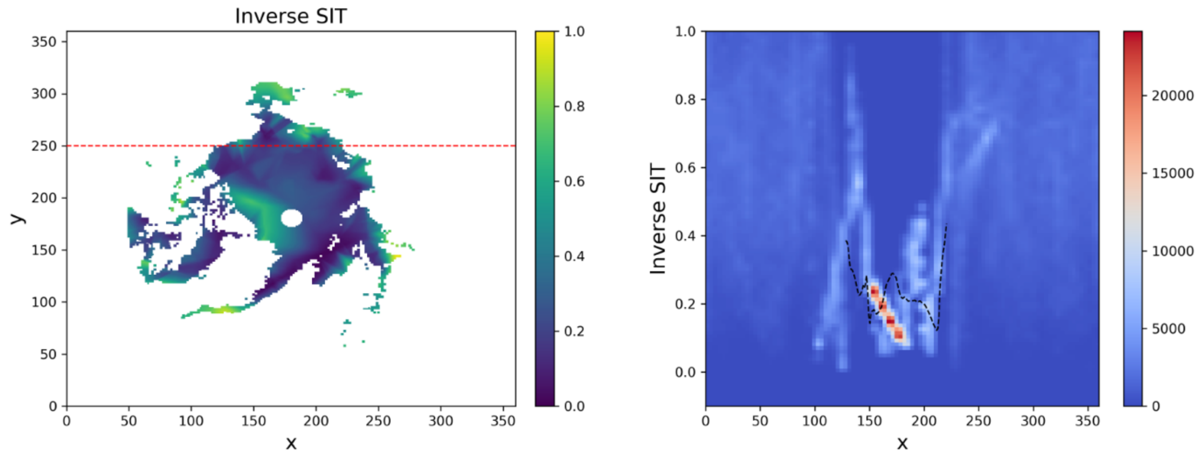


Figure 7: (right) Distribution of models on the horizontal line y=250. The black dashed line is the True SIT

The same kind of analysis can be performed for a specific location as depicted on Figure 8 and 9:

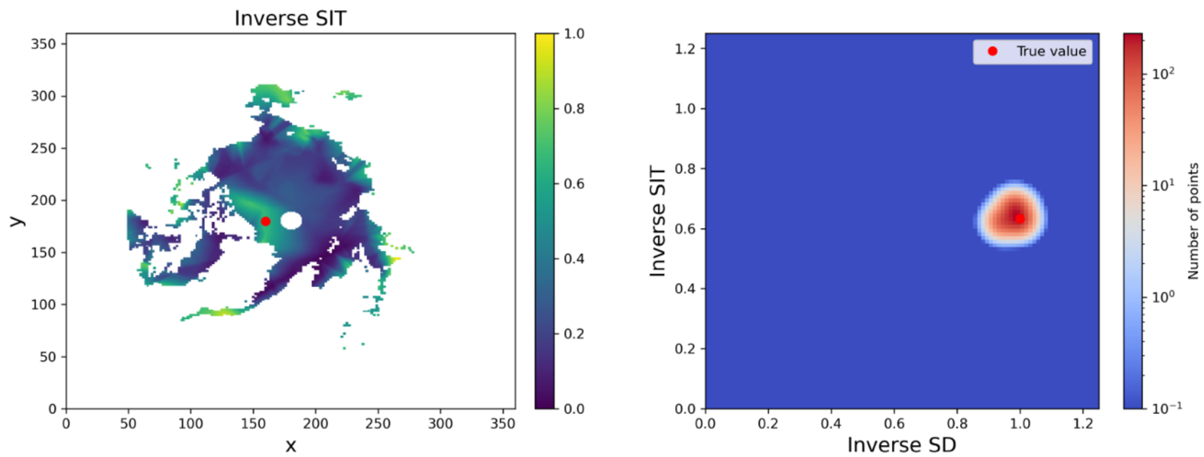


Figure 8: (right) Distribution of models on a specific location. The True SIT value is plotted with the red point

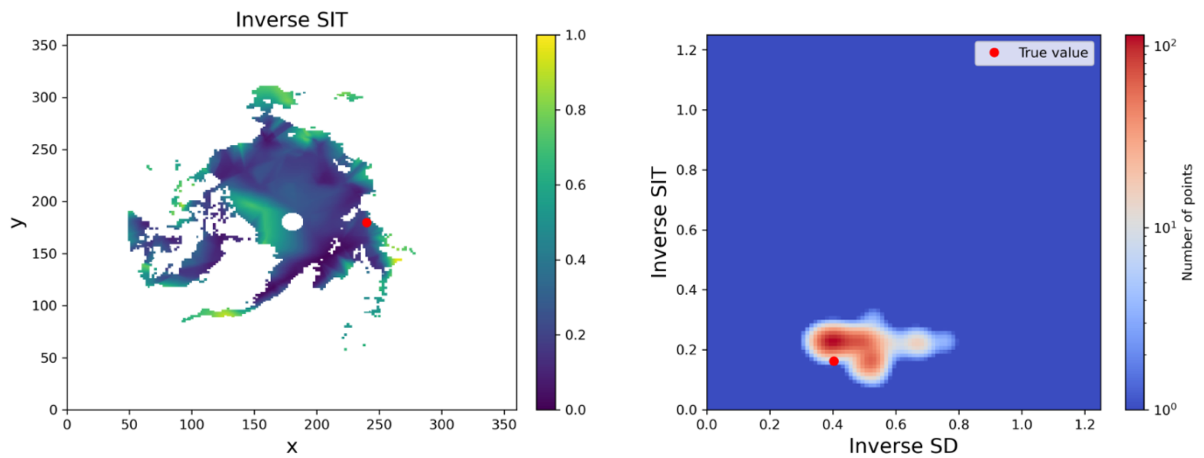


Figure 9: (right) Distribution of models on a specific location. The True SIT value is plotted with the red point

We hope that the replies above are addressing your concerns and we would like to thank you again for your very thorough and constructive criticism.

Regards
The authors